DEVELOPING GENETIC TOOLS FOR THE HYPERTHERMOPHILIC ARCHAEON,

*PYROCOCCUS FURIOSUS*

by

JOEL ANDREW FARKAS

(Under the Direction of Dr. Janet Westpheling)

ABSTRACT

*Pyrococcus furiosus* is a hyperthermophilic marine archaeon that has been extensively studied for more than 25 years. We previously identified an auxotrophic mutant, GLW101 (COM1 ΔpyrF), which is naturally and efficiently competent for DNA uptake. Replicating shuttle vectors have been constructed based on the chromosomal origin of replication (oriC). In the process of identifying the minimum replication origin required for autonomous plasmid replication in *P. furiosus*, we discovered that several previously predicted features were not essential for stable autonomous plasmid replication. A minimum region required to promote plasmid DNA replication was identified, and plasmids based on this sequence readily transformed *P. furiosus* GLW101. The plasmids replicated autonomously, existed in a single copy, were structurally unchanged after transformation and were stable without selection for more than 100 generations. We found that the combined transformation and recombination frequencies of the GLW101 strain allow marker replacement using linear DNA. We discovered that marker replacement was possible with as few as 40 nucleotides of flanking homology to the target region and adapted a strategy for selection of constructed deletions using PCR products with subsequent excision, or “pop-out,” of the selected marker. We used this method to construct a “markerless” deletion of the *trpAB* locus in the GLW101 background to generate a strain
(JFW02), which is a tight tryptophan auxotroph, providing a doubly auxotrophic strain. The utility of trpAB as a selectable marker was demonstrated using prototrophic selection of plasmids and genomic DNA containing the wild-type trpAB alleles. A deletion of radB was also constructed but had no obvious effect on either recombination or transformation, suggesting that it is not involved in the COM1 phenotype. Attempts to construct a radA deletion mutation were unsuccessful, suggesting that this may be an essential gene. The ease and speed of this procedure will facilitate the construction of strains with multiple genetic changes and deletion of virtually any nonessential gene. These advances will facilitate a number of research venues including hydrogen production, natural competence and CRISPR function. The development of multiply marked strains and a reporter gene system are remaining challenges to the advancement of the genetics effort.

INDEX WORDS: Archaea, Pyrococcus furiosus, hyperthermophile, genetic manipulation, natural competence, shuttle vector, oriC, homologous recombination, selectable markers
DEVELOPING GENETIC TOOLS FOR THE HYPERTHERMOPHILIC ARCHAEON, 

PYROCOCCUS FURIOSUS

by

JOEL ANDREW FARKAS

B.S., The University of Akron, 2006
M.S., The University of Akron, 2008

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA
2010
DEVELOPING GENETIC TOOLS FOR THE HYPERTHERMOPHILIC ARCHAEON,

PYROCOCCUS FURIOSUS

by

JOEL ANDREW FARKAS

Major Professor:  Janet Westpheling

Committee:  Michael W Adams
            Anna Karls
            Michael McEachern
            William Whitman

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2012
DEDICATION

I dedicate this work to my father, the smartest and strongest person I have ever known. He always supported me in my education and taught me so many things in life. My life will never be the same without him.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Jan Westpheling and members of our lab for guiding me through my doctoral training. I’d also like to thank Mike Adams and the members of his lab who were part of our collaboration, and laid the groundwork for *Pyrococcus* genetics. I’d particularly like to thanks Daehwan Chung for technical guidance in the lab, and Gina Lipscomb for all the scientific discussions over the years. Finally I’d like to thank my advisory committee for their guidance during the course of this work.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION AND LITERATURE REVIEW</td>
<td></td>
</tr>
<tr>
<td>Hyperthermophiles</td>
<td>1</td>
</tr>
<tr>
<td>Hyperthermophilic enzymes</td>
<td>3</td>
</tr>
<tr>
<td>Isolation and Characterization of Thermococcales</td>
<td>4</td>
</tr>
<tr>
<td>Context of <em>P. furiosus</em> research</td>
<td>8</td>
</tr>
<tr>
<td>Genetic manipulations in <em>Archaea</em>, anaerobes, and (hyper)thermophiles</td>
<td>10</td>
</tr>
<tr>
<td>Genetic Manipulations in <em>T. kodakarensis</em></td>
<td>14</td>
</tr>
<tr>
<td>Genetic Manipulations in <em>P. furiosus</em></td>
<td>22</td>
</tr>
<tr>
<td>2 DEFINING COMPONENTS OF THE CHROMOSOMAL ORIGIN OF REPLICATION OF THE HYPERTHERMOPHILIC ARCHAEON, <em>PYROCOCCUS FURIOSUS</em>, NEEDED FOR CONSTRUCTION OF A STABLE REPLICATING SHUTTLE VECTOR</td>
<td>24</td>
</tr>
<tr>
<td>Abstract</td>
<td>25</td>
</tr>
<tr>
<td>Introduction</td>
<td>25</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>30</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>34</td>
</tr>
<tr>
<td>Conclusions</td>
<td>42</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>43</td>
</tr>
<tr>
<td>References</td>
<td>44</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1.1. *Pyrococcus* species ........................................................................................................ 5
Table 1.2. Notable *Thermococcus* species .................................................................................... 7
Table 1.3. Selections for thermophilic Archaea ............................................................................... 13
Table 1.4. Selected *T. kodakarensis* strains essential for genetic tool development .................. 18
Table 1.5. Selected *T. kodakarensis* strains from applied genetic studies ................................. 21
Table 2.1. Plasmid transformation efficiencies .............................................................................. 33
Table 2.2. Maintenance of plasmids in *P. furiosus* ...................................................................... 35
Table 2.3. Conserved palindromic sequences within the *Pyrococcus oriC* region ....................... 41
Table 2.S1 Primer Sequences ......................................................................................................... 50
Table 3.1. *P. furiosus* strains used and constructed in this study ............................................... 58
Table 3.S1 Primer Sequences ......................................................................................................... 79
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Region of the <em>P. furiosus</em> chromosome predicted to contain the origin of replication, oriC</td>
</tr>
<tr>
<td>2.2</td>
<td>Construction of pJFW027</td>
</tr>
<tr>
<td>2.3</td>
<td>Determination of copy number for pJFW027 in <em>P. furiosus</em></td>
</tr>
<tr>
<td>2.4</td>
<td>Restriction analysis of plasmid DNA before and after transformation of <em>P. furiosus</em> and back-transformation to <em>E. coli</em></td>
</tr>
<tr>
<td>2.51</td>
<td>pJFW019 contains the pSC101 replication origin and aac gene cassette for replication and selection in <em>E. coli</em></td>
</tr>
<tr>
<td>2.52</td>
<td>Restriction analysis of pJFW019 plasmid DNA before and after transformation of <em>P. furiosus</em> and back-transformation to <em>E. coli</em></td>
</tr>
<tr>
<td>2.53</td>
<td>PCR confirmation of plasmid transformation of <em>P. furiosus</em></td>
</tr>
<tr>
<td>3.1</td>
<td>The wild-type pyrF region and PCR-amplified fragments (indicated as lines below the chromosomal region) with various lengths of flanking sequence used to transform GLW101</td>
</tr>
<tr>
<td>3.2</td>
<td>Pop-out marker replacement strategy</td>
</tr>
<tr>
<td>3.3</td>
<td>Construction of the <em>trpAB</em> pop-out markerless deletion</td>
</tr>
<tr>
<td>3.4</td>
<td>Construction of pJFW070</td>
</tr>
<tr>
<td>3.5</td>
<td>Characterization of JFW04</td>
</tr>
<tr>
<td>3.51</td>
<td>Screening of putative <em>trpAB</em> marker replacement mutants</td>
</tr>
<tr>
<td>3.52</td>
<td>Construction of pJFW051</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

*Pyrococcus furiosus* is a hyperthermophilic archaeon that has been studied extensively through bioinformatics and biochemistry for more than 25 years. Recent advances in genetics in the closely related *Thermococcus kodakarensis* have shown the enormous potential for applying genetic tools in *P. furiosus*. In order to achieve ongoing research goals to expand the potential for future research, our lab, in collaboration with the Adams lab, has invested in the development of genetic methodologies for *P. furiosus*. This collaboration, which took several years, has now come to its conclusion. These efforts were very successful, and the potential of applying genetic methodologies to *P. furiosus* research is already being realized.

The work presented in this dissertation describes my role in the development of genetic tools for *P. furiosus*, and subsequent biological outcomes, which were realized during the course of developing these tools.

**Hyperthermophiles**

Until the middle of the 20th century the prevailing opinion in microbiology had been that life could not exist at temperatures much higher than 60°C (25). The unexpected discovery of life in extreme environments, such as the identification of *Thermus aquaticus* from a geyser in Yellowstone National Park (26), started a revolution in microbiology. From that time on, microbiologists started looking for and finding life in a multitude of extreme and unusual environments. Those organisms, which thrive in high temperature environments, are called thermophiles. Over time, new organisms were discovered in even hotter environments. Those organisms which grow optimally at temperatures higher than 80°C are referred to as
hyperthermophiles (3, 158). The discovery of abyssal marine hydrothermal vents in the late 1970s (40) allowed for the isolation of many new hyperthermophilic species. These organisms are particularly interesting for their thermostable enzymes and lucrative biotechnology applications (20, 171) as well as their evolutionary significance (129, 157, 158).

The hyperthermophiles are an incredibly diverse group, comprised of both Archaea and eubacteria, from varied environments, but they share several notable similarities. Most hyperthermophiles are obligate anaerobes, a phenotype that likely derives from common ancestry. However there are some examples of hyperthermophiles that are aerotolerant and even some that utilize oxygen for metabolism (19, 69, 72, 133, 172). Terrestrial hot spring environments are usually freshwater, and the organisms found there grow optimally at low solute concentrations and normal atmospheric pressure. As many hyperthermophiles are found in deep-sea hydrothermal vents, they are normally subjected to extreme hydrostatic pressure. It comes as no surprise that many of these organisms are tolerant of high pressure. Some actually show better growth under pressure (105, 162, 163), and some even require high pressure for growth (18, 179). The marine habitat also explains the high salt tolerance and also requirements for specific trace elements such as iron (59, 163) and tungsten (27), that are found naturally in seawater.

The majority of hyperthermophiles, and the most thermophilic examples are members of the domain Archaea. The most thermophilic bacteria belong to the genera *Thermotoga* (70, 74), *Aquifex* (72), and *Thermocrinis* (69), that grow optimally between 80°C and 85°C. In contrast, there are several Archaeal genera containing hyperthermophilic species. Hyperthermophilic Archaea are represented in all major Archaeal lineages, suggesting that hyperthermophily may be an ancestral condition. Hyperthermophilic bacteria are among the most deeply rooted in the tree of life (157, 176), suggesting that the last universal common ancestor (LUCA) may have been a hyperthermophile (157).
Hyperthermophilic enzymes

All hyperthermophiles possess thermostable enzymes, which are adapted to withstand extreme heat and function optimally at the temperatures in their extreme environments. Such enzymes are potentially lucrative for biotechnology applications, as evidenced by the number of commercially available thermostable DNA polymerases, restriction endonucleases, and carbohydrate active enzymes. One prominent example is the DNA polymerase enzyme, Taq, isolated from the freshwater bacterium Thermus aquaticus (34). The ability of the enzyme to maintain its activity after prolonged periods of heating to 95°C made it ideal for use in the polymerase chain reaction (PCR) (114, 115, 132). The Taq polymerase has been a mainstay in molecular biology for more than two decades, though a number of polymerases from hyperthermophilic Archaea have been discovered that have specific advantages, such as increased survival during PCR cycles, processivity, elongation rate, and fidelity (58, 125, 155, 161). A number of (hyper)thermophilic organisms that are able to utilize complex carbohydrates for metabolism and growth produce enzymes that catalyze the efficient degradation of complex biomass. Such enzymes are promising prospects for converting biomass into biofuels (20).

In addition to biotechnological potential, hyperthermophilic enzymes have an intrinsic academic value because they provide insights into biological processes. The sequence and structure of hyperthermophilic enzymes provide insights into reaction mechanisms and their adaptation to high temperature (128, 129, 171). Novel hyperthermophilic enzymes also enhance our understanding of basic biological processes. The enzyme reverse gyrase is a prime example of a hyperthermophilic enzyme that performs a function common to all organisms, but in a novel way. Whereas mesophilic organisms utilize gyrase to maintain negative supercoiling in their genomic DNA, hyperthermophiles utilize reverse gyrase to induce positive supercoiling, which is believed to make DNA more stable at high temperatures (23, 39, 82).
Isolation and characterization of *Thermococcales*

The most commonly isolated hyperthermophiles belong to the order *Thermococcales*, which are regarded as the primary decomposers of organic matter in hydrothermal marine ecosystems (163). *Thermococcales* are regular to slightly irregular cocci and are motile by means of lophotrichous flagella. Members of the *Thermococcales* are widely distributed geographically and primarily utilize proteinaceous substrates, though some species can also metabolize carbohydrates. There are currently three recognized genera, *Pyrococcus*, *Thermococcus*, and *Palaeococcus*. There are only two reported species of *Palaeococcus* (7, 163), but many more species of *Pyrococcus* and *Thermococcus*.

*Pyrococcus* species are all heterotrophic, fermentative, sulfur-reducing cocci with optimal growth temperatures near 100°C. *P. furiosus* was the first isolated, and founding species of the genus. It was isolated from shallow marine sediments off the coast of Vulcano Island, Italy in 1986 (51). *P. furiosus* was not the first organism found to grow optimally above 100°C, however, it was one of the first to be thoroughly characterized in terms of its microbiological and biochemical properties (51). Its short generation time (51), relative ease of culturing, and growth to high cell densities without the need for elemental sulfur (127) made it an attractive study system among the hyperthermophilic Archaea. Soon after, *P. woesei* was reported as the second species of *Pyrococcus*. This species shows the interesting property of complete cell lysis upon reaching full cell density (181). Subsequent genomic data indicates that it is very similar to *P. furiosus* and may actually be a subspecies of *P. furiosus* (61, 80). The other *Pyrococcus* species were all isolated from deep-sea hydrothermal vent environments. *P. abyssi* was isolated from a fluid sample taken from 2000m deep in the North Fiji basin (43, 45). *P. horikoshii* was isolated at a depth of 1395m in the Okinawa Trough (56). *P. glycovarans* was isolated from a vent at a depth of 2650m along the East Pacific rise (14). Two more recent additions to the genus are *P. yayanosii* and *Pyrococcus* strain NA2, which were both isolated from deep-sea vents. Their complete genome sequences were recently determined (18, 77, 88).
Table 1.1. *Pyrococcus* species

<table>
<thead>
<tr>
<th><strong>Species</strong></th>
<th><strong>Location isolated</strong></th>
<th><strong>Growth Temperature range, (optimum)</strong></th>
<th><strong>Notable characteristics</strong></th>
<th><strong>Reference(s)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pyrococcus furiosus</em></td>
<td>Marine sediments, Vulcano, Porto di Levante, Vulcano, Italy</td>
<td>70-103, (100)</td>
<td></td>
<td>(51)</td>
</tr>
<tr>
<td><em>Pyrococcus abyssi</em></td>
<td>Deep-sea hydrothermal vent, North Fiji Basin</td>
<td>67-102, (96)</td>
<td>Endogenous plasmid pGT5</td>
<td>(43, 45)</td>
</tr>
<tr>
<td><em>Pyrococcus horikoshii</em></td>
<td>Deep-sea hydrothermal vent, Okinawa trough, NE Pacific Ocean</td>
<td>80-102, (98)</td>
<td>Unable to utilize carbohydrates for growth.</td>
<td>(56)</td>
</tr>
<tr>
<td><em>Pyrococcus glycovorans</em></td>
<td>Deep-sea hydrothermal vent, East Pacific Rise, 13 deg N</td>
<td>74-104, (95)</td>
<td>Able to utilize several carbohydrates for growth.</td>
<td>(14)</td>
</tr>
<tr>
<td><em>Pyrococcus woesei</em></td>
<td>Marine solfataras, Porto di Levante, Vulcano, Italy</td>
<td>ND-105, (100-103)</td>
<td>Cell lysis during stationary phase growth</td>
<td>(1)</td>
</tr>
<tr>
<td><em>Pyrococcus yayanosi</em></td>
<td>Deep-sea hydrothermal vent site on the Mid-Atlantic Ridge</td>
<td>80-108, (98)</td>
<td>Obligate piezophile</td>
<td>(18, 77, 179)</td>
</tr>
</tbody>
</table>
The former of the two is an obligate piezophile (barophile), requiring at least 20 megapascals of pressure for growth (18, 179).

Compared to *Pyrococcus*, the genus *Thermococcus* is much larger, containing dozens of species. *Thermococcus* species are metabolically and physiologically very similar to *Pyrococcus* species, but they grow at lower optimal temperatures. The type species, *T. celer* was isolated years before *P. furiosus* from a solfataric waterhole on the beach of Vulcano Island, Italy (180). *T. coalescens* has the unique property of cell fusion during growth, producing very large cells (85). *T. chitinophagous* grows efficiently on chitin, aided by multiple inducible chitinase enzymes (8, 71). *T. nautilus* is a species that carries an endogenous plasmid (154). *T. barophilus* was isolated from a deep-sea vent site at 3550m is barophilic, and shows improved growth under high (~40 MPa) hydrostatic pressure (105). But the most noteworthy, and best-studied species is *T. kodakarensis*. Originally reported as *Pyrococcus* sp. KOD1 (113), and subsequently as *Pyrococcus kodakaraensis*, it was the subject of several studies, mostly focused on unique enzymes (47, 48, 62-65, 73, 75, 102, 118, 120, 145, 160, 164), but also transcription (117, 165). It was later renamed *Thermococcus kodakaraensis* based on 16S rRNA sequence data (10).

Like most Archaea, Thermococcales do not have cell walls composed of murein (79), rather, they have a glycoprotein S-layer coat (10, 14, 45, 113). Cell membranes are composed of lipids with ether linkages, typical of archaea (104), in contrast to the ester linkages of bacterial membrane lipids (10, 14, 45, 113). *Pyrococcus* and *Thermococcus* species can utilize various proteinaceous substrates, growing fermentatively. Some species can also utilize carbohydrates. *T. hydrothermalis*, *T. stetteri*, and *P. woesei* can utilize a number of carbohydrate substrates (55, 112, 181). *P. glycovorans* and *P. furiosus* can utilize the most carbohydrate substrates, including starch, maltose, and cellobiose (14, 51). H₂ is produced in the absence of S⁰, but H₂S is produced in the presence of S⁰. S⁰ is required for growth on proteinaceous substrates but not for the fermentation of carbohydrates.
<table>
<thead>
<tr>
<th>Species</th>
<th>Location isolated</th>
<th>Temperature range, (optimum)</th>
<th>Notable characteristics</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Thermococcus celer</em></td>
<td>Marine solfatara, Vulcano, Italy</td>
<td>ND, (88)</td>
<td></td>
<td>(180)</td>
</tr>
<tr>
<td><em>Thermococcus kodakarensis</em></td>
<td>Solfatara on the shore of Kodakara Island</td>
<td>65-95, (85)</td>
<td></td>
<td>(10, 113)</td>
</tr>
<tr>
<td><em>Thermococcus barophilus</em></td>
<td>Deep-sea hydrothermal vent site on the Mid-Atlantic Ridge</td>
<td>48-95, (85)</td>
<td>Elevated hydrostatic pressure enhances growth</td>
<td>(105)</td>
</tr>
<tr>
<td><em>Thermococcus coalescens</em></td>
<td>Deep-sea caldera fluid in the Pacific south of Japan</td>
<td>57-90, (87)</td>
<td>Cells fuse to form larger cells</td>
<td>(85)</td>
</tr>
<tr>
<td><em>Thermococcus chitinophagus</em></td>
<td>Deep-sea hydrothermal vent site in the Pacific west of Mexico</td>
<td>60-93, (85)</td>
<td>Efficiently degrades chitin</td>
<td>(8, 71)</td>
</tr>
<tr>
<td><em>Thermococcus nautilus</em></td>
<td>Deep-sea hydrothermal vent, east Pacific Ocean</td>
<td>ND, (87.5)</td>
<td>Endogenous plasmid, pTN1</td>
<td>(90, 154)</td>
</tr>
</tbody>
</table>
Context of *P. furiosus* research

As a hyperthermophile, *P. furiosus* represents an excellent system for the study of adaptation to high temperature. As an archaeon, *P. furiosus* represents a suitable study system for basic biology of the Archaea. For more than a decade after its isolation, *P. furiosus* emerged as a prominent organism, with emphasis on its novel enzymes. The progress and scope of research was greatly facilitated and expanded by the availability of a whole genome sequence (103, 130). This enabled more sophisticated genomics and proteomics in *P. furiosus*. The availability of genome sequences from the closely related species *P. horikoshii* (103), *P. abyssi* (35), and *T. kodakarensis* (54) enabled comparative genomics and a much deeper understanding of the differences between these organisms (35, 54, 87, 103, 183). One of the most remarkable findings from the *P. furiosus* genome sequence is that so many of the proteins produced by *P. furiosus* are annotated as “hypothetical” proteins and do not have a known function (130). Through the use of whole genome microarrays, many of these hypothetical proteins have been implicated in specific cellular responses (146, 153, 159, 174, 175), though many of their specific functions remain unclear.

Hundreds of papers have been published related to enzymes purified or cloned from *P. furiosus*. Like other hyperthermophiles, *P. furiosus* encodes several novel and potentially lucrative thermostable enzymes, including a DNA polymerase. Like *Taq* polymerase, the *Pfu* polymerase is well suited for use in PCR. Unlike *Taq*, *Pfu* polymerase has a 3' to 5' exonuclease activity, which gives *Pfu* significantly higher fidelity (97). This makes *Pfu* polymerase better suited for applications such as cloning, sequencing, or protein expression, where PCR fidelity is critically important. However, *Pfu* polymerase has a slower elongation rate than *Taq* (161). In recent years, several recombinant DNA polymerases have become available, that retain beneficial characteristics of multiple native DNA polymerases, often by combining domains from different polymerases.
The stability of *P. furiosus* proteins makes its enzymes ideal for structural determination studies. For this reason, a massive structural genomics project was launched, which would potentially determine a protein structure for every gene encoded by the *P. furiosus* genome. In this project, genes from *P. furiosus* were cloned, expressed, and purified from *E. coli*, in order to determine their structure (4). Over the course of several years, this project produced structures for dozens of proteins, however the great majority of *P. furiosus* protein structures were not determined. This is perhaps not surprising, since expression in *E. coli* favors soluble monomeric proteins and is not well suited for membrane proteins, metalloproteins, proteins which are part of complexes, or proteins that are modified following translation (4, 149, 166). In addition to the crystal structures determined from this structural genomics project, many others have been purified directly from *P. furiosus*. To date, there are 285 entries in the Protein Data Bank (www.pdb.org) for structures of *P. furiosus* proteins (16).

*P. furiosus*, like other Archaeal species, is an attractive model system for the study of DNA replication and repair, because Archaeal information processing systems are similar to eukaryotic systems but with significantly reduced complexity (42, 121). While many of the replicative proteins are eukaryotic-like, the mechanisms of DNA replication are more bacterial in nature, and there are also Archaea-specific features (116, 131, 169). Studying such processes will provide a better understanding of basic biology and numerous human diseases that are caused or exacerbated by defects in these systems, most notably cancer. Given the wealth of genomic data, much of the earliest work on DNA replication in *P. furiosus* was bioinformatic in nature. Cumulative GC skew analysis (57) was developed as a method to identify the location of chromosomal replication origins. This method was used to identify the approximate location of the replication origin for *P. horikoshii* (96). Sequence comparison between *Pyrococcus* species identified a large intergenic space, neighboring a orc7/cdc6 homologue, predicted to be the location of the replication origin (116). Biochemical analysis subsequently confirmed the replication origin of *P. abyssi* to be contained within this region (108). The role of the Orc1/Cdc6
protein as the replication initiation protein was already suspected based on sequence and structural similarity to both eukaryotic Cdc6 and bacterial DnaA (30, 46, 95) and interaction with oriC (108). This function would be further supported by interaction with MCM helicase (81, 108, 152), and the identification of a DNA binding site found to be conserved in Archaeal origins (109). The physical structure of several well-studied Archaeal replication proteins has been determined using Pyrococcus proteins, including FEN-1 endonuclease (68), DNA primase (11), proliferating cell nuclear antigen (PCNA) (107) and replication factor C (RFC) small subunit (123), DNA ligase (119), and the Pfu DNA polymerase (83). Chapter 2 of this dissertation describes the cloning of the \textit{P. furiosus} oriC and an analysis of structures and open reading frames predicted to be important for DNA replication.

In addition to its other interesting features, \textit{P. furiosus} is an attractive organism for the production of hydrogen from peptides and carbohydrates (20, 78, 144). Like other members of the \textit{Thermococcales}, \textit{P. furiosus} preferentially ferments peptides, reducing sulfur to H$_2$S. However, \textit{P. furiosus} can utilize a variety of sugars and in the absence of elemental sulfur, produces H$_2$ (51, 144). Two soluble hydrogenases (100, 101) and a membrane-bound hydrogenase (139, 140) are involved with hydrogen metabolism in \textit{P. furiosus}. The availability of elemental sulfur influences not only the end products, but also induces a global regulatory shift in cellular metabolism (5, 36, 93). Both hydrogen metabolism and its regulation are the subjects of intense and ongoing study in \textit{P. furiosus}.

\textbf{Genetic manipulations in \textit{Archaea}, anaerobes, and (hyper)thermophiles}

There are several obstacles to developing genetic tools for the hyperthermophilic \textit{Archaea}. The most basic obstacle to developing genetic methods for any organism is the establishment of microbiological handling methods – specifically plating on a solid medium. In order to develop genetic tools, methods for culture, plating and selection are absolute prerequisites. The most important obstacle is the development of selectable markers for DNA
transformation. Handling protocols are largely a matter of innovation, but effective selections are an unavoidable matter of biology. Because Archaea have ribosomes and cell walls that differ from bacteria, most antibiotics are not suitable for selection even if the drugs and resistance markers are stable at high growth temperatures, leaving few drugs that are suitable for selection of transformants.

The first Archaea for which genetic tools were developed were the halophiles and mesophilic methanogens. Some early work in halophile genetics sought to establish protocols for DNA transfer. In the absence of selection, a phage plaque assay (38), and massive hybridization screening protocol (32), were used to show DNA transfer by transfection and transformation, respectively. The first selectable markers in archaea relied upon random mutagenesis to isolate nutritional auxotrophs (17, 111). This provided selective power to demonstrate genetic transfer between cells (111), and successful selection of transformation with exogenous DNA (17). Soon after, a genetic transformation system was developed for *Methanobacterium thermoautotrophicum* (177), which relied on the uracil prototrophy and 5-FOA resistance counterselection, first developed in yeast (21).

While isolation of random or spontaneous mutants can be useful first steps for the development of genetic methodologies, issues with spontaneous reversion and unintended mutations make this approach potentially treacherous for subsequent applications. Following the establishment of transformation protocols, the establishment of drug selections is often necessary to construct targeted mutants suitable for further genetic manipulations. In mesophilic Archaea, drugs like puromycin can be used for selection of transformants, with resistance conferred by a gene from mesophilic bacteria (168). Finding suitable selections for thermophiles is a challenge. Not only must the drug be thermally stable, but the gene that confers resistance must also be thermally stable. To overcome these challenges, drug selections often rely on the deletion or over-expression of endogenous genes or expression of genes from other thermophiles. Growth of *S. acidocaldarius* and *P. furiosus* is inhibited by certain antibiotics (see
table 1.3), and high concentrations of alcohols (2). The *Sulfolobus solfataricus* alcohol dehydrogenase (*adh*) gene was reported for use in transforming both *S. acidocaldarius* and *P. furiosus*, selecting for resistance to butanol and benzyl alcohol (9). Mevinolin is a drug that inhibits isoprenoid synthesis and was found to be inhibitory to *Halobacterium halobium* (28). Over-expression of 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA) was found to confer mevinolin resistance in *H. volcanii* (86), providing a potential selection that could be selected by transformation with a native gene.

Replicating shuttle plasmids are an invaluable genetic tool, capable of facilitating a variety of genetic manipulations. In order to construct a replicating shuttle plasmid, a replication origin is required in addition to a selectable marker. Naturally occurring plasmids and viral particles have been observed in a number of Archaeal species (15, 22, 31, 32, 44, 60, 182) and successfully used as the basis for developing replicating shuttle plasmids for species of haloarchaea (86), methanogens (167), and *Sulfolobus* (29). In theory, replicating plasmids could be constructed for any species, but it is often impossible to predict which plasmid or viral replication origins will function outside of their native host. Finding a native plasmid in a closely related organism provides the greatest chance for successfully developing a stable, replicating shuttle plasmid. One significant flaw of using a native plasmid to promote plasmid replication is the size of the native plasmid, which can be prohibitively large for many applications. Chapter 2 of this dissertation describes the construction and characterization of a replicating shuttle vector for *P. furiosus* based on oriC.

Another genetic tool that has been widely applied in mesophilic systems is a reporter gene expression system. Typically, a reporter gene’s expression will produce a product that is colored (or a precursor to a colored product), allowing visual inspection of colonies that express the reporter. This often relies on the cleavage of a synthetic substrate. Classic reporter gene systems also allow for assays for quantification of expression in a liquid medium. Depending upon the system, an exogenous gene may be introduced or a native gene may be placed under
<table>
<thead>
<tr>
<th>Drug / Nutrient(s)</th>
<th>Sensitivity / Auxotrophy</th>
<th>Mechanism of Resistance / Prototrophy</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uracil</td>
<td>pyrF deletion / inactivation</td>
<td>Inherent, pyrF complementation</td>
<td>(141, 142, 177)</td>
</tr>
<tr>
<td>5-FOA</td>
<td>Inherent, pyrF complementation</td>
<td>pyrF deletion / inactivation</td>
<td>(141, 142, 177)</td>
</tr>
<tr>
<td>Carbomycin, Celesticetin, Chloramphenicol, Thiostrepton</td>
<td>Inherent</td>
<td>Point mutations in 23S rRNA</td>
<td>(2)</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>trpE deletion / inactivation</td>
<td>Inherent</td>
<td>(141, 142)</td>
</tr>
<tr>
<td>Histidine</td>
<td>hisD deletion / inactivation</td>
<td>Inherent</td>
<td>(141)</td>
</tr>
<tr>
<td>Butanol, benzyl alcohol</td>
<td>Inherent</td>
<td>Expression of adh</td>
<td>(2, 9)</td>
</tr>
<tr>
<td>Mevinolin, simvastatin</td>
<td>Inherent</td>
<td>Overexpression of native hmg-CoA</td>
<td>(106, 136)</td>
</tr>
<tr>
<td>6-methyl purine</td>
<td>Inherent</td>
<td>Deletion of Hypoxanthine/guanine phosphoribosyltransferase</td>
<td>(137)</td>
</tr>
<tr>
<td>Agmatine</td>
<td>pdaD deletion / inactivation</td>
<td>Inherent, pdaD complementation</td>
<td>(53, 137)</td>
</tr>
</tbody>
</table>
the control of a promoter to be studied, sometimes by deletion and reintroduction. Developing this type of reporter system for hyperthermophilic anaerobes is particularly challenging. As with finding selectable markers, very few synthetic substrates are stable at high temperatures. A reporter system using lacS and x-gal was developed in S. solfataricus. In this system, a ΔpyrF ΔlacS mutant was transformed with a shuttle plasmid expressing both, selecting for uracil prototrophy, and assaying lacS expression by reaction of the LacS protein with x-gal (76). This system apparently circumvented the thermal instability of x-gal by spraying x-gal on plates after the growth of colonies and only allowing a short (1 hour) incubation at growth temperature (78°C). The requirement for anaerobicity is particularly challenging. The color change associated with classical reporter systems such as β-galactosidase and 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (x-gal) (67), luciferase (12, 52), and ade2 (33, 150, 151) all involve oxygen and/or other reactive oxygen species, making them incompatible with anaerobic systems. Clearly, a novel approach is needed to develop a reporter system for hyperthermophilic anaerobes.

*Genetic manipulations in T. kodakarensis*

The potential and need for genetics in *P. furiosus* is best illustrated by the recent advances made in the closely related *T. kodakarensis*. The direction of early research in *T. kodakarensis* was very similar to that of *P. furiosus* – mostly focused on novel and thermostable enzymes. The whole genome sequence of *T. kodakarensis* was not determined until 2005 (54), but the development of genetic tools was already well underway. In the absence of suitable drug selections for hyperthermophiles, the first genetic tools were developed using nutritional selections (141, 142) based on those developed for yeast in the mid 1980’s (21). *T. kodakarensis* cells were randomly mutagenized by exposure to UV light, and the toxic uracil analog 5-FOA was used to select for mutations in the pyrF gene (*ura3* in yeast). Such mutants are resistant to 5-FOA but are unable to grow in the absence of uracil, allowing for selection of
transformants by uracil prototrophic selection. One mutant, KU25, showed a single base deletion, early in the pyrF open reading frame, resulting in a truncated protein and complete loss of function. In a proof of principle experiment, they demonstrated direct marker replacement of the trpE gene with the pyrF gene, resulting in a KW4, a tryptophan auxotroph. This showed that *T. kodakarensis* could be transformed by exogenous DNA, and that genetic manipulation using prototrophic selections was possible. However recombination between the plasmid and chromosomal pyrF genes often restored the point mutation to wild type, resulting in a uracil prototroph without the intended plasmid integration event (142). This system was improved by constructing a targeted deletion of the pyrF gene, eliminating the potential for recombination between plasmid and chromosomal pyrF as well as simple reversion of the point mutation. This strain, KU216, also relieved concerns of unintended mutations resulting from UV exposure (141). Using this new genetic background, they recreated a trpE marker replacement mutant, KW128, and adapted a method from yeast (6) for making markerless deletions. This “pop-out” strategy, relies on transformation using the wild type pyrF allele, flanked by direct repeat sequences of the gene to be targeted and selection for uracil prototrophy. This intermediate strain contains the wild type pyrF gene, is a uracil prototroph, and 5-FOA sensitive. Excision of the pyrF marker and subsequent selection for 5-FOA resistance results in a makerless deletion and a strain which is a uracil auxotroph suitable for iterative manipulation and marker replacements. This method allowed the successful targeted deletion of trpE and hisD in the same strain, allowing for prototrophic selection with as many as three markers in a single strain (141). At the same time, they reported that *T. kodakarensis* could be efficiently transformed with either circular or linear DNA with as few as 500bp of homologous flanking regions. They also reported that transformation did not depend upon CaCl₂ treatment of cells, and suggested that transformation might be the result of natural competence (141).

While these nutritional selections provided all the essential tools for transformation and selections of mutants, they all had the intrinsic shortcoming that all manipulations must be
performed in defined media. Additionally, the pop-out deletion strategy developed by Sato et al, was reported to be unsuccessful by Santangelo et al (137). This disparity could potentially be a slight difference in method or a result specific to the target gene. Either way, an alternative method for selecting marker replacement and subsequent excision events would be beneficial to ongoing genetics efforts. In a screen for mutants resistant to nucleotide analogs, mutations inactivating TK0664 conferred resistance to 6-methyl purine (6MP) (137). The \( trpE \) gene, along with TK0664 upstream flanking genes, was integrated into the chromosome downstream of the TK0664 locus, selecting for tryptophan prototrophy. This intermediate strain, when plated onto medium containing tryptophan and 6MP, produced colonies resistant to 6MP, resulting from the excision of both \( trpE \) and TK0664. This strain, TS517, is a tryptophan auxotroph and resistant to 6MP. In this background strain (as well as derivative strains), a marker replacement cassette containing both \( trpE \) and TK0664 can be used similarly to select the direct integration of the cassette into the chromosome by tryptophan prototrophy and subsequent excision by 6MP resistance (137). The finding that deletion of \( pdaD \) results in a strict growth requirement for agmatine (53) presented an opportunity to develop a selectable marker that could be used in rich media. The deletion of \( pdaD \) in the TS517 background provides a multiple marked auxotrophic strain, TS559, which can be used for selections in either defined or rich media (137).

The development of replicating shuttle plasmids for \( T. kodakarensis \) was hampered by the lack of endogenous plasmids from closely related species. \( P. abyssi \) had long been known to contain a plasmid (43, 45) that likely replicates by a rolling circle mechanism (44). Shuttle plasmids based on this plasmid, pGT5, have been reported for use in \( Sulfolobus acidocaldarius \) and \( P. furiosus \) (2, 9), but constructs based on pGT5 proved to be unstable in \( T. kodakarensis \) (136). The discovery of an endogenous plasmid from the closely related species \( T. nautilus \), pTN1 (154), provided a basis for developing such a shuttle vector for \( T. kodakarensis \). The entire pTN1 sequence was cloned into a commercial cloning vector and found to be capable of
stable replication in both *E. coli* and *T. kodakarensis*. The utility of the shuttle vector was demonstrated by homologous expression of an RNA polymerase subunit, RpoL (136).

As discussed earlier, developing a reporter system for hyperthermophilic anaerobes is particularly challenging. The first reporter system reported for *T. kodakarensis* was based on the expression of TK1761, a nonessential β-glycosidase (48, 138). Using transforming plasmids with chromosomal homology, it was demonstrated that promoter sequences could be inserted upstream of the TK1761 gene, and transcript level and β-glycosidase activity of TK1761-1763 could be used to assay reporter gene expression (138). Subsequently, they reported that they identified another β-glycosidase, TK1827. Using their newly developed selection / counterselection protocol, they generated new deletion strains lacking both enzymes responsible for β-glycosidase activity (137).

While the development of genetic tools in *T. kodakarensis* has been impressive, the ultimate goal is the application of genetic tools to better understand the biology of the organism. With all the essential genetic tools, *T. kodakarensis* researchers have been able to do things that were never before possible. Perhaps the most straightforward application of genetics in *T. kodakarensis* is the validation of predicted gene functions and determination of functions for genes of interest. One of the first examples of applied genetics in *T. kodakarensis* was the identification of the enzyme responsible for the fructose-1,6-bisphosphatase (FBPase) activity during gluconeogenesis (143). At the time, two genes, *fbp* (126) and *imp* (156, 170) were candidates for the true hyperthermophilic FBPase. Generating knockout strains for both genes clearly showed that *fbp* was the FBPase, both phenotypically due to the inability of the ∆*fbp*-8J mutant to grow on pyruvate and by lack of FBPase activity in ∆*fbp*-8J cell extracts (143). A similar approach showed that 3-hexulose-6-phosphate synthase (*hps*) and 6-phospho-3-hexulosisomerase (*phi*) are required for synthesis of nucleosides through the ribulose monophosphate pathway (122).
Table 1.4. Selected *T. kodakarensis* strains essential for genetic tool development

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Phenotype(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOD1</td>
<td>Wild type</td>
<td>Wild type</td>
<td>(113)</td>
</tr>
<tr>
<td>KU25</td>
<td>ΔpyrF96T</td>
<td>5FOA&lt;sup&gt;R&lt;/sup&gt;, Ura&lt;sup&gt;+&lt;/sup&gt;</td>
<td>(142)</td>
</tr>
<tr>
<td>KU26</td>
<td>ΔpyrF96T, ΔtrpE::pyrF</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;</td>
<td>(142)</td>
</tr>
<tr>
<td>KW128</td>
<td>ΔpyrF; ΔtrpE::pyrF</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;</td>
<td>(141)</td>
</tr>
<tr>
<td>KuW1</td>
<td>ΔpyrF; ΔtrpE::3′ region of trpE-pyrF</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;</td>
<td>(141)</td>
</tr>
<tr>
<td>KuWH1</td>
<td>ΔpyrF; ΔtrpE; ΔhisD::3′ region of hisD-pyrF</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;, His&lt;sup&gt;-&lt;/sup&gt;</td>
<td>(141)</td>
</tr>
<tr>
<td>TS372</td>
<td>KW128; ΔPTK1761::TrpE-P&lt;sub&gt;prom&lt;/sub&gt;-TK1761</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>(138)</td>
</tr>
<tr>
<td>TS416</td>
<td>TS372; TK1761 stop mutation at codon 3</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>(138)</td>
</tr>
<tr>
<td>TS419</td>
<td>TS372; TK1761 disrupted TATA box</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;+&lt;/sup&gt;</td>
<td>(138)</td>
</tr>
<tr>
<td>PdaD</td>
<td>ΔpyrF; ΔTK0148-0149</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Agm&lt;sup&gt;-&lt;/sup&gt;</td>
<td>(53)</td>
</tr>
<tr>
<td>TS517</td>
<td>KW128; ΔTK0644</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;, 6MP&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(137)</td>
</tr>
<tr>
<td>TS538</td>
<td>TS517; ΔTK1761-1763</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;, 6MP&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(137)</td>
</tr>
<tr>
<td>TS541</td>
<td>TS517; ΔTK1827</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;, 6MP&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(137)</td>
</tr>
<tr>
<td>TS1079</td>
<td>TS517; ΔTK1827; ΔTK1761-1763</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;, 6MP&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(137)</td>
</tr>
<tr>
<td>TS559</td>
<td>TS517; ΔTK0149</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;, 6MP&lt;sup&gt;R&lt;/sup&gt;, Agm&lt;sup&gt;-&lt;/sup&gt;</td>
<td>(137)</td>
</tr>
<tr>
<td>TS561</td>
<td>TS517; ΔTK0148-0149</td>
<td>Ura&lt;sup&gt;+&lt;/sup&gt;, Trp&lt;sup&gt;-&lt;/sup&gt;, 6MP&lt;sup&gt;R&lt;/sup&gt;, Agm&lt;sup&gt;-&lt;/sup&gt;</td>
<td>(137)</td>
</tr>
</tbody>
</table>

*pyrF=TK2276, trpE=TK0254, hisD=TK0244, pdaD=TK0149.*
Genetic methodologies have also enhanced the understanding of transcription and translation in *T. kodakarensis*. Many Archaeal genomes encode more than one copy of the Archaeal homologue of the general transcription factor TFIIB (TFB), though the reason for this is not clear. One simple explanation is functional redundancy, though it has also been hypothesized that different TFBs may direct transcription initiation from different promoters (13). *T. kodakarensis* encodes two TFB homologues, TK1280 and TK2287. Transcription from various promoters *in vitro* shows that either TFB promotes transcription initiation. Deletion of either TK1280 or TK2287 results in no detectable growth phenotype, whereas a double knockout strain was never successfully constructed, supporting the functional redundancy hypothesis and that TFB is essential, (134).

A shuttle vector expression system was used to determine the consensus ribosome binding site (RBS) and the effect of alternative start codons on expression. Both GTG and TTG can function as start codons and code for methionine, however the expression level is significantly reduced by TTG as a start codon (136). A reporter system allowed for the demonstration of nonsense mediated polarity in an Archaeal operon. By creating stop codons in different positions within a transcribed gene, and monitoring expression of downstream genes, it was clear that an early stop codon caused transcriptional termination. This finding also suggests that some unidentified termination factor may be responsible for transcriptional termination (138).

The application of genetics has also been invaluable for identifying unknown components in DNA replication. Since Archaeal DNA replication is very well studied in many organisms, there are a number of known replisome components. In one study, 19 different *T. kodakarensis* strains were generated that express His-tagged replication proteins. From these strains, it was possible to isolate native complexes of replication proteins and identify specific interactions between proteins. Many of these interactions were already well established, but novel proteins were also identified, that were previously not known to play a role in replication.
One of these proteins is a novel nuclease associated with the GINS complex. Another interesting component of the replication machinery is the MCM helicase. Whereas most Archaea only have one MCM helicase, *T. kodakarensis* encodes three MCM homologs. Deletion of the TK0096 and TK1361 genes, both singly, and in combination resulted in no detectable growth defects in these strains, however the TK1620 gene was never successfully deleted suggesting that it might be essential and that TK1620 functions as the replicative MCM helicase. It is likely that the other two MCM homologs are the remnants of viral infections, and any potential functions in *T. kodakarensis* remain unknown.

With the recent focus on biofuels, hydrogen production by *T. kodakarensis* has become an intense area of study. Reported enzyme activities from *P. furiosus* (36, 37, 89, 93, 94, 98, 99, 101, 139, 140, 147, 178), as well as analysis of *T. kodakarensis* culture products (78) and the genome annotation (54), provided a substantial basis for the prediction of the mechanism and regulation of hydrogen metabolism in *T. kodakarensis*. Genetic analysis of several of the enzymes has allowed a direct test of these predictions and genes involved in hydrogen metabolism. This analysis has resulted in a refined picture of hydrogen production. Deletion of either the membrane-bound hydrogenase (Mbh) operon or the sulfur responsive transcriptional regulator SurR resulted in no hydrogen production, indicating that most (if not all) hydrogen produced by *T. kodakarensis* is produced by Mbh, and that transcriptional activation by SurR is required for its expression (135). Deletion of the cytosolic hydrogenase operon resulted in increased hydrogen production but reduced culture density, indicating that it is involved in hydrogen consumption, which is coupled to growth. Deletion of the membrane-bound oxidoreductase (Mbx) but not NADPH sulfur reductase (NSR) resulted in growth defects with sulfur (135), supporting the predicted necessity of Mbx for growth with sulfur (147). Deletion of NSR did increase hydrogen production, consistent with a role in channeling reductant flow to Mbx (135). Surprisingly, deletion of the alanine aminotransferase or either of two ferredoxin NADP oxidoreductase operons resulted in no detectable effect on growth or hydrogen
<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Phenotype(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δfbp-8J</td>
<td>KW128; Δfbp::trpE</td>
<td>No growth on pyruvate</td>
<td>(143)</td>
</tr>
<tr>
<td>Δimp-2A</td>
<td>KW128; Δimp::trpE</td>
<td>Undetermined</td>
<td>(143)</td>
</tr>
<tr>
<td>Δhps-phi-7A</td>
<td>KW128; Δhps-phi::trpE</td>
<td>Growth requires nucleosides</td>
<td>(122)</td>
</tr>
<tr>
<td>TS601</td>
<td>TS517; ΔTK0096</td>
<td>No detected growth defect</td>
<td>(124)</td>
</tr>
<tr>
<td>TS602</td>
<td>TS517; ΔTK1361</td>
<td>No detected growth defect</td>
<td>(124)</td>
</tr>
<tr>
<td>TS604</td>
<td>TS517; ΔTK0096; ΔTK1361</td>
<td>No detected growth defect</td>
<td>(124)</td>
</tr>
<tr>
<td>TS1100</td>
<td>TS517; ΔTK1094</td>
<td>No detected difference in growth or H₂ production.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1101</td>
<td>TS517; ΔTK1086</td>
<td>No H₂ production. Growth on pyruvate requires S⁰.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1102</td>
<td>TS517; ΔTK1260-1261</td>
<td>Slight reduction in H₂ production on pyruvate.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1103</td>
<td>TS517; ΔTK2080-2093</td>
<td>No H₂ production. Growth requires S⁰.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1104</td>
<td>TS517; ΔTK2080-2093; ΔTK1094</td>
<td>No H₂ production. Growth requires S⁰.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1105</td>
<td>TS517; ΔTK1214-1226</td>
<td>Slow growth and reduced cell density with S⁰.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1106</td>
<td>TS517; ΔTK2069-2072</td>
<td>Increased H₂ production. Reduced cell density on pyruvate.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1107</td>
<td>TS517; ΔTK1684-1685</td>
<td>No detected difference in growth or H₂ production.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1108</td>
<td>TS517; ΔTK1325-1326</td>
<td>No detected difference in growth or H₂ production.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1109</td>
<td>TS517; ΔTK1299</td>
<td>Increased H₂ production on pyruvate.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1110</td>
<td>TS517; ΔTK1299; Δ2069-2072</td>
<td>Slow growth and reduced cell density, but increased H₂ production with pyruvate.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1111</td>
<td>TS517; ΔTK1214-1226; ΔTK1299</td>
<td>Slow growth and reduced cell density with S⁰. Increased H₂ production on pyruvate.</td>
<td>(135)</td>
</tr>
<tr>
<td>TS1113</td>
<td>TS517; ΔTK1684-1685; ΔTK1325-1326</td>
<td>No detected difference in growth or H₂ production.</td>
<td>(135)</td>
</tr>
</tbody>
</table>
production. The explanation for this is not clear but could be due to redundant pathways for reductant recycling. The SipA and SipB proteins were predicted by microarray to be involved with growth on sulfur (36, 148). It is puzzling that deletion of SipA and SipB resulted in a slight increase in hydrogen production on pyruvate but no growth defects on sulfur (135). So while SipA and SipB may have a function during growth on sulfur, it is apparently not essential for growth, and its function remains unclear.

Genetic manipulations in *P. furiosus*

Bioinformatics and biochemistry have made enormous contributions and that work has set the stage for the use of genetics as a tool however many of the conclusions and predictions drawn from these studies can only be tested using genetic methods. The ability to make knockout mutants is critical for testing the role of uncharacterized proteins implicated in cellular pathways. The ability to express affinity tagged proteins will facilitate the expression of modified proteins and the determination of protein structures. It has been noted that *P. furiosus* proteins purified from *E. coli* do not always contain their proper metal ions (166) or those that require modification are not always processed properly. This issue is eliminated by homologous expression in *P. furiosus*. As demonstrated in *T. kodakarensis*, homologous expression could also facilitate the identification of novel proteins in well-known pathways. The ability to modify native proteins would be useful for elucidating catalytic mechanisms and protein interactions.

Recognizing the potential of genetics for the advancement of research in *P. furiosus*, a recent effort to develop genetic tools has been the focus of the collaboration between the Westpheling and Adams labs. A ΔpyrF mutant strain was generated, providing a suitable background strain for genetic manipulation (94). During the course of this work, a mutant, designated COM1, was isolated that is naturally and efficiently competent for DNA uptake. While the nature of this competence is not known, the COM1 strain is naturally competent for DNA uptake, and the combined frequencies of transformation of recombination allow marker
replacement using PCR products constructed to make deletions. Using a plasmid integration/excision method, the utility of this strain was first demonstrated in making deletions of the two soluble hydrogenases (94). Chapter 3 of this dissertation reports an analysis of the recombinogenic properties of COM1 and its use for the rapid construction of markerless deletions.
CHAPTER 2
DEFINING COMPONENTS OF THE CHROMOSOMAL ORIGIN OF REPLICATION OF THE
HYPERHEMOPHILIC ARCHAEON, *PYROCOCCUS FURIOSUS*, NEEDED FOR
CONSTRUCTION OF A STABLE REPLICATING SHUTTLE VECTOR¹

Environmental Microbiology*. 77:6343-6349.
Reprinted here with permission of the publisher.
Abstract
We report the construction of a series of replicating shuttle vectors that consist of a low-copy-number cloning vector for Escherichia coli and functional components of the origin of replication (oriC) of the chromosome of the hyperthermophilic archaeon Pyrococcus furiosus. In the process of identifying the minimum replication origin sequence required for autonomous plasmid replication in P. furiosus, we discovered that several features of the origin predicted by bioinformatic analysis and in vitro binding studies were not essential for stable autonomous plasmid replication. A minimum region required to promote plasmid DNA replication was identified, and plasmids based on this sequence readily transformed P. furiosus. The plasmids replicated autonomously and existed in a single copy. In contrast to shuttle vectors based on a plasmid from the closely related hyperthermophile Pyrococcus abyssi for use in P. furiosus, plasmids based on the P. furiosus chromosomal origin were structurally unchanged after transformation and were stable without selection for more than 100 generations.

Introduction
Pyrococcus species are hyperthermophilic marine archaea that grow anaerobically at temperatures near and above 100°C (19). Their interesting biology, evolutionary history, and potential commercial utility make them an important group to study (9, 47). We recently reported an efficient method for DNA transformation for Pyrococcus furiosus (27), which provides the basis for the development of methods for genetic manipulation. A markerless deletion of the orotidine monophosphate (OMP) decarboxylase (pyrF) locus was constructed in the P. furiosus genome, generating a mutant that is a uracil auxotroph and resistant to 5-fluoroorotic acid (5FOA). Complementation by the wild-type pyrF allele restored uracil prototrophy and 5FOA sensitivity to this strain. We took advantage of this marker to construct replicating shuttle vectors for use in P. furiosus.
Replicating vectors facilitate a variety of genetic manipulations, and vectors capable of shuttling between any host of interest and *Escherichia coli*, perhaps the most convenient biological host for DNA manipulation, are especially useful. Naturally occurring plasmids have been identified in archaeal species (7, 10–12, 16, 22, 51), and some of them have been used to develop shuttle vectors in the haloarchaea (25) and methanogens (49). A shuttle vector between *E. coli* and *Thermococcus kodakarensis*, a close relative of *P. furiosus* (4, 36), was reported, which combined a commercial *E. coli* cloning vector with a naturally occurring plasmid, pTN1, from *Thermococcus nautilus* (44, 46). This plasmid was used successfully to express a hemagglutinin (HA) epitope-tagged version of RpoL, a subunit of *T. kodakarensis* RNA polymerase, illustrating the utility of such vectors (44). Plasmid pGT5, a naturally occurring plasmid in *Pyrococcus abyssi* (17), was used to create a shuttle vector capable of replicating in both *P. abyssi* and *E. coli* (29), and most recently, a shuttle vector based on pGT5 was used to express a gene encoding His6-tagged subunit D of RNA polymerase in *P. furiosus* (50). Our attempts to construct a stable shuttle vector for *P. furiosus* based on pGT5 were unsuccessful.

*Pyrococcus* species are predicted to have a single origin of replication with both bacterial and eukaryotic features (28, 37). Other archaea (5, 30, 42) have multiple replication origins, and eukaryotes have hundreds or even thousands of replication origins (41). Eukaryotic replication origins are poorly defined in terms of sequence as well as the number and nature of auxiliary proteins that facilitate their function, and there is increasing evidence that origin maintenance in many eukaryotes is controlled by epigenetic factors whose function is also poorly understood (3, 45). *Saccharomyces cerevisiae* is a notable exception in that it has well-defined origins called autonomously replicating sequence (ARS) elements, generally 100 to 200 bp in length, containing multiple cis-acting sequence elements (38) and binding sites for the origin recognition complex (6). This complex is composed of origin recognition complex (Orc) proteins 1 to 6 (6) and a DNA binding protein, Cdc6, first identified in yeast as a gene whose product is involved in cell cycle control (15, 23). Most bacteria have a single origin sequence flanked by
polymerases and dnaA (35), which encodes a protein that binds DNA within the origin (20) and functions to recruit DnaB and DnaC to form the replication initiation complex. While DnaA is only distantly related to Cdc6, it provides the same function in nucleating the protein complex.

Replication origin regions in both bacteria and archaea are AT rich (5, 41).

The origin of replication in several archaeal genera was first predicted by the cumulative GC skew, an early bioinformatic method used to find prokaryotic replication origins (21). In P. abyssi the location of the chromosomal replication origin was predicted to be within an 80-kb region that contained a large intergenic space and genes for several putative replication proteins (28). This intergenic space, neighboring genes, and features are conserved in P. furiosus (Fig. 1). While the organization of putative protein binding sites and replicative proteins around the P. furiosus oriC is bacterial in nature, the proteins that likely bind the sequences are eukaryotic. In the location where dnaA often resides in bacteria, there is a single gene homologous to the cdc6/orc1 eukaryotic replication proteins (40). Previous studies of the function of the origin sequences of Halobacterium sp. strain NRC-1 in vivo showed that a 750-bp region containing oriC requires the adjacent orc7 gene in cis for autonomous plasmid replication (8). In Sulfolobus solfataricus, there are three origins of replication (30, 42), and for each origin of replication, a cdc6 gene is adjacent but is not required in cis for the origin to function in plasmid replication (14). Evidence for the location of the P. furiosus origin and for the role of putative replication proteins is based entirely on bioinformatic and in vitro analyses. Cells emerging from replication arrest were shown previously to incorporate radiolabeled nucleotides in the DNA sequence at the predicted origin (37). Chromatin immunoprecipitation of the Cdc6/Orc1 protein showed a significant enrichment of the predicted oriC DNA region (32), and DNA binding studies showed that the Sulfolobus Cdc6/Orc1 protein binds this region of P. abyssi as well as Sulfolobus DNA in vitro (33, 42). The sequence of the sites of protein binding (referred to as origin recognition boxes [ORBs]) suggested that inverted repeat elements flanking the predicted origin, conserved among archaea, were involved in replication complex
Figure 2.1. Region of the *P. furiosus* chromosome predicted to contain the origin of replication, oriC (A), with an expanded view of the intergenic space (B). ORBs are indicated in red, mini-ORBs are shown in pink, and the putative unwinding site is shown in green. The chromosomal regions that were cloned into pJFW017 to produce various plasmids are indicated by black lines below the diagrams. Conserved palindromes are marked by inverted black arrows. Inserts that resulted in plasmids capable of autonomous replication in *P. furiosus* are indicated by thick black lines.
formation and binding. The binding of the Cdc6/Orc1 protein to the intergenic space resulted in an unwinding of the DNA, supporting the notion that this was the location of the origin (34).

While the intergenic space is AT rich, the DNA sequence is not conserved across Pyrococcus species. Further evidence for the position of the origin comes from the fact that DNA at the origin contains a transient replication bubble produced by bidirectional DNA polymerization, and the structure may be isolated by displaying digested total genomic DNA, from actively dividing cells, on two-dimensional gels. Sequences containing a replication bubble were located to a 1-kb fragment that included the intergenic space, possibly overlapping the cdc6 gene (37).

Bioinformatic analyses identified several 13-bp mini-ORB repeats in and around the intergenic space suggested previously to be involved in origin recognition and potential Cdc6/Orc1 binding (26, 33).

Here we present an in vivo analysis of sequences at the origin of DNA replication in the P. furiosus chromosome and their function in autonomous plasmid replication. DNA fragments containing the putative chromosomal origin as well as predicted protein binding sites and the cdc6/orc1 gene were cloned into an E. coli plasmid vector and tested for the ability to promote autonomous DNA replication in P. furiosus. The cloning of the origin sequence without the cdc6/orc1 gene did not affect its ability to function, suggesting that this protein, while likely required for DNA replication, is not required in cis. The location of the origin predicted from in vitro analyses was confirmed. Only two of the predicted ORB sequences, however, are required for autonomous plasmid replication. We used this origin sequence in combination with the pSC101 origin from Salmonella enterica serovar Panama (13) for replication in E. coli to construct a replicating shuttle vector for P. furiosus that is stable in a single copy without selection for more than 100 generations and is structurally unchanged after transformation into P. furiosus and back-transformation into E. coli.
Materials and Methods

Strains, media, and growth conditions. *E. coli* strain DH5α was used for plasmid DNA constructions and preparations. Standard techniques for *E. coli* were performed as described previously (43). Apramycin was used for selection at 50 mg/ml. Wild-type strain *P. furiosus* DSM 3638 (19) and the *P. furiosus* COM1 ΔpyrF strain (27) were grown anaerobically in a defined medium with cellobiose as the carbon source (27) at 90°C for 16 to 20 h in 100-ml serum bottles containing 50 ml of liquid medium or on medium solidified with Phytagel (1%, wt/vol; Sigma) for 60 h. The *P. furiosus* COM1 ΔpyrF strain was used as a host for all DNA transformation experiments. *P. abyssi* strain GE5 (16, 18) was grown in a liquid base salts medium (1) containing 0.5% (wt/vol) casein hydrolysate and 0.2% (wt/vol) elemental sulfur for 40 to 48 h at 90°C under anaerobic conditions. Total genomic DNA was isolated as described previously (27), except that DNA was precipitated with isopropanol and resuspended with 50 µl TE buffer (10 mM Tris, 1 mM EDTA) containing RNase A (100 ng/ml).

Construction of vectors and transformation of *P. furiosus*. To construct pJFW027 and pJFW018, PCR products containing the indicated regions of the chromosome (Fig. 1) were ligated into a linear DNA fragment containing the entire pJFW017 plasmid (Fig. 2), also generated by PCR using primers JF266 and JF267. To generate plasmids pJFW031 to pJFW044, primers with restriction sites added to the 5′ ends were used to allow the digestion and subsequent directional cloning of origin-containing fragments into pJFW017. The 5′ end of each fragment contained a BamHI site, and the 3′ end contained a ClaI site. The PCR amplification of pJFW017 was done by use of primers JF266.2 and JF267.2 with the same restriction sites. Primers used in these constructions are listed in Table 1, and DNA sequences of the primers are shown in Table S1 in the supplemental material. *E. coli* strain DH5α cells were transformed by electroporation in a 2-mm-gap cuvette at 2.5 V. Plasmid DNA was isolated from liquid cultures by using QIAprep Spin Miniprep columns (Qiagen Inc.). For DNA transformations, the *P. furiosus* COM1 ΔpyrF strain was grown for 16 to 20 h in defined liquid
Figure 2.2. Construction of pJFW027. A linear DNA fragment containing the entire sequence of pJFW017 was generated by PCR amplification using primers JF266 and JF267 and ligated into the origin fragment indicated in Fig. 1, also generated by PCR amplification using primers JF268 and JF282. Plasmids containing the various origin fragments described in the legend of Fig. 1 were cloned into pJFW017 for testing.
medium containing 20 µM uracil. Plasmid DNA (100 to 200 ng) was added to 100 µl of culture and plated onto the defined medium without uracil. Prototrophic colonies were inoculated into liquid medium for DNA isolation. The presence of plasmid sequences in *P. furiosus* was confirmed by PCR amplification of the *aac* gene, present only on the plasmid, from *P. furiosus* total genomic DNA by using primers JF263 and JF264 (Table 1).

**Assessment of plasmid maintenance, stability, and copy number.** To assess plasmid maintenance, *P. furiosus* transformants were serially subcultured every 24 h for 10 days in selective and nonselective liquid media. After each passage, the culture was diluted 100-fold with base salts, and 30 µl of diluted culture was plated onto selective medium to determine the number of prototrophic colonies, i.e., those maintaining the plasmid. The cell density of the liquid culture was determined by direct cell counting using a Petroff-Hausser counting chamber. To assess the structural stability of the plasmid, total genomic DNA isolated from five independent *P. furiosus* transformants containing pJFW027 was used to back-transform *E. coli* for plasmid isolation and restriction digestion analysis. To determine plasmid copy numbers, total genomic DNA was isolated from *P. furiosus* plasmid transformants and digested twice with 10 U of HpaI for 120 min at 37°C. The restriction fragments were separated by electrophoresis in a 1.0% (wt/vol) agarose gel and transferred onto nylon membranes (Roche, Manheim, Germany). Primers GL021 and GL023 (27) were used to amplify the glutamate dehydrogenase (*gdh*) promoter from wild-type *P. furiosus* total genomic DNA to generate a digoxigenin (DIG)-labeled probe by random priming with DIG High Prime DNA Labeling and Detection Starter Kit I (Roche, Manheim, Germany). The membrane was incubated at 42°C and washed at 65°C. Band intensities were determined by using a Storm 840 PhosphorImager (GE Healthcare) equipped with ImageQuant v.5.4 software (Molecular Dynamics).
Table 2.1. Plasmid transformation efficiencies\(^a\).

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>oriC insert position</th>
<th>5' primer</th>
<th>3' primer</th>
<th>Transformation efficiency (no. of transformants/µg of plasmid DNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJFW017</td>
<td>None</td>
<td></td>
<td></td>
<td>8.2 \times 10^2</td>
</tr>
<tr>
<td>pJFW018</td>
<td>15382–16226</td>
<td>JF268</td>
<td>JF269</td>
<td>9.1 \times 10^5</td>
</tr>
<tr>
<td>pJFW027</td>
<td>15382–17576</td>
<td>JF268</td>
<td>JF282</td>
<td>5.8 \times 10^5</td>
</tr>
<tr>
<td>pJFW031</td>
<td>15382–16228</td>
<td>JF268.2</td>
<td>JF269.2</td>
<td>6.6 \times 10^5</td>
</tr>
<tr>
<td>pJFW032</td>
<td>15382–16187</td>
<td>JF306.2</td>
<td>JF269.2</td>
<td>8.0 \times 10^4</td>
</tr>
<tr>
<td>pJFW033</td>
<td>15382–16034</td>
<td>JF305.2</td>
<td>JF269.2</td>
<td>6.5 \times 10^5</td>
</tr>
<tr>
<td>pJFW034</td>
<td>15382–15890</td>
<td>JF304.2</td>
<td>JF269.2</td>
<td>5.7 \times 10^2</td>
</tr>
<tr>
<td>pJFW035</td>
<td>15382–15786</td>
<td>JF303.2</td>
<td>JF269.2</td>
<td>1.4 \times 10^3</td>
</tr>
<tr>
<td>pJFW037</td>
<td>15382–15705</td>
<td>JF301.1</td>
<td>JF269.2</td>
<td>1.4 \times 10^2</td>
</tr>
<tr>
<td>pJFW038</td>
<td>15492–16034</td>
<td>JF305.2</td>
<td>JF339</td>
<td>4.7 \times 10^1</td>
</tr>
<tr>
<td>pJFW039</td>
<td>15561–16034</td>
<td>JF305.2</td>
<td>JF345</td>
<td>&lt;4.7 \times 10^1</td>
</tr>
<tr>
<td>pJFW042</td>
<td>15746–16034</td>
<td>JF305.2</td>
<td>JF348</td>
<td>&lt;4.7 \times 10^1</td>
</tr>
<tr>
<td>pJFW043</td>
<td>15813–16034</td>
<td>JF305.2</td>
<td>JF349</td>
<td>&lt;4.7 \times 10^1</td>
</tr>
</tbody>
</table>

\(^a\) Genomic locations are based on the numbering convention of the *Pyrococcus furiosus* (accession number NC_003413.1) genome sequence deposited in GenBank. The detection threshold was 4.7 \times 10^2 transformants per µg of plasmid DNA.
Results and Discussion

The *P. furiosus* chromosomal replication origin functions for stable autonomous plasmid replication. Attempts to construct a stable replicating shuttle vector based on plasmid pGT5 from *P. abyssi* for use in *P. furiosus* were unsuccessful. Plasmids based on pGT5 exist in high copy numbers in *P. abyssi* (29, 50) but show a significantly reduced copy number in *P. furiosus* (50) and cannot be used for the selection of transformants in the closely related *Thermococcus kodakarensis* (44). In an attempt to construct a shuttle vector based on pGT5, the entire plasmid was cloned into pJFW017 that contained a pSC101 origin for replication in *E. coli*, an apramycin resistance gene for selection in *E. coli*, and a wild-type copy of the *P. furiosus* pyrF gene for the selection of transformants in the *P. furiosus* COM ΔpyrF strain. A fragment containing the entire pGT5 plasmid sequence was produced by PCR amplification with primers JF254 and JF270, linearizing the plasmid at a site within pGT5 previously shown not to interfere with replication functions (17, 31), to produce pJFW019 (see Fig. S1 in the supplemental material). This plasmid readily transformed *P. furiosus* but was rapidly lost without selection (Table 2) and showed internal deletions after transformation into *P. furiosus* and subsequent back-transformation into *E. coli* (Fig. S2). Other attempts to use pGT5 for the construction of shuttle vectors in *T. kodakarensis* were similarly unsuccessful (44).

To test whether the predicted *P. furiosus* chromosomal origin of replication could promote stable autonomous plasmid replication, a fragment of the chromosome containing the predicted origin sequence and the gene encoding Cdc6/Orc1 (Fig. 1) was cloned into an *E. coli* plasmid, pJFW017 (Fig. 2), to make pJFW027. We used the transformation efficiency as an assay for plasmid replication (8). As shown in Table 1, transformants of pJFW027 were observed at a frequency of $5.8 \times 10^5$ transformants per μg of plasmid DNA. No transformants were observed in the absence of added plasmid DNA, and while some transformants were obtained in experiments with pJFW017, which does not contain an origin sequence ($8.2 \times 10^2$ transformants per μg of plasmid DNA), this is most likely due to integration by homologous
Table 2.2. Maintenance of plasmids in *P. furiosus*\(^a\)

<table>
<thead>
<tr>
<th>Passage</th>
<th>pJFW018</th>
<th></th>
<th></th>
<th>pJFW027</th>
<th></th>
<th></th>
<th>pJFW019</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ura</td>
<td>-ura</td>
<td>+ura</td>
<td>-ura</td>
<td>+ura</td>
<td>-ura</td>
<td>+ura</td>
<td>-ura</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>187</td>
<td>256</td>
<td>190</td>
<td>194</td>
<td>73</td>
<td>112</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>132</td>
<td>217</td>
<td>97</td>
<td>203</td>
<td>54</td>
<td>268</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>112</td>
<td>232</td>
<td>132</td>
<td>143</td>
<td>11</td>
<td>138</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>146</td>
<td>117</td>
<td>154</td>
<td>165</td>
<td>6</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>87</td>
<td>276</td>
<td>100</td>
<td>113</td>
<td>0</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>138</td>
<td>197</td>
<td>151</td>
<td>201</td>
<td>0</td>
<td>104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>144</td>
<td>263</td>
<td>111</td>
<td>122</td>
<td>0</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>118</td>
<td>242</td>
<td>87</td>
<td>131</td>
<td>0</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>93</td>
<td>213</td>
<td>114</td>
<td>218</td>
<td>0</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>107</td>
<td>183</td>
<td>112</td>
<td>169</td>
<td>0</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Transformants containing each plasmid were serially passaged in liquid medium with uracil (+ura) or without uracil (−ura). Following each passage, a diluted culture was plated onto selective medium to determine the number of prototrophic cells remaining.
recombination between the *gdh* promoter region (283 bp), driving the transcription of the *pyrF* gene on the plasmid, and the *gdh* locus in the chromosome. In fact, we have observed the integration of nonreplicating plasmid DNA by homologous recombination at the same frequency (27). The transformation frequency of pJFW027 was a thousandfold greater than that of pJFW017, indicating that the plasmid was replicating autonomously. PCR amplification of the apramycin resistance gene, contained only on the plasmids, was used to confirm the presence of plasmid DNA in the transformants. A 950-bp product containing this sequence was obtained from transformant total genomic DNA but not from the wild-type or the *P. furiosus* COM1 Δ*pyrF* strain (see Fig. S3 in the supplemental material).

Attempts to isolate a significant quantity of pJFW027 plasmid DNA from *P. furiosus* were unsuccessful. This is perhaps not surprising, since quantities of plasmids based on the chromosomal origin would be expected to be low or 1 copy per chromosome. In lieu of direct plasmid isolation, we chose to rescue the plasmid by back-transformation to *E. coli*. That transformants contained a replicating copy of the plasmid was shown by using total genomic DNA isolated from 5 independent plasmid transformants to back-transform *E. coli* strain DH5α selecting for apramycin resistance. Back-transformants were obtained at frequencies of $10^4$ transformants per µg of DNA, an underestimate of the plasmid transformation, since this frequency is based on the number of transformants per microgram of total genomic DNA, and only covalently closed circular plasmid DNA is capable of transforming *E. coli* strain DH5α at this frequency (24). Plasmid DNA isolated from these back-transformants was indistinguishable from the pJFW027 plasmid DNA used to transform *P. furiosus* by using restriction digestion analysis, indicating that there were no gross rearrangements during transformation and replication in *P. furiosus* or subsequent back-transformation to *E. coli*. When the P*gdh* fragment, specific to plasmid pJFW027, was used as a probe for the Southern hybridization of total genomic DNA from the *P. furiosus* transformants with DNA digested with either EcoRV (data not shown) or HpaI (Fig. 3), which have a single cleavage site within the plasmid, a single band was
detected, showing that the plasmid DNA was not integrated into the chromosome and existed as an autonomously replicating molecule.

To examine plasmid maintenance, transformants of pJFW027 and pJFW018 were serially subcultured in liquid medium with or without selection (i.e., in the absence or presence of uracil) for more than 100 generations and then plated onto minimal medium without uracil to assay plasmid maintenance. No loss of plasmids with oriC was detected even without selection (Table 2). In addition, the restriction pattern of plasmid DNA isolated from E. coli after transformation into P. furiosus and subsequent transformation back into E. coli remained unchanged, indicating that no rearrangements of the plasmid DNA occurred (Fig. 4 [data for 1 of 10 transformants tested are shown]).

The cdc6/orc1 open reading frame is not required in cis for replication origin function. In bacteria, oriC is often, but not always, adjacent to dnaA. In E. coli, oriC is between gidA and mioC (which encodes another replication protein), approximately 43 kb from dnaA, and can function for the autonomous replication of plasmids without cis-acting replicating components (39, 48). In the chromosome of Halobacterium NRC-1, oriC requires the adjacent orc7 gene in cis for autonomous plasmid replication (8). In Sulfolobus solfataricus, there are three origins of replication, and for each origin of replication, a cdc6 gene is adjacent but is not required in cis for the origin to function in autonomous plasmid replication (14). In the sequenced Pyrococcus species P. furiosus, P. abyssi, and P. horikoshii as well as the closely related T. kodakarensis, there is a single oriC adjacent to a cdc6/orc1 homologue, but nothing is known about the requirement of this protein for oriC function. To test whether cdc6/orc1 was required in cis for autonomous plasmid replication in P. furiosus, a fragment containing only oriC was cloned into parent plasmid pJFW017 to generate pJFW18 (Fig. 1). As shown in Table 1, plasmids containing the fragment with only the oriC sequence without the cdc6/orc1 gene transformed P. furiosus as efficiently and were maintained as stably as plasmid pJFW27.
Figure 2.3. Determination of copy number for pJFW027 in *P. furiosus*. (A) Diagram of the chromosomal region, including the *gdh* open reading frame. Hpal sites are indicated, as are the locations of primers used to generate the *gdh* hybridization probe. (B) Southern blot of pJFW027 transformants. Lanes 1 to 10, DNA isolated from transformants and digested with Hpal; lanes 11 and 12, DNA from *P. furiosus* wild-type and COM1 ΔpyrF strains, respectively; lane 13, pJFW027 plasmid DNA purified from *E. coli*. 
Figure 2.4. Restriction analysis of plasmid DNA before and after transformation of *P. furiosus* and back-transformation to *E. coli*. Lanes M, 1-kb DNA ladder; lanes 1 to 4, pJFW018 plasmid DNA isolated from *E. coli* DH5α (lane 1) and digested with Accl (lane 2), AvaI (lane 3), and HindIII (lane 4); lanes 5 to 8, plasmid DNA isolated from *E. coli* DH5α back-transformed from *P. furiosus* transformants (lane 5) and digested with Accl (lane 6), AvaI (lane 7), and HindIII (lane 8).
carrying the \textit{cdc6/orc1} gene, suggesting that the \textit{cdc6/orc1} gene is not required in \textit{cis} for stable autonomous plasmid replication.

**Only two of the predicted ORB sequences and part of the predicted chromosomal origin sequence are required for plasmid replication.** The origin region was predicted previously to contain several ORB and mini-ORB sequences (33), suggested to be binding sites for the replication initiation protein Cdc6/Orc1, which is presumed to facilitate the nucleation of the replication complex. Our analysis using the BLASTN 2.2.24+ algorithm (2) identified three ORB repeats and several mini-ORB repeats by the self-alignment of the sequence of the genomic region containing \textit{oriC} and neighboring genes. These results are similar but not identical to those described previously by Matsunaga et al. (33), in that we found a clustering of mini-ORB repeats in and around \textit{oriC}, but the exact number and position of these mini-ORB repeats were different. In addition, we identified two conserved palindromic sequences (Table 3) conserved in all sequenced \textit{Pyrococcus} species. One of them contains compensating changes within the sequence that retain the perfect palindromic structure, suggesting that these are not random sequences within this highly repetitive region of DNA and may potentially be binding sites for other replication proteins or have a structural role in replication. These palindromes are not present in the \textit{oriC} region of the closely related \textit{Thermococcus} species, however, suggesting that if they have a function, it may be specific to \textit{Pyrococcus}. To test whether these sequences were required for autonomous plasmid replication, plasmids containing various portions of the region around the predicted origin were constructed and tested for the ability to replicate. The smallest insert able to promote autonomous plasmid replication was the 653-bp fragment cloned into pJFW033. As shown in Table 1, only two of the three ORB sequences, and only a part of the sequence predicted to contain the origin, were required for plasmid replication. The predicted unwinding site, for example, is apparently not required for autonomous plasmid replication.
Table 2.3. Conserved palindromic sequences within the *Pyrococcus* oriC region

<table>
<thead>
<tr>
<th>Species (GenBank accession no.)</th>
<th>Sequence</th>
<th>Genomic position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Palindrome 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. furiosus</em> (NC_003413.1)</td>
<td>ATATTTAAATAT</td>
<td>15641–15674</td>
</tr>
<tr>
<td><em>P. abyssi</em> (NC_000868.1)</td>
<td>TATTTAAATA</td>
<td>123223–123232</td>
</tr>
<tr>
<td><em>P. horikoshii</em> (NC_000961.1)</td>
<td>TATTTAAATA</td>
<td>111307–111316</td>
</tr>
<tr>
<td><strong>Palindrome 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. furiosus</em> (NC_003413.1)</td>
<td>ATTAagTTAAcTAAT</td>
<td>15809–15824</td>
</tr>
<tr>
<td><em>P. abyssi</em> (NC_000868.1)</td>
<td>ATTAagTTAAcTAAT</td>
<td>123072–123087</td>
</tr>
<tr>
<td><em>P. horikoshii</em> (NC_000961.1)</td>
<td>ATTAagTTAAcTAAT</td>
<td>111155–1111</td>
</tr>
</tbody>
</table>

*Base differences are indicated by lowercase type; underlining indicates a base that deviates from the palindrome consensus. Genomic locations are based on the numbering of the genome sequences deposited in GenBank.*
Replicating shuttle vectors based on the chromosomal origin exist in single copies. To determine the approximate copy number of the oriC-based plasmids, a PCR product generated from the \( P_{gdh} \) promoter was used as a probe in Southern hybridization experiments with total genomic DNA from \( P. furiosus \) wild-type cells and pJFW027 transformants. Since \( P_{gdh} \) is present in one copy both on pJFW027 and in the \( P. furiosus \) chromosome, a densitometry analysis of the amount of DIG-labeled probe hybridized to each one allowed an estimation of the number of plasmid copies per chromosome (Fig. 3). The relative intensities of the plasmid-derived hybridization signal to the chromosomally derived hybridization signal of \( P_{gdh} \) for the EcoRV and HpaI digests ranged from 1.4 to 1.8 for 10 transformants tested, indicating that the oriC-based plasmids exist in a single copy per chromosome.

Conclusions

The functional analysis of the replication origin of the \( P. furiosus \) chromosome reported here showed that only two of the three ORB sequences, those flanking an AT-rich sequence most conserved in arrangement and sequence among the Archaea (42), and no more than three of the mini-ORB sites are required for autonomous plasmid replication. In particular, the DNA-unwinding site, predicted by P1 endonuclease assays (34), is not required for autonomous plasmid replication, nor are any of the predicted ORB or mini-ORB sequences within the DNA polymerase small-subunit open reading frame. We emphasize that we have not ruled out the possibility that these sequences are important for chromosomal replication and that they may serve to promote additional Cdc6/Orc1 binding for chromosomal replication initiation. The open reading frame encoding the Cdc6/Orc1 protein present adjacent to the predicted origin sequence is not required in \( cis \) for autonomous plasmid replication. Vectors based on \( P. furiosus \) oriC were stably maintained for more than 100 generations without selection and showed no evidence of rearrangement after replication and transformation between \( E. coli \) and \( P. furiosus \). The smallest oriC fragment identified in this study capable of conferring autonomous
replication was 653 bp in length, and vectors based on the origin exist in a single copy per chromosome in the cell. Two conserved short palindromes were identified within the origin region that are conserved among *Pyrococcus* species but not in the closely related species *Thermococcus kodakarensis*, suggesting that if they have a function, it may be specific to *Pyrococcus* species. We anticipate that these vectors will have utility for homologous and heterologous gene expression, as well as providing a tool for the study of natural competence, and *in vivo* studies of replication and recombination in *P. furiosus*.

**Acknowledgements**

We are especially grateful to Gina Lipscomb and Karen Stirrett for discussions and advice throughout the course of this work and to Sidney Kushner for guidance in the design of experiments testing DNA replication. We also thank Jennifer Huddleston for critical review of the manuscript.

This work was supported by a grant to M.W.W.A. and J.W. from the BioEnergy Science Center (DE-PS02-06ER64304), administered by the Oak Ridge National Laboratory, and by the Office of Biological and Environmental Research (FG02-08ER64690) in the DOE Office of Science.
References


Supplemental Text

Plasmids based on the high copy pGT5 plasmid from *P. abyssi* have been reported (2, 5). These plasmids have a significantly reduced copy number in *P. furiosus* (5) and cannot be used for selection of transformants in closely related *Thermococcus kodakaraensis* (4). Our own constructs based on pGT5 were very unstable, rapidly lost without selection, and showed major internal deletions.

The pGT5 plasmid sequence encodes two open reading frames which cover 85% of the plasmid genome (1). The larger of the ORFs encodes a Rep75 protein involved in rolling-circle replication (1, 3). The smaller ORF does not have a clear function, and does not seem to be necessary for replication (2). A predicted replication origin is located 5’ of the Rep75 protein (1), and Erauso et al were successful in constructing a *P. abyssi* shuttle vector by opening the pGT5 plasmid 3’ of the Rep75 protein and cloning (1, 3) the plasmid into an *E. coli* plasmid. We, therefore, chose to amplify a linear fragment containing pGT5 with a break in this region. A linear pGT5 fragment was produced by PCR amplification from total *P. abyssi* DNA, using primers JF254 and JF270, and this fragment was cloned into pJFW017 to produce pJFW019.

The pJFW019 plasmid was used to transform *P. furiosus* COM1 ΔpyrF cells selecting uracil prototrophy. PCR screening of transformants (amplification of the *aac* gene on the plasmid) showed that transformants did, in fact, contain at least this portion of the plasmid but often DNA from these transformants used to back-transform *E. coli* did not yield transformants. Plasmid DNA isolated from *E. coli* back-transformants often showed significant internal deletions. These findings are consistent with the fact that plasmids based on this replicon were rapidly lost without selection perhaps because of rearrangements that lead to deterioration of the pJFW019 plasmid in vivo.
Supplemental References


<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JF254</td>
<td>AGAGAGGATTCACGGTGACCATCTTTTT</td>
</tr>
<tr>
<td>JF255</td>
<td>AGTTTCATCTTTTTATAACTCTCATG</td>
</tr>
<tr>
<td>JF263</td>
<td>AggtaccGGTTCATGTGCAAGCTCCATC</td>
</tr>
<tr>
<td>JF264</td>
<td>CTCCAACGTATCTCGTTCTC</td>
</tr>
<tr>
<td>JF266</td>
<td>TCACCAGCTCCCGGAAG</td>
</tr>
<tr>
<td>JF266.2</td>
<td>TCTCTCTatcgatTCACCAGCTCCGC</td>
</tr>
<tr>
<td>JF267</td>
<td>AGTACATCACCGACGAGCAAG</td>
</tr>
<tr>
<td>JF267.2</td>
<td>AAAAggatccAGTACATCACCGAGACGAGCAAG</td>
</tr>
<tr>
<td>JF268</td>
<td>ACCTTTGAGCTGCAAACCACC</td>
</tr>
<tr>
<td>JF268.2</td>
<td>AAAAatcgatACTTTGTAGCTGCAAACCACC</td>
</tr>
<tr>
<td>JF269</td>
<td>TCCATTGGAATATTGTGCCTCTAG</td>
</tr>
<tr>
<td>JF269.2</td>
<td>AAAAggatccTCCATTGGAATATTGTGCCTCTAG</td>
</tr>
<tr>
<td>JF270</td>
<td>AGGGAATGGCACAAGG</td>
</tr>
<tr>
<td>JF282</td>
<td>AGTATTCTCTCAAGAGATAGTAGGCAAG</td>
</tr>
<tr>
<td>JF301.2</td>
<td>AAAAggatccGACAACACGATCCTCCCATATTA</td>
</tr>
<tr>
<td>JF303.2</td>
<td>AAAAggatccTAACAGAAGTGAAAGTCCCCAG</td>
</tr>
<tr>
<td>JF304.2</td>
<td>AAAAggatccTTCCTGTTGACGACAAATGAAC</td>
</tr>
<tr>
<td>JF305.2</td>
<td>AAAAggatccTTTTATCCTTTGGCTTGAACAAACATG</td>
</tr>
<tr>
<td>JF306.2</td>
<td>AAAAggatccTTTCAGAGGAACATAACAC</td>
</tr>
<tr>
<td>JF339</td>
<td>AAAAatcgatTGTGAAGTGAAAGTGACATCCAG</td>
</tr>
<tr>
<td>JF345</td>
<td>AAAAatcgatCTAATGAAACATTTATTCAGTAATTGACC</td>
</tr>
<tr>
<td>JF348</td>
<td>AAAAatcgatGGTCCAGTGGAAATGAAACTC</td>
</tr>
<tr>
<td>JF349</td>
<td>AAAAatcgatTATCTAATGAAACTAAAAGGTGC</td>
</tr>
<tr>
<td>JF350</td>
<td>ATATatcgatCATTTGTCTCCACAGGAATCTG</td>
</tr>
<tr>
<td>GL21</td>
<td>GATTGAAAATGGAGTGAGCTGAG</td>
</tr>
<tr>
<td>GL23</td>
<td>GTTCATCCCTCACAATTAGGTG</td>
</tr>
</tbody>
</table>

Restriction sites used for cloning are indicated in lowercase letters.
Figure 2.S1. pJFW019 contains the pSC101 replication origin and $aac$ gene cassette for replication and selection in $E. coli$. The $pyrF$ gene, under control of the $gdh$ promoter provides for uracil prototrophic selection in $P. furiosus$. The entire pGT5 sequence was used to promote plasmid replication in $P. furiosus$. 
Figure 2.S2. Restriction analysis of pJFW019 plasmid DNA before and after transformation of *P. furiosus* and back-transformation to *E. coli*. 1kb DNA ladder (Lanes M). pJFW019 plasmid DNA isolated from *E. coli* DH5α (Lane 1), and digested with Accl (Lane 2), Aval (Lane 3), and HindIII (Lane 4). Plasmid DNA isolated from three independent *E. coli* DH5α back-transformants (Lane 5-8, 9-12, and 13-16), undigested and digested with these same enzymes in the same order.
Figure 2.S3. PCR confirmation of plasmid transformation of *P. furiosus*. 1kb DNA ladder (Lane M, NEB). PCR amplification of *aac* from pJFW018 (Lane 1), pJFW019 (Lane 2), and pJFW027 (Lane 3) plasmid DNA isolated from *E. coli* DH5α. PCR amplification of *aac* from genomic DNA from *P. furiosus* transformants of pJFW018 (Lane 4), pJFW019 (Lane 5), and pJFW027 (Lane 6). No product was amplified from *P. furiosus* wild type (Lane 7), COM1 Δ*pyrF* (Lane 8) or the no template control (Lane 9).
CHAPTER 3

THE RECOMBINOGENIC PROPERTIES OF THE *PYROCOCCUS FURIOSUS* COM1 STRAIN 
ENABLE RAPID SELECTION OF TARGETED MUTANTS

---


Reprinted here with permission of the publisher.
Abstract

We recently reported the isolation of a mutant of *Pyrococcus furiosus*, COM1, that is naturally and efficiently competent for DNA uptake. While we do not know the exact nature of this mutation, the combined transformation and recombination frequencies of this strain allow marker replacement by direct selection using linear DNA. In testing the limits of its recombination efficiency, we discovered that marker replacement was possible with as few as 40 nucleotides of flanking homology to the target region. We utilized this ability to design a strategy for selection of constructed deletions using PCR products with subsequent excision, or “pop-out,” of the selected marker. We used this method to construct a “markerless” deletion of the *trpAB* locus in the GLW101 (COM1 Δ*pyrF*) background to generate a strain (JFW02) that is a tight tryptophan auxotroph, providing a genetic background with two auxotrophic markers for further strain construction. The utility of *trpAB* as a selectable marker was demonstrated using prototrophic selection of plasmids and genomic DNA containing the wild-type *trpAB* alleles. A deletion of *radB* was also constructed that, surprisingly, had no obvious effect on either recombination or transformation, suggesting that its gene product is not involved in the COM1 phenotype. Attempts to construct a *radA* deletion mutation were unsuccessful, suggesting that this may be an essential gene. The ease and speed of this procedure will facilitate the construction of strains with multiple genetic changes and allow the construction of mutants with deletions of virtually any nonessential gene.

Introduction

Hyperthermophilic archaea are of special interest because of their evolutionary history and unique physiology, as well as several important biotechnology applications associated with their thermostable enzymes (4, 31). The development of genetic methods for this diverse group has presented many challenges, in part because of their extreme growth requirements. Recent progress has been made, however, in the ability to transform a variety of species by taking
advantage of the fact that some are highly recombinogenic and/or able to take up DNA via
natural competence. *Sulfolobus* sp., *Thermococcus kodakaraensis*, and *Pyrococcus furiosus*
can all be transformed by linear DNA fragments (8, 12, 19, 20, 28, 29), but the length of the
homologous flanking region needed for marker replacement varies. In *T. kodakaraensis*, which
is naturally competent, more than 100 bp of homologous flanking region are required for
homologous recombination (28), but in *Sulfolobus acidocaldarius*, which is transformed via
electroporation, 10 to 30 bp of homology is sufficient (19). Several methods that rely on
homologous recombination have also been developed to construct mutants in mesophilic
archaea, including in *Halobacterium* sp. (23, 36), *Haloferax volcanii* (2, 3), and *Methanosarcina
acetivorans* C2A (25).

The isolation of a mutant of *Pyrococcus furiosus*, previously designated COM1 (strain
GLW101 [COM1 ΔpyrF]), that is efficiently competent for DNA uptake was recently reported
(20). This strain is transformed by exogenous DNA without any chemical or physical treatment,
as opposed to the wild-type *P. furiosus*, in which transformants were not obtained under the
same conditions. Subsequently, this strain has enabled the construction of replicating shuttle
vectors based on the chromosomal replication origin (9), as well as the production of strains that
lack some key metabolic enzymes (5) and that overproduce affinity-tagged versions of the
native (6) and a subcomplex form (15) of the cytoplasmic hydrogenase.

In this study, we show that 1,000 bp efficiently direct integration into the chromosome and
as few as 40 bp allow efficient homologous recombination in the *P. furiosus* GLW101
chromosome using linear DNA fragments. Using this ability of GLW101 to recombine short
segments of homologous DNA, a strategy was developed for generating deletion mutants by
PCR amplification (without cloning) to select marker replacement events with subsequent
excision, or “pop-out,” of the selected marker. An important feature of this method is that it
allows direct selection of targeted mutants. We used this approach to generate a markerless
deletion of *trpAB*, generating a strain (Δ*trpAB ΔpyrF*) that allows simultaneous nutritional
selection of both markers, as demonstrated using a recently constructed replicating plasmid for *P. furiosus* (9). To examine the role of genes predicted to be important for recombination, we used this method to attempt to delete *radA* and *radB*, both homologues in *P. furiosus* of the eukaryotic recombinase *rad51*. Surprisingly, deletion of *radB* had no apparent effect on either recombination or DNA transformation in the GLW101 (COM1 ΔpyrF) background. Mutants with a deletion of *radA* were not obtained, suggesting that deletion of this gene may be lethal. Further analysis of the GLW101 strain suggests that the natural competence phenotype does not result from uracil starvation or the loss of a restriction enzyme as a barrier to DNA transformation, but the exact nature of this mutation remains unknown.

**Materials and Methods**

*P. furiosus* strains, media, and growth conditions. *P. furiosus* DSM3638 (10) wild type, GLW101 (COM1 ΔpyrF) (20), and other strains (Table 1) were grown anaerobically in a defined medium with cellobiose as the carbon source (20) at 90°C for 16 to 18 h in 100-ml serum bottles containing 50 ml of liquid medium or on medium solidified with Phytagel (1% wt/vol) for 64 h. For growth of GLW101 and other uracil auxotrophic strains, the defined medium contained 20 µM uracil. Transformation of GLW101 was performed as described previously (20). The GLW104 strain was generated by transforming GLW101 with a PCR product containing the wild-type *pyrF* allele and selecting for uracil prototrophy. Transformation of the JFW02 strain was performed similarly, but tryptophan prototrophic selection was performed on a medium with or without 20 µM uracil and lacking tryptophan. The transformation efficiencies reported here were calculated as the number of transformant colonies per µg of DNA added and do not take into account plating efficiencies, which are typically 1 to 5%. Transformation frequencies were calculated as the proportions of transformant colonies to total cells and do take into account cell viability.
Table 3.1. *P. furiosus* strains used and constructed in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Parent strain</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSM3638</td>
<td>Wild type</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>GLW101</td>
<td>COM1 ΔpyrF</td>
<td>DSM3638</td>
<td>21</td>
</tr>
<tr>
<td>GLW102</td>
<td>COM1 ΔpyrF trpF::P&lt;sub&gt;gdh&lt;/sub&gt;-pyrF</td>
<td>GLW101</td>
<td>This work</td>
</tr>
<tr>
<td>GLW103</td>
<td>COM1 ΔpyrF trpF::P&lt;sub&gt;gdh&lt;/sub&gt;-pyrF</td>
<td>GLW101</td>
<td>This work</td>
</tr>
<tr>
<td>GLW104</td>
<td>COM1 ΔpyrF::pyrF</td>
<td>GLW101</td>
<td>This work</td>
</tr>
<tr>
<td>JFW01</td>
<td>COM1 ΔpyrF trpAB::P&lt;sub&gt;gdh&lt;/sub&gt;-pyrF</td>
<td>GLW101</td>
<td>This work</td>
</tr>
<tr>
<td>JFW02</td>
<td>COM1 ΔpyrF ΔtrpAB</td>
<td>JFW01</td>
<td>This work</td>
</tr>
<tr>
<td>JFW03</td>
<td>COM1 ΔpyrF radB::P&lt;sub&gt;gdh&lt;/sub&gt;-pyrF</td>
<td>GLW101</td>
<td>This work</td>
</tr>
<tr>
<td>JFW04</td>
<td>COM1 ΔpyrF ΔradB</td>
<td>JFW03</td>
<td>This work</td>
</tr>
</tbody>
</table>
Purification of intermediate strains was performed by plating $10^3$ dilutions of transformant cultures onto selective plate medium (without uracil) and picking isolated colonies into selective liquid medium. JFW01 and JFW03 were both purified to homogeneity in two rounds of colony purification. We were never able to obtain a pure marker replacement mutant for radA. Pop-out recombination was accomplished by growing strains with the marker replacement cassette inserted into the chromosome in defined cellobiose medium containing 20 µM uracil from a 1% inoculum (~7 generations) and then plating onto defined medium containing 20 µM uracil and 3 mM 5-fluoroorotic acid (5-FOA).

**Restriction endonuclease assays.** Cell-free extracts (CFE) were prepared from 1-liter cultures as described previously (7). Endonuclease assays were performed in 10-µl reaction mixture volumes using 0.5 to 1 µg of pJFW051 DNA. Various amounts of CFE (from 0 to 20 µg total protein) were added to separate reaction mixtures that were incubated in NEBuffer 4 (20 mM Tris-acetate, pH 7.9, 50 mM potassium acetate, 10 mM magnesium acetate, and 1 mM dithiothreitol; New England BioLabs). Reaction mixtures were prepared on ice and incubated at 90°C for 15 min. Similar assays were performed to detect the type I or type III restriction endonucleases. These assays were performed as described above except with the inclusion of 1 mM ATP, 80 µM S-adenosylmethionine (SAM), or both.

**PCR amplification and transformation of the wild-type pyrF gene.** PCR amplification of the wild-type pyrF gene with flanking regions ranging in length from 1,000 to 0 bp (Fig. 1A) was performed using the following primer sets: GL055-GL058, pyrF500bpF-pyrF500bpR, pyrF250bpF-pyrF250bpR, pyrF150bpF-pyrF150bpR, pyrF100bpF-pyrF100bpR, pyrF50bpF-pyrF50bpR, pyrF40bpF-pyrF40bpR, pyrF30bpF-pyrF30bpR, pyrF20bpF-pyrF20bpR, pyrF10bpF-pyrF10bpR, and pyrF0bpF-pyrF0bpR (see Table S1 in the supplemental material). These products were purified using a DNA Clean & Concentrator-25 column (Zymo Research) and transformed into the GLW101 strain. Three biological replicates were performed for each
Figure 3.1. (A) The wild-type pyrF region and PCR-amplified fragments (indicated as lines below the chromosomal region) with various lengths of flanking sequence used to transform GLW101, selecting uracil prototrophy. (B) Transformation efficiencies using PCR products. Amounts of 1 µg of DNA were used to transform ~10^7 cells.
PCR product. *trpAB* pop-out marker replacement strategy. Sequence Manipulation Suite (32) was used to generate a random 40-bp sequence (5’ AAGTGAGCGTGTTACGCCGAGACCCGGTTTCGTCTCTCAT 3’) that was altered slightly at the 3’ end to prevent hairpin or self-annealing structures that could be problematic in PCR. This sequence was introduced into the pop-out PCR product using 5’ primer tails. Two primer sets (JF392-JF393 and JF394-JF395; see Table S1 in the supplemental material) were used to amplify 1-kb regions flanking *trpAB*. The P$_{gdh}$-*pyrF* marker cassette was amplified from pJFW017 (9) plasmid DNA using primers JF355.3 and JF356.3. The specific annealing regions of these primers were designed for melting temperatures at 55 ± 4°C (mean ± standard deviation). The overlap tails were 30 to 35 bases in length and designed so that the overlapping regions between PCR products would be 20 to 25 bases in length, melting at 62 ± 2°C. PCR was performed using PfuTurbo polymerase in a 50-µl reaction mixture volume according to the manufacturer’s instructions (Stratagene). Thermal cycling included 30 cycles with annealing at 58°C and a 70-s extension at 72°C. Products were purified using a DNA Clean & Concentrator-25 column (Zymo Research). Three fragments were put together by two rounds of splicing by overlap extension (SOE) PCR (16). The *trpAB* upstream flanking region was joined to the P$_{gdh}$-*pyrF* marker cassette, and in a separate reaction, the P$_{gdh}$-*pyrF* marker cassette was also joined to the *trpAB* 3’ flanking region. SOE PCR was performed using ~50 ng of each template DNA in a 50-µl reaction mixture. Prior to thermal cycling, the template was denatured without primers, allowed to anneal at 58°C, and extended for 10 min at 72°C. Thirty cycles of amplification were performed as described above, with the same end primers used to generate the template products but with the extension time increased to 120 s. These two products were purified and used as the template for another SOE PCR. In the second SOE PCR, the first annealing step was omitted and the two fragments were allowed to anneal and extend at 72°C for 10 min. Thirty cycles of amplification were performed as described above, with primers JF392 and JF395 and extension time increased to 180 s. The 3-kb PCR product was then purified and transformed into the GLW101 strain. Eight uracil
prototrophs were picked into liquid defined medium without uracil and grown overnight at 90°C. Putative transformants were screened for the marker replacement by PCR with JF392 and JF395, using the conditions described to generate the pop-out PCR product (see Fig. 2). The final deletion mutant was confirmed by sequencing of the \textit{trpAB} region, which contained the sequence as designed.

**Construction of plasmids.** To construct pJFW051, a 4.4-kb fragment was amplified by PCR from pJFW018 using primers JF264 and JF269. The P\textsubscript{gdh}-hmg cassette (21) was amplified from pGLW28 (20) using primers GL021 and GL022, treated with T4 polynucleotide kinase, and ligated into the 4.4-kb fragment. A cassette containing the wild-type \textit{trpAB} alleles under the transcriptional control of the phosphoenolpyruvate (PEP) synthase (PF0043) promoter (P\textsubscript{pep}-\textit{trpAB}) was constructed by SOE PCR. A 126-bp portion of the intergenic region upstream of the PEP synthase gene was amplified from wild-type genomic DNA using primers GL158 and WN008. The \textit{trpAB} genes (PF1705 and PF1706) were amplified using primers WN009 and WN010 and joined to the fragment containing the PEP regulatory region and a 12-bp terminator from the \textit{hpyA1} gene (PF1722) (30). The P\textsubscript{pep}-\textit{trpAB} cassette was treated with T4 polynucleotide kinase and ligated into the pJFW018 plasmid (9) that had been digested with EcoRV and treated with shrimp alkaline phosphatase. \textit{Escherichia coli} strain DH5\textalpha{} cells were transformed by electroporation in a 2-mm-gap cuvette at 2.5 V. The plasmid constructions were confirmed by restriction analysis. Plasmid DNA was isolated from liquid cultures by using QIAprep spin miniprep columns (Qiagen, Inc.).

**Results and Discussion**

**Fewer than 40 bp of homologous DNA allows selection of marker replacements in \textit{P. furiosus}**. To investigate the minimum homology required for recombination in \textit{P. furiosus} GLW101 (COM1 \textit{ΔpyrF}), PCR products containing the \textit{pyrF} gene with flanking DNA regions ranging in length from 0 to 1,000 bp were used to restore the \textit{ΔpyrF} locus in GLW101 to the wild
Figure 3.2. Pop-out marker replacement strategy. Six primers are used to construct a pop-out PCR product that is used to direct marker replacement and subsequent excision of the selected marker. (A) One-kilobase flanking regions are amplified from genomic DNA, and \( P_{\text{gdh}}-pyrF \) is amplified from pJFW017. Overlap tails for SOE PCR introduce the “pop-out scar” sequence and are indicated in red. (B) SOE PCR generates two overlap products. (C) A second SOE PCR generates the final pop-out construct. (D) Transformation into \( P. furiosus \) allows for selection of the marker replacement event. (E) 5-FOA selection of the pop-out cassette generates a markerless deletion.
type, selecting transformants for uracil prototrophy (Fig. 1A). The transformation efficiency was measured as the number of uracil prototrophic transformants per µg of DNA. The transformation efficiency increased exponentially with the increase in flanking region length, with up to 103 transformants per µg of DNA containing 1-kb flanking regions and a few transformants detectable for DNA with flanking regions as short as 20 nucleotides (Fig. 1B). It is important to note that the annotated pyrF open reading frame overlaps with the open reading frame of the downstream gene, so the constructed pyrF deletion retains the last 16 bases of pyrF to include the full-length downstream open reading frame. As a result, the 3’ flanking region contained 16 bases of additional homology to the pyrF gene (20). This may contribute little for larger fragments of homology but may well affect the interpretation of the data for the very short fragments. However, our results clearly show that recombination can occur between 40-base sequences.

In previous work, we routinely used high concentrations (2 to 10 µg/ml) of transforming DNA (20). To determine the relationship between transforming DNA concentration and transformation frequency, we transformed GLW101 with a range of DNA concentrations, using both a replicating shuttle vector, pJFW018 (9), and a linear wild-type pyrF-containing fragment with 1 kb of flanking homology. For both DNA types, transformants were detectable with DNA concentrations as low as 1 ng/ml. When the transforming DNA concentration is high (10 µg/ml), the transformation frequency of linear DNA is approximately 1 out of 100 to 500 viable cells. Transformation with linear DNA fragments requires both DNA uptake and integration into the genome. Given the long regions of homology and the overabundance of DNA, it is likely that DNA uptake is the more (though not necessarily the only) limiting factor. These results are consistent with a model of natural competence in which only a small subset of cells are competent but they are very efficiently transformed. Taken together, these data suggest that transformation is probably very efficient with ample regions of homology, even at very low DNA
concentrations. Transformation with very short regions of homology may be possible, but the efficiency may be prohibitively low for practical applications.

**Sequence homology within the PCR products used for selection of marker replacement allows pop-out of the selected marker.** Selection of marker replacements using the wild-type copy of pyrF results in strains that are uracil prototrophs not useful for further mutant construction using pyrF as a selectable marker. To overcome this, we adapted a strategy that had been used successfully in yeast (1) and *T. kodakaraensis* (28) for pop-out of the wild-type pyrF allele. An example of this strategy, targeting two genes involved in tryptophan biosynthesis, *trpAB*, is shown in Fig. 2. The transforming DNA fragment containing the deletion cassette contains pyrF under the control of the promoter for the gene encoding glutamate dehydrogenase (PF1602), P$_{gdh}$-pyrF (20), flanked by an additional 40-bp direct-repeat sequence with minimal homology to the *P. furiosus* chromosome and constructed using splicing by overlap extension (SOE) PCR (16). A direct-repeat sequence of 40 bp was sufficient to allow for pop-out of the P$_{gdh}$-pyrF marker, and these regions were introduced into the transformation construct using PCR primers shorter than 60 bases. Three separate PCRs were used to amplify the upstream and downstream flanking regions of *trpAB*, as well as the pyrF expression cassette (Fig. 2 and 3), and these products were then joined by two successive rounds of SOE PCR, using a total of six primers for the construction. A proofreading polymerase was used to minimize the potential for introducing changes during polymerization. It is interesting to note that, while larger regions of homology are necessary for efficient marker replacement, as little as 40 bp allows efficient pop-out of the selected marker, suggesting that recombination, apart from transformation, is also very efficient.

Transformation of the *trpAB* pop-out construction into the GLW101 strain resulted in hundreds of uracil prototrophic colonies, suggesting that marker replacement at this locus was also very efficient. Eight of these colonies were picked for PCR screening (see Fig. S1 in the supplemental material), and one was purified to homogeneity. This intermediate strain, JFW01,
Figure 3.3. Construction of the trpAB pop-out markerless deletion. (A) Lanes: M, 1-kb DNA ladder; 1, 1-kb 5’ flanking amplicon; 2, 1-kb 3’ flanking amplicon; 3, P_gdh-pyrF marker cassette; 4, overlapped 5’ flanking region to P_gdh-pyrF marker cassette; 5, overlapped 3’ flanking region to P_gdh-pyrF marker cassette; 6, trpAB pop-out marker replacement cassette. (B) PCR amplification of the genomic regions surrounding the trpAB loci showing the marker replacement and subsequent pop-out (confirmed by DNA sequencing of the PCR products). Lanes: M, 1-kb DNA ladder; 1, trpAB locus in GLW101; 2, trpAB locus in JFW01; 3, trpAB locus in JFW02.
was a uracil prototroph and a tryptophan auxotroph. To select the pop-out event at the trpAB locus, JFW01 was grown in liquid medium containing uracil and then (1 ml of culture) was grown on solid medium containing both uracil and 5-fluoroorotic acid (5-FOA). All 5-FOA-resistant colonies (10 total) were screened for pop-out of the pyrF cassette, and the frequency of pop-out was approximately $10^{-7}$. One of these strains was designated JFW02. JFW02 is a tight tryptophan auxotroph, and excision of the $P_{gdh}$-pyrF marker restored 5-FOA resistance and uracil auxotrophy. JFW02 is therefore a double auxotroph and suitable for further genetic manipulation.

This strategy has several important advantages over conventional deletion construction. It does not require cloning, and only six primers are needed to provide specificity for the target gene (see Table S1 in the supplemental material). Primers of 60 bases are significantly less expensive and eliminate the need for primer purification, making pop-out construction of deletions amenable to a high-throughput system. Since both the integration and excision of the Pgdh-pyrF cassette are selected, this method may be used to target any nonessential gene. The pop-out constructs, as described here, leave a 40-bp “scar” sequence that remains in the genome after pop-out of the $P_{gdh}$-pyrF marker cassette. If a scarless deletion is desired, this strategy could be modified so that only one 40-bp pop-out sequence is included in the construct, which would recombine with the native sequence on the other side of the $P_{gdh}$-pyrF marker cassette to generate a scarless deletion of the target gene. Alternatively, the 40-bp scar sequence provides flexibility for modifying genomic targets by introducing specific sequences, such as signal peptides or affinity tags for protein purification. The utility of such tags has been demonstrated (6, 15).

For the trpAB deletion mutant, two rounds of purification, selecting uracil prototrophy, were required to resolve merodiploids generated by the initial marker replacement event. If the deletion is viable and produces a small or no growth defect, segregation and allelic fixation should be random. In the case of trpAB, which we expected to have a mild phenotype with
tryptophan added to the medium, two rounds of colony purification were sufficient. The subsequent pop-out strain should have a neutral phenotype with both uracil and tryptophan added to the growth medium, and we found, in fact, that no additional purification was necessary after one round of selection on 5-FOA.

The ΔtrpAB strain is a tight tryptophan auxotroph but not resistant to 5-FAA. As with uracil biosynthesis, tryptophan biosynthesis allows for selection for both prototrophy and auxotrophy because the wild-type allele is counterselectable. First demonstrated in Saccharomyces cerevisiae, deletion of various genes in the biosynthetic pathway results in a tryptophan auxotroph that is resistant to 5-fluoroanthranilic acid (5-FAA) (35). This anthranilic acid analog is converted to 5-fluorotryptophan by the tryptophan biosynthetic pathway, and incorporation of 5-fluorotryptophan into proteins is toxic. In addition, 5-fluorotryptophan inhibits anthranilate synthase, thereby reducing the synthesis of tryptophan and increasing 5-fluorotryptophan toxicity (22). We found that P. furiosus is sensitive to 5-FAA on defined medium at a concentration of approximately 2 g/liter. In P. furiosus, the tryptophan biosynthetic pathway is predicted to be in an operon consisting of seven genes (trpA to -G) (34). Our first targets for deletion mutagenesis were trpF (PF1707), trpE (PF1709), and trpD (PF1710). Mutants with deletions of trpE were not obtained, most likely for technical reasons, and mutants with deletions of trpF (PF1707) and trpD (PF1710) were leaky auxotrophs not resistant to 5-FAA. Deletion of the trpAB locus (PF1705 and PF1706) (Fig. 2) resulted in a tight tryptophan auxotroph but also did not confer resistance to 5-FAA, suggesting that there are other mechanisms of 5-FAA toxicity in P. furiosus.

Complementation of the ΔtrpAB strain by the wild-type trpAB alleles restores tryptophan prototrophy. To test the utility of the trpAB deletion mutant for prototrophic selection, the wild-type trpAB alleles were cloned onto a replicating shuttle vector that also contained the pyrF expression cassette (9) to generate pJFW070 (Fig. 4). The COM1 ΔpyrF
Figure 3.4. Construction of pJFW070. (A) Primers GL158 and WN008 were used to amplify the PEP synthase promoter (P_{pep}). Primers WN009 and WN010 were used to amplify trpAB from wild-type genomic DNA. (B) These fragments were joined by SOE PCR to produce the P_{pep}-trpAB marker cassette, which was treated with T4 polynucleotide kinase (PNK) and ligated into the pJFW018 fragment produced by EcoRV digestion and shrimp alkaline phosphatase (SAP) treatment, producing pJFW070.
ΔtrpAB strain (JFW02) was readily transformed by this plasmid, selecting either uracil or tryptophan prototrophy. Since both markers are contained on the same plasmid, we were able to compare the transformation efficiencies of the two markers. The efficiencies (~10⁴ transformants per µg DNA) were similar to each other and comparable to that previously determined for pJFW018 (10⁴ transformants per µg DNA [9]). JFW02 could also be transformed to tryptophan prototrophy using wild-type genomic DNA. This strain will be important for applications that require multiple simultaneous selections in the same strain, such as maintaining a replicating shuttle vector with one marker and using the other marker to perform chromosomal manipulations. Selecting tryptophan prototrophy also provides an additional alternative to uracil prototrophy or a requirement of agmatine for growth (15).

A deletion of radB has no obvious effect on recombination in the GLW101 genetic background. To investigate the highly recombinant nature of GLW101 (COM1 ΔpyrF), we constructed a markerless deletion of radB (PF0021), a homologue of the eukaryotic rad51 gene (26). This protein has been implicated in recombination and repair in P. furiosus by its DNA binding affinity and interaction with other known recombination proteins (18). Its role as a recombination mediator rather than a true recombinase is supported by its weak ATPase and strand exchange activities in P. furiosus (18) and recombination and growth defects in deletion mutants of Haloferax volcanii (13, 14). The radB pop-out PCR product was constructed using the same approach used for trpAB. Hundreds of uracil prototrophic colonies were obtained, and of eight colonies screened, one was purified to homogeneity. This intermediate strain, designated JFW03, was grown in liquid medium containing uracil and plated on solid medium containing both uracil and 5-FOA. PCR amplification of the radB locus from the resulting colonies identified four that had pop-out of the pyrF allele, which was confirmed by DNA sequencing of the PCR products. One of these was designated JFW04. The fact that pop-out was readily selected in this radB marker replacement mutant suggests that deletion of radB had
no significant effect on recombination.

To further investigate recombination in the \textit{radB} deletion mutant, we transformed the JFW04 strain with several different DNA types: a replicating plasmid (pJFW018), a nonreplicating integrating plasmid (pGLW021) (20), integrating PCR products (\textit{trpAB::pyrF} pop-out construction and amplified wild-type \textit{pyrF} with 1-kb flanking regions), and wild-type (DSM3638) genomic DNA. All the DNA types transformed JFW04 to uracil prototrophy at equivalent frequencies (on the order of $10^2$ to $10^3$ transformants per viable cell count at a DNA concentration of 3 to 4 µg/ml, with plating efficiencies of approximately 1%).

Sensitivity to UV light was used to test recombination related to DNA repair. The survival of the \textit{radB} mutant after exposure to UV doses in the range of 0 to 10 mJ on a plate surface was indistinguishable from the survival of the wild type and GLW101 at all intensities tested (Fig. 5A). There was also no difference in the growth rates of the two strains in defined medium (Fig. 5B) or under conditions of oxidative shock (Fig. 5C), as measured by sensitivity to hydrogen peroxide (33).

The apparent wild-type phenotype of JFW04 is somewhat surprising given the severe phenotype of \textit{radB} mutants in the euryarchaeote \textit{Haloferax volcanii} (13, 14; T. Allers, personal communication). On the other hand, \textit{radB} is not present in any known member of the Crenarchaeota, including \textit{Sulfolobus} species. The apparent wild-type phenotype of JFW04 may result from differences in \textit{radB} functional divergence within the Euryarchaeota. It could also be the result of the GLW101 genetic background, but it is not possible to address this, as the wild-type strain, \textit{Pyrococcus furiosus} DSM3638, has proven to be genetically intractable in our hands. Future work with \textit{radB} in the closely related \textit{T. kodakaraensis} could address this issue.

\textbf{Attempts to construct a mutant with a deletion of \textit{radA} were unsuccessful.} We also attempted to construct a mutant with a deletion of \textit{radA} (PF1926), another \textit{rad51} homologue (27). In contrast to \textit{RadB}, \textit{RadA} is a true recombinase, with DNA binding, ATPase, and strand exchange activities in \textit{P. furiosus} (18). Deletion of \textit{radA} in \textit{Haloferax volcanii} results in
Figure 3.5. Characterization of JFW04. (A) Wild-type (light grey), GLW101 (grey), and JFW04 (dark grey) strains were exposed to UV radiation on a plate surface with doses ranging from 0 to 10 mJ. Relative survival was calculated as the proportion of colonies compared to the number for the unexposed control. (B) Growth curve of the JFW04 strain (circles) compared to the growth curves of GLW101 (squares) and the wild type (triangles). Culture growth was monitored by optical density at 660 nm. (C) Oxidative stress response of the JFW04 (circles) strain compared to the oxidative stress responses of GLW101 (squares) and the wild type (triangles). H$_2$O$_2$ was added at an OD$_{660}$ of 0.07 (7.5 h for wild type, 8.5 h for GLW101 and JFW04). Times of H$_2$O$_2$ addition are indicated by arrows. Each point represents the average of samples from two or more independent cultures, and the error bars show standard deviations.
recombination and growth defects (37). Transformation of a radA pop-out construction into GLW101 also produced hundreds of uracil prototrophic colonies. Eight were screened for the marker replacement event at the radA locus, and all contained both radA and the P_{gdh}pyrF marker replacement. Six rounds of colony purification, which were sufficient to purify other deletion mutants, failed to resolve these merodiploids, and repeated attempts to isolate a clean marker replacement strain were unsuccessful. In addition, the unresolved merodiploids showed a severe growth defect. Either in liquid medium or on plates, these strains took at least twice as long as the GLW101 parent to grow to a comparable cell density or colony size. These data suggest that a deletion of the radA gene results in a severe phenotype and that radA may, in fact, be essential for viability in P. furiosus, although further experiments would be required to prove this conclusively.

The exact nature of the mutation leading to the COM1 phenotype is unknown. The nature of competence in GLW101 is of considerable biological interest. We hypothesized that the initial transformation event, integrating pGLW021, selected using simvastatin resistance, was a rare event, never seen again, but the deletion of pyrF in this transformant resulted in a strain that was starved for uracil, and uracil starvation resulted in competence as a mechanism to take up DNA as a source of uracil. Competence would then depend on uracil starvation, i.e., cell growth on plates without uracil, as in prototrophic selection using the wild-type pyrF allele. To test this, we constructed a shuttle vector, pJFW051, which is similar to pJFW018 but contains the 3-hydroxy-3-methylglutaryl coenzyme A gene for simvastatin resistance selection (see Fig. S2 in the supplemental material). We found that GLW101 was readily transformed by this plasmid on defined medium containing uracil and simvastatin. We performed the same experiment with GLW104, a GLW101-derived strain with a restored pyrF gene, and transformation of this strain was equally efficient. Since uracil starvation is not an issue with the restored pyrF deletion, this would indicate that competence does not result from uracil starvation.
We also hypothesized that competence might be the result of a mutation in a restriction system, as restriction of heterologous DNA is often a barrier to transformation. To test this, cell extracts were prepared from wild-type *P. furiosus* and GL101 cultures and incubated with pJFW051 plasmid DNA, using conditions suitable for other restriction enzymes from *Pyrococcus* species (17, 24), reported for their commercial use. Under all conditions tested, no restriction activity was detected for either the wild type or GLW101. In addition, there are no annotated restriction endonucleases in the *P. furiosus* genome. Interestingly, there are also no homologues of competence genes. We also emphasize that many “naturally competent” organisms do not exhibit competence under all conditions, and it is possible that wild-type *P. furiosus* will be competent if appropriate conditions are used.

**Conclusions**

*P. furiosus* is an excellent model system for the study of DNA recombination, repair, and natural competence in the Archaea. The methods reported here will facilitate future studies by decreasing the time and expense required to generate marker replacement and deletion mutants. The *trpAB* deletion provides another selectable marker and will enable more sophisticated genetic analyses involving the maintenance of multiple selectable markers. In the GLW101 (COM1 ΔpyrF) background, we have found that the deletion of *radB* has no detectable phenotype but the deletion of *radA* has a severe, possibly lethal phenotype. It is not known if this is true for *P. furiosus* in general or is peculiar to the GLW101 strain, since the nature of competence in GLW101 is not yet understood. This is the subject of ongoing and future investigations, which will be further facilitated by the work presented here. The application of this method in *P. furiosus* will be especially valuable in elucidating the function of the ~600 genes which are unique to *P. furiosus* as compared to *T. kodakaraensis* (11). The increased transformation frequency associated with the GLW101 strain makes it particularly useful for the study of natural competence, CRISPR (clusters of regularly interspaced short palindromic
repeats) function, and homologous recombination in general. In addition, the pop-out strategy can be adapted to generate tagged proteins in two steps in vivo and will have other uses that make strain construction rapid.

Acknowledgements
We thank Daehwan Chung for technical advice throughout the course of the work and Jennifer Copeland for technical assistance. This work was supported by grants from the Bio-Energy Science Center (grant DE-PS02-06ER64304), administered by Oak Ridge National Laboratory, and from the Office of Biological and Environmental Research (grant FG02-08ER64690) and the Chemical Sciences, Geosciences and Biosciences Division (grant DE-FG05-95ER20175), Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy. J.F. was supported by a predoctoral Graduate Training In Genetics grant (grant NIH 5T32GM007103-30) to the Genetics Department of the University of Georgia.
References


27. Sandler SJ, Satin LH, Samra HS, Clark AJ. 1996. recA-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast


<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>pyrF500bpF</td>
<td>AAAACAGATATTCCGAAATACCTCGA</td>
</tr>
<tr>
<td>pyrF500bpR</td>
<td>GGTTATCTCCCAATCTTATCCCT</td>
</tr>
<tr>
<td>pyrF250bpF</td>
<td>ACGAGGCAATAAAGTTCGACGCA</td>
</tr>
<tr>
<td>pyrF250bpR</td>
<td>TAAAGATCTTCTTATAGAGAGCT</td>
</tr>
<tr>
<td>pyrF150bpF</td>
<td>GAGATGGCAAGAGACCTCTATGAT</td>
</tr>
<tr>
<td>pyrF150bpR</td>
<td>TCTGTGCTTATCTGAAACTCCGCA</td>
</tr>
<tr>
<td>pyrF100bpF</td>
<td>TACTTTGGCAATAAATTTGGGAAG</td>
</tr>
<tr>
<td>pyrF100bpR</td>
<td>AGCATCTTGGCTTGAATTTCCCT</td>
</tr>
<tr>
<td>pyrF50bpF</td>
<td>AGTCGAGCTAAGAGCAGTCGA</td>
</tr>
<tr>
<td>pyrF50bpR</td>
<td>TATCTTCTTTATCCCTCTGGGCT</td>
</tr>
<tr>
<td>GL021</td>
<td>GATGGAATGGAGTGAAGCTGAG</td>
</tr>
<tr>
<td>GL022</td>
<td>TCCATTTCCAGCAAATTTATGG</td>
</tr>
<tr>
<td>GL055</td>
<td>AGAGAGAGGCGCATGTCACCTCTCTCTATTTTGAGTTCCATG</td>
</tr>
<tr>
<td>GL058</td>
<td>GAGAGAGGCCGCCATCAGAGTGATGAGGTAGGTAGAGCC</td>
</tr>
<tr>
<td>GL158</td>
<td>TCCGTCAGCTCCACAAGCTCTTTATGGCT</td>
</tr>
<tr>
<td>JF264</td>
<td>CTCAAGCGTACATCTGCTT</td>
</tr>
<tr>
<td>JF269</td>
<td>TCCATTGGAAATTTGCTGCTT</td>
</tr>
<tr>
<td>JF307</td>
<td>AATGAGCATGCGGTATATAAACACATTTCC</td>
</tr>
<tr>
<td>JF311</td>
<td>AACGTCTAGAAAGATATGGGAGTGG</td>
</tr>
<tr>
<td>JF326</td>
<td>AGATGCTAGCGAAGGAGAAGTTG</td>
</tr>
<tr>
<td>JF357</td>
<td>TTAATATGCCAGGCCATCTCTCTCTCTGGCAGTACCACTGCCCATATGAGTGCTGAGTGC</td>
</tr>
<tr>
<td>JF358</td>
<td>AGGAGTGGCCCTGCTGAGTGCATGAGGAAGGATGACTGCTGACCCATGAGTGAGCAAGGAGTGAGGTGAACCT</td>
</tr>
<tr>
<td>JF355.1</td>
<td>AGGAGTGGCCCTGCTGAGTGCATGAGGAAGGATGACTGCTGACCCATGAGTGAGCAAGGAGTGAGGTGAACCT</td>
</tr>
<tr>
<td>JF356.1</td>
<td>TTAATATGCCAAGCCACTCTCCTCTCTGTGATGATTTTCTGAGCTCCATCTTTCAC</td>
</tr>
<tr>
<td>JF322</td>
<td>ACTGCCAGTGCTCAGT</td>
</tr>
<tr>
<td>JF360</td>
<td>TTAGGGCCATCATACAAAGAGAGGTAGAGTTCATCAGCCGGTACCCATGAGGACTGCTGAGTGGCC</td>
</tr>
<tr>
<td>JF355.2</td>
<td>AGCGGCTAGGCGCATCACAAAGAGAGGTAGAGTTCATCAGCCGGTACCCATGAGGACTGCTGAGTGGCC</td>
</tr>
<tr>
<td>JF356.2</td>
<td>TTAATATGCCAAGCCACTCTCCTCTCTGTGATGATTTTCTGAGCTCCATCTTTCAC</td>
</tr>
<tr>
<td>JF392</td>
<td>AAGAGAAGTCGCCAACATTCA</td>
</tr>
<tr>
<td>JF393</td>
<td>AGCGGGCAGGGTGGAGCCTCCCAAGGAGCAGGACCCATCCATGAGGACTGCTGAGTGGCC</td>
</tr>
<tr>
<td>JF394</td>
<td>TGGTACCCGAGACACCGGGCTTCCCCATCAGCCAGGACTGCTGAGTGGCC</td>
</tr>
<tr>
<td>JF395</td>
<td>TTAGGGCTATGTCATCTCGAGGAGGACTGCTGAGTGGCC</td>
</tr>
<tr>
<td>JF355.3</td>
<td>TTAGGAGCTGGCGAAGCAGGCCGGTGGTTGCTCGCATGATGAGGAAATGAGGTGAGCTGAGTGGCC</td>
</tr>
<tr>
<td>JF356.3</td>
<td>AAACCCGGGCTCAGCGGGTGAACAGCAGCAGCTGATGAGGACTGCTGAGTGGCC</td>
</tr>
<tr>
<td>WN008</td>
<td>CCAAAAAGCAATCTCATATGCTGAGGAGGACTGCTGAGTGGCC</td>
</tr>
<tr>
<td>WN009</td>
<td>GTGAAGCTGAAATGGTGGTTGCTGAGGAGGACTGCTGAGTGGCC</td>
</tr>
<tr>
<td>WN010</td>
<td>CTAAAAGAGATTTGGATCAGGCCTCAAAGGAGTGGCC</td>
</tr>
</tbody>
</table>
Figure 3.S1. Screening of putative trpAB marker replacement mutants. A 3.9kb product is produced from the wild type, while a 3.1kb band is produced by the targeted marker replacement.
Figure 3.S2. Construction of pJFW051. A 4.4kb fragment was amplified by PCR from pJFW018 using primers JF264 and JF269. The $P_{gdh}$-$hmg$ cassette (22) was amplified from pGLW28 (21) using primers GL021 and GL022, treated with T4 polynucleotide kinase and ligated into the 4.4kb fragment to produce pJFW051.
CHAPTER 4
CONCLUSIONS

The development of genetic methodologies for Pyrococcus furiosus was guided by previous genetic systems for archaeal anaerobes and hyperthermophiles, particularly the closely related Thermococcus kodakarensis. Following the example of Sato, et al, the first gene targeted for deletion in P. furiosus was pyrF (141, 142). The HMG-CoA marker was used to select transformants by simvastatin resistance, integrating a plasmid into the chromosome. Subsequent plasmid excision resulted in the deletion of the pyrF gene. The generation of the ΔpyrF mutant strain resulted in a strain that was also naturally competent for DNA uptake making genetic manipulation of P. furiosus possible. This strain, and its unexpected natural competence, enabled rapid progression in developing more sophisticated genetic tools (94). The COM1 strain is not only essential for genetic manipulation of P. furiosus, but its highly competent nature will allow the elucidation of the mechanism of natural competence as well as other biological processes, notably CRISPR function in P. furiosus. Both of these benefit from the highly competent nature of the COM1 strain (94) and the newly developed shuttle vectors (49).

During the course of the work described in chapter 2, we found that constructs based on pGT5, the native plasmid from P. abyssi, could transform P. furiosus, but were not stably maintained, and frequently rearranged in vivo (49). The unstable plasmids based on pGT5 are usable for protein expression (173), but are obviously undesirable for genetic applications. This is why we chose to use the chromosomal replication (oriC) origin to promote plasmid replication in our shuttle vectors. Replicating shuttle vectors facilitate a variety of genetic and biochemical applications in P. furiosus, including protein expression, without the need for chromosomal
integration. Replicating shuttle vectors allow the analysis of protein function using merodiploids, enabling a better understanding of protein functions through experimental determination of dominant and recessive mutations. Shuttle vectors will also be necessary tools for studying biological processes where DNA uptake and recombination are required.

In addition to their utility as a simple genetic tool, these shuttle vectors also provide an in vivo system for the study of origin recognition in Pyrococcus species. Since these vectors use oriC to promote plasmid replication, they are in essence, mini-chromosomes. The work described in chapter 2 shows that only two of the ORB sequences flanking the origin, and no more than three of the mini-ORB sites are required for autonomous plasmid replication. We showed that the putative DNA-unwinding site, predicted by P1 endonuclease assays (110), is not required for autonomous plasmid replication. We also showed the presence of conserved palindrome sequences in and around the predicted origin site. The necessity of these sequences for autonomous plasmid replication could easily be tested by site directed mutagenesis and a replication assay using an oriC-based plasmid.

The pop-out method described in chapter 3 has already been used to construct markerless deletions of several tryptophan biosynthetic genes, radB (50), as well as several genes involved with CRISPR function (Terns, personal communication). Because pop-out works well with pyrF selection and counterselection, there is no need to develop a selection protocol with two genes as performed by Santangelo, et al (137). The development of multiply marked strains will be useful for any application that requires multiple plasmids or plasmids in addition to chromosomal selections. The rapidity of deletion construction and known gene targets that confer agmatine auxotrophy (66, 137) and 6-methyl purine resistance (54, 137) will enable the construction of strains with even more markers, providing a single background strain suitable for nearly any purpose and the deletion of virtually any non-essential gene.

The mechanisms of natural competence have been well studied in bacteria, but to date only three Archaeal species have been reported to be naturally competent (17, 142, 177), and
nothing is known about the nature of natural competence in these species. Despite whole genome sequencing of the COM1 genome, the nature of natural competence remains unknown (24). Compared to *T. kodakarensis*, *P. furiosus* is able to recombine DNA fragments with much shorter regions of homology (50, 141). This may be partly responsible for the increased transformation frequencies observed with *P. furiosus*. The findings that *P. furiosus* is able to repair extensive DNA damage from ionizing radiation (41), integration of multiple plasmids in a single cell (94), and recombination with very short regions of homologous DNA (50) all suggest that recombination in *P. furiosus* is extremely efficient. The observation that very low DNA concentrations are sufficient for transformation, and that transformation frequency seems to reach a maximum frequency regardless of DNA concentration (50), suggests that there may be a stochastic nature to a cell developing competence. Since there are no obvious homologs of competence proteins in *P. furiosus* (94), the mechanism of DNA uptake is likely to be novel.

One missing and important tool for *P. furiosus* is a reporter gene system that allows visual detection of gene expression. As discussed in Chapter 1, developing a reporter system for hyperthermophilic anaerobes is challenging, and an innovative solution will be required. With a knockout methodology in place, a reporter system similar to that applied in *T. kodakarensis* (137, 138) could be adapted for *P. furiosus*, which would facilitate gene expression studies, but would require quantitative PCR or enzymatic assays. Extracellular chitinases or glycosidases could be expressed to produce zones of clearing surrounding colonies but this is not optimal as only a limited number of colonies could be examined on a plate. A color change with a pH indicator could potentially be produced which depends upon pyruvate utilization or acetate production. The formation of a colored precipitate could also provide for visual screening. Such a system might rely upon the production of an insoluble ion, such as sulfide from cysteine desulfurase, that would precipitate with iron in the medium, or sequestration of an ion like iron, preventing such a precipitate from forming. These are possibilities and may provide the basis for testing.
The investment in *P. furiosus* genetics has greatly facilitated potential future research, but also opened new avenues for research in *P. furiosus* and closely related organisms. With the wealth of literature on *P. furiosus*, and newly developed genetic capabilities, the potential for new research is enormous. Though our collaboration has come to a successful conclusion, this is indeed an exciting time for *P. furiosus*. 
REFERENCES


*kodakaraensis* KOD1 is composed solely of large subunits and forms a pentagonal structure. J. Mol. Biol. 293:57-66.


