ABSTRACT

Title of dissertation: REGULATION AND DISTRIBUTION OF TRANSPOSONS IN HYPERTHERMOPHILIC ARCHAEA

Wirojne Kanoksilapatham, Doctor of Philosophy, 2004

Dissertation directed by: Professor Frank T. Robb
Center of Marine Biotechnology,
University of Maryland Biotechnology Institute

The genus *Pyrococcus* consists of approximately 12 described species that are Archaea with optimal growth temperatures near to 100 ^{o}C. *Pyrococcus furiosus* and *Pyrococcus woesei* were both isolated from marine sediments in the same Italian Vulcano Island site. Insertion sequences (ISs) were identified in the genomes of *P. furiosus* and *P. woesei*. Specific primers to differentiate between the types of IS elements were developed. Nucleotide sequence of the rDNA operon (AY519654) from *P. woesei* was sequenced for the first time.

Extensive co-linearity between the genomes of the *P. woesei* and *P. furiosus* was demonstrated using radio-labeled IS element sequence probes. A recent transposition event that disrupted the *napA* gene of *P. woesei* was revealed. The presence of a type I IS element was observed in the same context in the genomes of *P. furiosus* and *P. woesei*. The strains were proposed to be sub sp. of *P. furiosus*, based on their identical rDNA operon sequence and the presence of IS element markers.

The ISs have putative archaeal promoters with a 5' TATA box "boxA". Full length IS mRNAs were detected. Evidence for programmed translational frameshift sites that might limit the translation of the full length transposase was found. Two functional programmed translational frameshift sites were identified in *E. coli*. The mechanism of frameshifting identified in context 79 in the *tnp-I* sequence appears to be a novel -2 mechanism, in that the shift site is part of a -1 frameshifting motif. The possible role of a putative ribosome binding site (rbs) observed in context 79, that correlates with high frameshift levels, is discussed.

REGULATION AND DISTRIBUTION OF TRANSPOSONS IN HYPERTHERMOPHILIC ARCHAEA

by

Wirojne Kanoksilapatham

Dissertation submitted to the Faculty of Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:

Professor Frank T. Robb, Chair
Associate Professor Harold J. Schreier
Assistant Professor Jocelyne DiRuggiero
Associate Professor Kevin R. Sowers
Professor Philip J. Farabaugh
Associate Professor Steven W. Hutcheson

©Copyright by

Wirojne Kanoksilapatham

2004

Acknowledgements

I would like to express my gratitude to the Royal Thai Government, specifically to the Ministry of University Affairs and Silpakorn University, for granting me support for the first five years to pursue my studies at the University of Maryland College Park. The support has given me not only opportunities to be exposed to higher graduate education, but also to work on a project that inspires me, "Molecular Biotechnology involving Hyperthermophiles". I am also pleased to acknowledge the Center of Marine Biotechnology, University of Maryland Biotechnology Institute, for facilities and warm hospitality during the seven years of my laboratory work to complete this dissertation.

I have greatly appreciated the wise guidance and supervision of my advisory committee, Associate Professor Harold J. Schreier, Assistant Professor Jocelyne DiRuggiero, Associate Professor Kevin R. Sowers, and Professor Philip J. Farabaugh. The committee provided insightful and constructive comments and suggestions and directed me back on track several times during the long process to complete my dissertation. Opportunities to participate in classes on "Research Seminar: Molecular Biology" by Dr. Schreier and "Scientific Seminar Presentation" by Dr. Sowers have improved my presentation delivery. The sharp observations and wise comments of Dr. Farabaugh have made this dissertation more complete. My special thanks are given to Associate Professor Steven W. Hutcheson, the Dean Representative who has generously sacrificed his time on my dissertation defense. I am immensely grateful to Professor Frank T. Robb, my advisor. From my first day in the U.S., Dr. Robb has

expressed his keen interest and patience as well as support and concern not only academics but also to the financial and social health of me and my family. Dr. Robb also has a wonderful teaching spirit that he used to help me polish my knowledge. His invaluable comments and remarkable insights have motivated me to do my best work and have contributed substantially to my dissertation.

I have been helped and encouraged by several people from the Robb Lab, Center of Marine Biotechnology, in some way or another. Although it would be difficult to acknowledge everyone who has contributed to the completion of this dissertation, I would like to acknowledge Dr. Elizabeth M. O'Connor and Ms. Deborah Howard (a Ph.D. candidate for MEES program) for their enthusiastic and sincere help to improve my dissertation preparation. Also my special appreciation is given to two former members of the Robb Lab, Dr. Dennis L. Maeder and Dr. Juan M. González. Dr. Maeder generously helped me acquire a solid background in Bioinformatics. The hands on experience and marvelous molecular and microbiological techniques passed on to me by Dr. González (IRNAS-CSIC, Sevilla, Spain), have helped me sharpen my abilities in order to conduct and complete my dissertation work independently.

I also want to express appreciation for my wife, Dr. Budsaba Kanoksilapatham, and my two children, Panu and Sinee Kanoksilapatham for their companionship, as well as nourishment of body and soul.

Finally, I would like to dedicate my dissertation to my mother, who recently passed away in Thailand while I was pursuing my Ph.D. degree. Her emotional

support, sympathy, admiration and desire to see me achieve the highest degree in life have been an inspiration to me since my childhood.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Literature review

1.1. The Order Thermococcales

The Domain Archaea comprises four Kingdoms: Euryarchaeota, Crenarchaeota, Korarchaeota and the newly discovered Nanoarchaeota (Huber, *et al.,* 2002). The order Thermococcales of the Euryarchaeota are at present composed of two genera, *Thermococcus* and *Pyrococcus*. All species identified so far are cocci about 1 μm in diameter occurring singly or in pairs and all species are hyperthermophiles with two temperature ranges for growth, from 75-85 ^{o}C (*Thermococcus*) and 85-104 ^{o}C (*Pyrococcus*). All known species of the Thermococcales are anaerobes, utilizing peptides, yeast extract and less frequently carbohydrates as carbon sources. Elemental sulfur is not strictly required for growth in most species but has been found to stimulate growth *via* formation of hydrogen sulfide by removal of a toxic hydrogen gas (Woese and Olsen, 1986; Adams, 1994).

1.2. General features of insertion sequences, transposons, retrotransposons and other related insertion sequences

Transposons (Tns) are mobile DNA elements that undergo non-reciprocal recombination. Their insertion into coding regions generates mutations. Since the discovery of the first transposons in maize (so-called "controlling elements" or "Ac and Ds elements") by McClintock in 1951, more than 500 transposons have been discovered in all three domains of life including Eukarya, Bacteria and Archaea.

1.2.1. Bacterial transposons and insertion sequences

Bacterial transposons, particularly, are often found associated with antibiotic or metal resistance markers and on transmissible plasmids. Minimal transposon is termed insertion sequence (IS). Transposition proteins are generally encoded by one or two open reading frames that often occupy most of the length of the element. Generally, these ORFs encode only the transposase that are required for mobility. Mutations that occur in these ORFs usually turn the IS elements from "autonomous" to "nonautonomous" transmissible elements (Lodish *et al.,* 1995). Most of the IS elements are flanked by terminal inverted repeat (IR) sequences (Mahillon and Chandler, 1998). IR sequences (10 to 40 bp long) are recognition sites for transposases (tnp) that specifically mediate transposition reactions. Direct repeats are short nucleotide sequences (2 to 10 bp) flanking the IS elements (or Tns) that are duplicated during the transposition process.

When two copies of an IS element flanking a DNA segment are able to act in concert, rendering the intervening region mobile, they are termed transposons (Lodish *et al.,* 1995). Transposons are categorized into two major groups: (i) non-composite drug resistant transposons (e.g., Tn3 and Tn7), and (ii) composite transposons (e.g., Tn*5* and Tn*10*).

Both composite and non-composite transposons are often associated with genes encoding antibiotic resistance. These are, for example, genes encoding ampicillin and streptomycin resistance in transposon Tn3 (Hedges *et al.,* 1977), kanamycin, bleomycin and streptomycin resistance in transposon Tn*5* (Mazodier *et al.,* 1985),

trimethoprim, streptomycin and spectinomycin resistance in Tn*7* (Barth *et al.,* 1978), and for tetracycline resistance in transposon Tn*10* (Foster *et al.,* 1981). Intervening regions found in composite transposons encoding other phenotypes have also been reported. These include transposons associated with pathogenicity, symbiosis, and the ability to degrade toxic wastes. One example is a gene encoding a heat stable toxin found in *E. coli* flanked by IRs of IS*1* (So *et al.,* 1979). IS*1* was found distributed in the chromosomes of various Enterobacteria, R factors and of the prophage P1. It participates in transferring antibiotic resistant genes to multiple antibiotic resistant clinical strains (Dumans *et al*., 1989; Rhodes *et al.,* 2000; Ramirez-Santos, *et al.,* 1992). Genes involved in symbiosis in *Rhizobium leguminosum* are surrounded by transposons that reside in pNGR234 (Freiberg *et al.,* 1997) and several genes coding for degradative and catalytic enzymes were transferred by several IS elements. The latter case is demonstrated by genes encoding the catabolism of toluene flanked by Tn*4651*/Tn*4653*, naphthalene in Tn*4655*, chlorobenzoate in Tn*5271*, and chlorobenzene in Tn*5280* (Wyndham *et al.,* 1994). Not surprisingly, clusters of insertion sequences including IS*1354,* IS*1355*, and IS*1357* were found in the vicinity of *dcm* genes of a strain of *Methylobacterium* sp. DM4 (a Gram-negative facultative methylotrophic bacterium), and *Pseudomonas* sp. DM1 (a Gram-negative aerobic bacterium). The *dcm* genes encode for dichloromethane dehalogenase (*dcmA*) and a regulatory protein (*dcmR*). The dichloromethane dehalogenase can catalyze dichloromethane (DCM) to inorganic chloride and formaldehyde, a central metabolite of methylotrophic growth. This versatility in DCM-utilization is likely to be

horizontally transferred among methylotrophic bacteria by transposons (Schmid-Appert *et al.,* 1997).

1.2.2. Archaeal transposons and insertion sequences

IS elements appear to be quite widely distributed in the Archaea. This is revealed by the completed archaeal genomic sequences of halophilic, methanogenic, and hyperthermophilic archaea (Bult *et al.,* 1996; Ng *et al.,* 1998; Klenk, *et al.,* 1997; Smith *et al.,* 1997). IS elements from archaea are classified on the basis of the sequence identities of their encoded transposases and many have been assigned to families like the system for classification of IS elements from bacteria. A minireview on "mobile elements in archaeal genomes" cited more than 300 IS elements found in the genomes of both Crenarchaea (termed "ISC") and Euryarchaea (termed "ISE") (Brugger *et al.,* 2002). The elements fall into two main types, named in the minireview as "autonomous insertion sequence elements" and "non-autonomous miniature inverted repeat element (MITE)-like elements". The latter term "MITE"-like elements refers to defective elements that lack the ORF encoding the transposase and are thus equivalent to the terms "nonautonomous transmissible elements" (mentioned above) and Ds (dissociation) elements commonly found in Eukarya including *Zea mays* (mentioned below) (Lodish *et al.,* 1995).

Insertion sequences found in extremely halophilic euryarchaea named "ISHs" are active in insertional mutations and DNA rearrangements of the *Halobacterium* (Charlebois and Doolittle, 1989). They are non-randomly distributed within the genome of *Halobacterium halobium.* They reside mostly on plasmids, on

bacteriophage ΦH, and in AT-rich chromosomal regions. In *Halobacterium* sp. NRC-1, an ISH was demonstrated to account for the movement of chromosomal genes to the 200-kbp mega-plasmid pNRC100, which seems to be evolving into a mini-chromosome (Ng *et al.,* 1998). The haloarchaeal transposon ThD28 was artificially constructed from two copies of ISH*28* isolated from *H. salinarum* (formerly *H. halobium*) flanking a mevinolin resistance marker (Mev^r) obtained from *Haloarcula hispanica* (Dyall-Smith and Doolittle, 1994).

Two hundred insertion sequences (sizes 774 to 1913 bp) were reported to cluster into two broad regions in the genome of *Sulfolobus solfataricus* P2, a hyperthermophilic crenarchaeon (She *et al.,* 1998). These elements constitute approximately 10% of the genome. Most of the IS elements (184 elements) belong to bacterial/eukaryal families, whereas the rest may be archaea-specific. Among these elements, ISC*774* (archaea-specific family), ISC*1212* (IS*5* family), and ISC*1229* (IS*110* family) contain two ORFs which overlap. ISC*1316*, ISC*1332*, and ISC*1913* are also found in the conjugative plasmids pNOB8 and pING of *Sulfolobus* (Stedman, *et al.,* 2000). Four different, nonhomologous ISC elements (ISC*1058*, ISC*1217*, ISC*1359,* and ISC*1439*) of the IS*4*-IS*5* families have been demonstrated to be active in spontaneous mutation of *S. solfataricus* (Martusewitsch *et al.,* 2000). They were identified at different positions within the chromosomal *pyrEF* locus (encodes for uracil-auxotrophic phenotype) of mutants resistant to 5-fluoroorotate. All 4 IS elements were found in high copy numbers not only in the genomes of *S. solfataricus* P1 and P2,

but also in natural strains. Thus are actively reshuffling genome of the genus *Sulfolobus*.

Two insertion sequences (ORF406 and ORF413) were identified in the first archaeal conjugative plasmid, pNOB8, which was isolated from *Sulfolobus* isolate NOB8-H2 (She *et al.,* 1998). ORF406 has similar nucleotide sequence (>75% identity) to several IS elements identified in the genome of *S. solfataricus* P2 (Sensen *et al.,* 1996), whereas ORF413 is a putative homolog of the IS*256* family of bacterial transposases, with highest identity to transposases from *Mycobacterium* and *Rhizobium*. The ORF406 lacked an inverted repeat; however, the ORF413 was flanked by 32 bp inverted repeat sequences and 5 bp direct repeat sequences.

Eleven members of IS*AMJ1* (approx. 300 to 700 bp long) distributed among three groups (A, B, and C) were detected in the complete genome sequence of *Methanococcus jannaschii* (a methanogenic euryarchaeon) (Bult *et al.,* 1996). Two copies of the the IS*AMJ1-A* (703 bp in length) are highly similar to each other containing an ORF coding (ORFs MJ0017 and MJ1466) for a putative transposase (214 amino acids in length) that is 27% identical to the IS*240* transposase of *Bacillus thuringiensis*. Eight copies of the family (IS*AMJ1-B*) range from 358 to 360 bp in length. They appear to be deletion mutants of the IS*AMJ1-A* family. Some members of group B have putative frameshifts and in-frame UGA codons. The remaining single copy of IS*AMJ1-C* is 265 bp in length and its internal sequence lacks part of the internal sequence of the IS*AMJ1-A*.

The genome sequence of *Archaeoglobus fulgidus* (a hyperthermophilic euryarchaeon) contains several copies of putative IS elements that are classified into 3 groups named LR-3, LR-4, and LR-6 (Klenk *et al.,* 1997). The LR-3 (ISA*1214*) has 14-bp inverted repeats and 2 genes-one of which has weak identity to a transposase from *Halobacterium salinarum*. A copy of LR-3 (IS*A1214-5*) was found interrupting ORF AF2090. The LR-4 and LR-6 encode putative transposases which were not identified in *M. jannaschii*.

The genome of *Pyrococcus furiosus* contains 23 copies of IS elements including the composite transposon flanking16 kb DNA encoding the ABC transport system for maltose and trehalose that was probably laterally transferred from a *Thermococcus sp.* (DiRugierro *et al.,* 2000). A copy of the IS element was also detected in the genome of *Thermococcus litoralis,* however, the genomes of *Pyrococcus abyssi* and *Pyrococcus horikoshii* contain none of the IS element.

1.2.3. Retrotransposons and other related transposons in eucaryotes

Many strains of *Saccharomyces cerevisiae* contain 50 to 100 copies of the viral retrotransposon DNA sequences called Ty elements. They transpose *via* a reverse transcription process, and their structures are flanked by two long terminal repeat (LTR) sequences (Boeke, 1989). Nucleotide sequences of Ty1 and Ty2 families differ mainly within the 2 open reading frames (TYA and TYB). However, the ends are relatively similar. TYA is analogous to the *gag* genes of retroviruses. Domains of protease, integrase, reverse transcriptase, and RNase H similar to Pol proteins in retroviruses are located in the TYB. The LTR is called δ for Ty1 and Ty2 (copia-like

retrotransposons), and σ for Ty3 (gypsy-like retrotransposons). Nucleotide sequences of LTRs are more complicated than IRs found in bacterial and archaeal transposons. LTR comprises 3 regions similar to U3, R, and U5, respectively in retroviruses. In brief, U3 contains short, usually imperfect, inverted repeats (conserved GT) at its 5' end and is followed by two sub-domains (transcriptional enhancer and promoter). R is a terminal repeat that is necessary for DNA strand transfer during reverse transcription; U5 contains the inverted repeat sequence (conserved AC) at its 3' end. Similar to IR (specifically recognized by tnp), the *attU*3 and *attU5* (at 3' and 5' ends of the retrovirus, respectively) are required for the integration process that is mediated by the enzyme integrase (Varmus and Brown, 1989).

Although most of the mobile elements in *Drosophila* function as retrotransposons, P elements function as bacterial transposons. Like IS elements, the P elements have inverted repeat sequences at each end. Moreover, both copies of this sequence are required for transposition. Furthermore, the elements are usually flanked by direct repeat sequences. The central portions of the P elements encode bacterial-like tnps, except that the gene encoding the P element tnp contains an intron (Lodish *et al.,* 1995).

The controlling elements in maize (*Zea mays*) were renamed as Ac or *activator*, and Ds or *dissociation* after the types of mutation they generate (McClintock, 1951). Mutations by Ac elements were often found to revert whereas mutations by Ds elements tend to be associated with chromosome breaks and do not revert. The Ac element encodes a tnp (5 exons) and is larger than a Ds element. Ac can transpose

autonomously; however, Ds cannot transpose autonomously because the Ds tnp is defective. Both Ac and Ds elements are flanked by terminal inverted repeats that can be recognized by the Ac tnp.

Long interspersed elements (LINES) and short interspersed elements (SINES) are two classes of abundant mobile elements found in mammals (approx. 10% of total human DNA) that transpose through reverse transcription. However, both classes contain no LTR and direct repeats usually flank the elements. The L1 LINE family is the most common LINE element (6- to 7-kb long), with approximately 50,000 copies in the human genome. A L1 contains two ORFs; one of the ORFs encodes a homolog of reverse transcriptase. SINES are called *Alu* sequences in humans. The name *Alu* is derived from the presence of a site for the restriction enzyme *Alu-I* in the repeat. They are shorter (approx. 300-bp long) but 10 times more abundant in the human genome compared to L1 elements. Like all other DNA mobile elements, *Alu* sequences are usually flanked by direct repeats. Unlike L1 elements, *Alu* sequences do not encode a reverse transcriptase. Similar to L1 elements, *Alu* sequences are transcribed by RNA polymerase III and contain an A/T-rich region at one end (Lodish *et al.,* 1995).

1.3. Bacterial insertion sequences that carry mobile promoter motifs

Several insertion sequences (such as many IS elements in the Tn3 family including Tn*1000* (γδ), Tn7, and IS*10*) carry promoter-like sequences near the IR ends. The ends of these IS elements contain outward facing -35 regions (Galas and Chandler, 1989). IS*1355* (a dichloromethane transposon) carry a short ORF153 in 0 phase and ORF179 in -1 phase. The ORF 153 was preceded by a potential -35/-10 promoter and

by a putative ribosome binding site. Integration of such sequences into appropriate regions can form new outwardly directed hybrid promoters at the insertion site. Indeed, these elements were shown to play an important role in activating or enhancing the expression of their flanking genes *in vivo*. It was shown that IS*2* (a member of the Tn3 family) provides a -35 promoter region if correctly spaced from the resident *gal*E -10 or *amp*C -10 regions in *E. coli* (Jaurin and Normark, 1983; Hinton and Musso, 1982). Similar results were reported with the insertion of IS*1* into the promoter region of the pBR322 beta-lactamase (*bla*) gene. IS*1* provided its -35 sequence located in each end to form a new hybrid promoter with the resident -10 of the *bla* gene (Prentki *et al.,* 1986). IS*1* was also shown to exhibit strong orientation bias at the *E. coli bgl* locus. The *bgl* operon encodes proteins for transport and utilization of aromatic β-glucoside; the operon is transcribed at low levels in wild type cells. The IS*1* insertions near the CAP binding site activate the *bgl* operon (Manna *et al.,* 2001). IS*3* (of the Tn*3* family) and IS*10* (from Tn*10*) transcriptions of flanking genes have been demonstrated to initiate from promoter sequences within the ends of the IS elements (44 nucleotides within one end of the IS*3* and pOUT of the IS*10*, respectively) (Charlier *et al.,* 1982; Kleckner, 1989). Formation of transient transposon circle of IS*911* can drive transcription of tnps (Duval-Valentin *et al.,* 2001; Ton-Hoang *et al.,* 1997). A well known mechanism regulating flagellar phase variations in *Salmonella typhimurium* involves a specific DNA recombination reaction catalyzed by an invertase (*hin*) encoded in a host-associated Tn (related to the Tn*3*) (Simon *et al.,* 1980; Silverman *et al.,* 1979). The Tn provides a partial promoter sequence upstream of the *fljB* (which

encodes an axial component of flagellin phase-2) and *fljA* (which encodes a phase 1 repressor protein) genes. Briefly, two structural genes, the *fliC* gene encoding an axial component of flagellin phase-1, and *fljB* are in distant locations on the chromosome. The promoter for the *fljB* operon is part of an inverted repeat (*hix*) sequence that flanks the *hin* operon (regulates expression of the Hin protein or the invertase). The Hin protein catalyzes a specific DNA inversion reaction at the *hix* site to invert the *hin* operon thus switching the *fljB* and *fljA* genes downstream on and off. The FljA represses expression of the FliC phase-1 flagellin when the *fljB* operon is expressed (Macnab, 1996).

1.4. Host factors associated with transposition

Transposition activity of bacterial IS elements and perhaps archaeal IS elements is frequently modulated by various host factors. These effects are generally specific for each element. Such factors include the DNA chaperone integration host factor (IHF), histone like proteins (HU and HNS), factor for inversion stimulation (FIS), the replication initiator DnaA, the protein chaperone/protease ClpX, ClpP, and ClpA, the SOS control protein LexA, and the Dam DNA methylase. In addition, proteins or antibiotics that govern DNA supercoil density *in vivo* may influence transposition.

The DNA chaperones may play a crucial role in correcting the three-dimensional structure of DNA-protein complexes necessary for productive transposition and in regulating the expression of *tnp*. For example, sites for the binding of the IHF defined earlier ((AorT)ATCAANNNNTT(AorG) or AAnnnnTTGAT)) were identified in the ends of IS*1* (Gamas *et al.,* 1985; Gamas *et al.,* 1987), of IS*10*

(Morisato and Kleckner, 1987), and of Tn*1000* (Wiater and Grindley, 1988). IHF is required for the site-specific recombination that leads to the integration of bacteriophage lambda into the bacterial genome. Sites for the binding of DnaA protein were also identified at the end of IS*50* (from Tn*5*). This site is believed to be involved in the modulation of IS*50* activity (Yin and Reznikoff, 1987; Reznikoff, 1993). Mutations in the *dnaA* gene in *E. coli* decreased the frequency of transposition of Tn*5* without affecting Tn*5* gene expression. Dam methylation sites (GATC) play an important role in regulating not only IS*50*, but also IS*10* transpositions. Both transposons, including IS*903*, carry methylation sites in the *tnp* promoter regions, and in each case, promoter activity is increased in a *dam* mutant. Excision frequency of the *Antirrhinum* transposon Tam3 is dependent on methylation in the ends of the element (Kitamura *et al.,* 2001). The Fis protein of *E. coli* was found to stimulate both Tn*5* and IS*50* transposition events and to inhibit IS*50* transposition in *Dam*$^-$ bacteria (Weinreich and Reznikoff, 1992). DNase I foot printing revealed that the inside end of IS*50* overlaps a Fis site. The Fis site at the inside end overlaps three Dam sites (GATCAGATCTTGATC). Fis binds efficiently only to the unmethylated substrate. Fis protein expression is growth phase dependent; thus, high levels of Fis present during exponential growth might stimulate transposition events. A site for a Rho-dependent terminator has been found at the end of the *insA* gene of IS*1*. InsA and InsB proteins are poorly expressed *in vivo* because the putative promoter and ribosome-binding sites appear to be weak. However, high levels of InsA from a strong promoter were observed. InsA and InsB are coupled through a Rho-dependent terminator.

Protein chaperone/proteases such as ClpX, ClpP, and Lon, play a vital role in proteolysis and transposition. Together with ClpP, ClpX can cleave Mu repressors and thus stimulate lytic growth of the Mu phage (Laachouch *et al.,* 1996). Lon protease is implicated in the proteolysis of the *cis*-acting *tnp* (*transposase*) from IS*903* (Derbyshire and Grindley, 1996), resulting in instability of the *tnp*.

IS*1*-based transposons Tn*9* and Tn*5* behave similarly with respect to the *recA* allele. Intermolecular transposition of both transposons is stimulated in the presence of RecA (Ahmed, 1986). Excision and transposition of Tn*5* are greatly enhanced by *recA* genes through repression of LexA upon expression of the *tnp* gene (Kuan and Tessman, 1992; Kuan and Tessman, 1991; Kuan *et al.,* 1991). A binding site for LexA on the end of the IS*50R* (AC CTT TCCCGTTTC CAG GA) in Tn*5* was demonstrated (Walker, 1984) and thus supports the expression of *tnp* enhanced by the SOS response. On the other hand, Weinreich and colleagues (1991) reported that transposition of Tn*5* is inhibited (5- to 10-fold) in cells that express SOS functions constitutively (e.g., *lexA*$^-$ strains).

Chi sites (GCTGGTGG) specifying sites of action for RecBCD nuclease were found in IS*136*. Transposition of Tn*5* has implications from super-coiling of donor DNA which is maintained *in vivo* by a gyrase (Isberg and Syvanen, 1982). DNA gyrase was found to be important in transposition of Mu bacteriophage (Pato and Banerjee, 2000; Sokolsky and Baker, 2003). Presumably, this phenomenon involved configuration of the transposon and phage DNAs so that they exposed the recognition sites to enzymes involved in the transposition process. DNA polymerase-I (Syvanen *et*

al., 1982) presumably repairs single-strand gaps that result from staggered cleavage of the target. This repair mechanism often generates a flanking direct repeat that is a hallmark of transposition.

Ty elements in budding yeasts exhibit preference sites for their selection of target to reduce the harmful effects caused by gene disruption. Copia-like (Ty1, 2, 4, and 5) and gypsy-like (Ty3) elements require cell-encoded factors to mediate position-specific integration. Ty1, Ty2, Ty3, and Ty4 require the presence of subunits of transcription factors (TFIIIB and TFIIIC) to target them to specific sites near the transcription initiation site of genes transcribed by RNA polymerase III (Ty1 and Ty3) and tRNA (Ty2 and Ty4) (Yieh *et al.,* 2000). Integration of Ty5 into regions of silent DNA including the silent mating type loci and telomeres requires factors involved in the establishment of silent chromatin (Zou and Voytas, 1997).

Mechanisms used by viral complexes and retrotransposons to enter the nucleus are diverse. For instance, some retroviruses access the chromatin of mitotic cells, in which the nuclear membrane is not intact. On the other hand, human immunodeficiency virus-1 (HIV-1) can enter the nucleus of infected cells independently of mitotic nuclear disassembly (Katz *et al.,* 2003). The mechanism of HIV nuclear localization is controlled by matrix, viral protein R, and integrase (Bukrinsky and Haffar, 1999; Fouchier and Malim, 1999). Nuclear localization of the Ty3 element is mediated by integrase (Lin *et al.,* 2001).

1.5. Transpositions and mechanisms

1.5.1. General modes of transpositions

Transposition reactions are a variant of DNA recombination that results in the translocation of transposable DNA from donor molecules to nonhomologous target sites. Transpositions are mediated by tnp enzymes usually encoded in each IS or Tn. Typically, the tnps recognize inverted repeat sequences specifically at the ends of the IS or Tn. The cleavage steps during transposition differentiate modes of transposition into nonreplicative and replicative modes (Mizuuchi, 1992). In nonreplicative mode (conservative mode or simple insertion), the element is excised from the donor site and reinserted into a target site without replication (Figure 1). In this mode, both strands of donor DNA are cleaved *via* a mechanism that generates hairpin (3'OH) to produce double strand breaks prior to the insertion of the donor DNA into the target DNA without duplication of the donor DNA. Examples of elements which adopt this strategy include Tn*7*, Tn*10*, Tn*5*, the Tc/mariner family, and the P element.

The replicative mechanism is rather complicated and best studied in Mu phage, Tn*3* family members, IS*911* and IS*2*, and the IS*6* family. Replicative transposition involves cleavage of only one strand at the proximal and distal ends of the transposon (Savilahti and Mizuuchi, 1996). At this stage, the transposon is still attached to the donor molecule. Transferring the donor transposon to the target occurs in such a way as to create a DNA replication fork. Unlike simple insertion, the 3' ends of the donor are presumed to act as primers in the replication folk, to duplicate a new copy of the donor

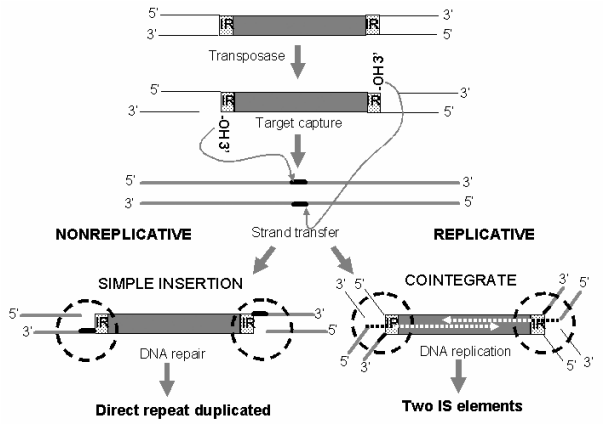


Figure 1. Modes of transposition. Top of the diagram shows cleavage of one strand on each inverted repeat end by transposase. Hydroxyl groups at the 3' ends (3' OH) attack target DNA in target capture step in staggered fashion (indicated by small arrow). In nonreplicative mode (bottom left), a 2nd cleavage releases the IS element from DNA followed by simple insertion in to the target DNA. Gaps generated by this step will be repaired by host enzymes to create direct repeats flanking the IS element. In replicative mode (bottom right), there is no 2nd cleavage step and the IS element remains attached to the donor molecule. Transferring the donor transposon to the target occurs in such a way as to create a DNA replication fork and cointegrate structure. DNA replication occurs from the 3' OH ends of target DNA primers to duplicate the IS element and followed by resolving of the cointegrate structure. Gray rectangles represent IS elements. Dotted box represents inverted repeat (IR).Thin double lines represent donor DNA. Thick double lines represent of target DNA. Short black lines represent short fragment of nucleotide (ca. of 2-10 bp long) on target molecule. Circles show gaps generated after strand transfer reactions. Dotted arrows represent directions of DNA replication.

molecule, and to form cointegrate structures that are resolved later either by resolvases encoded in some transposons or by host enzymes (Figure 1). In the case of IS*911*, IS*2* and IS*3*, tnp nicks a strand at one end of the transposon and transfers it to the same strand of the opposite end to form a transposon circle. The transposon circle then undergoes integration. An alternative mode of the IS*3* family including IS*911*using a cut and paste mechanism can be observed. In this mode, the transposon is excised as a linear molecule by double-strand cleavage at each end, followed by strand transfer into the target molecule (Ton-Hoang *et al.,* 1998).

Although the details on transposition mechanisms vary among IS elements and the Mu bacteriophage, their mechanisms are basically similar. The process begins when tnp binds an inverted repeat sequence at the end of donor DNA. Then, the end sequences are brought together *via* oligomerization of the enzyme to form a nucleoprotein complex structure (synaptic complex). Once a stable synaptic complex has been formed, tnp catalyzes cleavage at the 3' ends of the element by nucleophilic attack to expose a free 3' OH group (Peterson and Reznikoff, 2003). Unlike site-specific recombination reactions, cleavage by the transposase does not require ATP (Hallet and Sherratt, 1997). The hydroxyl group, in turn, acts as a nucleophile to attack the 5'-phosphate group in the target DNA in a transesterification reaction.

In Tn*5* transposition, the initial chemical steps result in the formation of a hairpin intermediates, followed by blunt end cleavage of the transposons at the IR. The transposon is released from the donor DNA (Bhasin *et al.,* 2000; Bhasin *et al.,* 1999). In other words, tnp catalyses a 3' hydrolytic nick to form free 3'-OH groups that then

attacks the phosphodiester bond on the opposite strand to form a hairpin at the transposon end. This intermediate is further hydrolyzed by tnp to expose another 3'-OH group and the transposon is subsequently released from donor DNA. Strand transfer occurs *via* one step transesterification in which the two 3'-OH groups (one from each transposon end) attack phosphodiester bonds in the target DNA in a staggered fashion. The transposon then inserts into this site by joining the 3' end of the transposon to the 5' end of the target DNA and leaves gaps flanking the other strands.

In vivo, transposon excision may introduce a potentially lethal double-strand (DS) break in the donor molecule that can be rescued by DS break repair or by gene conversion using a sister chromosome as a template. For IS*10*, passive replication of the element by host replication can activate transposition. This ensures duplication of IS*10* in which the nonreplicative mechanism occurs before DS break transposition (Kleckner *et al.,* 1996). Probably because of the potentially lethal event, transposon activity is usually regulated at a low level.

1.5.2. Transposition in Mu bacteriophage and Tn*7*

Mu bacteriophage exhibits two life cycles in its host: lysogenic and lytic stages. On injection of phage DNA into the host cell, it is integrated at random sites on the host chromosome by the phage-encoded transposition apparatus (including MuA and MuB proteins) to establish the lysogenic state by a conservative mechanism (Chaconas, and Surette, 1988). On the other hand, lytic development involves many replicative transposition events that exploit the host replication apparatus (including a host encoded HU protein), the transposase (MuA), and a transposition protein (MuB) to

form multiple integrated copies of Mu. HU aids the assembly of MuA into an oligomeric transposome that is tightly bound to both Mu ends (Watson and Chaconas, 1996). The transposome then promotes integration of Mu ends into the target DNA that is bound by MuB. In this process, the tetrameric core of the transposome produces a nick at each Mu end and promotes the transfer of the resulting 3'-OH ends to the target DNA. The resulting strand exchange product generates a DNA replication folk on host chromosome that becomes the initiation site for Mu DNA synthesis. At this stage, host replication proteins will duplicate Mu DNA and form a co-integrate product. However, the transposome remains tightly bound to both ends after strand exchange has been completed (Craigie and Mizuuchi, 1987; Surette *et al.,* 1987).

Tn*7* is a large (14-kb) non-composite transposon that contains several antibiotic resistance genes and transposition genes (tnsE, tnsD, tnsC, tnsB and tnsA, respectively from the left to the right end). Transposons that are similar to Tn*7*, include such as Tn*3*, Tn*1824*, Tn*1527*, and Tn*552*. Tn*7* transposition (in *E. coli*) occurs in two alternative, conservative modes: site-specific integration at a single chromosomal site (*att*T7), and random insertion (Craig, 1989). The proteins tnsA, tnsB, and tnsC are essential to all transposition events. Site-specific integration involves four Tn-encoded proteins; tnsA, tnsB, tnsC, and tnsD, in which tnsD involves target recognition (*att*T7 site) (Chakrabarti *et al.,* 2004). In the random pathway of integration, tnsA, tnsB, tnsC, and tnsE promote low-frequency transposition to sites unrelated to the *att*Tn7 (Craig, 1989).

1.5.3. Translocation mechanism of retroviruses

Translocation of retroviruses is carried out by the integrase enzyme encoded in the *pol* gene of viral RNA. Integrase recognizes the LTR that is analogous to the inverted repeats in insertion sequences. *In vivo*, the LTR elements undergo amplification by transcription from an integrated DNA copy into RNA copies. A free double-stranded cDNA copy, designated for integration, is then synthesized by reverse transcription (Varmus and Brown, 1989). This generally carries 2-bp extensions at each end, compared with the integrated copy. Single-strand cleavage occurs at both ends to remove these two terminal bases. The viral DNA then transfers into a target site in the host chromosome to give simple insertions. Although cleavages are limited to a single strand at each end, integration is conservative.

1.5.4. Mechanism of V(D)J recombination mimics transposition

V(D)J site specific recombination is the process by which functional immunoglobulin and T cell receptor genes are assembled from multiple gene coding segments. Recombination is directed by a specific recombination signal sequence (RSS) situated adjacent to every coding segment. An RSS is composed of conserved heptamer and monomer motifs separated by a relatively nonconserved spacer region of either 12-bp or 23-bp (Hiom *et al.,* 1998). The V(D)J recombination reactions are catalyzed by lymphoid-specific factors RAG1 and RAG2. Cleavage of the RSS by the RAG proteins occurs in two steps similar to the cleavage of IRs or LTR by tnp or integrase, respectively. First, the DNA is nicked at the 5' end of a signal heptamer adjacent to the coding DNA. This nick is then converted to a double-strand DNA break

with a blunt 5'phosphorylated signal end and a hairpin coding end. The hairpin formed by the RAG proteins occurs in a single transesterification step, similar to the strand transfer reactions mediated by transposases and HIV integrase (Engelman *et al.,* 1991). Hybrid joints formed by joining a cleaved signal end to a hairpin coding end are also mediated by RAG proteins through nucleophilic attack of the 3'-OH at the end of the signal joint.

1.6. Transposases

Transposases (tnp) belong to the DDE (triad of acidic amino acids with a characteristic spacing) protein family referring to a conserved protein sequence motif. Members of this family include the integrase proteins of retroviruses and retrotransposons, the tnp of bacterial insertion sequences of the IS*3* family, the MuA protein of Mu bacteriophage, the TnsB protein of Tn7, the tnp encoded by IS*10*R in Tn*10* and IS*50*R in Tn*5*, and RAG proteins of the V(D)J recombinase (Polard and Chandler,1995; Swanson, 2001; Fugmann *et al.,* 2000). Transposases and integrases are multifunctional enzymes that can locate the ends of an element, bind them, and bring them together into a synaptic complex. They recognize the precise phosphodiester bonds to be cleaved, ensure the incorporation into a target DNA molecule at an appropriate stage in the process of transposition, and direct strand transfer. These functions involve specific and nonspecific DNA binding, catalysis, the capability to form oligomeric proteins and, in some cases, the capability to interact with accessory proteins (Haren *et al.,* 1999).

Examples of protein-protein interactions were reported with respect to ORF AB (which codes for a transposase) and ORF A (encodes a regulatory protein) in IS*911 via* a leucine zipper motif in the ORF A (Haren *et al.,* 2000); and the MuA transposase and the target DNA activator MuB ATPase. Mizuuchi and Mizuuchi (2001) reported that the efficiency of MuA (a tetramer) to form a transpososome which subsequently integrates into target DNA is stimulated by HU (a host encoded protein) and by MuB.

Most tnp proteins can be divided into 3 domains. The N-terminal tnp domain can directly bind to the DNA ends of their cognate transposon. These specific binding activities have been observed in transposases from several insertion sequences including members of IS*3* family (IS*1*, IS*30*, IS*903*, IS*50*, IS*911*), Tn*3*, Tn*552* and one of the two Tn*7* enzymes (TnsB), but not the other (TnsA). The N-terminal 113 amino acids of the tnp encoded in IS*50R* can bind to a specific sequence (IR) at the end of the outside end (OE) of the transposon (Braam and Reznikoff, 1998).

The second domain is the catalytic core that has a characteristic highly conserved triad of acidic amino acids known as the DD(35)E motif and several additional conserved residues, particularly, a K or R residue six or seven amino acids downstream. The triad was also detected in the MuA protein of Mu bacteriophage (Baker and Luo 1994) and in the TnsA and TnsB proteins of Tn7 (Sarnovsky *et al.,* 1996). However, spacing between the first and the second aspartic acid (D) residues, and the second aspartic acid and the glutamic acid (E) residues can vary from 51 to 90 amino acids and 33 to over 100 amino acids, respectively (Haren *et al.,* 1999). This acidic amino acid triad is intimately involved in catalysis and plays a crucial role in

coordinating divalent metal cations (in particular Mg^{2+}) implicated in assisting the various nucleophilic attack groups during the course of the reaction. In retroviruses, this motif interacts with the terminal base pairs of the element, presumably contributing to correct positioning of the transposon end in the active site (Jenkins *et al.,* 1997). The core domain also may contain nonspecific binding activities since the isolated core domains of MuA and the tnps of IS*10* and Tc*1/3* show nonspecific DNA binding properties (Kwon *et al.,* 1995; Vos *et al.,* 1993).

The third, C-terminal domain of tnp from IS*50R* of Tn*5* (amino acids 441-476) is essential for synaptic complex formation. Presumably, the C-terminal domain is required for tnp to tnp homodimerization in the synaptic complex formation. This domain seems to be able to act together with an active site on the same monomer or another monomer in the complex (Davies *et al.,* 2000).

Crystal structures of the Tn*5* synaptic complex (Davies *et al.,* 2000) revealed that the N-terminal, core catalytic and C-terminal domains consist of 70, approximately 300, and approximately 100 amino acid residues, respectively. The N-terminal domain is made up entirely of α-helices and turns, and functions primarily to bind DNA. The catalytic domain consists of an α/β/α motif which is roughly twice the size of the catalytic domains of the retroviral integrases. The catalytic domain of the Tn*5* transposase also contains its own DNA binding motif. The C-terminal domain is composed entirely of α-helices and turns. This domain forms protein-protein interactions with the C-terminal domain on the second tnp molecule (the synaptic complex is composed of a dimer of the tnp) and this complex interacts with DNA. In

fact, all three domains participate in DNA binding and play roles in forming the structure of the synaptic complex. In contrast, studies of the Mu transposase and the retroviral integrases have indicated that these proteins form synaptic complexes that include a tetramer of transposase/integrase proteins (Baker and Luo, 1994; Craigie and Mizuuchi, 1987).

1.7. Translational regulation

Programmed translational frameshift is a control mechanism that is adopted in all three domains of life including also several retroviruses, retrotransposons, transposons, and insertion sequences. Translational frameshift mechanisms act during elongation rather than at initiation and are determined by sequences within the mRNA that constitute shift sites. Signals may be heptameric oligomers such as X XXY YYZ, upstream ribosome binding site (X, Y and Z represent any nucleotides), and downstream stem loops and/or pseudoknots. Rare codon usage has been shown to induce pausing of ribosomes during translation (Baranov *et al.,* 2003; Hansen *et al.,* 2003; Farabaugh, 1996; Hatfield and Oroszlan, 1990). A rare histidine codon (CAC) and a heptanucleotide sequence (A AAA AAT) at the junction of two open reading frames (SSO11867 and SSO3036 ORFs) in the genome of the thermophilic Archaeon *Sulfolobus solfataricus* were suggested as stimulators in a -1 frameshift mechanism to translate the enzyme α-L-fucosidase (Cobucci-Ponzano *et al.,* 2003). This is the only instance of regulated frameshifting known to date in the Archaea.

The three basic steps of translation are (1) selection and insertion of cognate aminoacyl-tRNA into the ribosomal A-site, (2) peptidyl transfer, and (3) translocation

of the nascent peptide chain from the A-site to the P-site. During translation, the non-programmed translational error rate has been estimated at 10^{-5} per codon (Kurland, 1992). Two peptidyl-transferase inhibitors (anisomycin and sparsomycin) altered the efficiency of -1, but not of +1, ribosomal frameshifting (Dinman *et al.,* 1997). Non-programmed frameshifting can also occur at very low frequency in the absence of ribosomal pausing. This frequency can be increased by changes in the anti-codon vicinity and physiological conditions such as starvation for specific amino acids (Farabaugh, 1996). Lysine-starvation of *E. coli* cells induced -1 (C UUC AAG) and +1 (GCC AAG C) frameshifting through the so-called "hungry" lysine codon, and the mechanism is dependent on the context of the codon at the juxtaposition of the AAG codon. The first mechanism involves the quadruplet (CUUC) to the left of the hungry codon that interacts with peptidyl-tRNA when the hungry codon is in the aminoacyl (A) site (Kolor *et al.,* 1993). Unlike the -1 mechanism, the +1 frameshifts induced by a "hungry" lysine codon involves $tRNA^{Ser}$ with the AGC codon. In this context, the C in position 3 is necessary for shifting: replacement of this base by A or G abolished shifting, and replacement by T reduced it markedly (Lindsey and Gallant 1993; Weiss *et al.,* 1988).

In most cases, programmed translational frameshifting resulted in the read through of two overlapping open reading frames to produce a protein fusion. The dsRNA ScV-L-A virus of the yeast *Saccharomyces cerevisiae*, the human immunodeficiency virus-1 and the coronavirus IBV produced gag-pol poly proteins by means of -1 frameshift mechanisms (Lopinski *et al.,* 2000; Yelverton *et al.,* 1994;

Brierley *et al.,* 1992). The Ty1 elements utilized a +1 frameshift mechanism to produce reverse transcriptase (Clare *et al.,* 1988). Release factor 2 (RF2) is produced by a +1 frameshift through the internal UGA codon of the *prfB* gene in *E. coli* (Craigen and Caskey, 1986; Craigen *et al.,* 1985). IS*911* and IS*150* from *E. coli* produced transposases as fusion proteins (ORF AB and InAB) from ORF A and ORF B, and *InsA* and *InsB*, respectively by means of -1 frameshift mechanisms (Loot *et al.,* 2002; Haas and Rak, 2002). *Vice versa*, frameshift could also result in termination of translation of an mRNA. In a well defined function of the γ subunit of DNA polymerase III, a truncated form of the τ subunit encoded by *dnaX* in *E. coli*, is produced by means of a frameshift mechanism in which the ribosome encounters a stop codon in a -1 frame downstream (Tsuchihashi and Kornberg, 1990). IS*1221* from *Mycoplasma hyorhinis* utilizes a -1 frameshifting mechanism to produce a truncated product from the upstream coding domain of a *transposase* open reading frame. Presumably, low-level production of the full-length transposase was ensured (Zheng and McIntosh, 1995).

Many IS elements encode their tnp in a single large ORF spanning almost the entire element, others (such as IS*1,* IS*150* and IS*911* a member of the IS*3* family) exhibit two consecutive overlapping ORFs placed in different reading frames. Slippage of the ribosome between the ORFs can lead to synthesis of the upstream ORF fused to the downstream ORF (Chandler and Fayet, 1993). The IS*1355* that forms a dichloromethane transposon contains a short ORF153 in 0 phase and ORF179 in -1 phase. The region where ORF153 and ORF179 overlap has a potential ribosomal

frameshift signal sequence (5'-AAAAAGGGGG-3') similar to the sequence detected in IS*1032* from *Acetobacter xylinum* (Schmid-Appert *et al.,* 1997).

IS*1* is 768-bp in length, carrying terminal inverted repeats of 23 bp and exhibiting two consecutive ORFs, *insA* and *insB'* in the 0 and -1 reading frames, respectively. Transcription normally occurs from a promoter located partially within the left inverted repeat. InsA protein and a fusion protein InsAB' (but no product from *ORF B'*) have been detected in a phage T7 expression system (Escoubas *et al.,* 1994). Two overlapping heptanucleotides (U UUA AAA, and A AAA AAC) are located at the 3' end of *InsA*. The frequency of frameshifting appears to be low (approximately 1%). InsA is not only a repressor of transcription but also an inhibitor of transposition by the transposase (InsAB'). InsA binds to both terminal IRs and protects 25 bp at each end (Zerbib *et al.,* 1990). Since the InsAB' shares the InsA domain at its N-terminal, the InsAB' can also bind to both ends of the IS.

Although IS*1* and IS*911* have superficially similar organization, the ORFs encoded on IS*1* have different functions from those on IS*911*. The ORF A encoded on IS*911* does not act as an inhibitor of transposition, as does InsA, but rather directs activity of ORF AB' into intermolecular transposition at the expense of intra-molecular circle formation. In ORF A of IS*911*, a slippery codon (A AAA AAG) allows the expression of 3 products (ORF A, ORF B, and ORF AB'). The product of ORF B uses an AUU initiation codon and ORF AB' is translated from a -1 frameshift to fuse ORF A and ORF B with a frequency about 15%. The ORF AB' itself can induce excision,

circularization, and intermolecular transposition of the IS*911* element (Ton-Hoang *et al.,* 1998).

Most retroviruses and retrotransposons encode group-specific antigens (capsid proteins) on the *gag* genes and reverse transcriptases and integrases (enzymes essential for viral integration) or the *pol* genes. Both genes are in different reading frames. Since products of the *pol* gene contain no packaging signal, to ensure incorporation of the products from both genes into viral capsids, approximately 2 - 25% of viral fusion proteins are translated due to a programmed translational frameshift mechanism. The gag-pol fusion protein benefits the virus in that Pol proteins share packaging signals with Gag proteins, thus, an appropriate amount of the Pol proteins can be incorporated so that the viral maturation process can be continued in the viral capsids. In most cases, this mechanism involves nucleotide sequences that allow backward shifting of the ribosome, so-called "slippery" codons or heptanucleotides of the type X XXY YYZ (Farabaugh and Bjork, 1999). However, such slippery sequences, though necessary, are not sufficient for frameshifting. A second feature that is believed to cause the ribosome to pause (for sufficient length of time to allow the frameshift to occur) is required (Farabaugh, 1996). In most described cases of IS, translational frameshift mechanisms appear to follow the retroviral gag-pol paradigm. Although this type of codon pattern is insufficient to allow -1 frameshifting, the sequence of the heptanucleotide itself can have a profound effect on frameshift frequency. One of the most efficient motifs is the sequence A AAA AAG found in several retroviruses and insertion sequences as well as in the *dnaX* gene of *E. coli* (Tsuchihashi and Brown, 1992). A single base change (in

the mouse mammary tumor virus (MMTV) gag-pol shift site from the normal A AAA AAC to A AAA AAG) dramatically increases -1 frameshifting from approximately 2 to 50% and a 10-fold decrease between A AAA AAG and A AAA AAA (Atkins *et al.,* 1990). Strings of only four identical nucleotides are sufficient to promote -1 frameshifting at low levels especially when located just before a stop codon (Weiss *et al.,* 1990). The frequency of the frameshift event may also be modulated by several other factors. For example, potential secondary structures in the mRNA located downstream of a slippery codon can stimulate frameshift. These include stem-loops and pseudoknots. The presence of an upstream ribosome-binding site also affects the efficiency of -1 frameshifting. These components presumably cause the ribosome to pause transiently, thus, increasing the probability of slippage. Other types of sequences that decrease the speed of ribosomal translocation, such as binding sites for specific proteins and codons corresponding to rare tRNA in the upstream reading frame, are also capable of stimulating frameshift (Chandler and Fayet, 1993). Mutations in tRNAs can induce high-frequency frameshifting in both the -1 and +1 directions (Atkins *et al.,* 1991).

For +1 frameshifting in a yeast Ty family of retroviral-like transposon, the shift site (CUU AGG CCA GGA AC) was identified in the 0 frame codon CUU that was stimulated by the adjacent rare arginine codon (Belcourt and Farabaugh, 1990). When the level of the minor $tRNA^{Arg}$ increased, the level of frameshifting decreased (Xu and Boeke, 1990).

1.8. Hypotheses and objectives

To date, few genetic techniques have been developed for *Pyrococcus* and other Thermococcales. The presence of quite high copy numbers of insertion sequence (IS) transition to the first hypothesis that "*IS elements in Pyrococcus species are actively mobile*." Since the IS elements are relatively simple and apparently lack a transcriptional control cassette to maintain sublethal limits of transposition frequency, this transition to the second hypothesis that "*programmed translational frameshift control mechanisms in the transposase gene down regulate expression of the IS elements*."

Analyses of direct repeat sequences may allow us to identify preferred target sites for transposition and predict the specificity of gene transfer. Finally, studies of a programmed translational frameshift mechanism in the *tnp* might allow us to increase the yield of recombinant transposase and to discover novel mechanisms of translational regulation by studying transposition *in vitro*.

Sequence analyses and comparative studies of the IS elements in the closely related species, *Pyrococcus furiosus* and *Pyrococcus woesei* may permit development of genetic tools for gene knockouts. Studies of the transposase from hyperthermophilic IS elements might allow development of a high temperature *in vitro* transposition system for the species *Pyrococcus abyssi* and *Pyrococcus horikoshii* which lack IS elements.

Chapter 2

Distribution of IS elements and expression of recombinant transposase genes

2.1. Introduction

2.1.1. History and genome sequences of four *Pyrococcus* species

To date, the genome sequences of three species of Pyrococcus are available in databases. The genome of *Pyrococcus furiosus* (GenBank Accession Number NC_003413) contains 23 copies of homologous insertion sequences (IS) that can be grouped into 3 types, namely: IS-I, IS-II, and IS-III. Interestingly, the genomes of *Pyrococcus horikoshii* OT3 (GenBank Accession Number NC_000961) and *Pyrococcus abyssi* (GenBank Accession Number NC_000868) contain no intact IS element (DiRuggiero *et al.,* 2000). *P. furiosus* (Fiala and Stetter, 1986) and *P. woesei* (Zillig *et al.,* 1987) were isolated from geothermal sediments collected from the Vulcano Island site in the Mediterranean Sea. *P. horikoshii* (González *et al.,* 1998) and *P. abyssi* (Erauso *et al.,* 1993) were isolated from deep-sea hydrothermal vents at Okinawa Trough in the North East Pacific Ocean (at depth ca. 1400 m), and North Fiji Basin in the South West Pacific Ocean (at depth ca. 2000 m), respectively. Little is known about the genome of *P. woesei* and the presence of the IS elements in the genome of *P. woesei* was examined in this study.

2.1.2. Archaeal promoters

Members of the genus *Pyrococcus* are hyperthermophilic archaea. Archaea are prokaryotic microorganisms that are phylogenetically distinct from eubacteria and

eukaryotes (Woese and Olsen 1986; Woese, 2002). The basal transcription machinery of Archaea is similar to the core components of the eukaryal RNA polymerase II (Bell and Jackson, 1998) apparatus rather than that of eubacteria. Archaeal promoters contain a TATA box "box A" [consensus TTTA(A or T)A] found 25-30 bp upstream of the transcription start site (Baumann *et al.,* 1995) and a "box B" motif [consensus (A or T)TG(A or C)] containing the transcription start site (Bell *et al.,* 1999). The TATA box is the area on the promoter where the formation of a ternary complex (TATA box-TATA binding protein-transcription factor B) occurs. The complex is required for the recruitment of the RNA polymerase to the transcription initiation site ("box B"), downstream from the TATA box. It appears that sequences flanking the TATA box define transcription polarity and the strength of the promoter (Bell *et al.,* 1999; Qureshi and Jackson, 1998).

2.1.3. Specific objectives

In this chapter, analyses of nucleotide sequences and structures, identification of deletion mutants, detection of the IS elements and tnp transcripts, and expression of recombinant transposases are addressed and discussed. The goals of these studies are to understand the IS elements and to express and purify a recombinant transposase.

2.2. Materials and methods

2.2.1. Organisms, cultivation, and DNA extraction

P. furiosus DSM 3638 and *P. woesei* DSM 3773 were grown under anaerobic conditions at 95 ^{o}C in Pf medium containing (per liter): 24 g NaCl, 4 g Na_2SO_4, 0.7 g KCl, 0.2 g $NaHCO_3$, 0.1 g KBr, 0.03 g H_3BO_3, 10.8 g $MgCl_2 \cdot 6H_2O$, 1.5 g $CaCl_2 \cdot 2H_2O$,

0.025 g $SrCl_2 \cdot 6H_2O$, 5 g tryptone, 1g yeast extract, 1 ml resazurin (0.2 g/l solution), 5 - 10 g sulfur powder and the pH was adjusted to 6.8 (González *et al.,* 1998). Cells were harvested by centrifugation and DNA was extracted as previously described (Charbonnier and Forterre, 1995). DNA pellets were washed 3 times with 70% ethanol at room temperature and dissolved in TE buffer (pH8.0). RNA was removed by RNase treatment (Ausubel *et al.,* 1987). Nucleic acid concentration was determined spectrophotometrically. Genomic DNAs of *P. horikoshii* and *P. abyssi* were gifts from Dr. González (IRNAS-CSIC, Sevilla, Spain).

2.2.2. Sequence analysis

Nucleotide sequences were identified using BLAST (Altschul *et al.,* 1990). Types of IS (I, II and III) were categorized based on the nucleotide sequence of inverted repeats (16 bp-long), spacing sequences (54 to 57 bp-long), and *tnp* (699 bp-long). Type D referred to an unclassified type that includes four truncated copies with multiple deletions. Nucleotide number 1 was assigned to the first base of the *tnps* to keep the number of nucleotides in the three major types equal in length (see Figure 2a). Inserts were sequenced at least two times on both strands and sequences were aligned using AssemblyLIGN™ (Oxford Molecular Group PLC, Oxford Science Park Medawar Centre Oxford, UK). Ambiguous nucleotides obtained from sequencing data were resequenced.

Analysis of frequency of direct repeat sequences observed in certain IS elements was performed using a Test goodness of fit (χ^2 df=1) (Sokal and Rohlf, 1996).

2.2.3. Primers and PCR conditions

All primers were designed based on sequences obtained from the genomic sequence of *P. furiosus* (http://www.genome.utah.edu/). Nucleotide numbers 1 and 702 are assigned to As of the initiation and termination codons of the putative transposase gene (*tnp*) respectively. Nucleotide sequences of primers specific for the *tnp* are 5'-ATG AAG GCT GAG AGC ATT CTA TAC TC-3', 5'-ATG AAG TCT GAA ACC ATT ATT TAC TGG-3', and 5'-TTA GGA TAA CTG GGG CAT CAC-3' (named primers 1, 2 and 3, respectively). These primers bind specifically to the *tnp* at nucleotide numbers 1 - 26, 1 - 27, and 682 - 702, respectively. Primer pairs 1-3 (primer type I) and 2-3 (primer type II) were employed to amplify *tnp-I* and *tnp-II*, respectively.

Polymerase chain reactions (PCRs) were performed in a 50 µl reaction volume either using PLATINUM® Taq polymerase High fidelity (Life Technologies, Inc., Gaithersburg, MD) or Cloned Pfu DNA Polymerase (STRATAGENE®, La Jolla, CA). Generally, a reaction contained 100 - 150 ng of genomic DNA, 1.3 mmoles of $MgCl_2$, 10 nmoles of dNTP, 20 pmoles of a forward primer, 20 pmoles of a reverse primer and 2.5 units of heat-stable DNA polymerase. Thermal cycling conditions for the primers type I and type II were 95 ^{o}C for 30s, 55 ^{o}C for 30s, and 72 ^{o}C for 60s for 30 cycles. For other primers, thermal cycling conditions varied: 95 ^{o}C for 30s, 52 ^{o}C for 40s, and 72 ^{o}C for 60 - 120s for 30 cycles depending on sizes of the products (approximately 60s for a fragment size of 1 kb).

2.2.4. Reverse transcription PCR

Total RNAs were extracted from cultures of *P. furiosus* growing at 95 ^{o}C for 5 and 16 hours in Pf medium using the S.N.A.P.™ Total RNA Isolation kit (Invitrogen®, San Diego, CA). The RNA extracts were digested with RNAse-free DNAase and isolated using mini-columns as recommended by the manufacturer. Total RNAs obtained were tested for contamination by genomic DNA using PCR. Negative results were obtained in control experiments with extracts used in reverse transcription PCRs. Reverse transcription PCRs were performed using Advantage® One-Step RT-PCR kit (CLONTECH Laboratories, Inc., Palo Alto, CA) and the primers type I and type II mentioned above in Section 2.2.3.

2.2.5. Cloning and sequencing

PCR products were analyzed by agarose gel electrophoresis. Gel purified DNA fragments were ligated into either a TA cloning vector or a Zero Blunt PCR cloning vector (Invitrogen®, Carlsbad, CA) depending on the DNA polymerase used during PCR amplification. Ligated products were transformed into *E. coli* Top10. Plasmid extraction and DNA clean up from agarose gels were performed using QIAprep® Spin Miniprep kit and QIAquick Gel Extraction kit (QIAGEN, Valencia, CA), respectively. DNA sequencing was performed at the Center of Marine Biotechnology, University of Maryland, Baltimore, MD. Sequencing of the *napA* region in *P. woesei* was performed using chromosome walk techniques starting from an IS-III clone.

2.3. Results

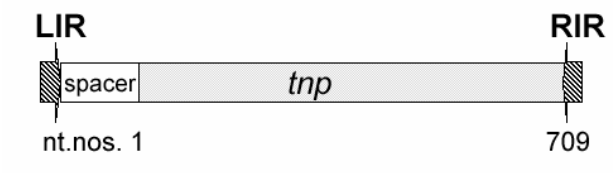
2.3.1. Classes and distrinbution of IS elements in *P. furiosus*

The 23 copies of IS elements in the genome of *P. furiosus* were assigned into 3 types according to their nucleotide identities (83 – 84% identity among each other) and named IS-I (11 copies), IS-II (8 copies), and IS-III (4 copies). Typically, the IS elements are rather conserved and simple, containing a short spacer (54-57 nt. long) and one transposase open reading frame (ORF) flanked by left and right inverted repeat (IR) sequences (Figure 2a). Nucleotide sequences of the left IR (LIR), right IR (RIR) and spacers are shown in Figures 1b and 1c.

The eleven copies of IS-I (size 782 bp) and eight copies of the IS-II (size 781 bp) are identical between types except for element number 17 in which a single base deletion is observed in the *tnp* sequence, and elements 10 and 6 in which a single base transition occurs in the LIR type II, and 8 variant bases located in between nucleotides 522 and 562 in the *tnp-II*, respectively. The four copies of IS-III (size 779 bp) are 99.9% similar in nucleotide sequence. In order to simplify subsequent discussions, element numbers are assigned and coordinates of the IS elements in the genome of *P. furiosus* are listed in Table 1.

Figure 2. Structure of the IS elements and alignments of their nucleotide sequences.
(a) Diagram of a typical IS element in the genome of *P. furiosus*. The striped block arrows represent left and right inverted repeats (LIR and RIR, respectively). The white box represents a spacer region (size 54-57 bp). Filled box represents transposase open reading frame (size 702 bp). Nucleotide number 1 is assigned to A of the initiation codon. Nucleotide number 709 is the last base of the RIR.
(b) Alignment of LIR and RIR sequences. Gray letters indicate differences. Copies indicate copy numbers observed in the genome. Blank and filled boxes represent *transposase* and spacer region, respectively. The sizes (bp) of tnps are indicated in parentheses. The type D assigned to element numbers 24-27 contains deletion tnp open reading frames.
(c) Alignment of spacer regions. Gray letters indicate differences. Dashes indicate missing bases. Copies indicate copy numbers observed in the genome. Negative nucleotide numbers indicate downstream nucleotides relative to the nucleotide number 1 (A of the initiation codon) defined in this study. A putative TATA box "boxA" is boxed.

(a).



(b).

Types:	LIRs			RIRs	Copies
I:	GATACTGTTAGGATAA-	Spacer	Transposase (699 bp)	TTATCCTGACAGTATC	11
II:	GATACTGTTAGGATAA-	Spacer	Transposase (699 bp)	TTATCCTAACAGTATC	7
II:	GATACTATTAGGATAA-	Spacer	Transposase (699 bp)	TTATCCTAACAGTATC	1
III:	GACACTGTCAGGATAA-	Spacer	Transposase (699 bp)	TTATCCTGACAGTATC	4
24D:	GATACTGTCAGGATAA-	Spacer	Transposase (500 bp)	TTATCCCGACAGTATG	1
25D:	GATACTGTTAGGATAA-	Spacer	Transposase (622 bp)	CTATCCTAACAGTATC	1
26D:	GATACTGTTTGGATAG-		Transposase (585 bp)	Not found	1
27D:	CAAAGTGTATAGAACAA	Spacer	Transposase (228 bp)	Not found	1

(c).

Types:	-57 -50	-40	-30 -20 -10 -1	Copies
I:	GCGGTAGGGCGTTAGG	TTTAAGT	TTCTGATAATTTTTCAGAAAAAGAAGGGAGAGAA	11
II:	GCTGTGGGGTGTCAGG	TTTAAGT	TACTGGCAGTACTTTAGAAAAAGAAGGGGGA-AC	8
III:	GCAGTGGGACGTTAGG	TTTAAGT	TTCTGATACCTCTTCAGAAAA-G--GGGAGAGAA	4
24D:	AGAGTGAGATGTTAGG	TTTAAGT	TTCTGATAATT-TTCAGAAAAAGA-G--A-A-AC	1
25D:	GCTGTGGGGTATCAGG	TTTAAAT	CCCAGACACTACCTCAGAAAAAGAAGG-AG-GAA	1
26D:			no detectable homology	1
27D:	GTAGTGAGGTAACAGG	TTTAAGT	TAGTGACAATACTTCAGAAAAAGAAGGGAGA-AC	1

Table 1. Types, element numbers and coordinates of IS elements in the genome of *P. furiosus*. Types are categorized by sequence identity. Element numbers are assigned arbitrarily. Element numbers 1-10 and 24 are located on the plus strand, and 11-23 and 25-27 are located on the minus strands relative to the LIR.

Types	Element numbers	Genome Coordinates		Types	Element numbers	Genome coordinates	
		5’	3’			5’	3’
I	11	72986	73767	II	7	909866	910646
	1	284983	285764		8	1161577	1162357
	12	361521	362302		23	1284310	1285090
	2	397627	398408		9	1612359	1613139
	13	463555	464336		10	1629433	1630213
	3	752270	753051	III	18	552546	553324
	4	767495	768276		5	771322	772100
	14	797130	797911		19	870640	871418
	17	1142040	1142820		20	977655	978433
	15	1720267	1721048	D			
	16	1877741	1878522		24	250231	250804
II	21	12566	13346		26	253075	253675
	22	414414	416194		25	437640	438339
	6	579456	580236		27	1443764	1444063

2.3.2. Non-autonomous miniature inverted repeat element (MITE)-like elements (Type D)

Blast searches of entire IS elements against the complete *P. furiosus* genome revealed an additional four truncated copies of the IS elements (approximately, 300-700 bp) which are classified as type D (element numbers 24-27). Coordinates are listed in Table 1. Their sequences show high identity (60-85%) to several localized patches of the complete element; these include spacers (Figure 2c), fragmented *transposase* (*tnp*) sequences, and at least one of the IR sequences (Figure 2b). Of the four deletion mutants identified, the spacer regions and defective transposase ORFs of element numbers 24D and 25D are flanked by two IR sequences at both ends such that their structures are similar to the non-autonomous miniature inverted repeat element (MITE)-like elements (Brugger *et al.,* 2002). Element number 26D is composed of an LIR and deletion tnp ORF (585 bp), but it lacks the spacer and the RIR sequences. Like element number 26D, element number 27D is composed of a deletion tnp ORF (228 bp) and a sequence similar to the LIR, but no RIR sequence was observed. Unlike element number 26D, element number 27D also contains a conserved spacer region between the LIR and tnp ORF.

2.3.3. Mixed types of inverted repeats

Inverted repeat (IR) sequences forming the presumed recognition site for transposases are shown in Figure 2b and can be grouped into three major types. An LIR containing a nucleotide transition (gray letter in LIRs line 3, Figure 2b) is observed in element 10. The LIR and RIR pairs are nearly perfect matches (in element

types I and III) and perfect matches (in element type II) inverted repeat sequences (Figure 2b).

Among the three major types, the IS elements are flanked by combinations of the LIR and RIR pairs. The IS elements type I (11 copies), type II (7 out of 8 copies), and type III (4 copies) are flanked by LIR type I (5'-GATACTGT TAGGATAA-3') and RIR type I (5'-TTATCCTG ACAGTATC-3'), by LIR type I and RIR type II (5'-TTAT CCTA ACAGTATC-3'), and by LIR type III (5'-GACACTGT CAGGATAA-3') and RIR type I, respectively.

This analysis also implies that the IS elements may be targets for cross transposition by the conserved transposase of all three classes of the IS elements.

2.3.4. Conserved spacer regions

IS types I, II and III contain a homologous sequence (57, 56 and 54 bp, respectively) spanning the LIR and the tnp ORF, named the spacer region. A+T contents of the spacers are 61, 57, and 56%, for types I, II, and III, respectively. Alignments of the spacer regions are shown in Figure 2c. In Figure 2c, negative numbers of nucleotide are assigned relative to the nucleotide number 1 (A of the initiation codon) mentioned above. A putative TATA box (25 copies of 5'-TTTAAGTT-3' and one copy of 5'-TTTAAAAT-3') centered at nucleotide number -38 is observed to be conserved in this region adjacent to the LIR end. Functions of the spacers are not known. They could serve as buffer regions or as weak promoters for the tnp ORFs downstream.

Table 2. Nucleotide sequences of direct repeats in the genomes of *P. furiosus* and *P. woesei.* Pentanucleotide sequences of a consensus 5'-CARRR-3' are bold. Element numbers are defined as in Table 1. R represents purine nucleotides. Strand indicates orientation of the IS in the genome of *P. furiosus*. Sequences of the direct repeats shown represent the sequences on the + strand adjacent to the LIR. A pair of nearly perfect inverted repeat sequences observed flanking element numbers 5 and 18 are underlined.

Element numbers	Strand	Direct repeat sequences (5'→3')
[a]IS-III	+	C**CAAGG**AT
1	+	TAG**CAAGG**
[b]9&10	+	**CAGGA**GGA
13	-	TC**CAGAA**C
16	-	C**CAGAA**TC
22	-	**CAAAA**GCA
Consensus:		5'-$N_{(0--3)}$**CARRR**$N_{(0-3)}$-3'
Element numbers.	Strand	Direct repeat sequences (5'→3')
14	-	AATGCGGG
11	-	ATGTGGGA
15	-	ATATTATG
6	+	GTGGAAGC
19	-	TTGATCGAC
20	-	GCTTGCCCT
5	+	<u>TCT</u>T<u>CCGAG</u>
18	-	<u>CTCGG</u>G<u>AGA</u>
21	-	TCCTCAAC
7	+	CTTTTTCA
25D	-	[c]AAATTTTTG(T orA)

[a] An IS-III detected in the *napA* locus in *P. woesei*.
[b] A composite transposon.
[c] AAATTTTTGT and AAATTTTTGA observed flanking the element 25D.

2.3.5. Direct repeats and a target preference consensus

Of the 27 elements, 16 flanked by 8-9 bp-long direct repeat sequences are observed in element numbers 1, 5-7, 9 and 10 (a composite Tn), 11, 13-16, 18-22, and 25D. The presence of direct repeats, hallmarks of transposition events, flanking about 60% of the elements implies that these elements were inserted into the genome of *P. furiosus* by transposition.

Nucleotide sequences of the 17 direct repeat sequences including a sequence from *P. woesei* (identified in Chapter 3) are listed in Table 2. It is observed that 6 repeat sequences {elements numbers 1, 13, 16, 22, a Tn (nos. 9 and 10) and the sequence flanking the IS-III in *P. woesei*}contain pentanucleotides with a consensus 5'-$N_{(0--3)}$ CARRR$N_{(0-3)}$-3' (bold letters in Table 2) regardless of spacing from the LIR. A goodness of fit analysis indicates that the observed frequency is much greater than the expected frequency (χ^2 significant at p-value <0.0001).

The results imply that this is a target preferred for transposition.

2.3.6. Transposases

Open reading frames encoding transposases begin right after the 3'end of the spacers. They are identical in length (699 bp) and nucleotide sequence from each type with a stop codon located in the center of the RIR. Alignments of the nucleotide sequences are shown in Figure 3. Predicted Tnp protein (233 amino acids) sizes are 28.7-28.8 KDa with high isoelectric points, around pIs 11.1-11.2. The nucleotide sequences are rich in lysine and arginine codons (bold letters in the Figure 3).

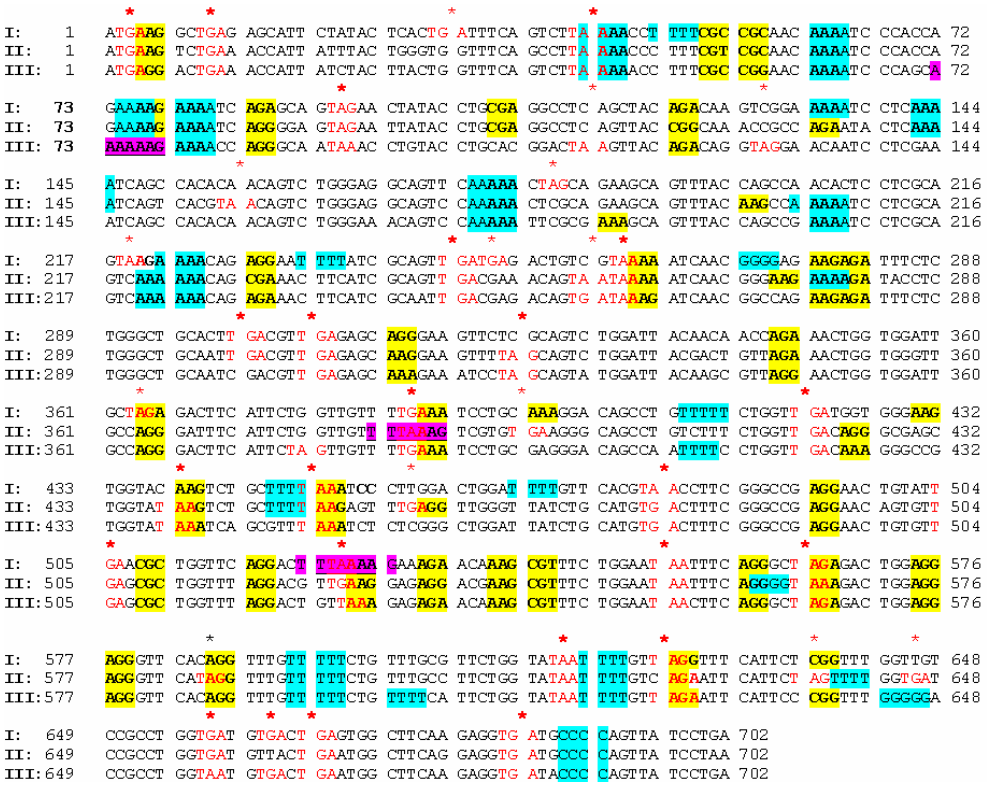


Figure 3. Alignment of nucleotide sequences of *tnps-I*, *-II* and *-III*. Bolded and yellow highlighted letters indicate lysine and arginine codons. Pink highlighted letters represent putative heptanucleotide frameshift sequences. Blue highlighted letters indicate runs of 4-5 nucleotides. Red asterisks and red letters represent stop codons in +1 and -1 phases. Roman numbers represent the IS type. Numbers indicate nucleotide numbers. Nucleotide number 1 is the A of the tnp initiation codon.

The nucleotide sequences contain many small open reading frames observed in +1 and -1 phases. The ends of the small ORFs are indicated by asterisks in Figure 3. Each type of sequence contains a putative -1 frameshift site (pink highlighted A AAA AAG, T TTA AAG and T TTA AAA in types III, II and I sequences in Figure 3) followed by stop codon in + and -1 phases located in short distances. In addition, the sequences contain many strings of 4-5 nucleotides (blue letters in Figure 3) that might create conditions favorable for frameshifting during translation. For instance, strings of 4As and 5As, and 5Ts are located at nucleotide numbers 42-45 and 175-179, and 408-412 and 593-597, respectively.

2.3.7. Detection of IS elements in the genus *Pyrococcus*

A PCR based technique used primer pairs (primers 1-3 and 2-3) that bind specifically to the proximal (primers 1 and 2) and distal portions of the *tnp* sequences (see also nucleotide sequences and PCR conditions in Section 2.2: Materials and Methods). Results of PCR amplifications of genomic DNAs isolated from *P. furiosus*, *P. woesei, P. horikoshii* and *P. abyssi* are shown in Figure 4.

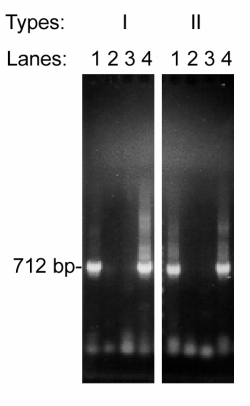


Figure 4. PCR products amplified from genomic DNAs of four species of *Pyrococcus*. Lanes 1-4 show products amplified from *P. furiosus*, *P. horikoshii*, *P. abyssi*, and *P. woesei*, respectively. Types I and II represent the primer pairs 1 and 3, and 2 and 3, respectively. Expected product size of 712 bp is indicated.

The results confirm the presence (lanes 1 and 4), and absence (lanes 2 and 3) of the IS elements in the genomes of *P. furiosus* (lane 1), and *P. horikoshii* (lanes 2) and *P. abyssi* (lane 3). In lane 4, the results reveal that the genome of *P. woesei* contains both IS types I and II. Cloning and sequencing of the PCR products obtained using specific primers (types I and II) confirm specificity of the primers, since all 22 clones contain sequences of the types I and II corresponding to the primer pairs employed. Moreover, when the type II primers (pair 2-3) were employed to amplify a plasmid harboring either a *tnp-I* or *tnp-II* sequence, no bands were detected on the type I template (data not shown). The results clearly confirm specificity of the type II primer under the standard PCR conditions.

Cloning and sequencing of the specific bands obtained from the template from *P. woesei* reveal that the nucleotide sequences of the *tnp-I* (GenBank Accession No. AF443788) are identical to the sequences reported in the genome of *P. furiosus*. Unexpectedly, 67% (4 out of 6) of the sequences from *tnp-II* amplified from *P. woesei'*s DNA are variant sequences (GenBank Accession No. AF420277) and the remaining 33% (2 out of 6) are identical to the prototype *tnp-II*. The variant nucleotides in the *tnp-II* obtained from *P. woesei* are ^{358}A, ^{369}C, and ^{378}A relative to the ^{1}A in the initiation codon. The ^{358}A is the first base of codon AUU (isoleucine) instead of codon GUU (valine).

2.3.8. Reverse transcription PCR

In order to detect transposase transcripts, *P. furiosus* cells were grown at 95 ^{o}C in Pf medium to exponential (5 hours) and stationary (16 hours) phases. Total RNA extracts were tested using PCR and the specific primers for type I and type II to ensure no DNA contamination (results not shown). Results obtained after RT-PCR amplifications using the specific primers for type I and type II are shown in Figure 5. Using the primer set for type I, tnp type I transcripts were detected in both stationary and exponential phases (band size ca. 700 bp in lanes 1 and 2). Unlike the reactions amplified using the primer set for type I, the anticipated RT-PCR product (ca. 700 bp) amplified using primer set for type II was observed in the 16 hour sample (lane 3), but not detected in the 5 hour sample (lane 4). These results indicate the tnp transcripts may be actively expressed from some elements in both exponential and stationary phases. This finding supports the putative TATA box "boxA" observed in the spacer region mentioned above in Section 2.3.4. However, the results do not exclude the possibility of transcription from outside the elements and can not be assigned to particular elements with in groups I, II and III.

(a). (b).

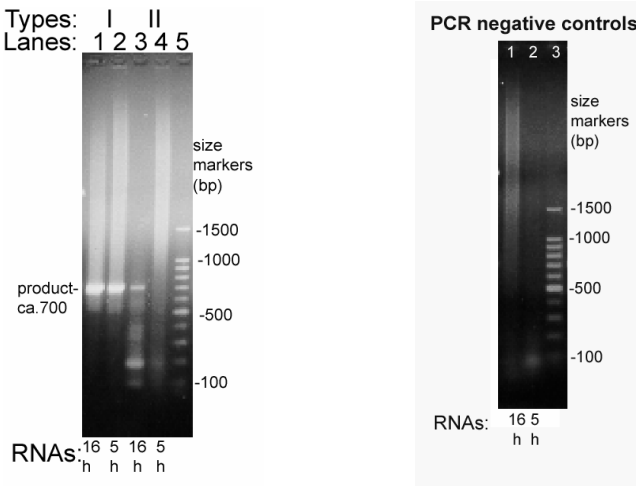


Figure 5. Detection of transposase transcripts in total RNA extracts from *P. furiosus*. (a) RT-PCR products amplified using specific primer types I and II. Lanes 1 and 3, and 2 and 4 represent lanes loaded with reactions on RNA extracts from cultures harvested at 16 and 5 hours, respectively. Lane 5 represents size markers. Types I and II specific primers employed on the reactions are indicated on top. (b). PCR negative controls on RNA extracts employed in the RT-PCR reactions mentioned in Figure 5a. Lanes 1 and 2 represent RNA extracts from cultures harvested at 16 and 5 hours, respectively. Lane 3 represents size marker. All samples show no product amplified using PCR and type I and II specific primers, indicating there was no DNA contamination in RNA samples.

2.3.9. Expression of recombinant transposases

The IS elements contain a single long open reading frame encoding a transposase (233 amino acids). It is anticipated that the enzyme from hyperthermophilic IS element is more stable than enzymes from transposons in mesophiles. The tnps-I and -II were cloned in two expression systems using pCR®T7/CT TOPO (under T7 promoter) and pBAD TOPO (under *ara*BAD promoter) vectors with and without histidine tags at the C-terminal of the protein sequences. Plasmids named pTnO11 (contains *tnp-I*) and pTnB1 (contains *tnp-II*) are derivatives of the pCR®T7/CT TOPO vector and plasmids named pOH1 (contain *tnp-I* in-frame fusion to the 6 histidine sequence) is a derivative of the pBADTOPO vector. These plasmids (pTnO11, pTnB1 and pOH1) were expressed for recombinant tnp proteins. Low production of the recombinant proteins was obtained (data not shown).

Plasmid pSJS1240 (Kim *et al.,* 1998) contains tRNA genes coding for $tRNA^{AGG}$, $tRNA^{AGA}$ and $tRNA^{AUA}$ cognates for rare arginine (AGG and AGA) and isoleucine (AUA) codons. Co-transformed strains containing pSJS1240 and the expression vectors mentioned above were tested to attempt to improve yield of transposase. The yields of the recombinant protein were not improved (data not shown).

2.4. Conclusions

The IS elements in the genomes of *Pyrococcus* are small and simple elements containing one open reading frame and a short spacer region flanked by 16 bp inverted repeats (IRs) (Figure 2a). IRs were observed among the three types of IS in *Pyrococcus* species implying that the transposases can recognize these IRs among the IS elements. A high percentage (ca. 60%) of the IS elements were flanked by direct repeat sequences that were duplicated during transposition events. Motif conservation is evident among the direct repeat sequences, the analysis of frequency suggests that the IS elements might prefer to attack the sequence CA(AorG)(AorG)(AorG) in target DNAs. The nucleotide sequences of the spacers are highly conserved within and across types, and a putative TATA box "boxA" is observed (Figure 2c). Although the function of the spacers is unknown, they might serve as buffer regions (reduce damage from non specific recombination reactions and mutations) and/or as transcriptional regulatory sequences for the IS. The latter possibility is also supported by the presence of transposase transcripts in the total RNAs extracted from cultures of *P. furiosus* (Figure 5). However, the transcripts may also be transcribed from promoters outside of the elements. The element type D identified in this study suggests that the IS elements are also targets for homologous DNA recombination and suggests a mechanism for the generation of miniature inverted repeat elements (MITE-like elements).

A PCR based strategy was successfully developed to detect and differentiate between the IS elements. Two types of *tnp* sequences (GenBank Accession Nos. AF443788 and AF420277 for types I and II, respectively) were identified in the

genome of *P. woesei* (Figure 4). Sequencing experiments reveal a variant type II with 99.5% identity to the prototype type II sequence reported in the genome of *P. furiosus*. The percentage obtained is close to the percentage obtained from the analysis of other homologous genes between the two species (Table 3 in Chapter 3). *P. furiosus* and *P. woesei* were both isolated from the Mediterranean Sea. The absence of the IS elements (Figure 4) in the two species, *P. horikoshii* and *P. abyssi*, isolated from Pacific Ocean supports a hypothesis made in this study that geographic aspects of the continents might limit the distribution of the elements. As yet, the numbers of isolates tested are small.

The expression of recombinant transposases in *E. coli* proved to be a difficult task. Analysis of the *tnp* sequences reveals that many sites along the sequences are slippery (Figure 3) with at least one -1 frameshift heptamer located in each type of *tnps*. Experiments on a programmed translational frameshift site and mechanism will be addressed in Chapter 4.

In conclusion, significant copy numbers of potentially active IS elements in the genome of *Pyrococcus* may play significant roles in genome rearrangement (Maeder *et al.,* 1999; Lecompte *et al.,* 2001; Zivanovic *et al.,* 2002) including gene disruption, translocation, inversion and insertion observed in the *P. furiosus* genome relative to the genomes of *P. abyssi* and *P. horikoshii*.

Chapter 3

IS elements as phylogenetic markers

3.1. Introduction

Pyrococcus species are hyperthermophilic members of the Thermococcales with optimal growth temperatures near 100 ^{o}C. All of the known *Pyrococcus* species grow optimally, and relatively rapidly, at temperatures above 90 ^{o}C, and produce H_2S from S^o (Zillig *et al,* 1987). Two *Pyrococcus* species, *Pyrococcus furiosus* (Fiala and Stetter, 1986) and *Pyrococcus woesei* (Zillig *et al.,* 1987) were isolated from marine sediments in the Vulcano Island beach site in Italy. They show similar physiological characteristics; both species are cocci, and are motile by means of a tuft of polar flagella. They are heterotrophic and grow optimally at 95 - 100 ^{o}C, utilizing peptides as major carbon and nitrogen sources. Both species also utilize several carbohydrates such as maltose, cellobiose, and starch. Growth of *P. furiosus* on a wide range of β-glucans such as chitin, laminarin, cellobiose and cellulose has been reported (Gao *et al.,* 2003; Gueguen *et al.,* 1997; Kengen *et al.,* 1993; Bauer *et al.,* 1999; Blamey *et al.,* 1999). Neither species requires S^o for growth, but S^o appears to stimulate growth via detoxification of H_2, the end product of metabolism. In contrast to *P. furiosus*, *P. woesei* may continue to grow at higher levels of H_2, produced during cultivation (Mukund and Adams, 1991). *P. furiosus* requires tungsten (10-25 μM Na_2WO_4) for optimal growth in S^o free medium with maltose (Bryant and Adams, 1989). The optimal pH and NaCl concentration for growth of *P. furiosus* may be slightly higher

and cover a broader range than *P. woesei* (pH 6-8 and 15-35 g NaCl/l for *P. furiosus,* and pH 6 and 30 g/l NaCl/l for *P. woesei*, respectively) (Fiala and Stetter 1986; Zillig *et al.,* 1987) although this may also reflect lab-to-lab variation.

Three *Pyrococcus* genome sequences (*P. furiosus, P. horikoshii* and *P. abyssi*) are available in the literature (Robb *et al.,* 2001). In contrast, limited numbers of sequences accessions from *P. woesei* are available in GenBank. The result obtained from the studies described in Chapter 2 reveal that the genome of *P. woesei* also contains IS elements (GenBank Accession Numbers AF420277 and AF443788). Comparative studies of the genomic content and positions of the IS elements may therefore provide a measure of relatedness between these two species.

3.2. Materials and methods

3.2.1. Organisms and growth in maltose

P. furiosus DSM 3638 and *P. woesei* DSM 3773 were used in this study. Cells were grown under anaerobic conditions at 95 ^{o}C in Pf medium as described in Section 2.2.2. DNA was extracted as previously described in Section 2.2.2. Growth of *P. furiosus* and *P. woesei* in maltose was performed in S^{o}-free medium. The medium composition was described (see Pf medium) except that elemental sulfur was omitted, and maltose (1% w/v) and sodium tungstate (10-100 μM) were added. Cell densities were determined (from triplicate cultures) using the acridine orange staining technique and direct counting by epifluorescent microscopy. Briefly, samples (0.1-1.0 ml) were mixed with 37% formalin saturated with sodium borate (10-25 μl), 0.01% acridine orange (0.2 ml) and diluted with Milli-Q water to obtain 1 ml final volume. The

solutions were filtered through 0.22 μm polycarbonate membranes (OSMONICS INC, San Diego, CA) and the membranes were mounted on microscopic slides. Cell counting was performed randomly on at least 50 unit squares (1 unit square is 0.01 mm^2). Cell numbers were determined using the following equation:

Cell numbers/ml = Number of cells per 10 unit squares x 2300/volume in ml.

3.2.2. 16S rDNA of *P. woesei*

Sequencing of 16S rRNA-$tRNA^{Ala}$-23S rRNA gene cluster in *P. woesei* was performed using primers 344F (5' ACGG GGC GCA GCA GGC GCGA 3'), 345R (5' CCCT TCC GCC CCG AGC 3'), 346F (5' AACG GGT CCG ACG GTG 3'), 347R (5' CCGC GAT CCG AAC TGA 3'), 348F (5' CCTC TTG CAT ACA TAA CTT CTC 3'), and 349R (5' TTC CCG CAT TGC GGA 3') and on both strands. Ambiguous nucleotide calls obtained from sequencing data were re-sequenced.

3.2.3. Southern blots for comparison of restriction patterns

Genomic DNAs (5 μg) were digested with the restriction enzyme *Pst-1*. Approximately 1 μg of restricted DNA per lane was loaded and separated in 0.9% agarose gels (11 cm x 14 cm). Linearized plasmids constructed in this study (Section 2.3.9) containing *tnp* sequences were used as size markers. Electrophoresis was performed at 30 V in TAE buffer (pH 8.5) for 14 to 15 hours in a cold room. Transfer of DNAs to a nylon membrane (Immobilon™-Ny^+ Transfer Membrane) was performed according to the manufacturer's recommendation (Millipore, Bedford, MA) using the perfusion technique. Briefly, DNA in the agarose gel was depurinated by soaking in 0.25 N HCl for 15 minutes with occasional agitation. The agarose gel was rinsed

briefly with MQ-water, denatured in 0.5 N NaOH+1.5 M NaCl for 30 minutes, and neutralized with 1M Tris·Cl (pH8.0)+1.5M NaCl for 30 minutes. DNAs were transferred to the nylon membrane using 20x SSC for 6 to 8 hours using standard blotting technique. DNA probes were prepared using 0.1 ng of purified *tnp-I* and *tnp-II* sequences as templates. A PCR reaction amended with 25 μCi of ^{32}P-α-labeled dATP was used to produce the labeled probes. The probes were boiled for 10 minutes and immediately added to the pre-warmed hybridization mixture. Hybridization was performed, following standard procedures as previously described (Ausubel *et al.,* 1987). Images were visualized using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). Band intensity was measured using program NIH Image 1.61 (http://www.tsc.udel.edu/macsoftdist/image.html). Relative intensity was estimated using the equation: Index (per unit area) = (Band intensity - Average intensity of background)/(Intensity of a band containing one IS - Average intensity of background).

Nucleotide sequences containing IS were extracted from the *P. furiosus* genome sequence using GeneMate (http://comb5-156.umbi.umd.edu/gene-mate/). *Pst-I* restriction fragments containing an IS sequence were sized from the genome sequence of *P. furiosus*.

3.2.4. Lambda library and screening for insertion sequences

Approximately 5 - 6 kb *Pst-I* restriction fragments from genomic DNA of *P. woesei* were gel purified. The fragments (1 μg) were partially digested with the restriction enzyme *Tsp509-I* (0.1 units) at 56^{o} C for 15 minutes to create short fragments with AATT sticky ends. The restriction enzyme was removed using a spin

column (QIAquick Gel Extraction kit, QIAGEN, Valencia, CA). Approximately 50 ng of the *Tsp509-I* fragments were ligated to λZIPLOX™, *EcoR* I Arms vector (Invitrogen™, Gaithersburg, MD) using T4 DNA ligase (Promega, Madison, WI) and packaged (Gigapack® II Gold Packaging Extract STRATAGENE®, La Jolla, CA). Propagation of λ phage and excision of plasmid pZL1 were performed following the methods recommended by the manufacturer (Invitrogen™, Inc., Gaithersburg, MD). Briefly, plaques obtained on *E. coli* Y1090(ZL) were transferred to duplicated nylon membranes (Millipore, Bedford, MA). The membranes containing phages were air dried and treated with 0.2 N NaOH+1.5 M NaCl for 2 minutes, 0.2 M Tris·Cl (pH7.6)+2x SSC for 2 minutes and 2x SSC for 2 minutes, respectively. The treated membranes were air dried and baked under vacuum at 80 ^{o}C for 2 hours. Preparation of DNA probes, hybridization of the probes to the membranes and visualization of positive plaques were performed as described above in the Southern blot (Section 3.2.3). Positive plaques were picked, excised re-purified to create pZL1 plasmids in *E. coli* DH10B(ZIP) and at least 5 clones from each plaque were re-plated and screened using standard colony hybridization techniques following the methods recommended for a nylon membrane (Immobilon™-Ny^{+} Transfer Membrane, Millipore, Bedford, MA).

3.2.5. Primers and PCR conditions

Nucleotide sequences of primers 1, 2, and 3 (primers type I and type II) and conditions for PCR reactions were as described in Section 2.2.3. Primers 4, 5, 6, 7, and 8 were designed to bind specifically to the *napA* locus in the genome of *P. furiosus*

(binding coordinates 286645 - 286663, 286694 - 286716, 286776 - 286797, 286922 - 286947, and 285765 - 285786, respectively). Nucleotide sequences of the primers 4 - 8 are listed in the following: 5'-TCC ATG CTA GTC TTT TTGC-3', 5'-CGA GTG AAC ATC CGA CAA TCT TA-3', 5'-TTA AGG CAA CTT CCA TCC TTGG-3', 5'-AGA ATC ATT AGA ACC AGA AGA GTA GC-3' and 5'-TAG CAA GGG GAT GAT AAC ATGG-3', respectively. The nucleotide sequence of primer 9 (5'-CTT CAA GAG GTG ATA CCC CAG TTA TCC-3'), an outward and conserved primer sequence, was designed to bind specifically to the 3' end of the *tnp*s (binding coordinates 672 - 699). PCR reactions were performed in a 50 µl reaction volume. Thermal cycling conditions varied: 95 ^{o}C for 30s, 52 ^{o}C for 40s, and 72 ^{o}C for 60 - 120s for 30 cycles depending on the sizes of the products (approximately 60s for fragment size of 1 kb).

Nested PCR reactions were performed using dilute samples (1:600). Primer pair 4-7 was employed in negative control reactions.

3.2.6. Phylogenetic analysis of *P. woesei* nucleotide sequences

Nucleotide sequences of *P. woesei* were compared by BLAST against genomic sequences of *P. furiosus, P. horikoshii* OT3 and *P. abyssi* using the GeneMate site (http://comb5-156.umbi.umd.edu/genemate/). All available nucleotide sequences of *P. woesei* were obtained from GenBank. The Accession numbers are listed as follows: X15329, AY443493, AF443788, AF420277, AF043283, U84155, Y09481, AF240464, AF177906, AF239672, X59857, X73527, X60161, X70668, X67205, U56247, U10285, M83988, M83987, and AF461062.

3.3. Results

3.3.1. Growth in maltose and tungsten

P. woesei shows similar physiological properties to *P. furiosus*. Cell yields (ca. $1x10^7$ cells/ml) of both species from over night cultures in the Pf medium were not significantly different. Although both species grow in S^o-free medium containing maltose and 10-100 μM sodium tungstate (Na_2WO_4), cultures (5 replicates each) of *P. woesei* appear less turbid than identical cultures *P. furiosus* after overnight incubation. In order to compare growth of both species, one ml of each culture was inoculated into three identical bottles of 100 ml S^o-free medium containing 1% maltose and 100 μM sodium tungstate (Na_2WO_4) to obtain cell densities of about $7\text{-}8x10^4$ cells/ml. The cultures were incubated anaerobically at 95 ^oC. Growth curves in the S^o-free medium containing maltose and 100 μM Na_2WO_4 of *P. furiosus* and *P. woesei* are shown in Figure 6a. During the first 6 hours of incubation, growth of *P. furiosus* (solid diamonds) and *P. woesei* (blank diamonds) were identical (cell yields at 6 hours are $1.1x10^6 \pm 3.8x10^5$ and $1.1x10^6 \pm 2.3x10^5$ cells/ml, respectively) and they both entered stationary phase after 10 hours of incubation. However, the maximal cell density obtained from *P. woesei* ($3.6x10^6 \pm 6.0x10^5$ cells/ml at 10 hours) was approximately 1 log cycle less than from *P. furiosus* ($1.6x10^7 \pm 2.7x10^6$ cells/ml at 10 hours). At stationary phase cell density increased in *P. furiosus* (to $3.0x10^7 \pm 6.6x10^6$ cells/ml at 19 hours). Unlike *P. furiosus*, the cell density of *P. woesei* declined and there was an approximate 2 fold decrease in the cell numbers ($1.6x10^6 \pm 4.9x10^5$ cells/ml) between 10 and 19 hours.

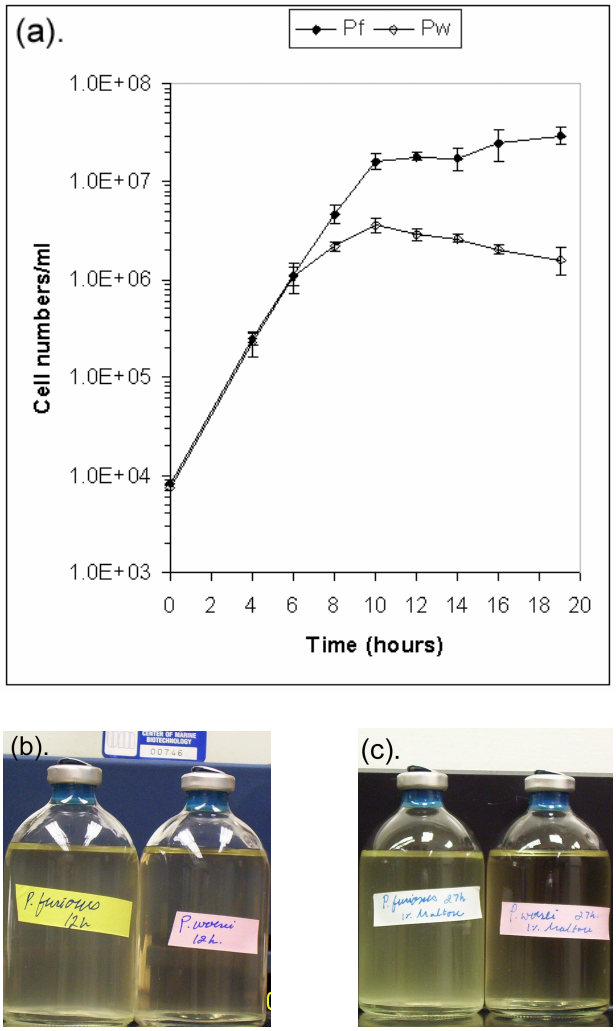


Figure 6. Growth of *P. furiosus* and *P. woesei* in S^o-free medium+1% maltose+100 μM sodium tungstate. (a) Representative growth curves. Triplicate cultures of *P. furiosus* and *P. woesei* were incubated at 95 ^oC anaerobically. Cells were counted randomly for at least 50 unit squares (100 unit squares is1 mm^2). Solid and blank diamonds represent cell densities (numbers/ml) of *P. furiosus* and *P. woesei*, respectively. (b) Cultures of *P. furiosus* (left) and *P. woesei* (right) at 12 hours. (c) Cultures of *P. furiosus* (left) and *P. woesei* (right) at 27 hours.

Figure 7. Micrographs of acridine orange stained cells of *P. furiosus* and *P. woesei* growing in S^o-free medium+1% maltose+100 μM sodium tungstate. The photographs were taken at identical magnification using an oil immersion objective (1000x). Bars represent size reference (10 μm). Numbers shown in parentheses represent volumes of cultures used to prepare the microscopic slides. (a) *P. woesei* at 6 hours. Cell density obtained corresponded to the 6 hour data point shown in Figure 6a (ca. $1.1x10^6$ cells/ml). Almost all of the cells appear as singlets and doublets. (b) *P. furiosus* at 6 hours. Cell density obtained corresponded to the 6 hour data point shown in Figure 6a (ca. $1.1x10^6$ cells/ml). Most cells are in singlet and doublet morphologies. (c) *P. woesei* at 19 hours. Cell density obtained corresponded to the 19 hour data point shown in Figure 6a (ca. $1.6x10^6$ cells/ml). Most of the cells are clumped in large aggregates and very rarely occur in singlet or doublet morphologies. (d) *P. furiosus* at 19 hours. Cell density obtained corresponded to the 19 hour data point shown in Figure 6a (ca. $3x10^7$ cells/ml). Most of cells were single, in pairs or in short chains and rarely in aggregates.

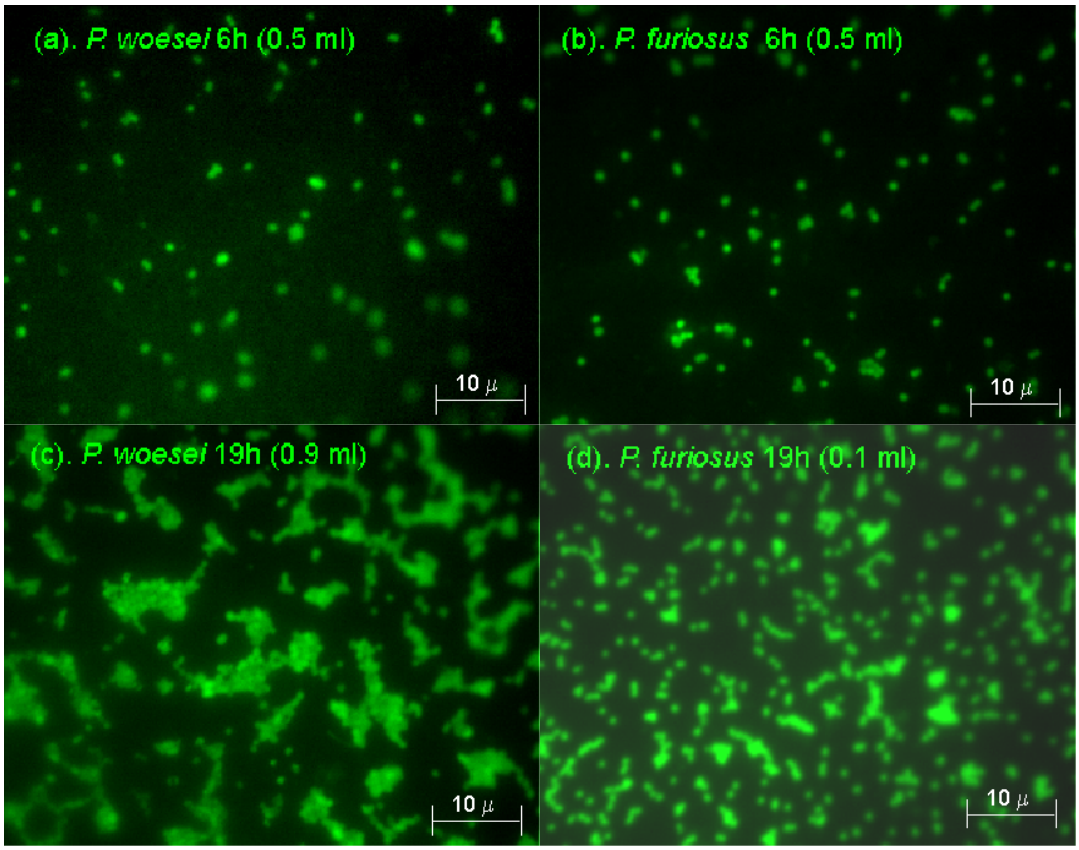

(a). P. woesei 6h (0.5 ml)
10 μ
(b). P. furiosus 6h (0.5 ml)
10 μ
(c). P. woesei 19h (0.9 ml)
10 μ
(d). P. furiosus 19h (0.1 ml)
10 μ

Figure 6b shows different turbidities of cultures of *P. furiosus* (left) and *P. woesei* (right). The results correspond to cell densities obtained from cell counts at t=12 h (mentioned above). Figure 6c shows the turbidities of *P. furiosus* (left) and *P. woesei* (right) corresponding to increase and decrease in cell densities in stationary phase, respectively (compare turbidities of Figure 6b and 6c). Figures 7a-7d show microscopic morphologies of *P. woesei* (Figures 7a and 7c) and *P. furiosus* (Figures 7b and 7d). Figures 6a and 6b show similar cell morphologies and arrangements of *P. woesei* (Figure 7a) and *P. furiosus* (Figure 7b) observed during exponential phase (6 hours). Most of the cells were arranged singly or in pairs. Unlike *P. furiosus* (Figure 7d), large aggregates of cells are observed in *P. woesei* (Figure 7c) in stationary phase (19 hours).

The results indicate that both *P. woesei* and *P. furiosus* utilize maltose in the presence of 100 μM Na_2WO_4 equally well during exponential phase. However, the cell density of *P. woesei* declined significantly in stationary phase (Figure 6). This result suggests *P. woesei* is more sensitive to stress induced autolysis than *P. furiosus* cells in the stationary phase.

3.3.2. Identical 16s rDNA sequences

Nucleotide sequences of the 16S rRNA, $tRNA^{Ala}$ and 23S rRNA (partial) cluster from *P. woesei* were sequenced (GenBank Accession Number AY519654) and the results revealed sequences identical to the homologous cluster in *P. furiosus* (GenBank Accession Number AE010139 including the intergenic spacer regions).

Table 3. BLASTn results of nucleotide sequences from *P. woesei* (*P.w.*) against homologous sequences in the genomes of *P. furiosus* (*P.f.*) , *P. horikoshii* OT3 (*P.h.*) and *P. abyssi* (*P.a.*).

Access. Nos. (*P.w.)*	Coordinates (*P.f.)*	Descriptions	Identity (%) to homologous sequences in		
			P.f.	*P.h.*	*P.a.*
AY443493	1844183-653	*dUTPase*	100	85	75
X15329	1387256-380	5S rRNA	97	96	96
[a]AF461062	[b]285776-6947	*napA* Na^+/H^+ antiporter gene (disrupted by an IS-III)	99	69	69
[a]AF461062	977655-8433	IS element type III	99	Not found	Not found
[a]AF420277	909866-10646	*transposase type II*	99	Not found	Not found
[a]AF443788	284983-5764	*transposase type I*	100	Not found	Not found
AF043283	1148005-9537	*β-galactosidase*	99	66	74
U84155	228589 -30933	*DNA-dependent DNA polymerase*	100	73-74	74
Y09481	1771227-86	*triosephosphate isomerase*	100	78	77
AF240464	491345-2905	*α-amylase*	99	Not found	Not found
AF177906	491384-2863	*α-amylase* (*amyA*)	100	Not found	Not found
AF239672	297558- 301754	*pyrolysin*	99	Not found	Not found
X59857	1289516- 92456	elongation factor-1α, ribosomal protein S10 and tRNA-Ser cluster	99%	73-83	70-79
X73527	1011378-2827	*3-phosphoglycerate kinase*	99	71	73-77
X60161	465583-7110	*glutamine synthetase*	99	73	75
X70668	1291998-3065	*transcription factor IIB*	100	79	79
X67205	1858886-61654	*translational elongation factor 2*	99	83	70-82
U56247	1838967-40285	cell division protein (*FtsZ*)	99	76	77
U10285	1222001- 576	TATA-binding protein	100	78	79

M83988	1728168- 9247	*glyceraldehyde 3-phosphate dehydrogenase*	100	80	75
M83987	1727120-479	putative ribosomal protein	99	76	76

[a]Sequences from this work, [b]No IS element disrupting the *napA* in *P. furiosus*

The two intergenic spacer regions are sensitive sites for differentiation of Archaea in the genus *Pyrococcus* as shown by DiRuggiero and colleagues (1995).

3.3.3. Phylogenetic analyses of nucleotide sequences

To date, there are 20 *P. woesei* sequences available as GenBank Accessions. These nucleotide sequences (28,857 bp) include 22 genes and 1 three gene cluster (X59857). BLAST results against genome sequences of *P. furiosus, P. horikoshii* OT3 and *P. abyssi* are summarized in Table 3. The results indicate that nucleotide sequences of *P. woesei* are highly similar or identical (99-100% identities) to the homologous sequences of *P. furiosus*. They have uniformly lower identity (approximately 75%) to the homologs in *P. horikoshii* and *P. abyssi*. It is apparent that the genomes of *P. horikoshii* and *P. abyssi* contain no amylase {*P. furiosus*' α-amylase genes (AF240464 and AF177906)} and no IS elements, confirming a previous study (DiRuggiero *et al.*, 2000). Three transition nucleotides (^{358}A, ^{369}C and ^{378}A) were observed in a *tnp-II* (AF420277) in *P. woesei* and the gene has 99.6% identity to the prototype sequence in *P. furiosus*.

3.3.4. Southern blots for comparison of restriction patterns

The genome of *P. woesei* contains IS elements (GenBank Accession Numbers AF420277 and AF443788) homologous to the IS elements in the genome of *P. furiosus*. In *P. furiosus*, the IS elements are grouped into three types according to their nucleotide sequences, respectively; IS types I (11 copies), type II (8 copies) and type III (4 copies) and type D (4 copies) (see also Section 2.3.1). The nucleotide sequences of the IS elements in all classes contain no *Pst-1* restriction sites. Element numbers

have been assigned previously (Table 1). The sizes of fragments containing an IS element between two *Pst-1* restriction sites were predicted from the genome sequence of *P. furiosus*. The predicted *Pst-1* restriction fragment sizes are summarized in Table 4. Southern blot analysis using *Pst-1* designation and DNA hybridization using a probe specific to *tnp* sequences (Figure 8) revealed that the genome of *P. woesei* (lane 1) contains multiple copies of these IS elements. In the genome of *P. furiosus* (lane 2), band numbers 1 - 7 and a cluster of large fragments (bracket 8) were mapped according to their relative intensities and predicted *Pst*-1 restriction fragment sizes (see Section 3.2.3).

Table 4. Element numbers (IS Nos.) and predicted *Pst-1* restriction fragment sizes (kb). Sizes of fragments containing a single IS element sequence were calculated from the two nearest *Pst-1* restriction sites flanking an IS element sequence. Band Nos. are numbers assigned to the bands observed in lane 2 of Figure 8.

Types	IS Nos.	Sizes (kb)	Band Nos.	Types	IS Nos.	Sizes (kb)	Band Nos.
I	1	5.8	5	I	15	10.8	8
I	2	5.8	5	I	16	10.7	8
I	3	3.6	2	I	17	5.2	4
I	4	15.8	8	III	18	11.8	8
III	5	15.8	8	III	19	6.7	6
II	6	18.1	8	III	20	2.2	1
II	7	22.7	8	II	21	25.9	8
II	8	12.9	8	II	22	6.1	5
II	9	12.6	8	II	23	12.4	8
II	10	16.8	8	D	24	4.6	3
I	11	6.8	6	D	25	10.4	8
I	12	11.7	8	D	26	4.5	3
I	13	5.2	4	D	27	5.8	5
I	14	2.1	1				

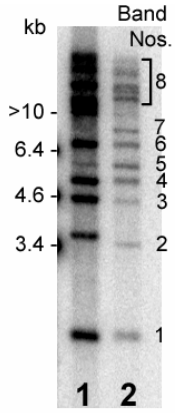


Figure 8. Restriction patterns from a Southern blot probed with ^{32}P-labeled *tnp* probes. DNAs from *P. woesei* (lane 1) and *P. furiosus* (lane 2) were completely digested with *Pst-1* restriction enzyme. Size markers are shown on the left of the Figure. Band numbers were assigned only to the bands from *P. furiosus* and are shown on the right of the Figure. Intensities of bands and indices were determined (see Section 3.2.3). Intensity of band 2 (size ca. 3.5 kb) represents fragments containing a single element and index one was assigned. Band 1 (index = 1.9) contains two IS elements (sizes 2.1 and 2.2 kb). Band 3 (index = 1.5) contains two type D elements (sizes 4.5 and 4.6 kb). Band 4 (index = 2.6) contains two IS elements (size ca. 5.2 kb). Band 5 (index = 3.2) contains three IS (including IS element 1 upstream of the *napA* mentioned below) and one deletion mutant elements (size ca. 5.8 – 6.1 kb). Band 6 (index = 3.0) contains two IS elements (size ca. 6.8). No predicted fragment size matches band 7 (size ca. 8 kb). Bracket 8 is assumed to contain the remaining IS elements in the fragment size larger than 10 kb (see also Table 4). The results imply that genomes of *P. furiosus* and *P. woesei* are not identical and are co-linear in many regions.

Generally, the mapping correlated well with band intensities obtained from bands 1 - 6. The cluster 8 (bracket) contained large unresolved fragments (>10kb) and may contain a mixture of fragments. This phenomenon was also observed in band 7 (size ca. 8 kb), since no IS element containing a fragment size around 8 kb is located in this band (see also Table 4). Thus, this area was not accounted for in the comparison. The restriction pattern observed in *P. woesei* (lane 1) revealed the same relative *Pst-1* fragments in both genomes (bands 1 and 3-6 of *P. furiosus*) except for the band corresponding to one element in band 2 in *P. furiosus* (lane 2) and relative low intensity band of *P. woesei* that corresponds to four elements in band 5 (element numbers 1, 2, 22 and 27D). These bands contained ten elements in bands 4-6 and two elements of type D in band 3 of *P. furiosus* and element 1 in band 5 (see also Table 4). The result implied that the genomes of *P. furiosus* and *P. woesei* are co-linear in the regions covered by 13 IS elements (a total of 64.4 kb). Interestingly, the band from *P. woesei* that corresponds to the band 5, which includes element number 1 and a Na^+/H^+ antiporter (restriction fragment size 5.8), of *P. furiosus* was observed to have low intensity relative to other bands in lane 2 (see also Section 3.3.6).

3.3.5. Lambda library and mapping of the *napA* locus in *P. woesei*

To search for the flanking DNAs at the ends of the IS in a λ library containing genomic DNA from *P. woesei* using labeled probes specific to the *tnp* sequences, approximately $1x10^{-5}$ positive plaques were obtained. Twenty plaques were rescued and inserts from pZL1 were sequenced. All positive clones contained short fragments (approximately 100 - 300 bp) of the IS sequences. This was because the nucleotide

sequence of the *Pyrococcus* species DNAs are AT rich and thus the AATT recognition site for the *Tsp509-I* occurred frequently. An alignment of the twenty clones carries a 78 nucleotide long sequence at the right end of the IS (corresponding to nucleotide numbers 632 – 779 of an IS-III) joining with a 22 nucleotide long sequence aligned perfectly to coordinates 286776 - 286797 in the genome of *P. furiosus*. In *P. furiosus*, the coordinates 286776 - 286797 are part of a putative Na^+/H^+ antiporter (*napA*) gene (coordinates 285783 - 286922) that had 69% homology to a putative *napA-3* (PAB1247) in *P. abyssi* and a hypothetical Na^+/H^+ antiporter open reading frame (PH0302) in *P. horikoshii* OT3.

In order to map the locus in *P. woesei*, primers 4 to 9 were employed to walk upstream and downstream of the IS sequences. The results of PCR amplification on genomic DNA of *P. woesei* and *P. furiosus* (control) are shown in Figure 9. The results revealed fragment sizes of ca. 904 bp and 210 bp amplified using primer pairs 4-3 and 9-7, respectively, from *P. woesei* DNA (indicated by arrows in lanes 1 and 5). From *P. furiosus*, a negative amplification reaction (lane 2) and a fragment size of 303 bp (indicated by an arrow in lane 3) were obtained using the primer pairs 4-3 and 4-7, respectively. The results corresponded to the presence and absence of an IS (mentioned above) in the *napA* locus of *P. woesei* and *P. furiosus*, respectively. Cloning and sequencing experiments of the *napA* locus in *P. woesei* confirmed the presence of an IS type III disrupting the relative coordinates 2828675-6 and revealed a direct repeat sequence (5'-CCAGGAT-3) flanking the type III element. These results indicated that an IS element disrupts the *napA* sequence in *P. woesei*.

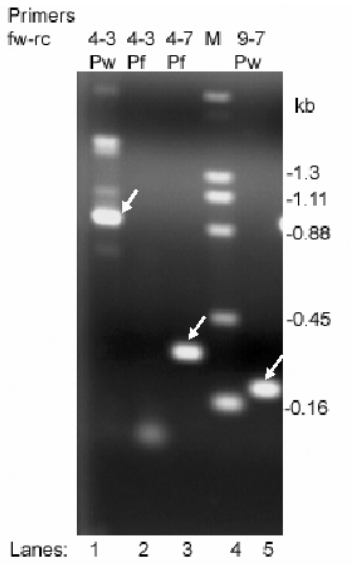


Figure 9. PCR products amplified using primer pairs 4-3, 4-7 and 4-9. Pw and Pf represent lane loaded with reactions using genomic DNA from *P. woesei* (lanes 1 and 5) and *P. furiosus* (lanes 2 and 3). M marks a lane (lane 4) loaded with size markers. Sizes of the markers are indicated at the right hand side of the Figure. Arrows indicate bands correspond to anticipated bands. A negative result in lane 2 confirms that no IS element disrupts the *napA* gene in *P. furiosus*. (see binding coordinates in Section 3.2.5 and Figures 12a and 12b below).

The nucleotide sequence of the IS-III is 99.5% similar to the type III element 20. The nucleotide sequence of the disrupted *napA* locus (total 1969 bp) was submitted to GenBank (Accession Number AF461062). Apart from the IS-III insert and the direct repeat sequences, the nucleotide sequence of the *napA* of *P. woesei* has 99.3% identity to the homologous sequence (coordinates 285783-286922) in *P. furiosus.* This level of divergence is typical of the genomic sequences analyzed in Table 3.

3.3.6. Identification of an IS-I element in exact context as a genetic marker

The genome of *P. furiosus* contains an IS-I element 1 located at coordinates 284983-285764 (Table 1) upstream of the *napA* locus. This raises an interesting question: Is this IS type I also present in *P. woesei*? A PCR based strategy with primer pair 9-6 was employed (see map in Figures 12a and 12b below) to amplify DNAs of *P. woesei* and a *P. furiosus* positive control. As anticipated, a fragment of ca. 1079 bp was obtained from the positive control (lane 2 in Figure 10). From *P. woesei*, two fragments (sizes 60 bp and 1858 bp) should have been amplified using primers 9 and 6. The results revealed only the short fragment (lane 1 in Figure 10) due to the blocking of PCR reactions by the distal distant primer pair 9-6 (see map in Figure 12b).

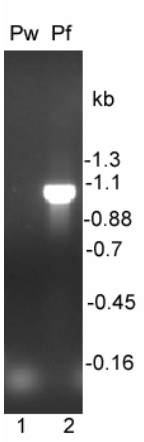


Figure 10. PCR products amplified using primer pair 9-6. Lanes 1 and 2 mark lanes loaded with reactions amplified from the genomic DNAs of *P. woesei* and *P. furiosus*, respectively. Size markers are shown to the right of the Figure.

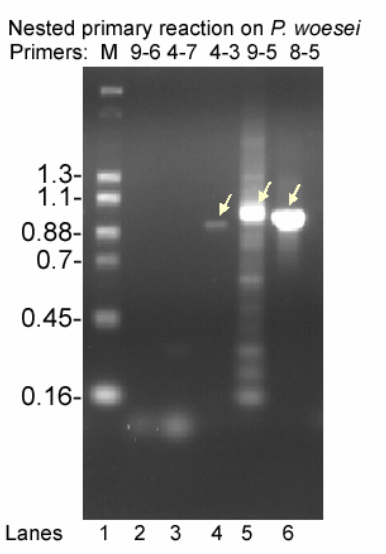


Figure 11. Nested PCR products amplified from a dilute (1:600) primary reaction from *P. woesei* genomic DNA. Primer pairs 9-6, 4-7 (negative control), 4-3, 9-5, and 8-5 employed as indicated on top of the Figure correspond to lanes 2-6, respectively (see also relative coordinates in Figures 12a and 12b below). Lane 1 (M) represents size standards (kb). Arrows indicate bands correspond to anticipated products.

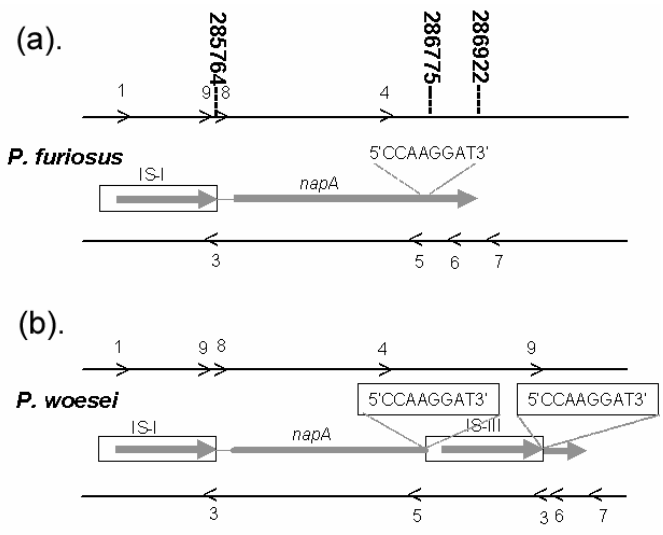


Figure 12. Maps of IS elements in the *napA* loci of (a) *P. furiosus* and (b) *P. woesei*. Parallel lines represent double stranded DNAs. Boxes with thick arrows represent IS elements. Types of the IS elements are indicated on top of the boxes. Thick arrows represent genes. *NapA* (or Na^+/H^+ antiporter gene) are indicated on top. Greater and lesser signs with numbers represent forward and reverse primers, respectively (see also binding coordinates in Section 3.2.5). Coordinates in the genome of *P. furiosus* are shown after vertical dotted lines. Direct repeat sequences are boxed. The *napA* in *P. woesei* contains two IS elements including the IS type I in the same context as in the genome of *P. furiosus*. The results indicated recent transposition event disrupted the *NapA* in *P. woesei*. This corresponds to band shift observed on band corresponded to band 5 to band corresponded to band 6 (see also Figure 8).

Nested PCR amplifications on a diluted (1:600) primary reaction from *P. woesei* using a set of primer pairs (4-7 as negative control, 4-3, 9-5, and 9-6) that bind internally and/or externally to the large fragment were employed to demonstrate the presence of a trace amount of PCR fragment containing the IS type I sequence. Results are shown in Figure 11. Amplification using the primer pair 4-7 (negative control) suggests that sample preparation technique was able to eliminate contamination of the genomic DNA template from the primary reaction (mentioned above) because negative amplification was obtained (lane 3 in Figure 11). On the other hand, amplification using the primer pairs 9-6, 4-3, 9-5, and 8-5, revealed the anticipated fragment sizes of ca. 60, 904, 990 and 952 bp (lanes 2 and 4-6, respectively). These results suggest there was a trace amount of the large fragment present in the primary amplification reaction using the primer pair 9-6 from the genomic DNA of *P. woesei*. This result implies that the genome of *P. woesei* contains the IS element type I (IS-I) in the exact context as identified in the *napA* locus in the genome of *P. furiosus*. Sequencing of the PCR products confirms the presence of the large fragment. Maps comparing the *napA* loci of *P. furiosus* and *P. woesei* are shown in Figures 12a and 12b.

The IS-I element 1and *napA* restriction fragment size of 5.8 kb from *P. furiosus* was mapped in band 5 in Figure 8 mentioned above (see also restriction fragment sizes of element numbers 1, 2, 22 and 27D in Table 4). This band in *P. woesei* shows relatively low intensity compared to the others. This observation thus points to a band shift observed due to an insertion disruption of the *napA* by the IS element type III (size of ca. 0.8 kb) in *P. woesei* (mentioned above).

3.4. Discussion and conclusions

P. woesei shows very similar physiological properties to *P. furiosus*. Both species were isolated from marine sediments from the same Vulcano Island beach site in Italy. Both species grow rapidly in the presence of 0.1% maltose and 10-100 μM Na_2WO_4 in a S^o-free medium. However, cell density of *P. woesei* cultures decreases faster than that of the *P. furiosus* in stationary phase, showing extensive autolysis. This observation is also observed in over night cultures of *P. furiosus* and *P. woesei* in the S^o-free medium (5 replicates each). Therefore, growth phenotypes in the S^o-free medium of both species are not identical. This observation might reflect a phenotypic difference between theses two isolates relating to differences in membrane integrity and osmotic functions possibly due in part to the defective *napA* gene in *P. woesei*. Na^+/H^+ antiporters are membrane proteins that exchange Na^+ and protons (H^+) and could play an important role in osmoregulation (Hellmer *et al.,* 2002; Waditee *et al.,* 2002.). Experiments to restore the lytic phenotype of *P. woesei* will be necessary.

The 16S rRNA, $tRNA^{Ala}$ and 23S rRNA (partial) gene cluster from *P. woesei* (GenBank Accession Number AY519654) were sequenced and the results reveal identical sequences to the homologous gene cluster in *P. furiosus* (GenBank Accession Number AE010139). Comparisons of *P. woesei* sequences available from public sources (Table 3) confirm highly similar (95-100%) sequences to the nucleotide sequences of *P. furiosus*. Restriction patterns of fragments containing IS element markers (Figure 8, at least 12 out of 13 elements) support co-linearity between the genomes of both species. The presence of two IS elements in the Na^+/H^+ antiporter

(*napA*) region in *P. woesei*, including one in the same context as the upstream IS-I in *P. furiosus*, not only suggests recent movement of the IS elements, but also provides crucial evidence to confirm recent movement following divergence of the two strains as well as serve as a genetic marker of the two siblings. The fragmented *napA* gene in *P. woesei* has the same DNA sequence as the *P. furiosus napA*, except for 8 nucleotides (99.3% identity). The variation observed corresponds to the variations of other active genes including the IS-III (99.5% identity) and the *tnp-II* (AF420277) gene (99.6% identity) (Table 3). The results indicate that the IS elements predated the divergence of *P. furiosus* and *P. woesei*.

In conclusion, the experiments support the hypothesis that the IS elements are active. The comparison of the *P. furiosus* and *P. woesei* genomes suggests that these strains are very closely related. Since genome of *P. furiosus* (Fiala and Stetter, 1986) has been sequenced and is a well known species of hyperthermophilic Archaea. On the other hand, *P. woesei* (Zillig *et al.,* 1987) was isolated later and is not as well known as *P. furiosus*. The designation of *P. woesei* to *P. furiosus* sub sp. *woesei* is proposed based on (i) both species have identical rDNA operon, (ii) both species shared highly nucleotide sequence identity (99-100%), and (iii) both species shared IS elements marker in particular the exact context of the IS type I element 1 in the *napA* locus.

Chapter 4

Identification of programmed translational frameshift sites and possible mechanisms

4.1. Introduction

4.1.1 Literature review

Programmed translational frameshifting (PTF) is a mechanism that occurs in eukaryotes, bacteria and Archaea, as well as in several retroviruses and transposons. Transposons and insertion sequences control levels of protein expression by PTF. In most cases, frameshifting results in the fusion of two proteins encoded by adjacent open reading frames. In *Sulfolobus solfataricus*, a hyperthermophilic archaeon, translation of α-L-fucosidase is accomplished by means of a translational frameshift utilizing a heptanucleotide sequence (A AAA AAT) in the junction between two open reading frames (SSO11867 and SSO3036 ORFs) and a rare histidine codon (CAC) (Cobucci-Ponzano *et al.,* 2003). Frameshifts can also result in premature termination of long open reading frames. In the γ subunit of DNA polymerase III, a truncated form of the τ subunit encoded by *dnaX* in *E. coli*, was produced by means of a frameshifting mechanism that positioned the ribosome so that it encountered a stop codon in the -1 frame downstream (Tsuchihashi and Kornberg, 1990). IS*1221* from *Mycoplasma hyorhinis* utilizes a -1 frameshifting mechanism to produce a truncated product corresponding to upstream coding region of a putative *transposase*. Presumably low-

level production of the full-length transposase is maintained to provide moderate rates of transposition (Zheng and McIntosh, 1995).

Programmed translational frameshift sites in mRNAs can also increase error rates of ribosomes during translation (see reviews in Farabaugh, 1996; Gesteland and Atkins, 1996). The non-programmed frameshift error rate of non-PTF sites was estimated at 10^{-5} per codon (Kurland, 1992). In contrast, up to 50% of the γ subunits of the DNA polymerase III holoenzyme in *E. coli* (Tsuchihashi and Kornberg, 1990) and *Thermus thermophilis* (Yurieva *et al.,* 1997) were produced by premature termination of the *dnaX* genes at the A_6G and A_9 shift sites, respectively.

Mechanisms of PTF differ among organisms from bacteria to higher Eucaryotes in response to particular sequences. A ribosome binding site upstream as well as downstream, secondary structures of mRNA (stem loops) (Rettberg *et al.,* 1999) and/or pseudoknots (Nixon *et al.,* 2002; Kim *et al.,* 2000) are usually involved in the process as additional stimulators adjacent to the shift sites. Spacing of the stimulators in relation to a shift site and surrounding nucleotide sequence in a frameshift context may also provide combined effects and determine the frequency and mechanisms at a particular frameshift site (review in Farabaugh, 1996). These stimulators may combine their effects by interacting with or pausing the ribosome during the elongation process and thus result in the slippage of the ribosome at the programmed frameshift site either one base backward (termed -1 frameshifting), or one base forward (termed +1 frameshifting), or two bases backward (with a similar outcome to +1 frameshifting). Many +1 frameshifts occur by means of single t-RNA slippage mechanisms (review in

Farabaugh and Bjork, 1999) that may be stimulated by pause-prone codons and weak pairing of near cognate codon-anticodon interactions at either the ribosomal A or P sites (Pande *et al.,* 1995; Vimaladithan and Farabaugh, 1994; Curran, 1995). The most common *cis*-acting signals for -1 frameshifting are motifs termed heptanucleotide slippery sequences (consensus X XXY YYZ) at the shift site, and the motifs were first identified in several human retroviruses and coronaviruses (Brierley *et al.,* 1992; Chamorro *et al.,* 1992). They have subsequently been found in a wide variety of organisms and genetic elements including bacteria, transposons, and IS elements in family 3 (Mahillon and Chandler, 1998).

Most -1 frameshifting mechanisms occur at the slippery heptamers by means of simultaneous slippage in which both aminoacyl-tRNA at the A site and peptidyl-tRNA at the P site slip backward one base (Weiss and Gallant, 1983; Farabaugh, 1996). A strong and well characterized -1 frameshift site (A AAA AAG) in the *dnaX* gene of *E. coli* is flanked by an upstream SD (with a 10 bp spacer) and a downstream stem loop (with a 6 bp spacer) which may combine their effects on the frameshift site (Tsuchihashi and Kornberg, 1990; Tsuchihashi and Brown, 1992; Blinkowa and Walker, 1990; Flower and McHenry, 1990).

Spacing between the SD and the frameshift site may affect the outcome of the frameshift frequency and mechanism. Moving the SD from the 10 bp spacing in the *dnaX* gene of *E. coli* away (19 -22 bp) or toward (4 – 6 bp) the frameshift site resulted in decreased frameshifting or inhibition of -1 frameshifting, respectively (Larsen *et al.,* 1994). Two frameshift contexts contain short spacing between SD stimulators and

frameshift sites, one is in the *prfB* gene (with 3 bp spacing) in *E. coli* (Craigen and Caskey 1986; Craigen *et al.,* 1985) and the other is in the homologous *dnaX* gene (with 1 bp spacing) in the extreme thermophile *Thermus thermophilis* (Yurieva *et. al.,* 1997) that generates a +1 frameshift outcome.

4.1.2. Specific objectives

Previous results revealed that full-length transcripts of the tnp-I and tnp-II mRNA are present in exponential and stationary phase cultures of *P. furiosus* (Section 2.3.8). Since transposases are enzymes that mediate transposition, unrestrained translation and consequent continuous transposition could be deleterious to the host. The analysis of the *tnp* sequence (Section 2.3.6) and fruitless attempts to express recombinant *tnp*s (Section 2.3.9) indicate that the genes are difficult to translate and might be regulated by programmed translational frameshifting mechanisms. This was supported by the measurements of mRNA in the *E. coli* cultures (results not shown). The goals of this study are (i) to identify programmed translational frameshift context(s) and shift site(s) in the *tnp-I* sequence, (ii) to verify the mechanisms of the frameshift, and (iii) to study the kinetics of frameshifting.

4.2. Materials and methods

4.2.1. Reporters

Plasmid pUC 18 was employed as a zero reporter (Z) in which *lacZ*α is expressed as an in-frame fusion with the preceding multiple cloning sites (mcs). Plus 1 (P) and minus 1 (M) reporters were constructed by the replacement of the nucleotide sequence (5' CTGCAG GCATGC AAGCTT 3') between *Pst 1* and *Hind III* of the mcs

with modified nucleotide sequences 5'CTGCAGG CAT T GCA AGCTT 3' and CTGCAGG CAT TA GCA AGCTTG (T and TA were inserted bases), respectively. White transformants on X-gal plates (LB agar + 100 μg/ml ampicillin + 40 μg of 5-Bromo-4-chloro-3-indolyl-β-D-galactoside) were obtained. The frameshifted *lacZα* constructs were sequenced to confirm the nucleotide sequences of the inserts and the *lacZα* open reading frames.

4.2.2. Screening primers and procedures

Primers named 1Ifw, 2Irc, 3Irc, 4Ifw, 5Ifw, 6Irc, 7Ifw, 8Irc, 9Irc, 10Ifw, and 11Irc were designed to contain either *Eco* RI sites (in forward primers) or *Pst-1* sites (in reverse complementary primers) for cloning purposes. Nucleotide sequences are listed in Appendix A. Combinations of the forward (suffix Ifw) and reverse (suffix Irc) primers were employed to amplify PCR fragments of the tnp-I template encoded in the pTnO11 plasmid constructed in this study. PCR fragments with size ranges from 66 to 699 bp long representing proximal, central, distal and complete tnp-I sequences were digested with *Eco* RI and *Pst-1* at 37 ^{o}C for 4-6 hours. Gel purified restriction fragments were ligated between *Eco* RI and *Pst-1* restriction sites of the Z reporter using T4 ligase, and then transformed into *E. coli* Top10. Two small fragments (66 bp long) that caused light blue colonies in the Z reporting strains on the X-gal plates were chosen for further studies.

These fragments (representing regions 1011 and 79) were examined for out of frame stop codon(s) both in +1 and/or -1 frames. Site-directed mutagenesis procedures were performed using modified primers to remove the out-of-phase stop codons (but

the amino acid sequences were conserved). The mutated primers named "7CIfw" (5'-AAAAAA *GAATTC* ACGCTG GTTCAG GACTCT CAAAG-3'), "9CIrc" (5'-AAAAAA *CTGCAG* CAGTCG CGAGCC CTGAAA TTG TTCCA-3'), "10CIfw" (5'-AAAAAA *GAATTC* CCTCAA ACCTTT TCG-3'), and "11CIrc" (5'-AAAAAA *CTGCAG* AGGTAT AGTTCG ACTGCT-3') were employed to amplify mutant fragments named "7C9C" and "10C11C", respectively. Restriction sites and mutations introduced into primers are shown in italic and underlined letters, respectively. The primers 7Ifw and 9CIrc, and 10Ifw and 11CIrc were employed to generate fragments named "79C" and "1011C", respectively (see Figures 13a and 13b).

The modified fragments were screened for frameshifting using the P and M reporters (mentioned above in Section 4.2.1). Strains (P or M) that produced green to light blue colonies on X-gal plates were chosen and the inserts, mcs and *lacZ*α were sequenced to confirm the nucleotide sequences and the frames of translation.

4.2.3. Oligonucleotide sequences and cloning procedures

Cloning of short DNA fragments was performed using pairs of synthetic oligonucleotides containing nucleotide sequences representing sub-regions 1011 and 79. The oligomeric pairs (suffixes Up and Lo in Appendices B and C) were annealed at 37 ^{o}C for 30 minutes prior to being ligated into reporters using T4 ligase at 4-10 ^{o}C overnight. Nucleotide sequences of the oligomeric pairs (331, 1031a and 1031b) employed to study region 1011 are listed in Appendix B. Nucleotide sequences of the oligomeric pairs (334, AGG, AGA, CGG, CGA, CGC, CGT, ThrA, ThrC, ThrG, NOAGG, and 79SH) employed to study region 79 are listed in Appendix C.

4.2.4. Nomenclature systems

PCR fragments were named after the primer pairs employed in the amplification reactions. For instance, fragments 12, 7C9C, and 1011C are the names of PCR products amplified using primer pairs 1Ifw-2Irc, 7CIfw-9CIrc, and 10Ifw-11CIrc, respectively. Strains were named after names of the fragments and a suffix Z, P, or M for zero, plus, or minus reporter employed, respectively. Thus strains 7C9C-Z, 7C9C-P, and 7C9C-M represent strains carrying a DNA fragment 7C9C in zero, plus, and minus reporters, respectively. Strains carrying short DNA sub-fragments were named after the names of the oligomeric pairs followed by a suffix (Z, P or M) otherwise mentioned in the body of the text. For example, strains 331-P and AGG-Z represent strains carrying inserts 331 and AGG in plus and zero reporters, respectively.

4.2.5. mRNA folding

Folding calculations for mRNA were performed using the mfold program (Zuker, 2003).

4.2.6. Screening on X-gal plates

Ligation reactions (2μl) were added to transform *E. coli* Top 10 competent cells (50 μl) in microcentrifuge tubes. The tubes were incubated on ice for 30 minutes, then heated at 42 ^{o}C for 45 seconds and finally cooled on ice for 2 minutes. SOB medium (300 μl) was added and the tubes were vigorously shaken at 37 ^{o}C for 1 hour. The cultures (100 μl) were spread on X-gal plates, and the plates were incubated at 37 ^{o}C for 18-24 hours. β-galactosidase production and the fusion lacZα peptides were indicated by blue/white colonies following incubation.

4.2.7. β-Galactosidase Assays

Enzyme activities were assayed using ONPG (2-Nitrophenyl β-D-galactopyranoside) as substrate in buffer Z, pH 7.3 (Miller, 1992). One unit of enzyme activity is defined as 1.0 OD_{420} change per minute per mg of protein at pH 7.3. Protein concentration was determined using a dye binding method (Bradford, 1976). Enzyme units (E.U.) were determined using the following equation:

E.U.s = 1000 x OD_{420}/volume (μl)/time (minutes)/protein (mg/ml).

4.2.8. Expression of +1 frameshift reporter proteins

Overnight cultures in LB+amp100 were used to inoculate 150 ml of LB+amp100 in 500-ml culture flasks to a 0.7% final concentration. The cultures were incubated in a shaker incubator at 37 ^{o}C, 220 rpm. Samples (10 ml) were taken at one hour intervals. Cells were collected by centrifugation and washed once with 100 mM phosphate buffer containing 1 mM PMSF, pH 7.0-7.3. The washed cells were resuspended in 1/10 volume of buffer. Cells were lysed using French Pressure (3 times, 800 p.s.i.). Quantities of fusion peptides were determined using the β-galactosidase activities mentioned above (Section 4.2.7). Growth was monitored using optical density at 600 nm. The experiments were repeated at least twice.

Other samples for β-galactosidase activity assays were collected from duplicate cultures growing in 100 ml of LB+amp100 under the conditions described above. Generally, samples obtained from strains carrying the P and Z reporters were collected hourly from 11 - 14 and 14 - 17 hours, respectively.

4.3. Results

4.3.1. Principle of enzyme assay and frameshift reporters

E. coli Top10 encoding *ZMΔ15* (N-terminal deletion of *lacZ*) produces low levels of repressor for lac promoter (*lacI*^{Q}), and expresses an inactive truncated form of β-galactosidase, the ω fragment. Plasmid pUC18 (a high copy number plasmid) expresses the *lacZ*α (encoding N-terminal portion of the β-galactosidase or the α fragment). The alpha (α) peptide together with the larger omega (ω) fragment thus can form an active tetrameric enzyme in the process referred to as alpha (intragenic) complementation.

The pUC18 plasmid is used as a zero reporter (Z) in which lacZ-alpha is expressed as an in-frame fusion with the preceding multiple cloning sites (mcs). Plus 1 (P) and minus 1 (M) reporters constructed in this study (section 4.2.1) produced no detectable activity both on X-gal plates (white colonies after incubation for 24-48 hours) and in liquid culture (10^{-4}-10^{-5} E.U.) due to the fact that the lacZα peptides are prematurely terminated (see also nucleotide sequences in Section 4.2.1).

4.3.2. Screening for DNA fragments containing frameshift sites

PCR products (size ranges from 66 to 699 nt. long) were cloned in-frame into zero reporters and screened for activity on X-gal plates. Two fragments (Figure 13) named 1011 (or region 1011) and 79 (or region 79) were chosen for further study. The DNA fragments 1011 and 79 contain sequences between nucleotide numbers 39-104 and 507-572 of the *tnp-I* gene; they contain two stop codons in the +1 phase of fragment 1011 and a + 1 and two -1 phase stop codons in fragment 79, respectively.

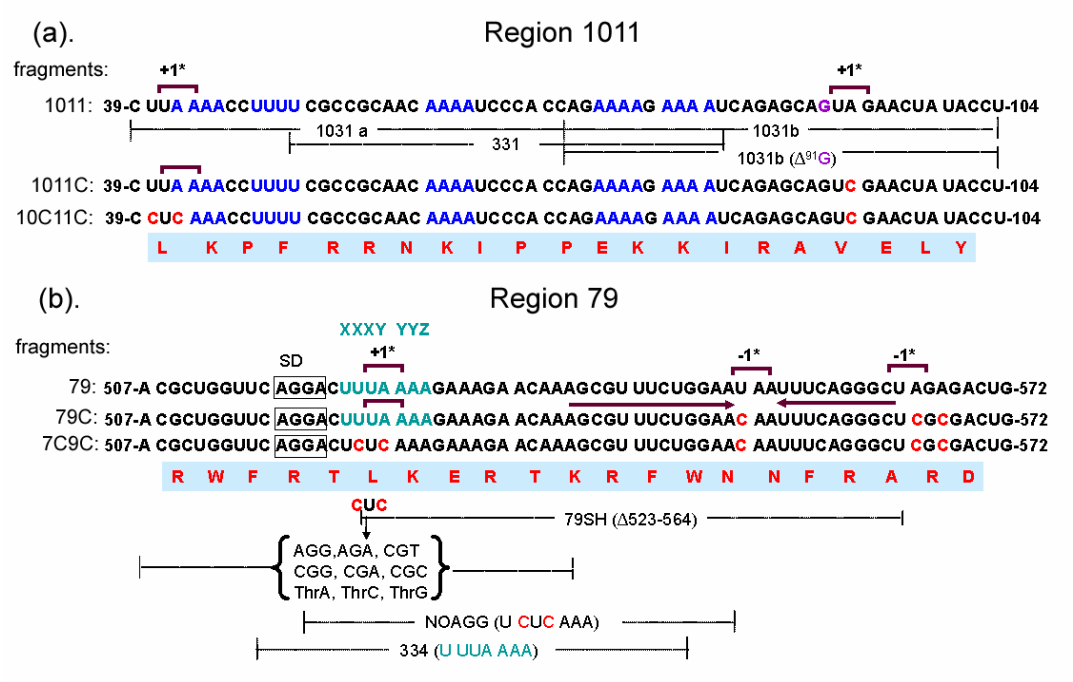


Figure 13. Predicted mRNA sequences of the regions 1011 (a) and 79 (b).

(a) Alignment of nucleotide sequences from fragments 1011, 1011C, and 10C11C. Nucleotide sequences of sub-fragments {1031a, 1031b, 331, and 1031b(Δ^{91}G)} are indicated by measurement lines (see details in Section 4.3.5). Asterisks represent stop codons in +1 and -1 phases. Blue letters represent nucleotides in stretches of 4Ns. Red letters indicate mutations of nucleotides that were mutated in order to remove out-of-phase stop codons (see also Sections 4.2.2 and 4.2.4). Letters in the blue rectangle represent the amino acid sequence decoded using the standard code.

(b) Alignment of nucleotide sequences from fragments 79, 79C, and 7C9C. Symbols are the same as described in (a). Shine Dalgarno (SD)-like sequences are boxed. Thick arrows represent areas with potential to form a stem loop. Putative heptanucleotide frameshift motif (XXXY YYZ) is shown in green letters.

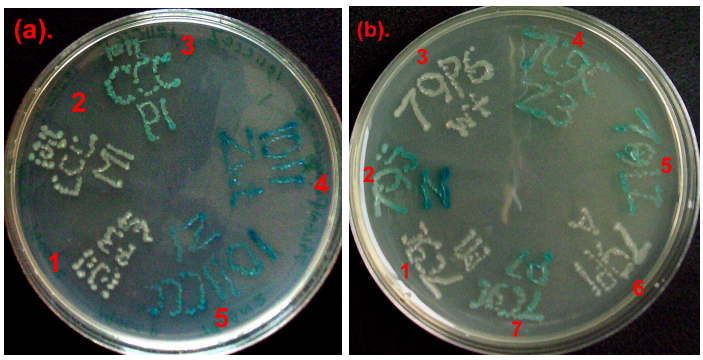


Figure 14. Colonies on X-gal plates at 20-22 hours of strains carrying wild type or mutant inserts in P, M and Z reporters.

(a) Photo taken at 20-22 hours shows phenotypes of strains carrying wild type (with +1 stop codons) fragments and mutant (without +1 stop codon) fragments of the region 1011. Sectors 1 and 4 show colonies of P and Z reporting strains (named "1011-P" and "1011-Z", respectively), which carry wild type fragment 1011. Sectors 2, 3 and 5 show colonies of M, P, and Z reporting strains (named "10C11C-M", "10C11C-P" and "10C11C-Z", respectively), which carry a mutant fragment 10C11C.

(b). Photo taken at 20-22 hours shows phenotypes of strains carrying fragment 79 insert. Sectors 1, 4 and 7 show colonies of M, P and Z reporting strains (named "7C9C-M", "7C9C-Z" and "7C9C-P", respectively), which carry a mutant fragment containing no out-of-phase stop codon. Sector 2 shows colonies of a Z reporting strain (named "79SH-Z"), which carries a deletion fragment {79SH (Δ523-564)}. Sectors 3 and 5 show colonies of P and Z reporting strains (named "79-P" and "79-Z", respectively), which carry a wild type insert 79 (with +1 and -1 stop codon).

In order to be able to detect frameshift products, the stop codons (+1 and -1 phases) were removed by point mutations that conserved the amino acid sequence. The fragments containing mutations are indicated by the suffix C in the fragment names. They were screened on X-gal plates. Screening results obtained from the regions 1011 and 79 are shown in Figures 14a and 14b, respectively. As predicted, +1 strains (named 1011-P and 79-P, respectively) carrying inserts containing wild type sequence (with a +1 stop codon(s)) show white phenotypes (sector 1 in Figure 14a, and sector 3 in Figure 14b, respectively). As observed, -1 frameshifting mechanisms from both regions (1011 and 79) were ruled out by white phenotypes of two M reporting strains named "10C11C-M" and "7C9C-M" (sector 2 in Figure 14a and sector 1 in Figure 14b, respectively). Sectors 4 and 5 in Figure 14a show blue colonies of two zero-strains named "1011-Z" and "10C11C-Z", respectively. Sectors 4 and 5 in Figure 14b show light blue colonies of two zero-strains named "7C9C-Z" and "79-Z", respectively. As expected, +1-strains (named "10C11C-P" and "7C9C-P", respectively) whose inserts contain no +1 stop codon showed blue phenotypes (sector 5 in Figure 14a, and sector 7 in Figure 14b, respectively).

Interestingly, a +1-strain named "79C-P" containing the +1 phase stop codon (524-TAA-526 in the fragment 79C in Figure 13b) showed a white color on the X-gal plate (sector 6 in Figure 14b). The phenotype observed on strain 79C is in contrast to the blue phenotype observed in strain 7C9C-P mentioned above. This result indicates that the sequence between the nucleotides 507-526 in the context of fragment 79 contains a functional, strong +1 frameshift site.

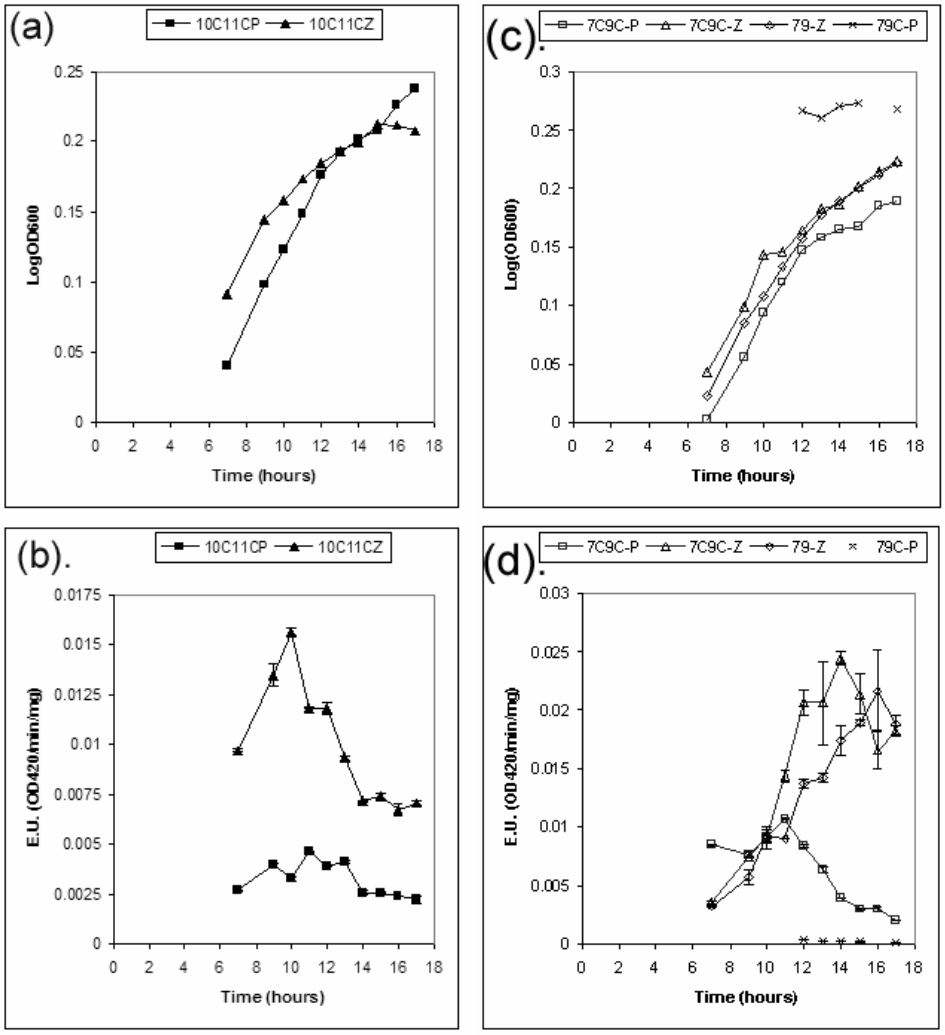


Figure 15. Bacterial growth (a, c) and enzyme production (b, d). Strains 10C11C-P (solid squares) and 10C11C-Z (solid triangles). Growth (c) and enzyme production (d) curves obtained with strains 79-Z (blank diamonds), 7C9C-Z (blank triangles), 7C9C-P (blank squares) and 79C-P (Xs), respectively. The data in Figure 15d were averaged from duplicate cultures and triplicate assays. Standard errors of the mean are shown as bars.

4.3.3. Growth phase dependence of +1 frameshift frequency

Growth and β-galactosidase fusions from P *versus* Z reporting strains growing in LB+amp100 medium are shown in Figures 15a-15d. Figures 15a and 15b compare growth and +1- and 0-frame fusion β-galactosidase production curves between strains 10C11C-P (solid squares) *versus* 10C11C-Z (solid triangles). The results revealed that the activities of +1 and zero-fusion β-galactosidase activities reached their maximal levels (average activities $3.9x10^{-3}\pm3x10^{-4}$ and $1.2x10^{-2}\pm1x10^{-3}$ E.U., respectively) at 9-12 hours when the cells were in the mid and late exponential phases. The enzyme activities then declined when cells entered deceleration and stationary phases. Approximately 30% of the frameshifting level was evident in the strain 10C11C-P relative to strain 10C11C-Z at 9-13 hours. Figures 15c and 15d show growth and β-galactosidase production curves of strains 79-Z (blank diamonds), 7C9C-Z (blank triangles), 7C9C-P (blank squares) and 79C-P (Xs), respectively. Growth curves obtained from all of the strains carrying inserts from context 79 are similar and parallel except at the highest cell density in strain 79C-P. Enzyme production curves of the zero strains 79-Z and 7C9C-Z show similar patterns, and enzyme concentrations are maximal at 14 and 16 hours with average 0-fusion activities of $1.2x10^{-2}\pm 3x10^{-3}$ and $1.6x10^{-2}\pm5x10^{-3}$ E. U. (calculated from 10-13 hours), respectively. As expected, enzyme production of the control +1 strain 79C-P which carries an insert with the +1 stop codon was observed at background level ($2.3x10^{-4}\pm3x10^{-5}$ E.U.) (Figure 15d). The activity level corresponds to the white colonies on the X-gal plates mentioned above (sector 6 in Figure 14b). Unlike strain 79C-P, the +1-fusion enzyme production in

strain 7C9C-P, which carries an insert without the +1 stop codon, is approximately 40 times higher in the exponential phase, and the curve shows a maximal level ($8.5x10^{-3}\pm6x10^{-4}$ E. U.) at 12 hours with an average +1-fusion activity of $7.8x10^{-3}\pm3x10^{-4}$ E. U. (calculated from 10-13 hours). The activity declined sharply to an average value of $2.7x10^{-3} \pm 7x10^{-4}$ E.U. (estimated from 14-17 hours), or about 10 times above the background. Approximately 66 and 20% of the frameshifting levels occurred in the exponential (7-12 hours) and stationary (14-17 hours) phases, respectively as calculated in strain 7C9C-P compared to strain 7C9C-Z.

4.3.4. Co-expression of reporters with pSJS 1240

The previous results mentioned above in Section 4.3.2 reveal a +1 frameshift site located between the nucleotide numbers 507-526 (507-A CGC TGG TTC **AGG** ACT CTC A-526) in region 79. The nucleotide sequence contains a rare Arg codon (bold AGG in the sequence shown above) as part of a consensus Shine-Dalgarno like sequence (underlined) (see also Figure 13b). The plasmid pSJS1240 (described in Section 2.3.9) containing genes coding for three tRNA genes ($tRNA^{Arg}_{CCU}$, $tRNA^{Arg}_{UCU,}$ and $tRNA^{Ile}_{UAU}$) that address the three important rare codons in *E. coli* including the AGG codon, was co-transformed into the strains 7C9C-P and 7C9C-Z to improve expression of normal translation (Kim *et al.,* 1998; Laksanalamai *et al.,* 2003; Laksanalamai *et al.,* 2001) on strain 7C9C-Z and to reduce frameshift translation on strain 7C9C-P. Strains named "7C9C-P+pSJS and "7C9C-Z+pSJS" were obtained (see the plasmid preparations observed in lanes 2 and 3 in Figure 16a).

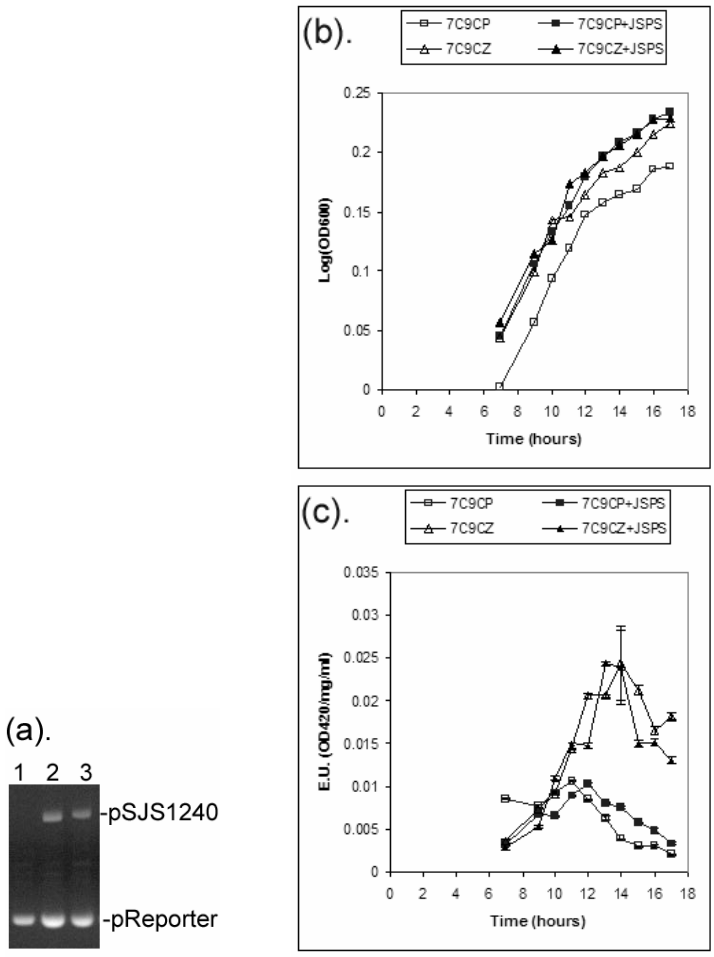


Figure 16. Growth and enzyme production curves in the presence of pSJS 1240.

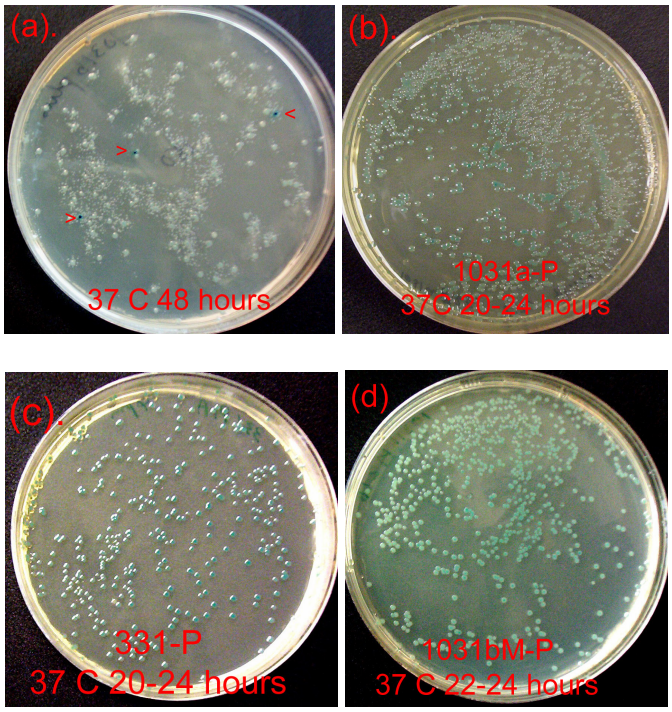
(a) Super-coiled plasmids isolated from strains 7C9C-Z (lane 1), 7C9C-P+pSJS (lane 2) and 7C9C-Z+pSJS (lane 3). The reporting and the pSJS 1240 plasmids are indicated on the right of the Figure. (b) Growth curves. (c) Enzyme production curves. Blank and solid squares represent data from the strains 7C9C-P and 7C9C-P+pSJS, respectively. Blank and solid triangles represent data from the strains 7C9C-Z and 7C9C-Z+pSJS, respectively. Strains containing only reporting plasmid were grown in LB+amp100 medium at 37 ^{o}C, 220 rpm. Strains containing both a reporting plasmid and plasmid pSJS 1240 were grown in LB+amp100+50 mg/ml spectinomycin at 37 ^{o}C, 220 rpm.

Figure 16a shows coexisting of pSJS 1240 and reporter plasmids in co-transformed strains 7C9C-P and 7C9C-Z. Figures 16b and 16c show growth and enzyme production curves of these strains, respectively. The growth and enzyme production curves are not significantly different (comparing strains 7C9C-P *versus* 7C9C-P+pSJS, and 7C9C-Z *versus* 7C9C-Z+pSJS in Figures 16b and 16c). In Figure 16c, enzymatic activities from the +1-reporting strains (7C9C-P and 7C9C-P+pSJS) were observed to be maximal ($1.0x10^{-2}\pm 2x10^{-4}$ E.U and $1.1x10^{-2} \pm 1x10^{-4}$ E.U, respectively) at 12 and 11 hours, respectively. These results indicate that co-expression of the pSJS 1240 does not increase the levels of +1- and 0-fusion β-galactosidase activities. These results also confirm that frameshifting was observed maximally during exponential phase. As a consequence, samples for determination of +1- and 0-fusion β-galactosidase activities in the following experiments were collected between 10-14 and 14-17 hours, respectively.

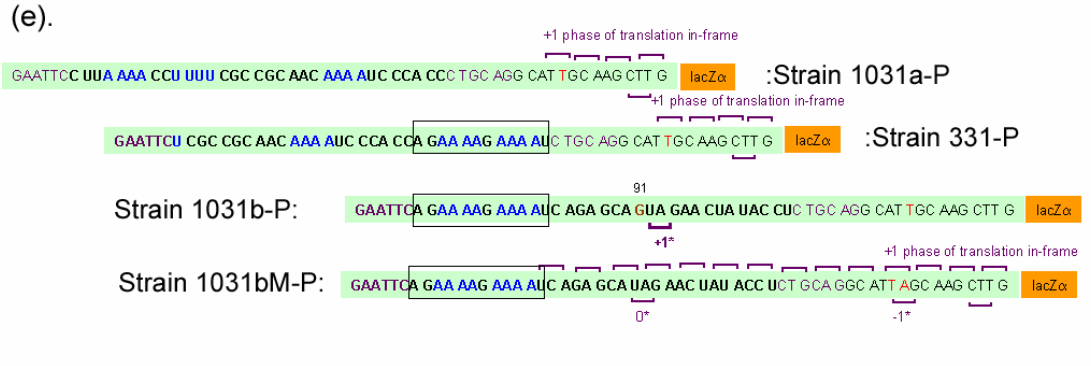
4.3.5. Genetic analyses of a +1 frameshift site in region 1011

Usually, pools of oligonucleotides obtained from chemical synthesis contain erroneous sequences that can be distinguished by cloning into appropriate reporters and sequencing to detect such errors. For example, Figure 17a shows three mutants (indicated by arrow heads), which carry a single nucleotide deletion on a sub-fragment 1031b (size 33 bp) compared with the majority of wild type sequence in the P reporter. The mutants (blue transformants) thus translated the fusion β-galactosidase in zero-frame. Synthetic error rate of 10^{-4} per base is estimated in this experiment.

Figure 17. Colonies on X-gal plates and nucleotide sequences of sub-region 1011. (a) Primary plates at 48 hours of white transformants containing sub-fragment 1031b in P reporter. Arrow heads indicate zero frame reporting strains carry a random single nucleotide deletion mutation in the inserts (made by oligo-synthesis errors). Satellite colonies observed on the plate were due to the growth of non-transformants after the depletion of the ampicillin antibiotic. (b-d) Secondary plates at 18-22 hours of light blue transformants were obtained after retransforming the *E. coli* Top10 with pure p1031a-P (b), p331-P (c) and p1031bM-P (d) plasmids, respectively. The p1031bM-P (derived from the M reporter) contains an insert with single deletion at guanine position 91 (d). (e) Nucleotide sequences of inserts and sequences between the *Pst-1* and *Hind III* restriction sites preceding the *lacZ*α ORF. The nucleotide sequences are shown using standard initials for triplet codes. Names of the strains are indicated on either the left or right hand side of the sequences. Red letters indicate the nucleotide (T and TA) markers between the *Pst-1* and *Hind III* restriction sites of the P and M reporters, respectively (see also Section 4.2.1). Purple letters indicate *Eco* RI and *Pst*-1 restriction sites. Blue letters indicate string of 4 nucleotides. Asterisks indicate stop codons. Facing downward staple pins indicate reading frame in +1 phase. The ^{91}G base is indicated by the number on top. Facing upward staple pins indicate zero reading frames. Orange rectangles indicate lacZα peptide ORF. Overlapped nucleotide sequences in the sub-fragments 331 and 1031b are boxed.


(a).
37 C 48 hours
(b).
1031a-P
37C 20-24 hours
(c).
331-P
37 C 20-24 hours
(d)
1031bM-P
37 C 22-24 hours

(e).


+1 phase of translation in-frame
GAATTCC UUA AAA CCU UUU CGC CGC AAC AAA AUC CCA CCC TGC AGG CAT TGC AAG CTT G
lacZα
:Strain 1031a-P
+1 phase of translation in-frame
GAATTCU CGC CGC AAC AAA AUC CCA CCA GAA AAG AAA AUC TGC AGG CAT TGC AAG CTT G
lacZα
:Strain 331-P
91
Strain 1031b-P:
GAATTCA GAA AAG AAA AUC AGA GCA GUA GAA CUA UAC CUC TGC AGG CAT TGC AAG CTT G
lacZα
+1*
+1 phase of translation in-frame
Strain 1031bM-P:
GAATTCA GAA AAG AAA AUC AGA GCA UAG AAC UAU ACC UCT GCA GGC ATT AGC AAG CTT G
lacZα
0*
-1*

The previous results (Section 4.3.2) indicated that context 1011 contains a +1 frameshift site(s). In order to identify a frameshift site, synthetic oligonucleotides (see also map in Figure 13a) named "1031a" (nt. nos. 39-71), "331" (nt. nos. 51-83) and "1031b" (nt. nos. 72-104) were tested in P and M reporters. Light blue transformants were obtained as the result of cloning experiments in the P reporter using the sub-fragments 1031a (Figure 17b) and 331 (Figure 17c). In contrast, white transformants (white colonies in Figure 17a) were obtained when the sub-fragment 1031b was cloned into the P reporter. Sequencing experiments confirm no mutation in the nucleotide sequences of the inserts or *lacZα* in the strains 1031a-P, 331-P and 1031b-P, respectively (Figure 17e).

The white phenotype observed in strain 1031b-P results from the presence of a +1 phase stop codon (92-TAG-94) (Figure 17e) located on the wild type fragment (1031b). When the sub-fragment 1031b was cloned into the M reporter, a few colonies with a light blue phenotype were observed on the X-gal screening plates. Sequencing experiment of the insert and *lacZα* from a mutant isolate (named "1031bM-P) reveals a deletion mutation at ^{91}guanine (Δ^{91}G) on the insert (Figure 17e) in the M reporter. Therefore, this strain is a P reporting strain that has an in-frame stop codon (91-TAG-94). Figure 17d shows the light blue phenotype of the +1-strain 1031bM-P on the X-gal plate. Average +1-fusion β-galactosidase activities of the strains 331-P ($3.0\text{x}10^{-3}$ $\pm 1\text{x}10^{-4}$ E.U.) and 1031bM-P ($2.7\text{x}10^{-3} \pm 7\text{x}10^{-4}$ E.U.) obtained from triplicate assays are not significantly different. Since the sub-fragments 331 and 1031b contain an overlapping region (nucleotide numbers 72-83) and both show equal strength of

frameshifts, they contain duplicate +1 frameshift sites as follows: “72- A GAA AAG AAA AU-83”. It should be noted here that the sequence 72-78 (72-A AAA AAG-78) in the *tnp-III* class is also a putative -1 frameshift motif as mentioned in Chapter 2 (see also pink highlighted in Figure 3).

4.3.6. Genetic analyses of a +1 frameshift site in region 79

Previously, a +1 frameshift site located between nucleotide numbers 507-526 was identified in fragment 79 (Section 4.3.2). Fragment 79 (size 66 bp) also contains a putative -1 frameshifting motif (522 T TTA AAA 528), a Shine-Dalgarno (SD)-like sequence (517-AGGA-520) and a secondary stem loop (nucleotide numbers 539 – 564). In addition, one +1 stop codon (524 UAA 526) and two -1 stop codons (552 UAA 554 and 564 UAG 566) are observed in this fragment (see also nucleotide sequences in Figure 13b).

Sub-fragments (size 24-33 bp), named “334”, “79SH”, “NOAGG”, “AGG”, “ThrA”, “ThrC” and “ThrG” (Figure 18a) that lack the stem loop and contain several types of mutations around the shift site were cloned into the P reporter. A +1-strain named “AGG-P” contains wild type codons (AGG ACU) decoding for ^{173}arginine (^{173}R) and ^{174}threonine (^{174}T), thus it was employed as a reference strain. The sequence AGG encoding the ^{173}R and the adenine from the first position of the ^{174}T codon is marked as a Shine-Dalgarno (SD)-like sequence (boxed in Figure 18a). A +1-strain named “ThrG-P” contains a mutated base at the third position of the codon decoding for the ^{174}T. A +1-strain named “NOAGG-P” contains an insert of 33 bp.

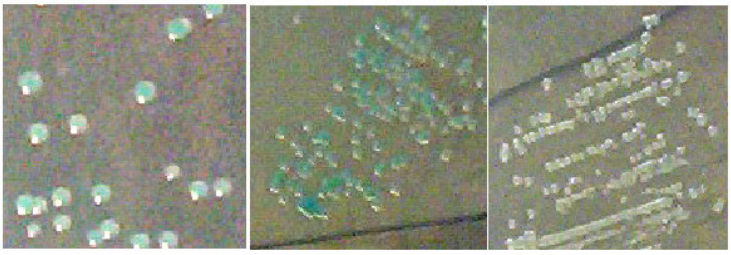
Figure 18. Alignments of mRNA sequences of sub-fragments from region 79, colonies on X-gal plate and ratios of enzyme activity. (a) Alignment of mRNA that lacks the stem loop. Boxes indicate Shine-Dalgarno (SD)-like sequences. A purple staple symbol with an asterisk represents the +1 stop codon (UAA). Sequences are typed in triplet codes corresponding to the codons in *tnp-I*. Orange staple symbol indicates +1-phase reading frames. Green letters represent nucleotides that belong to part of the putative -1 frameshift heptamer. Red letters indicate mutations. Underlined letters indicate the +1 frameshift site identified (see detail in text). A purple arrow represents partial nucleotide sequence that belongs to a partial sequence of the stem loop (see also two purple arrows in Figure 13b). Amino acid sequence is shown using the standard codon table in a blue box.

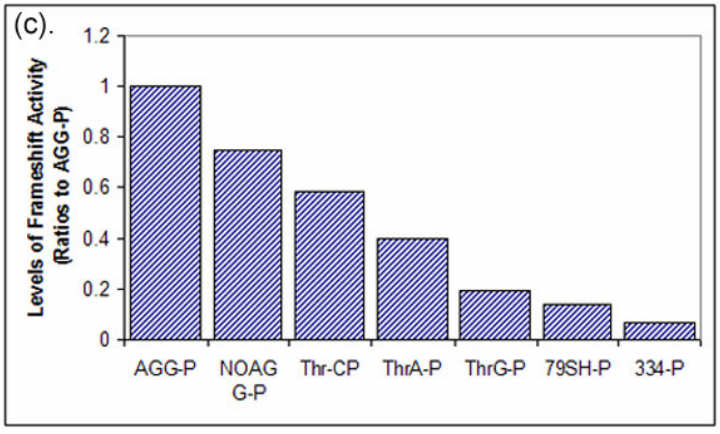
(b) Colonies of +1-strains AGG-P (left), NOAGG-P (center) and ThrG-P (right) on X-gal plates at 20-22 hours.

(c) A histogram comparing ratios of average +1-fusion β-galactosidase activities to average activity ($1.2x10^{-2}\pm3x10^{-4}$ E.U) of the reference strain, AGG-P. The suffix P indicates P reporter. The X axis represents strain names. The results indicate that the codons ^{174}T and ^{175}L (either ACT CTC or ACT TTA) are essential for frameshifting.


(a).
Sub-fragments:
334:
NOAGG:
ThrA:
ThrC:
ThrG:
AGG:
AGA:
CGG:
CGA:
79SH:
SD
+1*
516-C AGG ACU UUA AAA GAA AGA ACA AAG CGU UUC UG-548
519- G ACU CUC AAA GAA AGA ACA AAG CGU UUC UGG AA-551
507 A CGC UGG UUC AGG ACA CUC AAA GAA AGA ACA AA 539
507 A CGC UGG UUC AGG ACC CUC AAA GAA AGA ACA AA 539
507 A CGC UGG UUC AGG ACG CUC AAA GAA AGA ACA AA 539
507 A CGC UGG UUC AGG ACU CUC AAA GAA AGA ACA AA 539
507 A CGC UGG UUC AGA ACU CUC AAA GAA AGA ACA AA 539
507 A CGC UGG UUC CGG ACU CUC AAA GAA AGA ACA AA 539
507 A CGC UGG UUC CGA ACU CUC AAA GAA AGA ACA AA 539
507 A CGC UGG UUC AGG ACU [- - -
(Δ523-564)
- - -] CGCGACUG-572
R W F R T L A E R T R R F R D
Amino acid nos.: 170 171 172 173 174 175 176 177 178 179 180 181 182 ------ 189 190

(b).


AGG-P
NOAGG_P
ThrG-P


(c).
Levels of Frameshift Activity (Ratios to AGG-P)
1.2
1
0.8
0.6
0.4
0.2
0
AGG-P
NOAG G-P
Thr-CP
ThrA-P
ThrG-P
79SH-P
334-P

Its nucleotide sequence begins with the third base (519) of the ^{173}R codon to nucleotide number 551. Therefore, the strain NOAGG-P lacks the SD-like sequence as well as the stem loop. Strain 79SH-P (Δ523-564) carries a 24 bp long sub-fragment that lacks a putative frameshift site and downstream sequence.

Figure 18b show colonies on X-gal plates of the strains AGG-P, NOAGG-P, and ThrG-P from the left to right, respectively. Screening results on the X-gal plates show that mutations at the third position of the ^{174}T codon significantly decrease the intensity of the blue phenotype in the strain ThrG-P. The results in Figure 18c show decreases in expression ratios observed in the strains NOAGG-P (1.25 fold), ThrC (1.7 fold), ThrA-P (2.5 fold), ThrG-P (5.2 fold), 79SH-P (7.2 fold), and 334-P (15 fold). As expected, the +1-strain 334-P (where the insert includes a +1 stop codon) shows no detectable activity (ca. 10^{-4} E.U.). The +1-strain NOAGG-P (the insert includes codons 174 and 175, but not the codon 173) shows no significant decrease compared to the reference strain. The strains ThrA-P, ThrC-P and ThrG-P which contain a mutated base at the third position of the ^{174}T codon show significantly decreased activities ($4.8x10^{-3}\pm2x10^{-4}$, $7.1x10^{-3}\pm1x10^{-4}$ and $2.2x10^{-3}\pm2x10^{-4}$ E.U., respectively). In particular the guanine base in the third position of the ^{174}T codon (ThrG-P) strongly inhibits expression. Although the +1-strain 79SH-P (deletion Δ523-546 including the codon ^{175}L) shows a low level of frameshifting ($1.7x10^{-3}\pm3x10^{-4}$ E.U.), it was significantly decreased compared with the wild type (reference sequence). The results imply that the codons ^{174}T and ^{175}L are essential for frameshifting, and I propose that this is the probable frameshift site. The mRNA sequence coding for the amino acids ^{174}T and ^{175}L

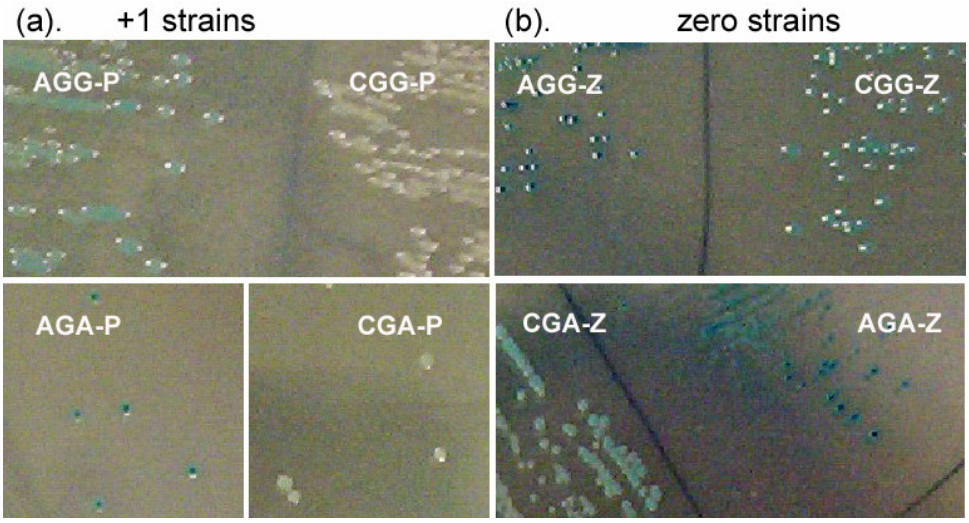
(ACU CUC) is evident as a probable major frameshift site in context 79. However, it is possible that modifications made on the TL site might create new artificial shift site downstream. Although the +1-strains containing wild type codons of the amino acids ^{174}T, ^{175}L, and ^{176}K (ACU UUA AAA in fragments 79 and 79C and sub-fragment 334) (see nucleotide sequences in Figures 13b and 18a, respectively) show white phenotypes and background levels of β-galactosidase activity, the ACU UUA wild type motif may also be a shift site. This is because the motif is difficult to test due to the termination of the frameshifting product by the UAA stop codon in the +1 phase (underlined). The low frameshift frequency (a decrease of 7.2 fold) of strain 79SH-P (deletion of the ^{175}T as well as internal downstream sequence) implies that the downstream mRNA sequence might also play a role. On the other hand, the result can not rule out a second shift site that might be located in the deleted sequence.

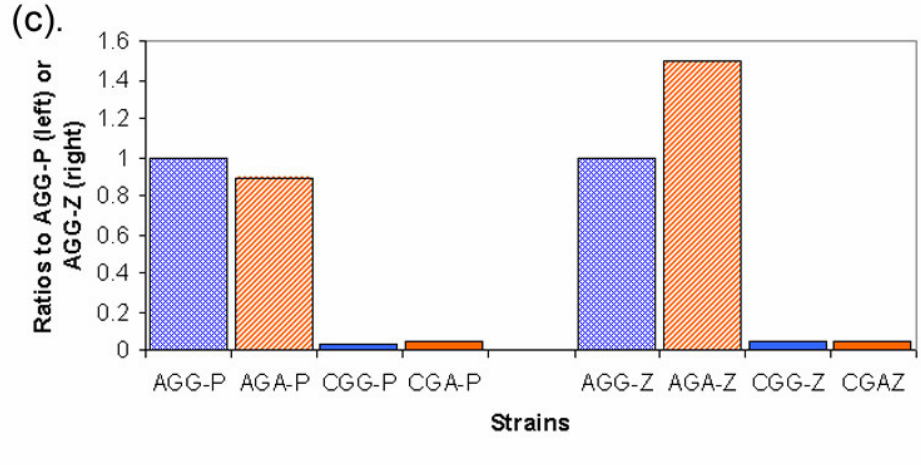
It worth noting here that a strain (named "ThrG-Z") carrying the sub-fragment ThrG in the Z reporter shows significant decreased zero-fusion β-galactosidase activity ($3.5x10^{-2}\pm5x10^{-3}$). The activity is approximately 3.8 times lower compared with the strain named "AGG-Z" that carries the reference sequence. The result implies that alteration at the third base, namely ^{174}T to G, has the strongest effect on frameshifting by decreasing zero frame translation.

4.3.7. Shine-Dalgarno-like sequences ensure frameshifting through increased translation

The consensus *E. coli* Shine-Dalgarno (SD)-like sequence adjacent to the +1 frameshift site consists of a rare arginine codon (AGG for ^{173}R) and the first nucleotide (adenine) of the ^{174}T (boxed in Figure 18a). As discussed previously in Section 4.3.4, co-expression of the pSJS 1240 does not inhibit frameshifting. Mutations made on the first and third positions of the codon encoding ^{173}R resulted in the following sub-fragments, named "CGG", "CGA", "CGT", "CGC" and "AGA", respectively. Nucleotide sequences of the sub-fragments CGG, CGA and AGA are shown in Figure 19a. The sub-fragments were employed to test the ability of ribosomes to translate these sequences in response to a ribosome binding site (rbs) upstream (1 base spacing) of the frameshift site. Screening results obtained in the P and Z reporting strains growing on X gal plates are shown in Figures 19a and 19b, respectively. The results reveal that the +1-strains named "AGG-P" and "AGA-P" (top and bottom left in Figure 19a, respectively) which carry the rbs in their inserts as mentioned above, are able to produce blue phenotypes. In contrast to the strains AGG-P and AGA-P, the +1-strains named "CGG-P" and "CGA-P" (top and bottom right in Figure 19a, respectively), and CGT-P and CGC-P (photographs are not shown) which lack the rbs in their inserts have white phenotypes on the X-gal plates. Ratios of β-galactosidase activity are shown in the left panel of Figure 19c.

Figures 19. Colonies of +1 and zero reporting strains on X-gal plates at 37 ^{o}C, 22-24 hours. (a) Photographs of colonies of +1 reporting strains named "AGG-P", "CGG-P", "CGA-P" and "AGA-P" (indicated on the pictures) show blue, white, white and blue colonies, respectively. (b) Photographs of colonies of zero reporting strains named "AGG-Z, CGG-Z, CGA-Z and AGA-Z (indicated on the pictures) show dark blue, light blue, dark blue and light blue, respectively. (c) Histograms comparing ratios of β-galactosidase activities to activities of corresponding reference strains (AGG-P and AGG-Z). Ratios of +1 (left) and zero fusion β-galactosidase were calculated using average activities $1.2x10^{-2}\pm3x10^{-4}$ and $1.4x10^{-1}\pm5x10^{-3}$ E.U. of the strains AGG-P and AGG-Z, respectively. Dotted (blue) and striped (orange) bars indicate sub-fragments AGG (containing an AGGA wild type rbs) and AGA (containing an AGAA mutant rbs). Solid blue and orange bars represent sub-fragments CGG (containing the sequence CGGA, i.e. lacking the rbs sequence) and CGA (containing the sequence CGAA, also lacking a rbs sequence), respectively.


(a). +1 strains
(b). zero strains
AGG-P
CGG-P
AGG-Z
CGG-Z
AGA-P
CGA-P
CGA-Z
AGA-Z


(c).
Ratios to AGG-P (left) or AGG-Z (right)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
AGG-P
AGA-P
CGG-P
CGA-P
AGG-Z
AGA-Z
CGG-Z
CGAZ
Strains

The results reveal no significant difference in the ratio of β-galactosidase activity obtained from the +1-strain AGA-P (ratio = 0.9), and confirm that significantly decreased ratios of β-galactosidase activity were obtained from the +1-strains CGG-P (28 fold), and CGA-P (23 fold) relative to the reference +1-strain AGG-P. These results imply that either the consensus rbs (AGGA) or the divergent rbs (AGAA) can play a role as a stimulator of the frameshift process.

Similar ratios were obtained from the zero-reporting strains AGA-Z (increased 1.5 fold), CGG-Z (decreased 21 fold), and CGA-Z (decreased 24 fold) relative to the reference zero-strain AGG-Z (right panel in Figure 19c, comparing color intensities of colonies in Figure 19b). The results obtained are opposite to our expected results in the event that pausing is due to rare codons (AGG and AGA) in *E. coli*. These results thus reveal an additional role of the rbs in ensuring a high rate of zero-frame translation such that high frequency frameshifting follows.

4.3.8. Proposed -2 frameshift mechanism in context 79

The findings that cytosine and guanine bases in the first and third positions of the codons ^{173}R and ^{174}T inhibit both +1 and zero translation (mentioned in the previous Section 4.3.7) imply that the molecular mechanism may involve interactions among the nucleotide sequences of the mRNA (upstream and down stream), ribosome, and/or anti-codons of the peptidyl- and aminoacyl-tRNAs. These interactions thus provide forces to pause the translating ribosomes. Assuming that the strength of the net force and direction that are generated depend on the spacing of the rbs upstream and the sequence of the mRNA downstream, both factors determine the duration of the

ribosomal pause. The pause duration may thus affect the translocation step of the aminoacyl- and/or peptidyl tRNAs such that disruption of some of the ternary complexes (ribosome, tRNA and mRNA) results, and a shift of the reading frame occurs. The level of frameshift product observed is the result of the ability of the frameshift site to stabilize the shifted frame peptidyl- and/or aminoacyl-tRNAs such that translation can go on in a shifted phase. In the case of context 79, inhibition of zero-frame translation observed in the strains CGG-Z and CGA-Z (mentioned in the previous Section) supports the hypothesis that disruption of the translating ternary complexes precludes frameshifting. A model of frameshifting observed in the context studied and the ACT CTC motif fit a hypothesis that "frameshifting involves a force that retards the ribosome disrupting nucleotide pairing between mRNA and the anti-codon of t-RNA", if the frameshifting occurs using a minus 2 mechanism after the leucine-t-RNA$^{leu}_{GAG}$ has entered the A site of the ribosome. A cartoon explaining the proposed -2 mechanism is shown in Figure 20.

Figure 20. Proposed novel minus two (-2) frameshifting mechanism. The left cartoon shows pairing of 5'AGU 3' and 5'GAG 3' anti codons of peptidyl-tRNA^{thr} and acyl-t-RNA^{leu} with their complementary codons at P and A sites, respectively. The right cartoon shows pairing of the anticodon (5'GAG3') after translocation of the peptidyl-tRNA to the P site with the second and third bases (5'CU 3') of the ^{174}T resulting in a 2 nucleotide backward slippage and exposing a serine codon (5'UCA3') instead of a lysine codon (5'AAA3'). Gray rectangles represent a ribosome. Letters P and A in white boxes represent P and A sites on the ribosome, respectively. An inversed letter L (⌋) with an anticodon represents a tRNA. Dots represent hydrogen bonds.

4.4. Discussion and conclusions

Two programmed translational frameshift contexts (66 bp-long, representing regions 1011 and 79) were identified in the *tnp-I* sequence. Genetic analyses using translational fusions to *lacZα* in *E. coli* indicated that the first context in region 1011 contains a +1 frameshift site narrowed down to nucleotide number 72-83 (5'-A GAA AAG AAA AU-3') as evident from strains 1031bM-P and 331-P (Section 4.3.5). However, details of the shift site and mechanism in this context 1011 were not addressed in this study. The frameshifts in strain 1031a-P (Figure 17c) indicate that this context may contain more than one frameshift sites. Moderate frameshift frequency (ca. 20-30%) of the 66 bp-long context 1011 was estimated relative to the 10C11C-Z zero.

The second context in region 79 is a strong (ca. 50% level of frameshifting) programmed translational +1 frameshifting context evident at the ^{174}T and ^{175}L codons (520-ACT TTA-525 or the 520-ACT CTC-525). This is supported by high frameshift frequencies observed in the strain NOAGG-P and no detectable frameshift product observed in strain 334-P (due to the +1 stop codon in the sub-fragment). In addition, strains ThrA-P, ThrC-P and ThrG-P that carried fragments containing mutation made at the third base of ^{174}T show unpredictable frequency. Although the wild type nucleotide sequence coding for the ^{174}T ^{175}L amino acids (520-ACT TTA-525) in the strains 334-P and 79C-P (comprises a +1 stop codon as part of the shift site) show no detectable frameshift frequency, the context is difficult to test for frameshifting that may occur through the upstream sequence due to the presence of the +1 stop codon,

keeping in mind that mutations made at the shift site might introduce new shift sites into the context. Although strain NOAGG-P which lacks the sequence upstream also support the notion that ^{174}T ^{175}L is the shift site, a new shift site might be created at downstream due to the change of the context.

The putative stem loop evidence in the 66 nucleotide long of the context 79 may not be involve in the frameshifting since all sub-fragments (33 bp-long) that lack the putative stem loop can cause frameshift frequencies as high as frameshift frequency that was observed in the full context (context 79). However, the mRNA sequence downstream may play a role in frameshifting frequency since a dramatic decrease in frameshifting frequency was evident in the strain 79SH-P (Δ523-564). Again deletion of the nucleotide sequence (Δ523-564) made in the sub-fragment 79SH may remove other stimulators, and as yet unidentified enhancing factors.

Mutations made on the sub-fragment AGG to create sub-fragments AGA, CGG, CGA, CGT and CGG thus represent two parameters tested in Section 4.3.7, either SD-like sequence or rare codon usage in *E. coli* in sub-fragments AGG and AGA, and lack of SD-like sequence or normal codon usage in *E. coli* in sub-fragments CGG, CGA, CGT and CGG, respectively. The strains AGG-P and AGA-P as well as AGG-Z and AGA-Z showed no significantly different ratios of activity (Figure 19). In contrast the strains CGG-P and CGA-P as well as CGG-Z and CGA-Z revealed dramatically decreased ratios of activity (Figure 19). The results imply rare codon usage in *E. coli* parameter may not be the cause. On the other hand, the experiment tested for effects of SD-like sequences (or ribosome binding sites). The results revealed

that frameshifting frequency corresponds to normal translation frequency. Thus the result implies that the ribosome binding site (rbs) upstream of the ^{174}T ^{175}L codon shift site may enhance frameshift frequency through increasing normal translation frequency. This correlation was also observed in the strains ThrG-P (decrease 5.2 fold to AGG-P) and ThrG-Z (decrease 3.8 fold to AGG-Z). It worth noting that a correlation of frameshifting and growth phase (Figures 15c and 15d, and 16b and 16c) was observed that frameshifting declined when cells were in stationary phase. The result implies that other unknown intrinsic factors in *E. coli* may also play roles.

The model -2 frameshifting mechanisms (Figure 20) proposed based on ACT CTC codons as shift site best fit the results obtained, assuming that interactions occur between m RNA sequence and ribosome. Net forces that might be generated from the interaction between the rbs upstream and the mRNA sequence downstream may be important for pausing. The model also portrays the frameshift site as playing a role as complex stabilizer by forming codon anticodon pairings and thus protecting the complex from disruption and thus allowing translation to continue in shifted phase.

In conclusion, the *tnp-I* sequence from the hyperthermophilic IS elements contains many sites that are slippery. Two frameshift sites were demonstrated in the sequences named as context 1011 and 79. The sequence appears to be difficult to translate due to premature termination of the nascent peptide by frameshifting mechanisms. The results support the second hypothesis of this study (mentioned in Section 1.8 in Chapter 1) that "programmed translational frameshift control mechanisms in the *tnp* gene down regulate expression of the IS elements." Bearing in

mind that these studies were done in *E. coli*, and not in *P. furiosus*, the experiments in the strains containing the plasmid pSJS 1240 are significant. Frameshifting was not altered in these strains, suggesting that the mechanism was not artifact of codon usage discrepancies involving arginine codons.

Chapter 5

Conclusions and research impact

5.1. Conclusions

The genomes of *Pyrococcus furiosus* and *Pyrococcus woesei* both contain IS elements that were categorized into 3 types according to their nucleotide sequences. Specific primers to differentiate types of IS were developed in this study. Co-linearity between the genomes of *P. woesei* and *P. furiosus* was demonstrated using Southern blots and radio-labeled IS element sequence probes. A recent transposition event that disrupted the *napA* gene of *P. woesei* was revealed. This finding may relate to the phenotypic differences between these two species observed during growth in S^o-free medium containing maltose and sodium tungstate demonstrated in Chapter 3. The presence of a *P. woesei* IS element type I in the exact context as observed in the genome of *P. furiosus* was demonstrated. These results indicate that the IS elements are actively participating in genome rearrangements and reshuffling in these two closely related strains.

The IS open reading frames are transcribed as revealed by RT-PCR. The IS elements have putative promoters that may provide internal transcription from a putative TATA box "boxA" observed in this study. In addition, the IS elements may use programmed translational frameshifts to limit the expression of the full length transposase. The nucleotide sequences of the *tnp* gene contain at least two strong

programmed translational frameshift sites identified in Chapter 4 using *lacZα* fusions in *E. coli* as a probe for translation anomalies.

Interestingly, the -2 mechanism of frameshifting evidenced in context 79 in the *tnp* from hyperthermophilic IS element type I is a novel mechanism in that the shift site is part of a -1 frameshifting motif. This study also addresses the role of a ribosome binding site (rbs) with a short spacer (1 bp) upstream observed in context 79 which may ensure high frameshift levels by increasing normal frame translation. The role of the rbs in frameshifting suggested in this study may be linked to the finding that maximal levels of frameshift products were usually obtained during exponential growth phase. However, other as yet unknown intrinsic factors in *E. coli* may also play a role.

In conclusion, this study reveals that the IS elements are active and that they may be subject to translational regulation of transposition.

5.2. Research impact

To date, techniques for the genetic manipulation and mutation of *P. furiosus* have not been effective. A *napA* gene disrupted by an IS element in the two closely related strains, *P. furiosus* and *P. woesei*, revealed in this study provides a good model for comparative studies of function of the putative *napA* gene and of the membranes from hyperthermophilic Archaea. Due in part to the fact that *Pyrococcus* strains lack good selective markers, the phenotypes (low cell density and cell clumping) observed in the strain with the mutated *napA* gene may also be effective for negative selection. The active IS elements identified in the genomes of *P. furiosus* and *P. woesei* have

simple structures and show potential specificity to insert certain sequences, thus they may have the potential for developing new genetic tools. The specific primers developed in this study will also provide a simple PCR technique to detect and map mutations. Unlike the genus *Sulfolobus*, which are hyperthermophilic Crenarchaeota that contain multiple families of insertion sequences and viruses, the genus *Pyrococcus* contains species such as *P. horikoshii* and *P. abyssi* that contain no IS elements. This study revealed IS elements in limited numbers and diversity in the two species isolated from Vulcano Island. In addition, the genomes of *P. horikoshii* and *P. abyssi* lack the amylase genes that have been used as genetic markers in *Sulfolobus solfataricus* (Worthington *et al.,* 2003). Consequently *P. horikoshii* and *P. abyssi* may be potential hosts for IS elements with selectable markers such as maltose or amylase genes. In addition, these comparative studies supported the following proposal:- "To reclassify *P. woesei* as *P. furiosus* sub sp. *woesei* based on their identical rDNA operon sequences, highly sequence identity and shared IS element markers."

Appendix A

A list of primers for PTF screening (see Section 4.2.2).

Primers	Nucleotide Sequences (5' → 3')	Binding Coordi-nates
1Ifw	AAAAAA*GAATTC*GAAGGCTGAGAGCAT	2-16
2Irc	AAAAAA*CTGCAG*GATAACTGGGG	685-698
3Irc	AAAAAA*CTGCAG*ACTGTTGTGTGGC	149-164
4Ifw	AAAAAA*GAATTC*CTGCAAAGGACA	391-403
5Ifw	AAAAAA*GAATTC*AGTCTGGGAGGCA	158-171
6Irc	AAAAAA*CTGCAG*TCTCTAGCAATCCACC	353-370
7Ifw	AAAAAA*GAATTC*ACGCTGGTTCAGG	507-519
7CIfw	AAAAAA*GAATTC*ACGCTGGTTCAGGA<u>C</u>T<u>C</u>TCAAAG	507-529
8Irc	AAAAAA*CTGCAG*AGGGATTTAAAAGC	445-458
9Irc	AAAAAA*CTGCAG* CAGTCTCTAGCCCTG	558-572
9CIrc	AAAAAA*CTGCAG*CAGTC<u>G</u>C<u>G</u>AGCCCTGAAATT<u>G</u>TTCCA	547-572
10Ifw	AAAAAA*GAATTC*CTTAAAACCTTTTCGCC	39-55
10CIfw	AAAAAA*GAATTC*C<u>C</u>T<u>C</u>AAACCTTTTCG	39-53
11Irc	AAAAAA*CTGCAG*AGGTATAGTTCTACTGCT	86-104
11CIrc	AAAAAA*CTGCAG*AGGTATAGTTC<u>G</u>ACTGCT	86-104

Restriction sites and mutations made on primers are shown in italic and underlined letters, respectively.

Appendix B

A list of oligomeric pairs matching the region 1011 (see Section 4.2.3 and Figure 13a).

Oligomers	Nucleotide Sequences (5' → 3')
331-Up	aattcTCGCCGCAACAAAATCCCACCAGAAAAGAAAATctgca
331-Lo	gATTTTCTTTTCTGGTGGGATTTTGTTGCGGCGAg
1031-Up	aattcCTTAAAACCTTTTCGCCGCAACAAAATCCCACCctgca
1031-Lo	gGGTGGGATTTTGTTGCGGCGAAAAGGTTTTAAGg
1031b-Up	aattcAGAAAAGAAAATCAGAGCAGTAGAACTATACCTctgca
1031b-Lo	gAGGTATAGTTCTACTGCTCTGATTTTCTTTTCTg

Lower case letters represent sticky restriction sites.

Appendix C

A list of oligomeric pairs matching the region 79 (see Section 4.2.3 and Figure 13b).

Oligomers	Nucleotide Sequences (5' → 3')
334-Up 334-Lo	aattcC**AGGA**CTTTAAAAGAAAGAACAAAGCGTTTCTGctgca gCAGAAACGCTTTGTTCTTTCTTTTAAAGTCCTGg
AGG-Up AGG-Lo	aattcACGCTGGTTC**AGGA**CTCTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGAGTCCTGAACCAGCGTg
AGA-Up AGA-Lo	aattcACGCTGGTTC**AGAA**CTCTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGAGTTCTGAACCAGCGTg
CGG-Up CGG-LO	aattcACGCTGGTTCCGGACTCTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGAGTCCGGAACCAGCGTg
CGA-Up CGA-Lo	aattcACGCTGGTTCCGACTCTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGAGTTCGGAACCAGCGTg
CGC-Up CGG-LO	aattcACGCTGGTTCCGCACTCTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGAGTGCGGAACCAGCGTg
CGT-Up CGT-LO	aattcACGCTGGTTCCGTACTCTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGAGTGCGGAACCAGCGTg
ThrA-Up ThrA-Lo	aattcACGCTGGTTC**AGGA**CACTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGTGTCCTGAACCAGCGTg
ThrC-Up ThrC-Lo	aattcACGCTGGTTC**AGGA**CCCTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGGGTCCTGAACCAGCGTg
ThrG-Up ThrG-Lo	aattcACGCTGGTTC**AGGA**CGCTCAAAGAAAGAACAAActgca gTTTGTTCTTTCTTTGAGCGTCCTGAACCAGCGTg
NOAGG-Up NOAGG-Lo	aattcGACTCTCAAAGAAAGAACAAAGCGTTTCTGGAActgca gTTCCAGAAACGCTTTGTTCTTTCTTTGAGAGTCg

79SH-Up	aattcA CGC TGG TTC AGG GCTAGA GAC TGctgca
79SH-Lo	gCAGTC TCT AGC CCT GAACCA GCGTg

Mutations made on the sub-fragments are underlined only in the upper strand (suffix Up). Lower case letters represent sticky restriction ends. Bold letters indicate SD-like sequences.

References

1. Adams, M. W.. 1994. Biochemical diversity among sulfur-dependent, hyperthermophilic microorganisms. *FEMS Microbiol Rev*. 15(2-3):261-77.

2. Ahmed, A.. 1986. Evidence for replicative transposition of Tn*5* and Tn*9*. *J Mol Biol.* 191(1):75-84.

3. Altschul, S. F., Gish, W., Miller, W., Mayers, E. W. and Lipman, D. J.. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-10.

4. Atkins, J. F., Weiss, R. B. and Gesteland, R. F.. 1990. Ribosome gymnastics degree of difficulty 9.5, style 10.0. *Cell.* 62(3):413-23.

5. Atkins, J. F., Weiss, R. B., Thompson, S. and Gesteland, R. F.. 1991. Towards a genetic dissection of the basis of triplet decoding, and its natural subversion: programmed reading frame shifts and hops. *Annu Rev Genet*. 25:201-28.

6. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K.. 1987. *In* Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley-Interscience, John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore.

7. Baker, T. A. and Luo, L.. 1994. Identification of residues in the Mu transposase essential for catalysis. *Proc Natl Acad Sci USA*. 91(14):6654-8.

8. Baranov, P. V., Gurvich, O.L., Hammer, A. W., Gesteland, R.F. and Atkins, J.F.. 2003. RECODE. *Nucleic Acids Res.* 31:87-9.

9. Barth, P. T., Grinter, N. J. and Bradley, D. E.. 1978. Conjugal transfer system of plasmid RP4: analysis by transposon 7 insertion. *J Bacteriol*. 133(1):43-52.

10. Bauer, M. W., Driskill, L. E., Callen, W., Snead, M. A., Mathur, E. J. and Kelly, R. M.. 1999. An endoglucanase, *EglA*, from the hyperthermophilic archaeon *Pyrococcus furiosus* hydrolyzes beta-1,4 bonds in mixed-linkage (1→3),(1→4)-beta-D-glucans and cellulose. *J. Bacteriol*. 181:284-290.

11. Baumann, P., Qureshi, S. A. and Jackson, S.P.. 1995. Transcription: new insights from studies on Archaea. *Trends Genet*. 11(7):279-83.

12. Belcourt, M. F. and Farabaugh, P. J.. 1990. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell.* 62(2):339-52.

13. Bell, S. D. and Jackson, S. P.. 1998. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol.* 6(6):222-8.

14. Bell, S. D., Kosa, P. L., Sigler, P. B., and Jackson, S. P.. 1999. Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A*. 96(24):13662-7.

15. Bhasin, A., Goryshin, I. Y. and Reznikoff, W. S.. 1999. Hairpin formation in Tn*5* transposition. *J Biol Chem.* 274(52):37021-9.

16. Bhasin, A., Goryshin, I. Y., Steiniger-White. M., York, D. and Reznikoff, W. S.. 2000. Characterization of a Tn*5* pre-cleavage synaptic complex. *J Mol Biol.* 302(1):49-63.

17. Blamey, J., Chiong, M., Lopez, C. and Smith, E.. 1999. Optimization of the growth conditions of the extremely thermophilic microorganisms *Thermococcus celer* and *Pyrococcus woesei*. *J Microbiol Methods*. 38(1-2):169-75.

18. Blinkowa, A. L. and Walker, J. R.. 1990. Programmed ribosomal frameshifting generates the *Escherichia coli* DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucleic Acids Res.* 18: 1725-9.

19. Boeke, J. D.. 1989. Transposable Elements in *Saccharomyces cerevisiae*. *In* Mobile DNA. Eds. D. E. Berg and M. M. Howe. American Society of Microbiology, Washington, D.C. pp. 335-374.

20. Braam, L. A. and Reznikoff, W. S. 1998. Functional characterization of the Tn*5* transposase by limited proteolysis. *J Biol Chem.* 273(18):10908-13.

21. Bradford, M. M.. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 72:248-54.

22. Brierley, I., Jenner, A. J. and Inglis, S. C.. 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol.* 227:463-79.

23. Brugger, K., Redder, P., She, Q., Confalonieri, F., Zivanovic, Y. and Garrett, R. A.. 2002. Mobile elements in archaeal genomes. FEMS *Microbiol Lett.* 206(2):131-41.

24. Bryant, F. O. and Adams, M. W..1989. Characterization of hydrogenase from the hyperthermophilic archaebacterium, *Pyrococcus furiosus*. *J. Biol. Chem.* 264:5070-79.

25. Bukrinsky, M. I. and Haffar, O. K.. 1999. HIV-1 nuclear import: in search of a leader. *Front Biosci*. 4:D772-81.

26. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. and Venter, J. C.. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science.* 273(5278):1058-73.

27. Chaconas, G. and Surette, M. G.. 1988. Mechanism of Mu DNA transposition. *Bioessays.* 9(6):205-8.

28. Chakrabarti, A., Desai, P. and Wickstrom, E.. 2004. Transposon Tn*7* protein TnsD binding to *Escherichia coli att*Tn*7* DNA and its eukaryotic orthologs. *Biochemistry.* 43(10):2941-6.

29. Chamorro, M., Parkin, N. and Varmus, H. E.. 1992. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc Natl Acad Sci U S A.* 89:713-7.

30. Chandler, M. and Fayet, O.. 1993. Translational frameshifting in the control of transposition in bacteria. *Mol Microbiol.* 7(4):497-503.

31. Charbonnier, F. and Forterre, P.. 1995. Purification of plasmids from thermophilic and hyperthermophilic archaea. *In* Archaea: A Laboratory Manual –Thermophiles. Eds. F.T. Robb and A.R. Place. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY. pp. 87-90.

32. Charlebois, R.L. and Doolittle, W. F.. 1989. Transposable elements and genome structure in Halobacteria. *In* Mobile DNA. Eds. D. E. Berg and M.M. Howe. American Society of Microbiology, Washington, D.C. pp 297-307.

33. Charlier, D., Piette, J. and Glansdorff, N.. 1982. IS*3* can function as a mobile promoter in *E. coli*. *Nucleic Acids Res*. 10(19):5935-48.

34. Clare, J. J., Belcourt, M. and Farabaugh, P. J.. 1988. Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon. *Proc Natl Acad Sci U S A*. 85:6816-20.

35. Cobucci-Ponzano, B., Trincone, A., Giordano, A., Rossi, M. and Moracci, M.. 2003. Identification of an archaeal α-L-fucosidase encoded by an interrupted gene. Production of a functional enzyme by mutations mimicking programmed -1 frameshifting. *J Biol Chem*. 278:14622-31.

36. Craig, N. L.. 1989. Transposon T*7*. *In* Mobile DNA. Eds. D. E. Berg and M. M. Howe. American Society of Microbiology, Washington, D.C. pp. 211-225.

37. Craigie, R. and Mizuuchi, K.. 1987. Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell.* 51(3):493-501.

38. Craigen, W. J., Caskey, C. T.. 1986. Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature*. 322: 273-5.

39. Craigen, W. J., Cook, R. G., Tate, W.P. and Caskey, C. T.. 1985. Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *Proc Natl Acad Sci U S A.* 82(11):3616-20.

40. Curran, J. F.. 1995. Decoding with the A:I wobble pair is inefficient. *Nucleic Acids Res.* 23:683-8.

41. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S. and Rayment, I.. 2000. Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate. *Science.* 289(5476):77-85.

42. Derbyshire, K. M. and Grindley, N. D.. 1996. *Cis* preference of the IS*903* transposase is mediated by a combination of transposase instability and inefficient translation. *Mol Microbiol*. 21(6):1261-72.

43. Dinman, J. D, Ruiz-Echevarria, M. J., Czaplinski, K. and Peltz, S. W.. 1997. Peptidyl-transferase inhibitors have antiviral properties by altering programmed -1 ribosomal frameshifting efficiencies: development of model systems. *Proc Natl Acad Sci U S A.* 94(13):6606-11.

44. DiRuggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J. Horlacher, R., Robb, F.T., Boos, W. and Weiss, R.B.. 2000. Evidence of recent lateral gene transfer among Hyperthermophilic Archaea. *Mol. Microbiol.* 38:684-93.

45. DiRuggiero, J., Tuttle, J. H. and Robb, F. T.. 1995. Rapid differentiation of hyperthermophilic Archaea by restriction mapping of the intergenic spacer regions of the ribosomal RNA operons. *Mol Mar Biol Biotechnol.* 4(2):123-7.

46. Dumans, A. T., Carvalho, M. F. and Coelho, A.. 1989. Tn*4527*, a Tp Sp/Sm transposon related to Tn*7* and flanked by IS*1*. *Plasmid.* 22(3):256-9.

47. Duval-Valentin,G., Normand, C., Khemici, V., Marty, B. and Chandler, M.. 2001. Transient promoter formation: a new feedback mechanism for regulation of IS*911* transposition. *EMBO J*. 20(20):5802-11.

48. Dyall-Smith, M. L. and Doolittle, W. F.. 1994. Construction of composite transposons for halophilic Archaea. *Can J Microbiol*. 40(11):922-9.

49. Engelman, A., Mizuuchi, K. and Craigie, R.. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell.* 67(6):1211-21.

50. Erauso, G. A., Reysenbach, A. L., Godfroy, A., Meunier, J. R., Crump, B., Partensky, F., Baross, J. A., Marteisson, V., Barbier, G., Pace, N. R. and Prieur, D.. 1993. *Pyrococcus abyssi*, sp. Nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch. Microbiol*. 160:338-49.

51. Escoubas, J. M., Lane, D., and Chandler, M.. 1994. Is the IS*1* transposase, InsAB', the only IS*1*-encoded protein required for efficient transposition? *J Bacteriol.* 176(18):5864-7.

52. Farabaugh, P. J.. 1996. Programmed translational frameshifting. *Annu Rev Genet*. 30:507-28.

53. Farabaugh, P. J. and Bjork, G. R.. 1999. How translational accuracy influences reading frame maintenance. *EMBO J.* 18:1427-34.

54. Fiala, G. and Stetter, K. O.. 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 ^{o}C. *Arch. Microbiol*. 145:56-61.

55. Flower, A. M. and McHenry, C. S.. 1990. The gamma subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. *Proc Natl Acad Sci U S A.* 87: 3713-7.

56. Foster, T. J., Davis, M. A., Roberts, D. E., Takeshita, K. and Kleckner, N.. 1981. Genetic organization of transposon Tn*10*. *Cell.* 23(1):201-13.

57. Fouchier, R. A. and Malim, M. H.. 1999. Nuclear import of human immunodeficiency virus type-1 preintegration complexes. *Adv Virus Res.* 52:275-99.

58. Freiberg, C, Fellay, R, Bairoch, A., Broughton, W. J, Rosenthal, A. and Perret, X.. 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature.* 22:394-401.

59. Fugmann, S. D., Villey, I. J., Ptaszek, L. M. and Schatz, D. G.. 2000. Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. *Mol Cell.* 5(1):97-107.

60. Galas, D. J. and Chandler, M.. 1989. Bacterial Insertion Sequences. *In* Mobile DNA. Ed. D.E. Berg, and M.M. Howe. American Society of Microbiology, Washington, D.C. pp. 109-162.

61. Gamas, P., Chandler, M.G., Prentki, P., and Galas, D.J.. 1987. *Escherichia coli* integration host factor binds specifically to the ends of the insertion sequence IS*1* and to its major insertion hot-spot in pBR322. *J Mol Biol*. 195(2):261-72.

62. Gamas, P., Galas, D. and Chandler, M.. 1985. DNA sequence at the end of IS*1* required for transposition. *Nature.* 317:458-60.

63. Gao, J., Bauer, M. W. Shockley, K. R., Pysz, M. A. and Kelly, R. M.. 2003. Growth of hyperthermophilic archaeon *Pyrococcus furiosus* on chitin involves two family 18 chitinases. *Appl. Environ. Microbiol*. 69:3119-28.

64. Gesteland, R. F. and Atkins, J. F.. 1996. Recoding: dynamic reprogramming of translation. *Annu Rev Biochem*. 65: 741-68.

65. González, J. M., Robb, F.T., Ammerman, J. W., Maeder, D. L., Yanagibayashi, M., Tamaoka, J. and Kato, C.. 1998. *Pyrococcus horikoshii* sp. Nov., a hyperthermophilic Archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles*. 2:123-30.

66. Gueguen, Y., Voorhorst, W. G., van der Oost, J. and de Vos, W. M.. 1997. Molecular and biochemical characterization of an endo-beta-1,3-glucanase of

the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* 272:31258-31264.

67. Haas, M. and Rak, B.. 2002. *Escherichia coli* insertion sequence IS*150*: transposition *via* circular and linear intermediates. *J Bacteriol*. 184:5833-41.

68. Hansen, T. M., Baranov, P. V., Ivanov, I. P. and Gesteland, R. F.. 2003. Maintenance of the correct open reading frame by the ribosome. *EMBO Rep*. 4:499-504.

69. Hallet, B. and Sherratt, D. J.. 1997. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev*. 21(2):157-78.

70. Haren, L., Normand, C., Polard P., Alazard, R. and Chandler, M. 2000. IS*911* transposition is regulated by protein-protein interactions *via* a leucine zipper motif. *J Mol Biol.* 296(3):757-68.

71. Haren, L., Ton-Hoang, B. and Chandler, M.. 1999. Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol*. 53:245-81.

72. Hatfield, D. and Oroszlan, S.. 1990. The where, what and how of ribosomal frameshifting in retroviral protein synthesis. *Trends Biochem Sci*. 15:186-90.

73. Hedges, R. W., Matthew, M., Smith, D. I., Cresswell, J. M. and Jacob, A. E.. 1977. Properties of a transposon conferring resistance to penicillins and streptomycin. *Gene.* 1(3-4):241-53.

74. Hellmer, J., Patzold, R. and Zeilinger, C.. 2002. Identification of a pH regulated $Na^{(+)}/H^{(+)}$ antiporter of *Methanococcus jannaschii*. *FEBS Lett.* 527(1-3):245-9.

75. Hinton, D. M. and Musso, R. E.. 1982. Transcription initiation sites within an IS*2* insertion in a Gal-constitutive mutant of *Escherichia coli*. *Nucleic Acids Res*. 10(16):5015-31.

76. Hiom, K., Melek, M. and Gellert, M.. 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell.* 94(4):463-70.

77. Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., Stetter, K. O.. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature.* 417(6884):63-7.

78. Isberg, R. R. and Syvanen. M.. 1982. DNA gyrase is a host factor required for transposition of Tn*5*. *Cell*. 30(1):9-18.

79. Jaurin, B. and Normark, S.. 1983. Insertion of IS*2* creates a novel *ampC* promoter in *Escherichia coli. Cell.* 32(3):809-16.

80. Jenkins, T.M., Esposito, D., Engelman, A. and Craigie, R.. 1997. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.* 16(22):6849-59.

81. Katz, R. A., Greger, J. G., Boimel, P. and Skalka, A. M. 2003. Human immunodeficiency virus type 1 DNA nuclear import and integration are mitosis independent in cycling cells. *J Virol.* 77(24):13412-7.

82. Kengen, S. W., Luesink, E. J., Stams, A. J. and Zehnder, A. J.. 1993. Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus. Eur. J. Biochem*. 213:305-12.

83. Kim, R., Sandler, S. J., Goldman, S., Yokota, H., Clark, A. J. and Kim, S-H.. 1998. Overexpression of archaeal proteins in *Escherichia coli. Biotechnol. Letters.* 20(3):207-210.

84. Kim, Y. G., Maas, S., Wang, S. C. and Rich, A.. 2000. Mutational study reveals that tertiary interactions are conserved in ribosomal frameshifting pseudoknots of two luteoviruses. *RNA.* 6:1157-65.

85. Kitamura, K., Hashida, S. N., Mikami, T. and Kishima, Y.. 2001. Position effect of the excision frequency of the *Antirrhinum* transposon Tam3: implications for the degree of position-dependent methylation in the ends of the element. *Plant Mol Biol.* 47(4):475-90.

86. Kleckner, N.. 1989. Transposon Tn*10*. *In* Mobile DNA. Eds. D. E. Berg and M. M. Howe. American Society of Microbiology, Washington, D.C. pp. 227-268.

87. Kleckner, N., Chalmers, R. M., Kwon, D., Sakai, J. and Bolland, S.. 1996. Tn*10* and IS*10* transposition and chromosome rearrangements: mechanism and regulation *in vivo* and *in vitro*. *Curr Top Microbiol Immunol.* 204:49-82.

88. Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty B. A., McKenney, K., Adams, M. D., Loftus, B., Venter, J. C., *et al.,*. 1997.

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*. 27:390(6658):364-70.

89. Kolor, K., Lindsley, D. and Gallant, J. A.. 1993. On the role of the P-site in leftward ribosome frameshifting at a hungry codon. *J Mol Biol*. 230(1):1-5.

90. Kuan, C. T. and Tessman, I.. 1992. Further evidence that transposition of Tn*5* in *Escherichia coli* is strongly enhanced by constitutively activated RecA proteins. *J Bacteriol*. 174(21):6872-7.

91. Kuan, C. T. and Tessman, I.. 1991. LexA protein of *Escherichia coli* represses expression of the Tn*5* Transposase gene. *J Bacteriol*. 173(20):6406-10.

92. Kuan, C. T., Liu, S. K. and Tessman, I.. 1991. Excision and transposition of Tn*5* as an SOS activity in *Escherichia coli*. *Genetics.* 128(1):45-57.

93. Kurland, C. G.. 1992. Translational accuracy and the fitness of bacteria. *Annu Rev Genet.* 26:29-50.

94. Kwon, D., Chalmers, R. M. and Kleckner, N.. 1995. Structural domains of IS*10* transposase and reconstitution of transposition activity from proteolytic fragments lacking an interdomain linker. *Proc Natl Acad Sci USA*. 92(18):8234-8.

95. Laachouch, J. E., Desmet, L., Geuskens, V., Grimaud, R. and Toussaint, A.. 1996. Bacteriophage Mu repressor as a target for the *Escherichia coli* ATP-dependent Clp Protease. *EMBO J*. 15(2):437-44.

96. Laksanalamai, P., Jiemjit, A., Bu, Z., Maeder, D.L. and Robb, F.T. 2003. Multi-subunit assembly of the *Pyrococcus furiosus* small heat shock protein is essential for cellular protection at high temperature. *Extremophiles*. 7(1):79-83.

97. Laksanalamai, P., Maeder, D.L., and Robb, F.T. 2001. Regulation and mechanism of action of the small heat shock protein from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol.* 183(17):5198-202.

98. Larsen, B., Wills, N. M., Gesteland, R. F. and Atkins, J. F.. 1994. rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. *J Bacteriol.* 176: 6842-51.

99. Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., Thierry, J. C. and Poch, O.. 2001. Genome evolution at the genus level:

comparison of three complete genomes of hyperthermophilic archaea. *Genome Res.* 11(6):981-93.

100. Lin, S. S., Nymark-McMahon, M. H., Yieh, L. and Sandmeyer, S. B.. 2001. Integrase mediates nuclear localization of Ty*3*. *Mol Cell Biol*. 21(22):7826-38.

101. Lindsley, D. and Gallant, J.. 1993. On the directional specificity of ribosome frameshifting at a "hungry" codon. *Proc Natl Acad Sci U S A*. 90(12):5469-73.

102. Lodish, H., Baltimore, D., Berk, A., Zipursky, S. L., Matsudaira, P. and Darnell, J.. 1995. The Molecular Anatomy of Genes and Chromosomes. *In* Molecular Cell Biology. 3rd edition. Scientific American Books New York. pp. 307-364.

103. Loot, C., Turlan, C., Rousseau, P., Ton-Hoang, B. and Chandler, M. 2002. A target specificity switch in IS*911* transposition: the role of the OrfA protein. *EMBO J.* 21:4172-82.

104. Lopinski, J. D., Dinman, J. D. and Bruenn, J. A.. 2000. Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Mol Cell Biol.* 20:1095-103.

105. Macnab, R. M. 1996. Flagella and Mobility. *In Escherichia coli* and *Salmonella:* Cellular and Molecular Biology, Vol. I. Ed. F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. ASM Press, Washington, D.C.. pp. 123-57.

106. Maeder, D. L., Weiss, R. B., Dunn, D. M., Cherry, J. L., González, J. M., DiRuggiero, J. and Robb, F. T.. 1999. Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics.* 152(4):1299-305.

107. Mahillon, J. and Chandler, M.. 1998. Insertion sequences. *Microbiol Mol Biol Rev.* 62:725-74.

108. Manna, D., Wang, X. and Higgins, N. P.. 2001. Mu and IS*1* transpositions exhibit strong orientation bias at the *Escherichia coli* bgl locus. *J Bacteriol*. 183(11):3328-35.

109. Martusewitsch, E., Sensen, C.W. and Schleper, C.. 2000. High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J Bacteriol*. 182(9):2574-81.

110. Mazodier, P., Cossart, P., Giraud, E. and Gasser, F.. 1985. Completion of the nucleotide sequence of the central region of Tn*5* confirms the presence of three resistance genes. *Nucleic Acids Res*. 13(1):195-205.

111. McClintock, B.. 1951. Chromosome Organization and Genic Expression. Cold Spring Harbor Symp. *Quant. Biol*. 16:13-47.

112. Miller, J. H.. 1992. *In* A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria. Cold Spring Harbor Laboratory Press. pp. 71-74.

113. Mizuuchi, K.. 1992. Polynucleotidyl transfer reactions in transpositional DNA recombination. *J Biol Chem.* 267(30):21273-6. Review.

114. Mizuuchi, M. and Mizuuchi, K.. 2001. Conformational isomerization in phage Mu transpososome assembly: effects of the transpositional enhancer and of MuB. *EMBO J.* 20(23):6927-35.

115. Morisato, D. and Kleckner, N.. 1987. Tn*10* transposition and circle formation *in vitro. Cell*. 51(1):101-11.

116. Mukund, S. and Adams, M. W.. 1991. The novel tungsten-iron-sulfur protein of the hyperthermophilic archaebacterium, *Pyrococcus furiosus*, is an aldehyde ferredoxin oxidoreductase. Evidence for its participation in a unique glycolytic pathway. *J. Biol. Chem.* 266:14208-16.

117. Ng, W. V., Ciufo, S. A., Smith, T. M., Bumgarner, R. E., Baskin, D., Faust, J., Hall, B., Loretz, C., Seto, J., Slagel, J., Hood, L. and DasSarma, S. 1998. Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? *Genome Res*. 8(11):1131-41.

118. Nixon, P. L., Rangan, A., Kim, Y. G., Rich, A., Hoffman, D. W., Hennig, M. and Giedroc, D. P.. 2002. Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J Mol Biol.* 20:322:621-33.

119. Pande, S., Vimaladithan, A., Zhao, H. and Farabaugh, P. J.. 1995. Pulling the ribosome out of frame by +1 at a programmed frameshift site by cognate binding of aminoacyl-tRNA. *Mol Cell Biol.* 15:298-304.

120. Pato, M. L. and Banerjee, M.. 2000. Genetic analysis of the strong gyrase site (SGS) of bacteriophage Mu: localization of determinants required for promoting Mu replication. *Mol Microbiol*. 37(4):800-10.

121. Peterson, G. and Reznikoff, W.. 2003. Tn*5* transposase active site mutations suggest position of donor backbone DNA in synaptic complex. *J Biol Chem.* 278(3):1904-9.

122. Polard, P. and Chandler, M.. 1995. Bacterial transposases and retroviral integrases. *Mol Microbiol*. 15(1):13-23.

123. Prentki, P., Teter, B., Chandler, M. and Galas, D. J.. 1986. Functional promoters created by the insertion of transposable element IS*1*. *J Mol Biol*. 191(3):383-93.

124. Qureshi, S. A. and Jackson, S. P.. 1998. Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol Cell*. 1(3):389-400.

125. Ramirez-Santos, J., Alvarez, G., Cisneros, E. and Gomez-Eichelmann, M. C.. 1992. Distribution of insertion sequence IS*1* in multiple-antibiotic resistant clinical Enterobacteriaceae strains. *FEMS Microbiol Lett.* 72(2):189-93.

126. Rettberg, C. C., Prere, M. F., Gesteland, R. F., Atkins, J. F. and Fayet, O.. 1999. A three-way junction and constituent stem-loops as the stimulator for programmed -1 frameshifting in bacterial insertion sequence IS*911*. *J Mol Biol.* 286(5): 1365-78.

127. Reznikoff, W. S.. 1993. The Tn*5* transposon. *Annu Rev Microbiol*. 47:945-63.

128. Rhodes, G., Saunders J. R., and Pickup, R. W.. 2000. Detection and distribution of insertion sequence 1 (IS*1*)-containing bacteria in the freshwater environment. *FEMS Microbiol Ecol.* 34(1):81-90.

129. Robb, F. T., Maeder, D.L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B. and Dunn, D. M.. 2001. Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods Enzymol*. 330:134-57.

130. Savilahti, H. and Mizuuchi, K.. 1996. Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. *Cell.* 85(2):271-80.

131. Sarnovsky, R. J., May, E. W. and Craig, N. L.. 1996. The Tn*7* transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* 15(22):6348-61.

132. Schmid-Appert, M., Zoller, K., Traber, H., Vuilleumier, S. and Leisinger, T.. 1997. Association of newly discovered IS elements with the dichloromethane utilization genes of methylotrophic bacteria. *Microbiol.* 143(Pt 8):2557-67.

133. Sensen, C. W., Klenk, H. P., Singh, R. K., Allard, G., Chan, C. C., Liu, Q. Y., Penny, S. L., Young, F., Schenk, M. E., Gaasterland, T., Doolittle, W. F., Ragan, M. A. and Charlebois, R. L.. 1996. Organizational characteristics and information content of an archaeal genome: 156 kb of sequence from *Sulfolobus solfataricus* P2. *Mol Microbiol*. 22(1):175-91.

134. She, Q., Phan, H., Garrett, R. A., Albers, S. V., Stedman, K. M. and Zillig, W.. 1998. Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon. *Extremophiles*. 2(4):417-25.

135. Silverman, M., Zieg, J., Hilmen, M. and Simon, M.. 1979. Phase variation in *Salmonella*: genetic analysis of a recombinational switch. *Proc Natl Acad Sci U S A*. 76(1):391-5.

136. Simon, M., Zieg, J., Silverman, M., Mandel, G. and Doolittle, R.. 1980. Phase variation: evolution of a controlling element. *Science.* 209(4463):1370-4.

137. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J.N., *et al.*. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics. *J Bacteriol.* 179(22):7135-55.

138. So, M., Heffron, F. and McCarthy, B. J.. 1979. The *E. coli* gene encoding heat stable toxin is a bacterial transposon flanked by inverted repeats of IS*1*. *Nature.* 277:453-6.

139. Sokal, R. R. and Rohlf, F. J.. 1996. Analysis of Frequencies. In: Introduction to Biostatistics. (2nd edition). W. H. Freeman and Company, New York pp. 294-313.

140. Sokolsky, T. D. and Baker, T. A.. 2003. DNA gyrase requirements distinguish the alternate pathways of Mu transposition. *Mol Microbiol.* 47(2):397-409.

141. Stedman, K. M., She, Q., Phan, H., Holz, I., Singh, H., Prangishvili, D., Garrett, R. and Zillig, W.. 2000. pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. *J Bacteriol*. 182(24):7014-20.

142. Surette, M. G., Buch, S. J. and Chaconas, G.. 1987. Transpososomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. *Cell.* 49(2):253-62.

143. Swanson, P. C.. 2001. The DDE motif in RAG-1 is contributed in trans to a single active site that catalyzes the nicking and transesterification steps of V(D)J recombination. *Mol Cell Biol.* 21(2):449-58.

144. Syvanen, M., Hopkins, J. D. and Clements, M.. 1982. A new class of mutants in DNA polymerase I that affects gene transposition. *J Mol Biol*. 158(2):203-12.

145. Ton-Hoang, B., Betermier, M., Polard, P. and Chandler, M.. 1997. Assembly of a strong promoter following IS*911* circularization and the role of circles in transposition. *EMBO J.* 16(11):3357-71.

146. Ton-Hoang. B., Polard, P. and Chandler, M.. 1998. Efficient transposition of IS*911* circles *in vitro. EMBO J.* 17(4):1169-81.

147. Tsuchihashi, Z. and Brown, P. O.. 1992. Sequence requirements for efficient translational frameshifting in the *Escherichia coli dnaX* gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. *Genes Dev.* 6: 511-9.

148. Tsuchihashi, Z. and Kornberg, A.. 1990. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc Natl Acad Sci U S A*. 87:2516-20.

149. Varmus, H. and Brown, P.. 1989. Retroviruses. *In* Mobile DNA. Eds. D. E. Berg and M.M. Howe. American Society of Microbiology, Washington, D.C. pp. 53-108.

150. Vimaladithan, A. and Farabaugh, P.J.. 1994. Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. *Mol Cell Biol*. 14:8107-16.

151. Vos, J.C., van Luenen, H. G. and Plasterk, R. H.. 1993. Characterization of the *Caenorhabditis elegans* Tc1 transposase *in vivo* and *in vitro. Genes Dev.* 7(7A):1244-53.

152. Waditee, R., Hibino, T., Nakamura, T., Incharoensakdi, A. and Takabe, T.. 2002. Overexpression of a Na+/H+ antiporter confers salt tolerance on a freshwater cyanobacterium, making it capable of growth in sea water. *Proc Natl Acad Sci U S A.* 99(6):4109-14.

153. Walker, G. C.. 1984. Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli. Microbiol Rev.* 48(1):60-93.

154. Watson, M. A. and Chaconas, G.. 1996. Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site. *Cell.* 85(3):435-45.

155. Weinreich, M. D. and Reznikoff, W. S.. 1992. Fis plays a role in Tn*5* and IS*50* transposition. *J Bacteriol.* 174(14):4530-7.

156. Weinreich, M. D., Makris, J. C. and Reznikoff, W. S.. 1991. Induction of the SOS response in *Escherichia coli* inhibits Tn*5* and IS*50* transposition. *J Bacteriol*. 173(21):6910-18.

157. Weiss, R. B., Dunn, D. M., Atkins, J. F. and Gesteland, R. F.. 1990. Ribosomal frameshifting from -2 to +50 nucleotides. *Prog Nucleic Acid Res Mol Biol.* 39:159-83.

158. Weiss, R. and Gallant, J.. 1983. Mechanism of ribosome frameshifting during translation of the genetic code. *Nature*. 302:389-93.

159. Weiss, R., Lindsley, D., Falahee, B. and Gallant, J.. 1988. On the mechanism of ribosomal frameshifting at hungry codons. *J Mol Biol*. 203(2):403-10.

160. Wiater, L. A. and Grindley, N. D.. 1988. Gamma delta transposase and integration host factor bind cooperatively at both ends of gamma delta. *EMBO J.* 7(6):1907-11.

161. Woese, C. R.. 2002. On the evolution of cells. *Proc Natl Acad Sci U S A*. 99(13):8742-7.

162. Woese, C. R. and Olsen, G. J.. 1986. Archaebacterial phylogeny: perspectives on the urkingdoms. *Syst Appl Microbiol*. 7:161-77.

163. Worthington, P., Hoang, V., Perez-Pomares, F. and Blum, P.. 2003. Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus. J Bacteriol*. 185(2):482-8

164. Wyndham, R. C., Cashore, A. E., Nakatsu, C. H. and Peel, M. C.. 1994. Catabolic transposons. *Biodegradation.* 5(3-4):323-42.

165. Xu, H. and Boeke, J. D.. 1990. Host genes that influence transposition in yeast: the abundance of a rare tRNA regulates Ty1 transposition frequency. *Proc Natl Acad Sci USA*. 87(21):8360-4.

166. Yelverton, E., Lindsley, D., Yamauchi, P. and Gallant, J. A.. 1994. The function of a ribosomal frameshifting signal from human immunodeficiency virus-1 in *Escherichia coli. Mol Microbiol*. 11:303-13.

167. Yieh, L., Kassavetis, G., Geiduschek, E. P. and Sandmeyer, S. B.. 2000. The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the gypsy-like element, Ty3. *J Biol Chem*. 275(38):29800-7.

168. Yin, J. C. and Reznikoff, W.S.. 1987. *dnaA*, an essential host gene, and Tn*5* transposition. *J Bacteriol.* 169(10):4637-45.

169. Yurieva, O., Skangalis, M., Kuriyan, J. and O'Donnell, M.. 1997. *Thermus thermophilis dnaX* homolog encoding gamma- and tau-like proteins of the chromosomal replicase. *J Biol Chem*. 272:27131-9.

170. Zerbib, D., Polard, P., Escoubas, J.M., Galas, D. and Chandler, M.. 1990. The regulatory role of the IS*1*-encoded InsA protein in transposition. *Mol Microbiol.* 4(3):471-7.

171. Zheng, J. and McIntosh, M. A.. 1995. Characterization of IS*1221* from *Mycoplasma hyorhinis*: expression of its putative transposase in *Escherichia coli* incorporates a ribosomal frameshift mechanism. *Mol Microbiol*. 16:669-85.

172. Zillig, W., Holz, I, Klenk, H-P., Trent, J., Wunderl, S., Janekovic, D., Imsel, E. and Haas, B.. 1987. *Pyrococcus woesei* sp. Nov., an ultra-thermophilic marine Archaebacterium, representing a novel order, Thermococcales. *Syst. Appl. Microbiol.* 9:62-70.

173. Zivanovic, Y., Lopez, P., Philippe, H. and Forterre. P.. 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res.* 30(9):1902-10.

174. Zou, S. and Voytas, D. F.. 1997. Silent chromatin determines target preference of the *Saccharomyces* retrotransposon Ty5. *Proc Natl Acad Sci USA*. 94(14):7412-6.

175. Zuker, M.. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406-15.