J Microbiol Immunol Infect
2002;35:269-271

EJ072200200269

# Analysis of the baseplate region of phage AR1 that specifically infects *Escherichia coli* O157:H7

**Chung-Ping Liao, Wan-Jr Syu**

*Institute of Microbiology and Immunology, National Yang-Ming University, Taipei, Taiwan, ROC*

Received: April 9, 2002 Revised: May 15, 2002 Accepted: May 27, 2002

Phage AR1 that specifically infects *Escherichia coli* O157:H7 has been characterized previously for its primary host contact molecule, the long tail fiber. To strengthen its attachment to the host bacteria, the phage may use its baseplate for a second contact. This study investigated the characteristics of AR1 baseplate by deducing the responsible genes including *wac* and genes *9* to *15*. The wedge pin *g11* and the short tail fiber *g12* varied the most, whereas the others were highly conserved compared with their counterparts in T4. The degrees of gene conservation correlated well with the known functions of individual gene products.

**Key words:** Baseplate, *Escherichia coli*, O157:H7, phage AR1

Bacterial phages have specificity of infection with defined host ranges. This property is useful in typing pathogens of different origins. Coliphage AR1 isolated from cow manure has a high host specificity [1] when tested on various strains of *Escherichia coli*. It infects strains of enterohemorrhagic *E. coli* O157:H7 but not other *E. coli* strains including the K-12 strains. Previous studies indicated that genes encoding capsid proteins are highly conserved between AR1 and T4, whereas *segD* that encodes a nonessential protein in T4 is not conservatively observed in the AR1 genome [2]. Moreover, while T4 infects K-12 strains, AR1 does not. Reciprocally, T4 does not infect the AR1-susceptible O157:H7 strains.

Phage infections must result from a specific recognition between the bacterial surface structure and the phage molecules. Initial identification of the molecules involved in the specific host interaction with AR1 has highlighted the differences of the host outer membrane protein OmpC [2] and the phage long tail fiber gene products gp37/gp38 [3]. OmpC of the O157:H7 origin differs from that of K-12 by amino acid substitutions at 15 positions, a 5-residue deletion, and a 4-residue insertion in a total of 366 amino acids. In a strain of O157:H7 whose chromosomal *ompC* was disrupted, the host sensitivity to AR1 was completely lost [3]. When complemented with the authentic gene in a plasmid, this mutant strain fully regained permissibility to AR1. On the other hand, an OmpC-mutated K-12 strain remained resistant to AR1 when complemented with an OmpC from O157:H7. These observations suggest that additional interaction other than that between OmpC and the tail fiber may contribute to the specific infection. In T4, the phage recognition has been delineated into 2 stages of adsorption. The first stage is reversible adsorption that occurs through the function of the tail fiber whereas the second stage is irreversible adsorption that is mediated by the baseplate. Previous data have shown that the tail fiber varies grossly between AR1 and T4 [3]. The baseplate region might vary as well. To clarify this possibility, we investigated the baseplate genes covering *g9*, *g10*, *g11*, *g12*, *wac*, *g13*, *g14*, and *g15*, and their variations.

*Corresponding author: Dr. Wan-Jr Syu, Institute of Microbiology and Immunology, National Yang-Ming University, 155, Section 2, Li-Long Street, Pai-Tao, Taipei, 112, Taiwan, ROC. E-mail: wjsyu@ym.edu.tw*

## Materials and Methods

Strains of *E. coli* O157:H7 from the Felix d'Herelle Reference Center for Bacterial Viruses, Canada were obtained from the Culture Collection and Research Center (Hsinchu, Taiwan). Preparation of AR1 and extraction of phage DNA were performed as previously described [4]. A strategy of long polymerase chain reaction (PCR) using LA Taq polymerase (Takara, Shiga, Japan) was used to clone the baseplate-coding region. Primers were designed according to partial sequences of *g9* and *g15* previously deduced (C.P.L. and W.J.S., unpublished data). These oligonucleotides were AR1-62F (5'-CCGGATCCCGACTAGATGGGG-GTATGAAG) and AR1-13R (5'-CCGGATCCTGT-CTCAAGAAAAATCTTGGC), custom-synthesized by MWG Biotech (Munich, Germany). By following the

©2002 Chinese Society of Microbiology, Chinese Society of Immunology, and Infectious Diseases Society of the Republic of China

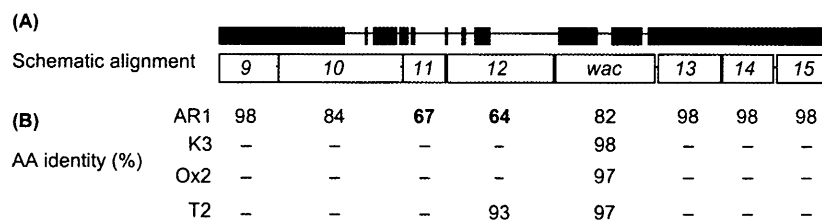
PCR conditions suggested for LA Taq, a 9.6-kb fragment was amplified directly from the AR1 DNA. This amplified DNA product was then cloned into pGEM-T Easy (Promega, Madison, WI, US). To facilitate autosequencing (serviced by Mission Biotech, Taipei, Taiwan), the 9.6-kb insert was fragmented and subcloned into pBluescript II SK (+) (Stratagene, Amsterdam, Netherlands). Common primers derived from pBluescript were used for the sequencing. When inserts were too large to be covered by the common primers, specific primers were synthesized according to the newly deduced sequences within the insert. Thereby, a total sequence of 9578 bp was deduced and is available in Genbank under accession no. AF406556.

## Results and Discussion

We have previously reported that AR1 is a T-even-like phage [3], with homology in essential genes but substantial differences in the long distal tail fibers. To explore possible differences in the phage structure, we cloned and sequenced the DNA segment responsible for the phage baseplate to which the tail fibers are attached. The deduced sequence was subjected to analysis with DNASIS 3.7 (Hitach, Tokyo, Japan) for open reading frames (ORFs). We then searched for homologues to the translated amino acid sequences using the Blast program in the database of Genbank. Counterparts of these ORFs are available from phage T4, and the results of comparison with this sequence are shown in Figure 1. Among the 8 deduced genes, *g9*, *g13*, *g14*, and *g15* are highly conserved between AR1 and T4, and individually they all have a 98% amino acid identity. Further comparison with other phages in these genes is prohibited by a lack of information in the databases. However, *wac* sequences from T2, K3, and Ox2 of the T2 family are available for comparison. These sequences are highly conserved and share 97% or more amino acid identity with that of T4. In contrast, the *wac* ORF of AR1 shares only 82% amino acid identity with that of T4 (Fig. 1). The wac protein is responsible for the rate-enhancing effect of attaching the tail fiber to the baseplate [5], and the tail fibers of AR1 and T4 vary greatly [3]. It is then not surprising to see a marked variance between the wac proteins of AR1 and T4.

Gp12 confers the phage interaction with host LPS [6] and is the least conserved protein among the 8 deduced ORFs (Fig. 1). Besides T4, T2 also has the sequence of gp12 available in Genbank (accession no. X56555). While the gp12 molecules of T4 and T2 share 93% identity amino acids, those of AR1 and T4 (Fig. 1) have only 64% identity. This observation suggests that gp12-interacting LPS structures are closely related between the hosts susceptible to T4 and T2 but less similar between those for T4 and AR1. Indeed, strains of O157:H7 have a distinct R3 type core polysaccharide [7]. Figure 1 also shows that gp11 has a similar degree of variation as gp12. In T4, this protein forms the tip of the tail pin [8] and is the required substrate for gp12 assembly. A possibility of gp11 involvement in cell contact has also been suggested [9]. This possibility is consistent with the gp11 sequence variations observed between AR1 and T4. The product of *g10* is a component forming the middle part of the tail pin and forms a complex with gp11. These structural properties are consistent with the observation that the degree of gene conservation of gp10 (84%) is higher than that of gp11 (67%) but lower than that of the highly conserved genes (greater than 97%).

To summarize the variations of gene products that are involved in determining the phage specificity, gp38 varies the most, with a maximum matching reaching only 21%. Gp37 is second to gp38 and has conservation




| | 9 | 10 | 11 | 12 | wac | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|
| AR1 | 98 | 84 | **67** | **64** | 82 | 98 | 98 | 98 |
| K3 | – | – | – | – | 98 | – | – | – |
| Ox2 | – | – | – | – | 97 | – | – | – |
| T2 | – | – | – | 93 | 97 | – | – | – |

**Fig. 1.** Comparison of genes between phages AR1 and T4. **(A)** Schematic representation of gene conservation between AR1 and T4. DNA sequence deduced from AR1 was compared with that of T4 using the Blast program, and the regions with alignment scores over 200 points [10] are represented by black bars. **(B)** Amino acid (AA) identity comparison. The amino acid sequences deduced from different phages were compared with the T4 counterparts by DNASIS 3.7 (Hitachi). Degrees of maximum matching less than 80% are indicated in bold. Dash marks indicate portions of the sequence of the phage, which are unavailable from databases.

limited to the N-terminal region in about 60 of its 1103 residues [3]. Overall, a maximum of 34% sequence matching was observed with gp37. In contrast to gp37 and gp38, gp11 and gp12 vary only moderately with 67% and 64% identities, respectively. These observations are consistent with the notion that these 2 proteins have a role of tuning the phage-host interaction after the initial contact of the long tail fiber with the OmpC receptor [6].

## Acknowledgments

This work was supported by a grant (NSC 90-2320-B-010-053) from National Science Council, Taiwan, ROC.

### References

1. Ronner AB, Cliver DO. Isolation and characterization of a coliphage specific for *Escherichia coli* O157:H7. J Food Prot 1990;53:944-7.
2. Yu SL, Ding HC, Seah JN, Wu KM, Chang YC, Chang KSS, Tam MF, Syu WJ. Characterization of a phage specific to hemorrhagic *Escherichia coli* O157:H7 and disclosure of variation in host outer membrane protein OmpC. J Biomed Sci 1998;5:370-82.
3. Yu SL, Ko KL, Chen CS, Chang YC, Syu WJ. Characterization of the distal tail fiber locus and determination of the receptor for phage AR1, which specifically infects *Escherichia coli* O157:H7. J Bacteriol 2000;182:5962-8.
4. Sambrook J, Fritsch EF, Maniatis T, eds. Molecular Cloning: A Laboratory Manual, 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989.
5. Conley MP, Wood WB. Bacteriophage T4 whiskers: a rudimentary environment-sensing device. Proc Natl Acad of Sci USA 1975;72:3701-5.
6. Reide I. Receptor specificity of the short tail fibres (gp12) of T-even type *Escherichia coli* phages. Mol Gen Genet 1987; 206:110-5.
7. Amor K, Heinrichs DE, Frirdich E, Ziebell K, Johnson RP, Whitfield C. Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. Infect Immun 2000;68:1116-24.
8. Crowther RA, Lenk EV, Kikuchi Y, King J. Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. J Mol Biol 1977;116:489-523.
9. King J. Assembly of the tail of bacteriophage T4. J Mol Biol 1968;32:231-62.
10. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-402.