The Genome of Archaeal Prophage Ψ M100 Encodes the Lytic Enzyme Responsible for Autolysis of *Methanothermobacter wolfeii*

YONGNENG LUO, PETER PFISTER, THOMAS LEISINGER, AND ALAIN WASSERFALLEN*

Institute of Microbiology, Swiss Federal Institute of Technology Zürich, CH-8092 Zu¨rich, Switzerland

Received 6 February 2001/Accepted 6 July 2001

Methanothermobacter wolfeii **(formerly** *Methanobacterium wolfei***), a thermophilic methanoarchaeon whose cultures lyse upon hydrogen starvation, carries a defective prophage called** C**M100 on its chromosome. The nucleotide sequence of** Ψ **M100 and its flanking regions was established and compared to that of the previously sequenced phage** C**M2 of** *Methanothermobacter marburgensis* **(formerly** *Methanobacterium thermoautotrophicum* **Marburg). The VM100 genome extends over 28,798 bp, and its borders are defined by flanking 21-bp direct repeats of a pure-AT sequence, which very likely forms the core of the putative attachment site where the** crossing over occurred during integration. A large fragment of 2,793 bp, IFa, apparently inserted into Ψ M100 but is absent in the genome of Ψ M2. The remaining part of the Ψ M100 genome showed 70.8% nucleotide sequence identity to the whole genome of VM2. Thirty-four open reading frames (ORFs) on the forward strand and one ORF on the reverse strand were identified in the Ψ M100 genome. Comparison of Ψ M100-encoded ORFs to those encoded by phage $\Psi M2$ and to other known protein sequences permitted the assignment of putative functions to some ORFs. The ORF28 protein of Ψ M100 was identified as the previously known **autolytic enzyme pseudomurein endoisopeptidase PeiW produced by** *M. wolfeii***.**

Cultures of the thermophilic methanoarchaeon *Methanothermobacter wolfeii* (formerly *Methanobacterium wolfei*) (20, 21) spontaneously lyse upon hydrogen limitation (10). However, no phage-like particles have been detected in the culture supernatant or autolysate. A lytic enzyme was purified from the autolysate, and this enzyme has been shown to be a pseudomurein endoisopeptidase (9). *M. wolfeii* proved insensitive to the virulent phage $\Psi M1$ and its deletion mutant $\Psi M2$ of *Methanothermobacter marburgensis* (formerly *Methanobacterium thermoautotrophicum*) Marburg (P. Pfister, unpublished observation). The results of Southern hybridization suggested that there is a prophage in the chromosome of *M. wolfeii*, Ψ M100, which is homologous to phages Ψ M1 and Ψ M2 (14, 19; P. Pfister, unpublished data). The approximate location of CM100 in the *M. wolfeii* chromosome was determined, and most of the prophage was shown to be located on a ca. 30-kb *Not*I-*Nhe*I fragment (19). Since attempts to detect free phage particles had failed, $\Psi M100$ was assumed to be defective. In contrast, infection of *M. marburgensis* Marburg with phages Ψ M1 and Ψ M2 consistently led to lysis of the host. The complete nucleotide sequence of $\Psi M2$ was established, and some of its open reading frames (ORFs) and the proteins they encode were characterized (15). In order to explore the relationships between the defective prophage Ψ M100, the autolysis phenomenon, and the $\Psi M1$ and $\Psi M2$ phages, we determined and analyzed the sequence of $\Psi M100$ and its flanking regions.

Cloning, PCR amplification, and nucleotide sequence determination of Ψ M100 and its flanking regions. Portions of YM100 and its flanking regions were obtained from the *M. wolfeii* chromosome as overlapping clones and PCR fragments

(data not shown). The methods used included the following: shotgun cloning, construction, and screening of a SuperCos1 based cosmid library; screening of a λ -ZAP Express genomic library (7); PCR amplification of either nonclonable portions or portions that were difficult to clone on both sides of $\Psi M100$; and PCR amplification of the regions across the restriction sites of fragments obtained by shotgun cloning. The nucleotide sequences of both strands were determined by primer walking and then assembled.

Properties of $\Psi M100$ **DNA.** The border of $\Psi M100$ is defined by the flanking 21-bp AT-only direct repeats, which probably represent the core of the putative attachment site (see below). The length of the Ψ M100 genome extends over 28,798 bp with an overall GC content of 45.4%. This is somewhat lower than the 48.3% GC determined for *M. wolfeii* by a melting point analysis (11). Sequence alignment of Ψ M100 and Ψ M2 (Fig. 1) reveals that, other than point mutations and small deletions or insertions, there is a large fragment of 2,793 bp, IFa, inserted into Ψ M100 between coordinates 5272 and 8066. The GC content of the IFa element is 33.4%, which is significantly lower than those of Ψ M100 and Ψ M100 without IFa (46.7%). Like Ψ M2, the GC content of the Ψ M100 sequence is not evenly distributed, with five low-GC ($\leq 40\%$) DNA regions of at least 300 nucleotides (nt) extending over parts of ORFs *orf6*, *orf28*, *orf29*, and *orf30* and over the entire ORFs *orf5*, the IFa-encoded *orfB*, -*C*, and -*D*, and the putative attachment sites attL and attR (Fig. 1).

Some direct repeats are clustered and prominent in their length and locations. For example, there are three contiguous copies of a direct repeat of 125 nt in the region between coordinates 5081 and 5455, and a short copy is duplicated between coordinates 5456 and 5481. This is apparently due to the insertion of IFa, since there are only two direct repeats of 67 nt in the corresponding region of Ψ M2, which, based on its

^{*} Corresponding author. Mailing address: Institute of Microbiology, Swiss Federal Institute of Technology Zürich, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland. Phone: 41 1 632 4488. Fax: 41 1 632 1148. E-mail: wasserfallen@micro.biol.ethz.ch.

FIG. 1. Schematic representation of the defective prophage with its flanking regions and comparison to that of phage VM2. The ORFs are represented by boxes numbered as in Table 1, and their vertical placement indicates the gene location in one of the six possible reading frames. Homologous ORFs carry the same numbers, and the assigned functions for some ORFs of VM2 are reported. MTW1215 and MTW1216 represent the two ORFs encoded by the chromosome sequences flanking VM100. *orfA* to *orfE* are unique to VM100. The experimentally determined *pac* locus for VM2 is shown as a solid black triangle. The IFa fragment (black bar) and the DNA regions with at least 200 bp with >98% identity in VM100 and VM2 (grey bars) are indicated. Three arrowheads represent three contiguous copies of a direct repeat of 125 nt in the vicinity of the left end of the IFa fragment.

similarity to *oriC* of *Escherichia coli*, was proposed to be the most-probable replication origin (Pfister, unpublished). Moreover, the direct repeats do not overlap any coding regions of Ψ M100 and Ψ M2.

After the IFa element was removed, the remaining region of Ψ M100 (26,005 bp) could be aligned with that of Ψ M2 (26,111 bp) over their full length with the GAP program of the Genetics Computer Group (GCG). It shows 70.8% identity in a 26,752-nt overlap. Eight stretches of sequence longer than 200 bp in Ψ M100 and Ψ M2, including the regions harboring the putative origin of replication and the experimentally determined *pac* site (8) , have $>98\%$ identity (Fig. 1).

Coding capacity of Ψ M100 and comparison to that of Ψ M2. Putative ORFs were defined based on the following assumptions: an ORF should code for a polypeptide of at least 90 amino acids (aa) and be preceded by a ribosome binding site with at least 45% identity to the proposed consensus sequence 5'-AGGAGGTGATC-3' (2). Two exceptions were made; these exceptions were *orf2*, encoding a 85-aa polypeptide with a homologue of 95 aa in Ψ M2, and *orf29*, encoding a putative integrase, the ribosome binding site of which has only 36% identity to the consensus sequence (as in Ψ M2). Consequently, 34 genes were identified on the forward strand and 1 gene was identified on the reverse strand (Fig. 1 and Table 1).

Except for *orfA* and the IFa-encoded *orfB*, -*C*, -*D*, and -*E*, all other ORFs are similar to their counterparts encoded by phage WM2. The ORFs of WM100 and WM2 exhibit similarities in size, location in the genome, and sequence identity, which ranges from 39.9 to 100% at both the nucleotide and amino acid levels. The homologue of $\Psi M2$ *orf20* is missing in $\Psi M100$ (Fig. 1 and Table 1). The ORFA protein (155 aa) apparently has no homologue of similar length in $\Psi M2$. The nucleotide sequence of $\Psi M2$ encodes a peptide of 86 aa, homologous to the N terminus of ORFA (probability calculated using BLAST, 10^{-24}). Sequence alignment suggests that the C-terminal part of ORFA was created by insertion of three noncontiguous DNA stretches into the Ψ M100 genome. Database searches with BLAST at the National Center for Biotechnology Information (1) also revealed that 16 of the proposed proteins are similar to other protein sequences (Table 1).

Integration site of $\Psi M100$ **in the chromosome of** *M. wolfeii***.** Screening for ORFs encoded by regions of the *M. wolfeii* chromosome flanking VM100 revealed two genes encoding MTW1216 on the forward strand upstream of the puta-

TABLE 1. General features of the putative ORFs encoded by Ψ M100 and comparison to the corresponding ORFs of CM2 and to other proteins in databases

^a F, reading frame in which an ORF was defined.

b HTH, helix-turn-helix motif.

^c NA, not available.

^{*d*} The percentages of amino acid (aa) identity were calculated using the GAP program of GCG. The numbers in parentheses are the sizes (in amino acids) of the alignment overlap. NA, not available.

 \overline{P} The database sources are shown before the accession numbers as follows: sp, SWISSPROT; pir, PIR database; gb, GenBank.

tive attL site and MTW1215 on the reverse strand downstream of the putative attR site (Fig. 1). The two host genes share high similarity to their counterparts (i.e., *mth*1215 and mth 1216) in the *Methanobacterium thermautotrophicus* ΔH

genome (18); therefore, the same numbers were used to designate the *M. wolfeii* homologues. At the amino acid level, MTW1216 exhibits 77% identity to MTH1216 (pantothenate metabolism flavoprotein) and MTW1215 is 81.9%

 $\mathbf C$

The core of Lambda att sites

TTAGTATAAAAAAGC

FIG. 2. (A) Sequences of the attR, attL, attP, attB, and the N-terminal part of the putative integrase ORF29. Sequences originating from the *M. wolfeii* chromosome are shown in italics. The crossed lines represent the sequences of attP and attB derived from the sequences of attR and attL. The core of the putative attachment sites (grey box), the stop codon of MTW1215 (CTA on the reverse strand) (grey box), and the start codon of ORF29 (ATG on the forward strand) (black box) are indicated. The numbers $-38\sim38$ and $39\sim85$ indicate the coordinates of the prophage VM100 sequence. (B) DNA sequences of the corresponding regions in the phage VM2 and *M. thermautotrophicus* ΔH genomes as well as of another region in the *M. thermautotrophicus* ΔH genome. The numbers refer to the coordinates of the sequences in the phage VM2 or *M. thermautotrophicus* ΔH genome. (C) DNA sequence of the core of coliphage λ attachment sites. In panels B and C, nucleotides identical to those of the core of the VM100 attachment sites are also highlighted.

identical to MTH1215 (fibrillarin-like pre-rRNA processing factor). MTW1215 and MTW1216 are transcribed convergently, as are MTH1215 and MTH1216. The MTH1215 gene in the *M. thermautotrophicus* ΔH genome is located at nucleotides 1117886 to 1118560 on the direct strand, and the gene encoding MTH1216 is located at nucleotides 1119977 to 1118817 on the complementary strand (18). Remarkably, integration of $\Psi M100$ occurred exactly within the intergenic regions between these two host genes and between prophage genes *orf28* (*peiW*) and *orf29*. Neither the host genes nor ORFs of the defective prophage were disrupted in terms of their transcription potential. Thus, the integration mode of $\Psi M100$ resembles that of the major integration pattern of coliphage λ $(4).$

Putative attachment sites of Ψ M100. Sequence analysis revealed that Ψ M100 is flanked by direct repeats of a 21-bp pure-AT nucleotide sequence. This sequence is very likely the core of putative attachment sites where the crossing-over occurred during integration and/or excision. Therefore, the flanking regions of the core might be defined as hybrid attL and attR sites. Conversely, the sequences of attP [(pro)phage-encoded attachment site] and attB [chromosomal attachment site] can be derived, although their exact lengths remain to be determined (Fig. 2A). In the region between orf28 and orf29 of phage YM2 (i.e., in the putative attP site of phage $\Psi M2$) and between MTH1216 and MTH1215 of *M. thermautotrophicus* ΔH (i.e., in the putative attB site), similar AT-rich sequences with four and three mismatches, respectively, were identified (Fig. 2B). Interestingly, one stretch of sequence which is 100% identical to the core sequence but located in a different context was present in another region of the M . thermautotrophicus ΔH genome (Fig. 2B). It is not known whether additional core-like sequences occur in the M . wolfeii genome. For phage λ , the core of the attachment sites consists of 15-bp AT-rich sequences (3). There are some similarities between the sequences of the λ and $\Psi M100$ cores (Fig. 2C).

Experimental identification of the autolytic enzyme pseudomurein endoisopeptidase PeiW. The deduced N-terminal sequence of the protein encoded by $orf28$ of prophage Ψ M100 is identical to that of the experimentally determined N-terminal sequence (18 aa) of the autolytic enzyme produced by *M. wolfeii* (15). This strongly suggests that the ORF28 protein is the pseudomurein endoisopeptidase PeiW that has been known for more than 10 years (9). The predicted molecular mass (33.4 kDa) is consistent with the observed 33 kDa for the autolytic enzyme (9).

The structural gene of PeiW was cloned and successfully overexpressed in *E. coli* BL21(DE3) grown under aerobic conditions. The cell wall-degrading activity of PeiW was confirmed in the activity assay described previously (15; data not shown).

Conclusions. The complete nucleotide sequences of phage YM2 of *M. marburgensis* and of the defective prophage VM100 of *M. wolfeii* and its flanking regions now allow us to make a thorough comparison between these two elements. In contrast to two other archaeal defective prophages in *Halobacterium halobium* (17) and in *Sulfolobus* sp. strain B12 (16), the sequence of Ψ M100 contains all the information necessary for synthesis of structural proteins homologous to those of Ψ M2. In addition to small insertions, deletions, and point mutations, downstream of the putative replication origin, $\Psi M100$ carries the inserted fragment IFa of 2,793 bp, which apparently originated from another source. The IFa element has none of the distinct

characteristics usually found in IS sequences (13). It encodes three putative ORFs, of which two short ones (i.e., *orfB* and *orfC*) are similar to hypothetical ORFs of *M. thermautotrophicus* ΔH . The largest ORF, *orfD*, has no homologues in the database, although PropSearch at the EMBL (6) yields some hits related to DNA metabolism. One of these hits is the potential transposase for IS*1151* of *Clostridium perfringens* (5). Using the GAP program of GCG, the two proteins show 18% identity in a 522-aa overlap. Only the 5' part of the IFa DNA sequence, which is a duplication of the sequence harboring the putative origin of replication, shows traces of terminal redundancy, possibly derived from the packaging mechanism of phage Ψ M2 (8). Therefore, the rearrangement due to the insertion of this IFa in the DNA of Ψ M100 might have led to this defect. However, there is no evidence explaining the order of insertion of IFa into the Ψ M100 genome and integration of the Ψ M100 ancestor(s), if any, into the *M. wolfeii* chromosome. Alternatively, $\Psi M100$ might be defective due to the lack of a functional excisionase, which is usually required to excise (pro) phage genomes from their host chromosomes, while VM2 might recruit the putative excisionase function of the ORF5 encoded by plasmid pME2001 present in the host strain (12).

Together with the finding that the ORF29 proteins of both Ψ M2 and Ψ M100 are highly similar to members of the Int family of site-specific recombinases, the presence of a 21-bp pure-AT direct repeat in the flanking regions of Ψ M100 supports the view that $\Psi M100$ is really a $\Psi M2$ -related prophage. Such direct repeats are typically derived from integrative recombination of phage DNA with the host chromosome (3). Since genes encoding site-specific integrases are not found in genomes of virulent phages, phages $\Psi M1$ and $\Psi M2$ might be temperate phages.

Nucleotide sequence accession number. The sequence is available in GenBank under accession number AF301375.

We thank R. Hedderich for providing the λ -ZAP Express genomic library of *M. wolfeii* and U. Kües for providing *E. coli* strain NM554 and cosmid SuperCos1.

This work was supported in part by grant 3100-50593.97 from the Swiss National Foundation for Scientific Research.

REFERENCES

- 1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
- 2. **Brown, J. W., C. J. Daniels, and J. N. Reeve.** 1989. Gene structure, organization, and expression in archaebacteria. Crit. Rev. Microbiol. **16:**287–338.
- 3. **Campbell, A. M.** 1992. Chromosomal insertion sites for phages and plasmids. J. Bacteriol. **174:**7495–7499.
- 4. **Campbell, A. M.** 1993. Thirty years ago in genetics: prophage insertion into bacterial chromosomes. Genetics **133:**433–437.
- 5. **Daube, G., P. Simon, and A. Kaeckenbeeck.** 1993. IS*1151*, an IS-like element of *Clostridium perfringens*. Nucleic Acids Res. **21:**352.
- 6. **Hobohm, U., and C. Sander.** 1995. A sequence property approach to searching protein databases. J. Mol. Biol. **251:**390–399.
- 7. **Hochheimer, A., R. Hedderich, and R. K. Thauer.** 1998. The formylmethanofuran dehydrogenase isoenzymes in *Methanobacterium wolfei* and *Methanobacterium thermoautotrophicum*: induction of the molybdenum isoenzyme by molybdate and constitutive synthesis of the tungsten isoenzyme. Arch. Microbiol. **170:**389–393.
- 8. **Jordan, M., L. Meile, and T. Leisinger.** 1989. Organization of *Methanobacterium thermoautotrophicum* bacteriophage Ψ M1 DNA. Mol. Gen. Genet. **220:**161–164.
- 9. **Kiener, A., H. König, J. Winter, and T. Leisinger.** 1987. Purification and use of *Methanobacterium wolfei* pseudomurein endopeptidase for lysis of *Methanobacterium thermoautotrophicum*. J. Bacteriol. **169:**1010–1016.
- 10. **Ko¨nig, H., R. Semmler, C. Lerch, and J. Winter.** 1985. Evidence for the occurrence of autolytic enzymes in *Methanobacterium wolfei*. Arch. Microbiol. **141:**177–180.
- 11. **Kotelnikova, S. V., A. Y. Obraztsova, K.-H. Blotevogel, and I. N. Popov.** 1993. Taxonomic analysis of thermophilic strains of the genus *Methanobacterium*: reclassification of *Methanobacterium thermoalcaliphilum* as a synonym of *Methanobacterium thermoautotrophicum*. Int. J. Syst. Bacteriol. **43:**591–596.
- 12. **Luo, Y., T. Leisinger, and A. Wasserfallen.** 2001. Comparative sequence analysis of plasmids pME2001 and pME2200 of *Methanothermobacter marburgensis* strains Marburg and ZH3. Plasmid **45:**18–30.
- 13. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
- 14. **Meile, L., U. Jenal, D. Studer, M. Jordan, and T. Leisinger.** 1989. Characterization of Ψ M1: a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. Arch. Microbiol. **152:**105–110.
- 15. **Pfister, P., A. Wasserfallen, R. Stettler, and T. Leisinger.** 1998. Molecular analysis of Methanobacterium phage VM2. Mol. Microbiol. 30:233-244.
- 16. **Reiter, W. D., and P. Palm.** 1990. Identification and characterization of a defective SSV1 genome integrated into a tRNA gene in the archaebacterium *Sulfolobus* sp. B12. Mol. Gen. Genet. **221:**65–71.
- 17. **Schnabel, H.** 1984. An immune strain of *Halobacterium halobium* carries the invertible L segment of phage ΦH as a plasmid. Proc. Natl. Acad. Sci. USA **81:**1017–1020.
- 18. **Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, H. Safer, D. Patwell, S. Prabhakar, S. McDougall, G. Shimer, A. Goyal, S. Pietrokov**ski, G. M. Church, C. J. Daniels, J.-I. Mao, P. Rice, J. Nölling, and J. N. **Reeve.** 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH : functional analysis and comparative genomics. J. Bacteriol. **179:**7135–7155.
- 19. **Stettler, R., C. Thurner, D. Stax, L. Meile, and T. Leisinger.** 1995. Evidence for a defective prophage on the chromosome of *Methanobacterium wolfei*. FEMS Microbiol. Lett. **132:**85–89.
- Wasserfallen, A., J. Nölling, P. Pfister, J. Reeve, and E. Conway de Macario. 2000. Phylogenetic analysis of 18 thermophilic *Methanobacterium* isolates supports the proposals to create a new genus, *Methanothermobacter* gen. nov., and to reclassify several isolates in three species, *Methanothermobacter thermautotrophicus* comb. nov., *Methanothermobacter wolfeii* comb. nov., and *Methanothermobacter marburgensis* sp. nov. Int. J. Syst. Evol. Microbiol. **50:** 43–53.
- 21. Winter, J., C. Lerp, H.-P. Zabel, F. X. Wildenauer, H. König, and F. Schin**dler.** 1984. *Methanobacterium wolfei* sp. nov., a new tungsten-requiring, thermophilic, autotrophic methanogen. Syst. Appl. Microbiol. **5:**457–466.