The gp38 Adhesins of the T4 Superfamily: A Complex Modular Determinant of the Phage's Host Specificity

Sabrina N. Trojet‡, Anne Caumont-Sarcos‡, Elsa Perrody, André M. Comeau† and H. M. Krisch*

Laboratoire de Microbiologie et Génétique Moléculaires, Centre National de la Recherche Scientifique-UMR 5100, Université Paul Sabatier, Toulouse, France

†Present address: Department of Biology, Institut de Biologie Intégrative et des Systémes/Québec-Océan, Université Laval, Québec, Canada

‡These authors contributed equivalently to this work

*Corresponding author: E-mail: krisch@ibcg.biotoul.fr.

Abstract

The tail fiber adhesins are the primary determinants of host range in the T4-type bacteriophages. Among the indispensable virion components, the sequences of the long tail fiber genes and their associated adhesins are among the most variable. The predominant form of the adhesin in the T4-type phages is not even the version of the gene encoded by T4, the archetype of the superfamily, but rather a small unrelated protein (gp38) encoded by closely related phages such as T2 and T6. This gp38 adhesin has a modular design: its N-terminal attachment domain binds at the tip of the tail fiber, whereas the C-terminal specificity domain determines its host receptor affinity. This specificity domain has a series of four hypervariable segments (HVSs) that are separated by a set of highly conserved glycine-rich motifs (GRMs) that apparently form the domain's conserved structural core. The role of gp38's various components was examined by a comparative analysis of a large series of gp38 adhesins from T-even superfamily phages with differing host specificities. A deletion analysis revealed that the individual HVSs and GRMs are essential to the T6 adhesin's function and suggests that these different components all act in synergy to mediate adsorption. The evolutionary advantages of the modular design of the adhesin involving both conserved structural elements and multiple independent and easily interchanged specificity determinants are discussed.

Key words: T4-like phages, gp38, adhesins, tail fibers, deletion analysis.

Introduction

Viruses infecting the microbial fauna have an enormous impact on the biosphere. This vast population of nano-predators is a major determinant of the planet's microbial ecology (Wommack and Colwell 2000; Suttle 2005). Such viruses also play an important role in evolution by mediating horizontal gene transfer (Comeau and Krisch 2005). Eubacterial and Archaeal viruses are enormously diverse and hence only an infinitesimal fraction has been studied. Among these are the virulent T-even phages (T2, T4, T6, etc.) that infect *Escherichia coli*. Recently, it has become evident that these classical phages belong to a ubiquitous virus superfamily (Desplats and Krisch 2003; Filée et al. 2005; Comeau and Krisch 2008; Comeau et al. 2010). The T4-like phages all share a common morphology based on an elongated icosahedral head that packages a 160–260 kb genome (Ackermann and Krisch 1997). The phage head attaches via one of its 20 vertices to a contractile tail structure that is terminated by a hexagonal base plate with two sets of six retractable tail fibers (Leiman et al. 2003). The long tail fibers (LTFs) reversibly anchor the virion to specific sites on the surface of bacteria via an adhesin protein located at their tip (Riede et al. 1987). It is the recognition specificity of these adhesins that largely defines the phage's host range. The binding of the phage's six LTFs to the receptors on the bacterial surface distorts the base plate structure and causes the release of the short tail fibers (STFs). This allows the STFs to bind irreversibly to a second set of bacterial receptors and triggers the injection of the phage genome into the host (Rossmann et al. 2004).

Although the adhesin is responsible for much of the phage's adsorption specificity, this protein's structure and function has not been studied in depth. Such knowledge is of considerable interest for both applied and fundamental research. For example, therapeutic use of phages for antibacterial treatment would be more effective if a phage's host range could be easily modified. The features of the adhesin

that permit this protein to recognize a variety of different targets but yet allow it to easily change its recognition specificity are interesting from a fundamental point of view but also potentially exploitable in nanotechnology applications. The genes that encode these adhesins have a curious mixture of conserved and hyperplastic sequences. Recombinational shuffling of these sequences can give rise to chimeric adhesins with novel combinations of the recognition determinants; interestingly, the genetic exchanges that do this often occur within or near the most conserved motifs of the adhesins (Tétart et al. 1996, 1998). This suggests that the adhesins may have an intriguingly simple and effective evolutionary strategy to generate diversity selectively within their recognition specificity domains.

Here, we report on the modular design of the gp38 adhesins. Their intricately modular structure allows them to resolve their apparently conflicting design constraints, permitting them to target bacterial receptors with a high specificity that can be changed easily.

Materials and Methods

Bacterial and Bacteriophage Strains

The T4 superfamily phage strains came from the Toulouse collection of myoviruses (the original sources are described in Tétart et al. 1998). *Escherichia coli* B^E (*su*$^-$, Epstein et al. 1963) was the host bacterium used to prepare stocks of all of these phages. Derivatives of *E. coli* B^E were isolated in our laboratory that were resistant to phage RB5 (B^E/RB5). The T6 strains containing amber mutations (see below) were propagated in *E. coli* CR63 (*su1*$^+$, Epstein et al. 1963) using standard phage techniques as described by Carlson and Miller (1994). The Keio collection mutants used in this study—JW3619 (Δ*yicC*), JW0940 (Δ*ompA*) JW0401 (Δ*tsx*), JW2341 (Δ*fadL*), JW0912 (Δ*ompF*), JW2203 (Δ*ompC*), JW3605 (Δ*waaP*), JW3606 (Δ*waaG*), and JW0146 (Δ*fhuA*)—were constructed by Baba et al. (2006) and contain precise deletions of genes of interest in *E. coli* K12 BW25113. *Escherichia coli* strains DH5α or JM109 (Promega) were used for cloning and as hosts for homologous recombination experiments. *Yersinia pseudotuberculosis* strain NCTC10275 was a gift from Dr Grimont of the Pasteur Institute in Paris, France. All strains were grown at 37 °C in Luria–Bertani (LB) medium. Where required kanamycin and ampicillin were added to liquid or solid medium at a concentration of 50 and 50–100 µg/ml, respectively. Fresh log-phase cultures were made in the same medium from overnight cultures.

Preparation and Titration of Phage Stocks

Phages stocks were prepared by infecting the appropriate *E. coli* strains using the standard phage techniques described in Carlson and Miller (1994).

Table 1

Spot-tests of the plating efficiencies of T4 Superfamily Phages on *Escherichia coli* K12 Strains with Single-Gene Deletions in Various known Phage Receptors

Phage	Mutated *E. coli* K12 Strains[a]							
	ompA	ompF	ompC	tsx	fadL	fhuA	waaG	waaP
T6	+++	+++	+++	−	+++	+++	+++	+++
RB3	+++	+++	+++	−	+++	+++	+++	+++
RB5	+++	+++	+++	−	+++	+++	+++	+++
RB6	+++	+++	+++	−	+++	+++	+++	+++
RB7	+++	+++	+++	−	+++	+++	+++	+++
RB14	+++	+++	+++	−	+++	+++	+++	+++
RB27	+++	+++	+++	−	+++	+++	+++	+++
PST	−	+++	+++	−	+++	+	+	+
Ac3	−	+++	+++	+++	+++	+++	+++	+++
K3	−	+++	+++	+++	+++	+++	+++	+++
JSE	−	+++	+++	+++	+++	+++	+++	+++
RB49	−	+++	+++	+++	+++	+++	+++	+++
phi1	−	+++	+++	+++	+++	+++	+++	+++
Mi	−	+++	+++	+++	+++	+++	+	+
RB32	−	+++	+++	+++	+++	+++	+	−
RB33	−	+++	+++	+++	+++	+++	+	−
Ox2	−	+++	+++	+++	+++	+++	+	−
T2	+++	+++	+++	+++	−	+++	+++	+++
SV76	+++	−	+++	+++	+++	−	−	−
RB51	+++	−	+++	+++	+++	+	−	−
RB69	+++	−	+++	+++	+++	+	−	−
T4	+++	+++	−	+++	+++	+++	+++	+++

NOTE.—The gray boxes indicate clearly modified infections when compared with the *E. coli* reference strain. Note that the extent of "normal" lysis is not identical for all the phages; some phage strains infect the K12 reference host less well than others. Normal infection is indicated by +++, some detectable infection by +, and lack of infection by −.

[a] WaaG and WaaP correspond to mutants that have different compositions and lengths of their core oligosaccharides of the LPS.

Host Range Determination

To determine their host range and the nature of their bacterial receptors, a set of representative T4 superfamily coliphages (table 1) were tested against *E. coli* K-12 strains containing precise deletions of genes encoding outer membranes proteins or enzymes involved in lipopolysaccharide (LPS) core biosynthesis (Keio collection mutants; Baba et al. 2006). A series of exponential day cultures were prepared for each host strain. One-hundred microliters of 10× concentrated cultures were mixed with 3 ml of liquefied Htop medium (10 g tryptone, 8 g NaCl, 2.4 g sodium citrate dihydrate, 3 g glucose, 7 g agar, and 1 ml 0.5 N NaOH for 1 l), the mixture was vortexed, and then poured onto individual LB plates. Once the Htop medium had solidified, 3 µl of serial dilutions of parental phage stocks were spotted onto the surface of the plates. The plates were inverted, incubated at 37 °C overnight, and then examined and photographed to detect the zones of lysis.

Gp38 Sequence Analysis

In order to prepare phage genomic DNA, aliquots of samples (100 µl) were subjected to a hot-cold treatment as described by Comeau et al. (2004). Polymerase chain reactions (PCRs)

in a volume of 25 μl (for analysis) or 50 μl (for cloning or sequencing) were performed using *Taq* DNA polymerase (New England Biolabs). The gene *38* of phages T6, RB3, RB5, RB6, RB7, RB14, RB27, FSα, PST, Mi, RB32, and RB33 was amplified using the forward primer FT6.80.11 (the sequences of all the oligos used as primers are listed in supplementary table S1, Supplementary Material online) and the reverse primer FR81; phage Ac3 was amplified with forward primer AC84 and reverse primer FR81; and phages T2 and SV76 with forward primer FT2.g37 and reverse primer FR81. PCR products were purified (QIAGEN) and cloned into the pGEM-T vector (Promega) as per manufacturer's instructions or sequenced commercially (MWG AG; Germany). The final versions of the *g38* sequences were assembled and analyzed using SeqMan II (DNASTAR) and Bio-Edit v.7 (www.mbio.ncsu.edu/bioedit/bioedit.html).

Isolation of RB42 and RB43 Mutants

Both of the closely related PseudoT-even phages RB42 and RB43 had always been propagated in Toulouse on the *E. coli* host strain B[E] because the original isolates were unable to grow on standard laboratory *E. coli* K12 strains. However, P. Genevaux (personal communication) informed us that the RB43 strain variant in the Geneva Laboratory of Prof. C. Georgopoulos grew apparently normally on some *E. coli* K12 strains. We used different methods to isolate similar variants of these phages that could grow on K12 hosts. For example, we isolated an RB42 host range mutant by first mixing an aliquot of an RB42 stock made on B[E] with an exponentially growing culture of *E. coli* K12 W3110. After 24 h, this culture was treated with chloroform and plated for PFUs on W3110. These independently isolated host range mutants were genetically repurified on *E. coli* K12 W3110 and were named: RB42 K1 and RB42 K2. In a second host range mutant isolation procedure, strains of *E. coli* B[E] and S/6 were isolated that were resistant to phage RB42. A direct plating of the RB42 stock was done on the RB42-resistant host strain and rare plaque-forming host range mutants were obtained (named RB42 B21 and RB42 S21). These were genetically purified on the resistant host strains for the sequence analysis of their adhesin genes.

The *g38*-like adhesin sequences of the parental RB42 and RB43 phages, and their host range mutants, were sequenced using the protocol described above. The PCR primers RB43g38F and RB43g38R were used to amplify an ~770-bp fragment flanking all of these closely related g38 sequences. These adhesin sequences were then aligned and compared with identify the location(s) of the host range mutations.

Isolation of RB5 × RB32 or RB5 × RB33 *g38* Recombinant Phages

RB32 or RB33 *g38* PCR inserts were cloned into a pGEM-T plasmid, then overnight stocks of RB5 wild-type phage were made on JM109 carrying either of these plasmids. The resulting lysates containing potential *g38* recombinant sequences were then plated on an RB5-resistant *E. coli* strain (B[E]/RB5) and a series of plaques with normal morphology were isolated from each lysate. These potentially recombinant phages were genetically purified by two successive passages through the B[E]/RB5 host.

Construction of T6 C-terminal Specificity Domain Amber Mutants

To introduce amber mutations into the GRM3 or GRM5 motifs, the T6 *g38* sequence was first amplified using primers FT6.80.11 and FR81 and cloned into the pGEM-T vector. Nonsense mutations were introduced by site-directed mutagenesis employing the QuikChange Site-Directed Mutagenesis Kit (Stratagene/Agilent Technologies). Mutagenic oligonucleotides T6*g38amGRM3F* and T6*g38amGRM3R* or T6*g38amGRM5F* and T6*g38amGRM5R* allowed us to introduce a TAG amber mutation at the beginning of the GRM3 or GRM5 sequences, respectively. JM109 strains carrying pGEM-T plasmids with the amber-mutated gene *38* inserts were infected with wild-type T6 phage. The resulting lysates, containing a small fraction of recombinant phage with the *g38* amber mutation, were plated on CR63 (*su*[+]) host strain. Mutants were then identified by replica plating the random plaques on B[E] (*su*[−]) and CR63 (*su*[+]) host strains. The *g38* of the phage identified as being potential gene *38* amber mutants were sequenced to verify that they contained either the T6*g38amGRM3* or T6*g38amGRM5* mutations.

Construction of Plasmids Containing T6 Adhesin Deletion Sequences

The various deletion mutants in the C-terminal coding region of T6 *g38* were created by the gene shuffling method (Boubakri et al. 2006), which uses mutagenic oligonucleotides to create the desired deletions by PCR. For each deletion, two overlapping mutagenic forward (F) and reverse (R) oligonucleotides, generically named T6*g38*Δdomain-F or T6*g38*Δdomain-R (where Δdomain stands for the HVS1, HVS2, HVS3, GRM2, GRM3, or GRM4 deleted domains) were designed to be complementary of upstream and downstream regions bordering the sequence to be deleted. Thus, the PCR primers did not contain the sequence to be deleted and amplified outward from the deletion point. Two separate PCR reactions were then carried out using T6*g38*Δdomain-F/FR81 and T6*g38*Δdomain-R/FT6.80.11 primer sets, respectively. PCR fragments were purified and reassembled by a self-priming PCR. The amplification protocol consisted of an initial denaturation at 95 °C for 1 min; followed by two cycles of denaturation at 95 °C for 30 s, annealing at 35 °C for 1 min, and extension at 72 °C for 2 min; followed by a final extension step at 72 °C for 9 min. Finally, sufficient quantities of the chimeric fragment were generated

via a standard PCR using the FT6.80.11 and FR81 outer primers. These final products were sequenced and cloned into pGEM-T plasmids which were designated as p*g38*ΔHVS1, p*g38*ΔHVS2, p*g38*ΔHVS3, p*g38*ΔGRM2, p*g38*ΔGRM3, and p*g38*ΔGRM4; each harboring T6 gene *38* with HVS1, HVS2, HVS3, GRM2, GRM3, or GRM4 deletions, respectively. The plasmid containing the wild-type T6 gene *38* sequence was designated as p*g38*WT.

Isolation and PCR Screening of the Potential T6 Adhesin Deletion Mutants

Escherichia coli strains DH5α or JM109 carrying pGEM-T plasmids with wild-type or mutated T6 gene *38* inserts (HVS or GRM deletion) were then infected by either T6*g38am*GRM3 or T6*g38am*GRM5 phages (Multiplicity of Infection [MOI] between 2 and 50). Resulting lysates containing potential recombinants, revertants, or amber mutations of the *g38* sequence were plated on host *su⁻* bacteria, thus allowing for selection of either recombinant or revertant phage. Turbid or clear plaques were then isolated and chimeric phages were analyzed by PCR. Each selected potential chimeric phage was used as a template for PCRs using different strategies. Primers HVS1-F, HVS2-F, GRM2-F, and GRM3-F, targeting sequences in the HVS1, HVS2, GRM2, and GRM3 domains specifically, were used as forward primers in combination with FR81 reverse primer. Absence of amplification should reveal the deletion in the targeted domain. All chimeric T6 *g38* were amplified using the FT6.80.11/FR81 primer set to amplify the entire *g38* sequence. Deletions in T6 *g38* HVS3 and GRM4 domains were then tested by BamHI and SacII digestion, respectively, of the PCR fragments as domain deletions should lead to the loss of the restriction site. For each cross, a set of full-length *g38* PCR products of representative phages were sequenced.

Isolation of T6 × PST Recombinant Phages

The *E. coli* strain DH5α carrying a pGEM-T plasmid with an insertion of the wild-type PST *g38* sequence was infected with an MOI > 10 by either T6*g38am*GRM3 or T6*g38am*GRM5 phage. The resulting lysates were then plated on host *su⁻* bacteria to select for *am+* recombinants. Those progeny with an altered host range were identified by replica plating random plaques on the *E. coli* Bᴱ and *Yersinia* host strains. Recombinant phages with the ability to form normal plaques on *Yersinia* were analyzed by PCR amplification and sequencing of their *g38* adhesins.

Nucleotide Sequence Accession Numbers

The previously unknown wild-type *g38* sequences presented in this paper were deposited in GenBank under accession numbers JF491355 to JF491364. The mutant and recombinant phage sequences are available upon request.

Results and Discussion

The Plasticity of the LTF Locus in the T4 Superfamily

Modular shuffling plays a crucial role in phage evolution (Botstein 1980; Campbell and Botstein 1983). In the T4 superfamily phages, the loss, acquisition, and exchange of modules is frequent (Comeau et al. 2007; Arbiol et al. 2010). Heteroduplex mapping (Kim and Davidson 1974) of the closely related T-even phage genomes (T2, T4, and T6) demonstrated that the locus encoding the LTFs and their associated adhesins were often occupied by novel sequences. Certain pairs of identically positioned genes (e.g., T4 *g38* and T6 *g38*) in this locus do not complement each other (Stahl and Murray 1966; Russell and Huskey 1974), indicating that their functions are not even interchangeable.

The genomes of a dozen phages representative of the nonmarine portion of the T4 superfamily have now been sequenced (Nolan et al. 2006; Petrov et al. 2006; Comeau et al. 2007). A comparative genomic analysis of these (Comeau et al. 2007) confirmed the suggestion that the LTF loci undergo frequent modular shuffling. The most striking example of the genetic plasticity of the locus is found in the genome of T4. In this phage, which is the archetype of the T4 superfamily, the C-terminal part of an ancestral version of distal tail fiber gene (*g37*) and adhesin gene (*g38*) have been replaced by sequences originating from the unrelated temperate phage λ (George et al. 1983; Hendrix and Duda 1992; Henning and Hashemolhosseini 1994).

Figure 1 presents a compilation of all the LTF sequence data available for the T4 superfamily. Many of these phages have LTFs composed of two similarly sized rigid half fibers (the proximal and distal half fibers) connected by a hinged joint (or knee). The overall length of the mature LTF is ~150 nm and the synteny of the genes in the LTF module, with the exception of the adhesin gene, is generally conserved (fig. 1). The locus' most conserved gene is *g34* that encodes the proximal half fiber. This large fibrous protein (~1300 aa) has a globular domain at its N-terminus which attaches it to the base plate (Wood et al. 1978). Most of the remainder of the gp34 sequence is folded into a three-stranded β-helical fibril that is attached at its C-terminus to a monomer of gp35 (~380 aa), the hinged joint. This hinge protein is connected to a trimer of gp36 (~220 aa), the short distal tail fiber subunit, which is further extended by another longer trimeric β-helical fibril encoded by *g37*. In many of the T-even phages such as T6 and T2, the C-terminus of the gp37 fiber is proteolytically cleaved and then capped by an ~260 aa gp38 phage adhesin which also appears to be trimeric. However, in phage T4, the distal portion of the tail fiber has a completely different structure: the gp37 fiber has an extension of the C-terminal domain that provides the adhesin function (Tétart et al. 1996). This T4
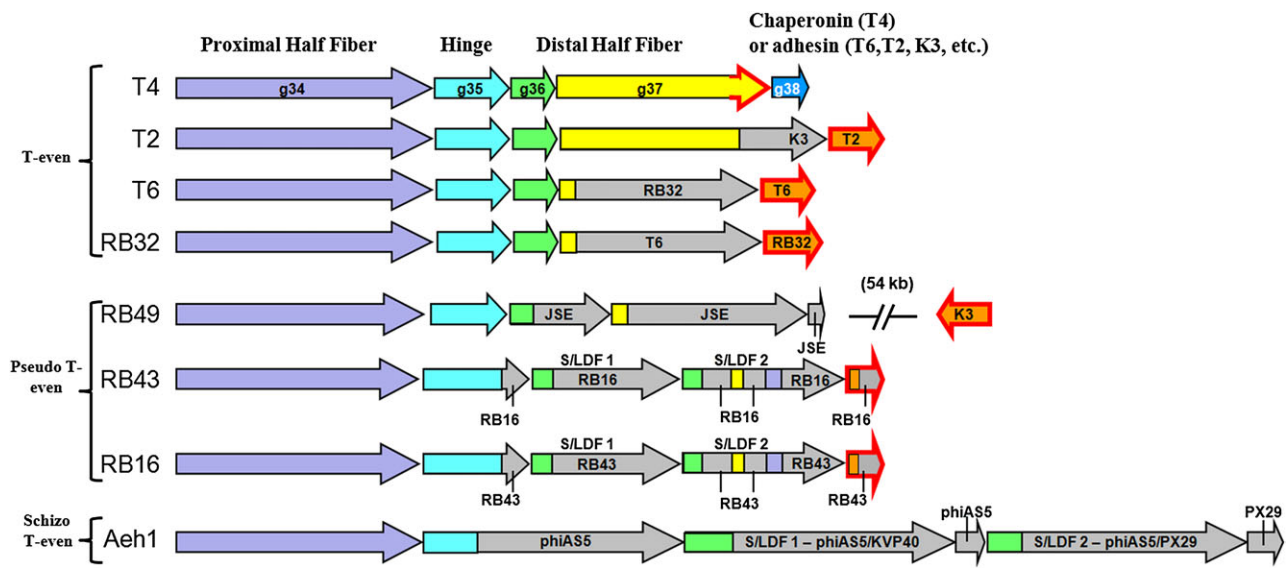
FIG. 1.—Hyperplasticity of the tail fiber modules of various members of the T4 superfamily compared with the classical T-even phages T2, T4, and T6. For each gene, regions of T4 homology (or for the adhesins, homology to T6 [AF052605]) are indicated by the same colors, regions of similarity to other phages are marked by gray, The best phage BLASTp homologies are indicated in the gray sections. Sequences with known adhesin function and their homologues are indicated by red arrows. RB49's *g38* adhesin homologue has been displaced to a distant locus on this genome. Some phages do not have identifiable adhesin genes (e.g., Aeh1). The length of the LTF module in the T4 genome is 10 kb.

gp37 fusion protein is folded into a trimeric fiber with the aid of two specialized chaperones, one encoded by the adjacent T4 *g38* and the second by the distant T4 *g57* (Hashemolhosseini et al. 1996).

The fibrous components of the LTF (gp34, gp36, and gp37) are constrained by the necessity to form the rigid three-stranded β-helical structure. However, the sequences compatible with such β-helical structures (that include multiple short α-helical domains) are widespread among phages genomes (Tétart et al. 1998) and hence swapping of these interchangeable constituent fibrous motifs probably explains much of the LTF's plasticity. Nevertheless, other conserved motifs can also mediate genetic shuffling (Tétart et al. 1998). For example, the gp38 adhesins are also quite plastic, but the sequence constraints on their modular shuffling appear to be more restrictive than those imposed on the fibrous genes.

Most of the LTF loci depicted in figure 1 have a similar size, but phylogenetically distant phages such as RB43 (belonging to the Pseudo T-even subgroup) and Aeh1 (Schizo T-even subgroup) have duplications of *g37*. The function of these duplications is unclear, the size of the virion's LTFs indicate that they contain only a single trimer of gp37. This could be a random mixture of the two gp37 subunits or it could be assembled from just a single expressed version of the two g37 sequences.

In many of the LTF loci in figure 1, there are small open reading frames (ORFs) located just beyond gp37. For such phages as T2, T6 and Ox2, these gene *38* sequences encode adhesins (Riede et al. 1987; Tétart et al. 1998). The location of

such adhesin genes is not always conserved. In the genome of the distant Pseudo T-even phage RB49, the *g38*-like sequence has been displaced elsewhere in the genome and, in the place usually occupied by *g38*, there is a small ORF of unknown function (Desplats 2002). In the more distant Schizo T-even phage Aeh1, where the *g37* has been tandemly duplicated (fig. 1), each of the duplicated *g37* sequences is followed by a small ORF of unknown function. Sequences homologous to these ORFs have been found in diverse phage genomes, including other T4-like *Aeromonas* phages. Although they could be novel adhesins, they may also be specialized LTF chaperons (Hashemolhosseini et al. 1996).

The Modularity of the Phylogenetically Distant RB43 *g38* Adhesin

RB43 is a coliphage that belongs to the Pseudo T-even subgroup of the T4 superfamily whose genomes have diverged substantially from the classical T-even coliphages. In the RB43 genome (fig. 1), at the place in LTF locus usually occupied by *g38*, there is an 174 aa ORF whose first ~50 aa is homologous to T6 gp38, but the C-terminus of this sequence is unrelated to the T6 adhesin specificity domain. Based on this gene's partial homology to T6 gp38 and its context next to a T6-like *g37*, it seemed likely to be the RB43 adhesin function with a novel C-terminal ~120 aa sequence that encodes its specificity domain. The closely related phages RB42 and RB16 both have an RB43 *g38* homologue at the same location in their genomes.
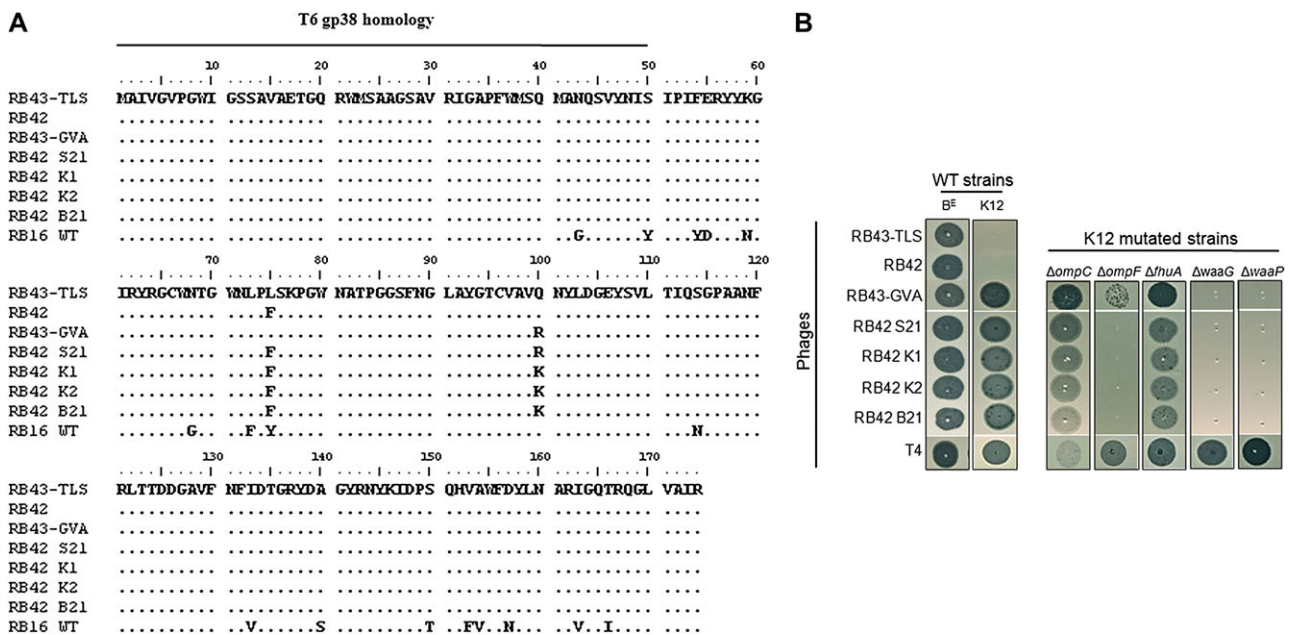
Fɪɢ. 2.—Analysis of the phylogenetically distant RB43 *g38* adhesin. (*A*) Comparison of gp38 protein sequences of RB42, RB43, RB16, and host range mutants. Homologies in the alignment are indicated by a dot and the differences by a letter. Above the sequences there is a bar indicating the region of homology with the T6 gp38 adhesin. (*B*) Spot tests of high titer stocks (5 μl of ~10^8 PFU/ml) of phages on *Escherichia coli* B^E, *E. coli* K12, and K12 strains with phage receptor deletions (outer membrane proteins OmpC, OmpF, and FhuA; and LPS synthesis proteins WaaG and WaaP). RB43-GVA shows better growth on *E. coli* K12 than on *E. coli* B^E (approximately one log). T4 was used as control. WT, wild type.

We have a variant of RB43 that has a clearly distinguishable host range: our original RB43 (now called RB43-TLS) spawned a host range derivative called RB43-GVA. RB43-TLS was propagated in Toulouse on *E. coli* B^E because it grew so poorly on *E. coli* K12, whereas RB43-GVA had been propagated in Geneva on various *E. coli* K12 host strains (Georgopoulos C, Ang D, and Genevaux P, personal communication). Considering how the phages had been maintained for 20 years, their difference in host range is probably the consequence of a spontaneous mutation of RB43-GVA that allows propagation on *E. coli* K12. The most likely location for such a mutation would be in the gene *38*-like sequence. Consequently, we sequenced this gene in both RB43-TLS and RB43-GVA and found a mutation (residue 100 (Q→R); fig. 2A) in the C-terminus of RB43-GVA. We also sequenced *g38* of phage RB42, which previous sequence analysis had shown to be closely related to RB43 (Desplats and Krisch 2003). The host range of RB42 appears to be identical to RB43-TLS, infecting *E. coli* B^E well and being unable to plaque on *E. coli* K12. The amino acid sequences *g38* of RB42 and RB43-TLS were identical (fig. 2A), except for a single seemingly silent mutation at position 75 (F in RB42 and L in both RB43-TLS and RB43-GVA).

To confirm the role of the gp38 C-terminal domain in RB43 host range determination, we selected RB42 or RB43-TLS mutant phages that, like RB43-GVA, made normal plaques on the *E. coli* K12 W3110. In addition, *E. coli* B^E host strains were isolated that were resistant to wild-type RB43-TLS or RB42

phages and these hosts were used to isolate phage mutants that grew normally on them. Sequencing of these *g38* mutants (RB42 B21, RB42 S21, RB42 K1, and RB42 K2) revealed that, in each case, the Q100 residue had been invariably altered, usually to a lysine, but occasionally to an arginine, as in RB43-GVA (fig. 2A).

RB16 is another coliphage in the RB series that has an LTF locus similar to RB42 and RB43, but its gp38-like sequence has diverged further, with 17 altered residues all but two of them being located in the C-terminal domain. The skewed distribution of sequence divergence in RB16 *g38* is compatible with the C-terminal segment encoding the adhesin module. The simple fact that the gp38 proteins of T6 and RB43 have no sequence homology in their C-terminal domain, but similar functions, is the strongest evidence available for the modular structure of gp38-like adhesins.

Studies employing hosts with deletions of various bacterial receptors should allow us to identify the bacterial receptors recognized by the RB43-like adhesins. Unfortunately, the best-characterized sets of such receptor mutants are in the *E. coli* K12 genetic background, and these host strains are resistant to wild-type RB42 and RB43-TLS. It is clear from figure 2B, however, that RB43-GVA which grows on K12 strains requires the LPS receptors because it fails to grow on Δ*waaG* and Δ*waaP* which are defective for core LPS synthesis. RB43-GVA also requires OmpF as a secondary receptor. All four of the RB42 mutants selected for growth on K12 also require the core LPS and OmpF, but in their case in

conjunction with OmpC and FhuA. This complex pattern of recognition of multiple host receptors will be further discussed in the succeeding sections on the targets of gp38-like adhesins.

The Bacterial Receptors Used by Coliphages of the T4 Superfamily

The receptors located on the bacteria's outer surface include many types of molecules: diverse proteins, LPS, teichoic acids, pili, flagella, and capsular components (Heller 1992). Such molecules play different roles in the cell, as structural components, membrane porins, chemical receptors, cellular adhesins, etc. In the T4 superfamily, the phages T6 and K3, for example, use as their receptors the *E. coli* Tsx and OmpA outer membrane proteins, respectively (Hantke 1976; Manning and Reeves 1976). Some phages can recognize more than one host receptor, such as T4 that recognizes both OmpC and LPS or T2 that targets both OmpF and FadL (Datta et al. 1977; Hantke 1978; Yu and Mizushima 1982; Riede et al. 1985; Morona and Henning 1986; Black 1988).

We have identified the host receptors recognized by a number of the phages belonging to the T4 superfamily (table 1). Many of phages belong to the T-even subgroup that infects *E. coli* and closely related enterobacteria. Thus, it is easy to characterize their receptors using a set of *E. coli* K12 strains having deletions of the genes encoding outer membrane proteins or LPS components that are known receptors (Keio collection, Baba et al. 2006). Quantitative spot dilution tests allowed us to determine which phages were unable to propagate on the various mutant host strains. Although many of the T-even phages require only a single bacterial receptor, others such as RB32, Ox2, RB69, and PST have more complex behavior (table 1). These phages require at least two host receptors to be fully infective. For example, phage RB32 requires OmpA but is also sensitive to the LPS composition. Similarly, infection by the phages RB69, RB51, and SV76 requires OmpF as well as the FhuA outer membrane protein and the full-length core oligosaccharide of the LPS. PST phage infection has a yet more complex recognition specificity, requiring OmpA, Tsx, and FhuA as well as the LPS. However, the LPS composition can have significant indirect effects on the cell surface protein composition (Ried et al. 1990).

The Organization of the Sequence Plasticity within the gp38 Adhesins

Sequence alignments of the gp38 adhesins of a few T-even phages (Tétart et al. 1998) indicated that the sequence plasticity was not uniform—the C-terminal domain being more variable than the relatively conserved N-terminus. A compilation of the sequences of 27 gp38 adhesins (fig. 3) clearly defines both the most and the least conserved segments of the protein. This comparison confirmed that the gp38

N-terminal attachment domain (fig. 3) has only a few major well-conserved variant forms: T6-, T2-, and K3-like. Drexler et al. (1986) and Tétart et al. (1998) have argued that the function of this domain is to attach the gp38 adhesin to the tip of the LTF. The gp37 C-terminus has a similar set of subtypes, and each of these is invariably associated in the genome with only one of the gp38 N-terminal subtypes, as to be expected for a lock-and-key interaction between the two domains.

Although there is considerable variability in the gp38 C-terminal domain (fig. 3), this plasticity is not random: it is strikingly organized. This domain contains four hypervariable segments (HVS1, 2, 3, and 4) that are separated by five conserved glycine-rich motifs (GRM1 to GRM5). The only other universally conserved motif in this segment is the small sequence at the C-terminal extremity. Drexler et al. (1989) proposed a structural model for the gp38 adhesin in which the five GRMs aggregate into a glycine-rich core with the HVSs between them radiating outwards in the form of omega loops. Thus, each omega loop would be a HVS and the roles of the intervening GRMs would be to form the core structure that presents the recognition motifs in the HVSs.

The gp38 sequence alignment reveals a strict conservation of both the sequence and synteny of the GRMs. In contrast, the HVSs are highly variable, although not randomly so—each one seems to have a restricted pallet of motifs (fig. 3). With a few exceptions (see below), gp38s with similar receptor recognition specificities also share common HVSs indicated by a common shading color in figure 3. Some sets of HVSs are frequently associated together. For example, nearly all of the gp38 adhesins that recognize the Tsx receptor have essentially the same set of four HVSs, perhaps because they recognize a series of conserved Tsx epitopes. The adhesins which recognize multiple receptors might be expected to have a chimeric set of HVSs, but this is generally not the case. For example, PST differs from T6 by only a few residues in HVS2 and 3 (fig. 3), but it has gained the capability to recognize OmpA (and potentially FhuA and LPS; table 1). Similarly, SV76 differs from T2 by only single residues, in both HVS2 and in HVS3, yet it recognizes OmpF, FhuA, and LPS instead of only FadL as in T2. The phage Mi does have a chimeric structure, having a T6-like HVS1 and novel sequences for the remaining HVSs. It does not, however, recognize the Tsx receptor.

Genetic Analysis of the *g38* Sequences that Determine Phage Host Range

The pair of phages RB5 and RB32 has significantly different C-terminal sequences and they target different receptors (Tsx for RB5 and OmpA/LPS for RB32). Thus, they were interesting candidates for creating chimeric *g38* adhesins. To do this, one of these phages was used to infect a cell with
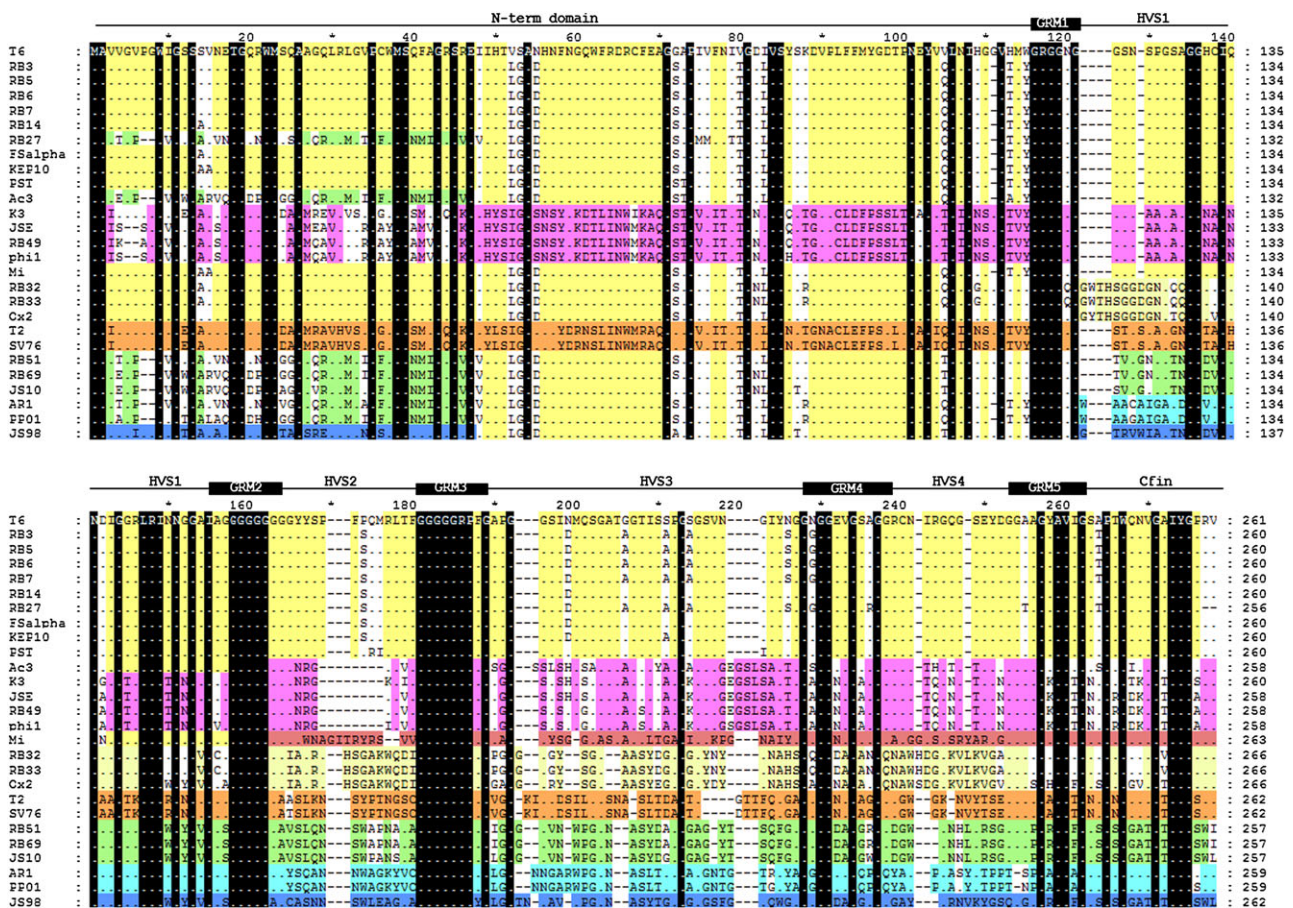
Fig. 3.—Comparison of the gp38 adhesin sequences of 27 T4 superfamily phages. Residues that are conserved in all these aligned sequences are in black boxes, and nonconserved residues are indicated in white. A series of subtypes can be identified primarily on the basis of the sequence in the C-terminal domain; these groupings are indicated by the different colored horizontal boxes. The last two codons of RB27 have not been determined. Above the sequences there is a schematic diagram indicating the identity/function of the adhesin's different segments: GRMs and HVSs.

a plasmid containing the other phage's version of the *g38* sequence. Such an infection will produce some recombinant phages with a *g38* sequence that had acquired *g38* sequences from the plasmid. Although there is no extended sequence homology in the g38 C-terminus, recombinants could be generated by crossovers within the small conserved GRM motifs (Tétart et al. 1998). Using a selection for such exchanges (see Materials and Methods), we obtained six RB5 phages that had recombined with the RB32 plasmid-borne sequence and could grow on a host resistant to parental RB5 (*E. coli* B[E]/RB5). Similarly, eight recombinant phages were obtained with the RB33 sequence (which has a *g38* sequence identical to RB32) and all 14 of these phages were confirmed by DNA sequencing as having replaced the sequences in their HVSs with RB32/33-derived sequences (fig. 4A). Among these, the smallest segment of RB32 sequence was in the chimeric phages ST403 and ST404. Both of these chimeric phages recognize the OmpA and LPS receptors targeted by the

RB32 adhesin (fig. 4B). Thus, the boundaries of the adhesin specificity determinant were delimited to the 400-bp segment containing the four HVS. Surprisingly, among the 14 chimeras we analyzed, none had an exchange between GRM1 and GMR5 that would have generated a novel assortment of the HVSs. Hence, we tentatively conclude that the phenotype we used to select these recombinants (growth either on an RB5- or T6-resistant hosts) requires a full (or nearly so) set of HVSs derived from RB32. Alternatively, the sequences in the gp38 C-terminal domain of RB5 and RB32/RB33 may have diverged so much that the much preferred site for exchange is in the conserved sequences that flank this specificity domain. Another curious aspect of these results is that some of the chimeras analyzed contained point mutations in the HVSs, including silent ones (data not shown). Among the explanations for this is that the recombination events that generated the chimeric phages also caused highly localized error-prone repair (Goodman 2002).
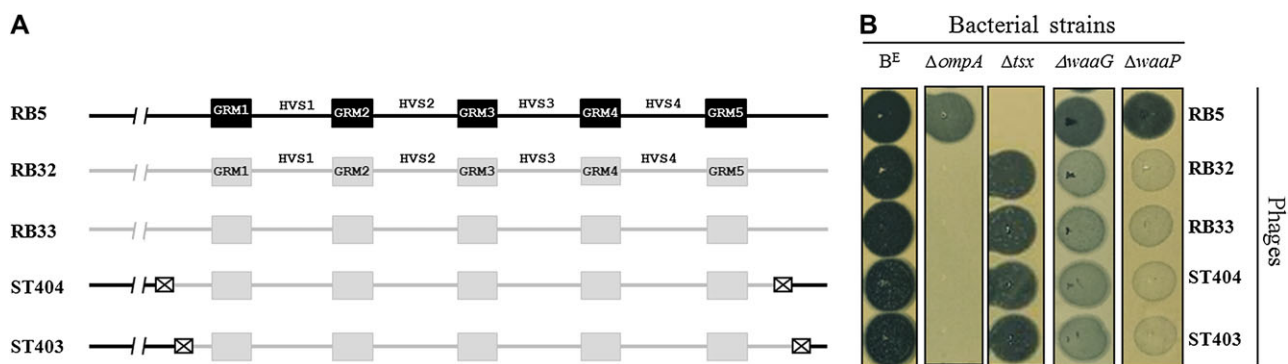
Fig. 4.—Adhesin domain swapping between gene *38* sequences of phages RB5 and RB32. (*A*) Sequences of parental phages and recombinants resulting from RB5 × RB32 or RB5 × RB33 *g38* crosses. The gene *38* coding region of phage RB5 and RB32 (or RB33) are indicated by either black or gray lines and boxes; the five GRM are indicated by rectangular filled boxes and the 4 HVSs are between them. The N-terminal coding region of gene *38* is partially represented, as indicated by dashes upstream of GRM1. The recombinant phages ST404 and ST403 have the smallest segment of RB32 sequence we obtained that confers the host range of RB32. The approximate positions of the genetic exchanges that created the modified host range are indicated by horizontal boxes with crosses within. The RB32/33-derived sequences in the chimeras are indicated in gray, whereas the RB5 sequence is in black. (*B*) Spot tests of high-titer stocks of the same phages as in (*A*) on *Escherichia coli* B[E] (control) and K12 strains with phage receptor deletions (outer membrane proteins Tsx and OmpA and LPS synthesis proteins WaaG and WaaP).

A Set of gp38 Adhesin Sequences that can Recognize Either *E. coli* or *Y. pseudotuberculosis* Receptors

To characterize further the roles of the different C-terminal components of the gp38 adhesins in receptor-binding specificity, we focused on a subset of the T4-type phages whose adhesins have a dual recognition specificity that allows them to infect both *E. coli* and *Y. pseudotuberculosis*, a serious human and animal pathogen (Naktin and Beavis 1999). These two bacterial species diverged ~375 to 500 Mya (Deng et al. 2002), but some phages with a T6-like gp38, such as PST and Mi (supplementary fig. S1, Supplementary Material online), have essentially equivalent plating efficiencies on both hosts. The sequence of PST's *g38* differs from that of T6 by only two nucleotides in HVS2 (resulting in Q167R, M168I) and by a pair of adjacent nucleotide changes in HVS3 (resulting in G208I). Consequently, it was not surprising that many of our *E. coli* phages with T6-like adhesin sequences such as RB3, RB5, RB6, RB7, RB14, and RB27 made plaques at a low, but detectable, frequency on *Yersinia* (supplementary fig. S1, Supplementary Material online): for example, RB14, whose gp38 C-terminal specificity domain differs from the PST sequence by just three residues in HVS2 and two in HVS3 or the phage RB5, whose C-terminal gp38 domain sequence differs from PST by 11 aa, mostly located in HVS2 and HVS3. In *E. coli*, all of these phages with T6-like adhesin sequences recognize the Tsx receptor (table 1). Interestingly, the PST adhesin targets both the Tsx and OmpA proteins of *E. coli* and to a lesser extend the LPS. Apparently, only a few point mutations in the T6-like HVS2 and HVS3 sequences suffice to expand the adhesin's recognition specificity to target an outer membrane protein or LPS of *Yersinia*. There are no homologs of *E. coli* Tsx in any of the *Y. pseudotuberculosis* genomes

that have been sequenced, but these *Yersinia* strains do have other outer membrane proteins, such as OmpA, with similar β-barrel structures having exposed extracellular loops like the Tsx protein (fig. 5). Such structurally related molecules could be weakly recognized by a gp38 adhesin targeting Tsx. Minor modifications in the adhesins Tsx recognition domain might be able shift recognition to another related *Yersinia* β-barrel protein (e.g., OmpA). The Tsx structure of *E. coli* has been solved (Ye and van den Berg 2004), as has that of *E. coli* OmpA protein (Pautsch and Schulz 1998). Although their sequences have only small patches of homology, it is evident from structural comparison that the two proteins share common features (fig. 5). The sequence of the Tsx surface motifs that are the binding sites for phage T6 adhesin are located in the extracellular loops (Schneider et al. 1993; Ye and van den Berg 2004), but the sequence of these Tsx loops are not obviously conserved in the *Yersinia* OmpA sequence. Hence, a priority objective of future research will be to identify the *Yersinia* protein(s) and their features that are the targets of the Tsx-type *g38* adhesins.

One plausible explanation for the ability of a phage like PST to switch hosts between *Yersinia* and *E. coli* is that some of its HVSs recognize β-barrel porin epitopes that are common to the *E. coli* and *Y. pseudotuberculosis* versions of this protein (fig. 5), but other epitopes could differ more between the two versions of the porin. To test this hypothesis, we isolated a set of T6*g38am* phage that had a genetic exchange with the wild-type PST *g38* sequence cloned into a plasmid. We then sequenced a series of such phage with recombinant *g38* which had been selected for a significantly improved ability to plaque on *Yersinia*. We found three different types of T6 × PST chimeric phages capable of efficiently infecting both strains. These had either exchanges in the HVS2 or
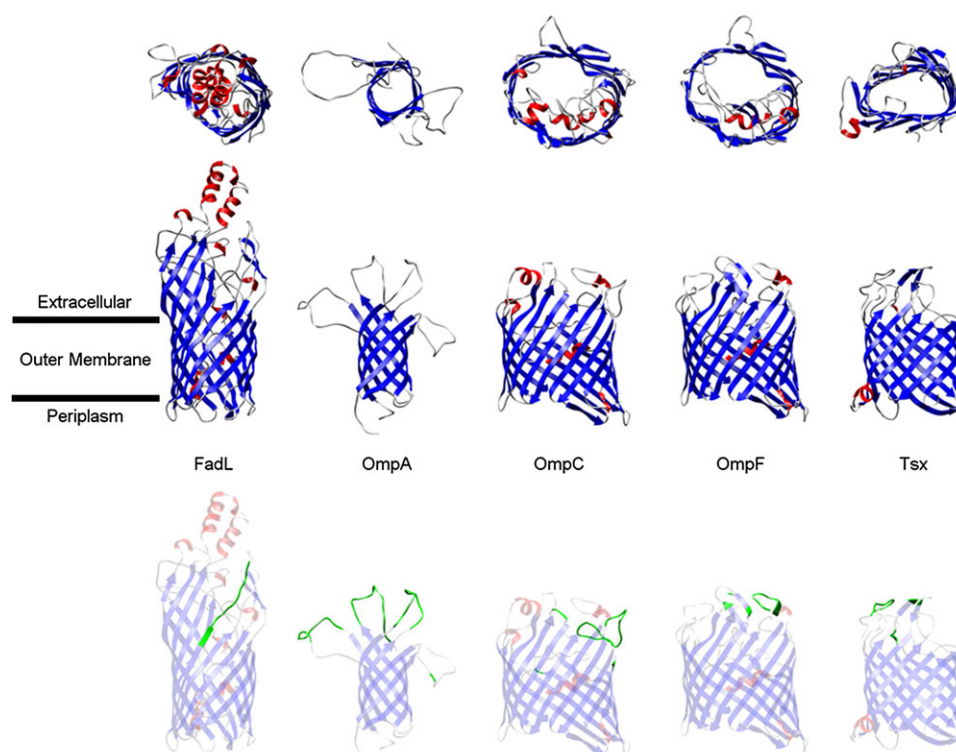
Fig. 5.—Structures of the various porin proteins utilized by T4-like phages as receptors. Top line: top view through the pores. Middle Line: profile view showing orientation through the outer membrane. Bottom line: the residues highlighted in green are those known, via biochemical/mutational analyses, to be involved in phage attachment to the receptor (i.e., their deletion abolished phage attachment). Note that, in all cases, the proteins are monomers as minimal functional units, but some associate into pseudo-dimers (simply two independent porin molecules side-by-side) or pseudo-trimers in the membrane.

HVS3 sequences or both (fig. 6; phages 322, 321 and 527, respectively). Interestingly, although recombinants with both the HVS2 and HVS3 sequences of PST grew best, either of these HVS alterations by themselves significantly improves absorption on *Yersinia*. Apparently, the HVS2 and HVS3 determinants involved in the recognition of *Yersinia* behave in a co-operative fashion. Finally, as a control, we sequenced a recombinant phage with poor plating efficiency on *Yersinia* (similar to T6): this phage (329; fig. 6) contained none of the PST sequence alterations in HVS2 and HVS3.

Although the phage PST requires both the OmpA and Tsx receptors for efficient recognition of *E. coli*, the T6 chimera having the PST HVS2 and HVS3 sequences primarily recognizes *E. coli* Tsx, but other receptors appear to be implicated as well (e.g., OmpA and LPS). Thus, PST's recognition of *E. coli* OmpA must involve altering more than the 3 aa in the HVSs (167R, 168I, and 208I). A simple explanation for this would be that the sequence differences in the gp38 C-terminus clearly influences PST's ability to recognize OmpA, but other components are also involved, for example, the sequences of the gp38 N-terminus.

The phage Mi also grows well on both *Yersinia* and *E. coli*, but it shares only the HVS1 sequence with PST; all of the other HVSs are novel (fig. 3). In *E. coli*, the Mi adsorption strongly depends on the OmpA receptor and to a lesser

extent on the LPS receptors (table 1), but as in PST, this could be an indirect consequence of the interplay between the expression of these receptors. Regardless, it is clear that at least some of the adhesins with dual *E. coli/Yersinia* host range recognize the OmpA receptor. The *E. coli* and *Y. pseudotuberculosis* OmpA sequences are surprisingly close, so some phage gp38s (e.g., Mi and PST) may be able to recognize both versions equally well, but others could require point mutations to switch recognition specificity between them.

Deletion Analysis of the GRMs and HVSs

The HVSs play a leading role in the adhesin's specificity determination, but there may be more HVSs than are required for adsorption. In such a case, the supernumerary HVSs would have been retained to provide evolutionary flexibility by facilitating host range switching. To investigate such design redundancy, we attempted to create a set of HVS deletion mutants, expecting, if this hypothesis was correct, some of these deletions would be viable. This approach required phages with C-terminal *g38* nonsense mutations and these were constructed by mutagenic PCR of a T6 *g38* cloned on a plasmid (see Materials and Methods). We obtained T6 *g38* strains with amber mutations located just upstream of either
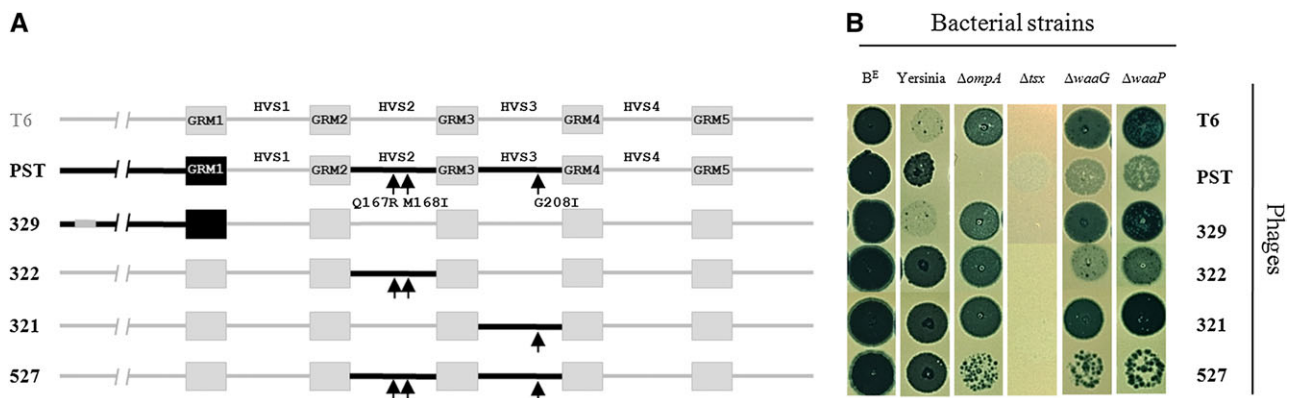
Fig. 6.—Adhesin domain swapping between gene *38* sequences of phages T6 and PST. (*A*) Sequences of parental phages and recombinants resulting from T6 × PST *g38* crosses. The C-terminal coding regions of *g38* that are specific to phage PST are indicated by black lines, whereas their common regions are indicated by gray lines; the 5 GRMs are indicated by rectangular filled boxes and the 4 HVSs are between them. T6 and PST C-terminal domains differ by only three amino acids whose exact positions and sequence differences are indicated by vertical arrows. Recombinant phages 332, 321, and 527 but not 329, were selected selected on their ability to infect *Yersinia*. (*B*) Spot tests of high titer stocks (3 μl of 10^8–10^9 PFU/ml) of the same phages as in (*A*) on *Yersinia pseudotuberculosis*, *Escherichia coli* B^E (control), and K12 strains with phage receptor deletions (outer membrane proteins Tsx and OmpA and LPS synthesis proteins WaaG and WaaP).

GRM3 (T6*g38am*GRM3) or GRM5 (T6*g38am*GRM5). Phages making these aborted versions of gp38 adhesin had normal virion morphology but were defective for adsorption and hence did not make plaques (data not shown).

Gene *38* hybrid plasmids were constructed that contained a series of deletions in the individual GRMs and HVSs. These were used to attempt to make viable phages containing the different *g38* deletions by genetic exchange between the T6*g38am* phages and the *g38* deletion plasmids. To do this, each of these plasmid strains was infected with a T6*g38am* and a set of progeny that made plaques on a host without an amber suppressor were isolated and their *g38* sequences were determined. Most of recombinant progeny produced by such infections would be expected to have the *g38* amber mutation replaced by the *g38Δ* sequence on the plasmid, but these can only make plaques if the deleted sequence is nonessential. In situations where the *g38am* mutation and the *g38Δ* sequences actually overlap, the recombinant progeny can only be *g38Δ*. In this case, a failure to obtain any recombinants indicates that the *g38* deleted sequence is essential. When we analyzed progeny phages corresponding to such situations (e.g., phage T6*g38am*GRM3 and plasmid p*g38Δ*GRM3), we obtained extremely few plaques. In order to discriminate phages bearing a nonessential deleted sequence from the pseudo wild-type progeny generated by the rare reversion of the *g38am* mutation, we employed a PCR screening strategy using a forward primer located in the deleted sequence and a reverse primer located outside of the deletion (see Materials and Methods). Among 48 phages thus analyzed in this experiment, we found none carrying the deletion mutants (table 2). DNA sequence analysis indicated that all of the viable progeny recovered had a reversion in their *g38am* mutation.

In the situation where the *g38Δ* and the *g38am* mutations are nonoverlapping, both the wild-type and *g38Δ* recombinant progeny can be produced by the infection. Nevertheless, using the PCR screening procedure, we did not find any deletions among the 220 phages that we screened from infections with plasmids bearing deletions in the different GRM and HVS motifs. Hence, we conclude that the various sequence deletions that were tested (ΔGRM2/3/4 and ΔHVS1/2/3) are all required for efficient adsorption.

Conclusion

The view of the gp38 adhesin that emerges here is that of a protein designed to sophisticatedly exploit a set of modular recognition sequences. This kind of arrangement provides the adhesins with the flexibility to alter their target specificity. The elegance of this design is that the recognition operates by a series of interchangeable modular-binding sequences (HVSs) that provide the adhesin with the crucial plasticity in its targeting because swapping HVSs creates new combinations of host-specificity determinants. Nevertheless, because some of the specificity determinants may remain unchanged, both the old and the new adhesins will often share at least some recognition features. The host is continually changing its surface topology to avoid phage infection. Hence, a phage stratagem of relying on both conserving previously successful elements for host recognition and on random recruitment of similar elements from a much wider population seems well adapted to creating the required diversity in a semi-intelligent (or at least not totally random) fashion. There is an enormous library of interchangeable adhesin-binding domains within diverse T4 superfamily genomes and recombination in these phages is both efficient and promiscuous. This adhesin design provides a powerful "networked" solution to the problem of successful switching of adhesin specificities—a problem that is probably too large and too complicated to be solved with the genetic resources of an individual phage.

Table 2

Characterization of Recombinant T6g38am+ Phages

Phage[a]	Infected strain	Lysate Titer on B^E log$_{10}$	PCR Analysis Fraction of Δg38 phages	Fraction of g38am+ revertants	Codon at *am* position
T6g38amGRM3	CR63	5.2	N/A[b]	5/5	TAT (am173Tyr)
	DH5α pg38ΔHVS1	7.5	0/29	0/4	WT[c]
	DH5α pg38ΔHVS2	4.4	0/32	4/4	3 × TAT (am173Tyr)
					1 × TCG (am173Ser)
	DH5α pg38ΔHVS3	7.7	0/28	0/4	WT
	JM109 pg38ΔGRM2	7.6	0/16	0/2	WT
	JM109 pg38ΔGRM3	4.2	0/16	2/2	1 × TAT (am173Tyr)
					1 × TAC (am173Tyr)
	JM109 pg38ΔGRM4	8.5	0/16	0/2	WT
	JM109 pg38WT	8.7	N/A	0/5	WT
T6g38amGRM5	CR63	4.9	N/A	5/5	1 × TAT (am235Tyr)
					1 × TCC (am235Ser)
					3 × TGG (am235Trp)
	DH5α pg38ΔHVS1	7.3	0/27	0/3	WT
	DH5α pg38ΔHVS2	7.6	0/32	0/7	WT
	DH5α pg38ΔHVS3	7.0	0/25	0/4	WT
	JM109 pg38ΔGRM2	7.7	0/16	0/2	WT
	JM109 pg38ΔGRM3	7.7	0/16	0/2	WT
	JM109 pg38ΔGRM4	7.9	0/15	0/3	WT
	JM109 pg38WT	8.5	N/A	0/5	WT

[a] T6g38amGRM3 or T6g38amGRM5 phages stocks were made in CR63 and the titer estimated at 10.79 and 10.76 log$_{10}$, respectively. T6g38amGRM3 and T6g38amGRM5 titers on DH5α are, respectively, 4.7 and 10.5 log$_{10}$. The titers of wild-type T6 obtained on *Escherichia coli* CR63, B^E, and DH5α are approximately 11 log$_{10}$.

[b] Not applicable.

[c] Wild type.

Finally, it should be noted that the structure of the receptor-binding domain of the trimeric phage T4 gp37 adhesin has just recently been solved (Bartual et al. 2010). Although the relevance of this T4 structure to the functionally comparable, but phylogenetically unrelated, T6 adhesin structure is unclear, both of them bind to similar target molecules (e.g., porins and LPS). At the distal tip of T4 g37 adhesin, there is bulb-like appendage that contains many of the residues indentified by mutations as being host range determinants. The outer diameter of this bulb and the inner diameter of the cavity formed by the T4 receptor are about the same size (25 Å) and computerized docking simulations indicate that the bulb snugly fits in the exposed cavity of the OmpC porin (Bartual et al. 2010). Interestingly, there are dockings possible where the 3-fold axes of both the trimeric bulb domain and the porin are aligned. It would hardly be surprising if the T6-like adhesins also docked to their receptors in a related fashion.

Supplementary Material

Supplementary figure S1 and table S1 are available at Genome Biology and Evolution online (http://www.gbe.ox-fordjournals.org/).

Acknowledgments

Literature Cited

Ackermann H-W, Krisch HM. 1997. A catalogue of T4-type bacteriophages. Arch Virol. 142:2329–2345.

Arbiol C, Comeau AM, Kutateladze M, Adamia R, Krisch HM. 2010. Mobile regulatory cassettes mediate modular shuffling in T4-type phage genomes. Genome Biol Evol. 2:140–152.

Baba T, et al. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2:2006.0008.

Bartual SG, et al. 2010. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proc Natl Acad Sci U S A. 107:20287–20292.

Black PN. 1988. The *fadL* gene product of *Escherichia coli* is an outer membrane protein required for uptake of long-chain fatty acids and involved in sensitivity to bacteriophage T2. J Bacteriol. 170:2850–2854.

Botstein D. 1980. A theory of modular evolution for bacteriophages. Ann N Y Acad Sci. 354:484–491.

Boubakri H, Beuf M, Simonet P, Vogel TM. 2006. Development of metagenomic DNA shuffling for the construction of a xenobiotic gene. Gene. 375:87–94.

Campbell A, Botstein D. 1983. Evolution of the lambdoid phages. In: Hendrix RW, Roberts JW, Stahl FW, Weisberg RA, editors. Lambda II. New York: Cold Spring Harbor Laboratory Press. p. 365–380.

Carlson K, Miller ES. 1994. Experiments in T4 genetics. In: Karam JD, editor. Molecular biology of bacteriophage T4. Washington (DC): ASM Press. p. 421–483.

Comeau AM, Arbiol C, Krisch HM. 2010. Gene network visualization and quantitative synteny analysis of more than 300 marine T4-like phage scaffolds from the GOS metagenome. Mol Biol Evol. 27:1935–1944.

Comeau AM, Bertrand C, Letarov A, Tétart F, Krisch HM. 2007. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. Virology. 362:384–396.

Comeau AM, Krisch HM. 2005. War is peace—dispatches from the bacterial and phage killing fields. Curr Opin Microbiol. 8:488–494.

Comeau AM, Krisch HM. 2008. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. Mol Biol Evol. 25:1321–1332.

Comeau AM, Short S, Suttle CA. 2004. The use of degenerate-primed random amplification of polymorphic DNA (DP-RAPD) for strain-typing and inferring the genetic similarity among closely related viruses. J Virol Meth. 118:95–100.

Datta DB, Arden B, Henning U. 1977. Major proteins of the *Escherichia coli* outer cell envelope membrane as bacteriophage receptors. J Bacteriol. 131:821–829.

Deng W, et al. 2002. Genome sequence of *Yersinia pestis* KIM. J Bacteriol. 184:4601–4611.

Desplats C. 2002. Characterization of the genetic diversity of the T4-type bacteriophages.[PhD thesis]. [Toulouse (France)]: Université Paul Sabatier.

Desplats C, Krisch HM. 2003. The diversity and evolution of the T4-type bacteriophages. Res Microbiol. 154:259–267.

Drexler K, Riede I, Henning U. 1986. Morphogenesis of the long tail fibers of bacteriophage T2 involves proteolytic processing of the polypeptide (gene product 37) constituting the distal part of the fiber. J Mol Biol. 191:267–272.

Drexler K, Riede I, Montag D, Eschbach ML, Henning U. 1989. Receptor specificity of the *Escherichia coli* T-even type phage Ox2. Mutational alterations in host range mutants. J Mol Biol. 207:797–803.

Epstein RH, et al. 1963. Physiological studies of conditional lethal mutants of bacteriophage T4D. Cold Spring Harb Symp Quant Biol. 28:375–394.

Filée J, Tétart F, Suttle CA, Krisch HM. 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proc Natl Acad Sci U S A. 102:12471–12476.

George DG, Yeh L-Su L, Barker WC. 1983. Unexpected relationships between bacteriophage lambda hypothetical proteins and bacteriophage T4 tail-fiber proteins. Biochem Biophys Res Commun. 115:1061–1068.

Goodman MF. 2002. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. Annu Rev Biochem. 71:17–50.

Hantke K. 1976. Phage T6-Colicin K receptor and nucleoside transport in *Escherichia coli*. FEBS Lett. 70:109–112.

Hantke K. 1978. Major outer membrane proteins of *E. coli* K-12 serve as receptors for the phages T2 (protein Ia) and 434 (protein Ib). Mol Gen Genet. 164:131–135.

Hashemolhosseini S, Stierhof YD, Hindennach I, Henning U. 1996. Characterization of the helper proteins for the assembly of tail fibers of coliphages T4 and lambda. J Bacteriol. 178:6258–6265.

Heller KJ. 1992. Molecular interaction between bacteriophage and the gram-negative cell envelope. Arch Microbiol. 158:235–248.

Hendrix RW, Duda RL. 1992. Bacteriophage λPaPa: not the mother of all λ phages. Science. 258:1145–1148.

Henning U, Hashemolhosseini S. 1994. Receptor recognition by T-even-type coliphages. In: Karam JD, editor. Molecular biology of bacteriophage T4. Washington (DC): ASM Press. p. 291–298.

Kim J-S, Davidson N. 1974. Electron microscope heteroduplex study of sequence relations of T2, T4, and T6 bacteriophage DNAs. Virology. 57:93–111.

Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F, Rossmann MG. 2003. Structure and morphogenesis of bacteriophage T4. Cell Mol Life Sci. 60:2356–2370.

Manning PA, Reeves P. 1976. Outer membrane of *Escherichia coli* K-12: Tsx mutants (resistant to bacteriophage T6 and colicin K) lack an outer membrane protein. Biochem Biophys Res Commun. 71:466–471.

Morona R, Henning U. 1986. New locus (*ttr*) in *Escherichia coli* K-12 affecting sensitivity to bacteriophage T2 and growth on oleate as the sole carbon source. J Bacteriol. 168:534–540.

Naktin J, Beavis KG. 1999. *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. Clin Lab Med. 19:523–536.

Nolan JM, Petrov V, Bertrand C, Krisch HM, Karam JD. 2006. Genetic diversity among five T4-like bacteriophages. Virol J. 3:30.

Pautsch A, Schulz GE. 1998. Structure of the outer membrane protein A transmembrane domain. Nat Struct Biol. 5:1013–1017.

Petrov VM, et al. 2006. Plasticity of the gene functions for DNA replication in the T4-like phages. J Mol Biol. 361:46–68.

Ried G, Hindennach I, Henning U. 1990. Role of lipopolysaccharide in assembly of *Escherichia coli* outer membrane proteins OmpA, OmpC, and OmpF. J Bacteriol. 172:6048–6053.

Riede I, Degen M, Henning U. 1985. The receptor specificity of bacteriophages can be determined by a tail fiber modifying protein. EMBO J. 4:2343–2346.

Riede I, Drexler K, Schwarz H, Henning U. 1987. T-even-type bacteriophages use an adhesin for recognition of cellular receptors. J Mol Biol. 194:23–30.

Rossmann MG, Mesyanzhinov VV, Arisaka F, Leiman PG. 2004. The bacteriophage T4 DNA injection machine. Curr Opin Struct Biol. 14:171–180.

Russell RL, Huskey RJ. 1974. Partial exclusion between T-even bacteriophages; an incipient genetic isolation mechanism. Genetics. 78:989–1014.

Schneider H, Fsihi H, Kottwitz B, Mygind B, Bremer E. 1993. Identification of a segment of the *Escherichia coli* Tsx protein that functions as a bacteriophage receptor area. J Bacteriol. 175:2809–2817.

Stahl FW, Murray NE. 1966. The evolution of gene clusters and genetic circularity in microorganisms. Genetics. 53:569–576.

Suttle CA. 2005. Viruses in the sea. Nature. 437:356–361.

Tétart F, Desplats C, Krisch HM. 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. J Mol Biol. 282:543–556.

Tétart F, Repoila F, Monod C, Krisch HM. 1996. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. J Mol Biol. 258:726–731.

Wommack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev. 64:69–114.

Wood B, Patricia M, Lyle L, Dickson RC. 1978. Attachment of tail fibers in bacteriophage T4 assembly. J Biol Chem. 253:2437–2445.

Ye J, van den Berg B. 2004. Crystal structure of the bacterial nucleoside transporter Tsx. EMBO J. 23:3187–3195.

Yu F, Mizushima S. 1982. Roles of lipopolysaccharide and outer membrane protein OmpC of *Escherichia coli* K-12 in the receptor function for bacteriophage T4. J Bacteriol. 151:718–722.

Associate editor: Purificación Lopez-Garcia