# Intragenic Suppression of a Capsid Assembly-Defective P22 Tailspike Mutation

Patricia A. Maurides, John J. Schwarz[1] and Peter B. Berget

*Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890*

## ABSTRACT

The tailspike protein of bacteriophage P22 assembles with mature capsids during the final reaction in phage morphogenesis. The gene *9* mutation hmH3034 synthesizes a tailspike protein with a change at amino acid 100 from Asp to Asn. This mutant form of trimeric tailspike protein fails to assemble with capsids *in vivo*. By using *in vitro* quantitative tailspike-capsid assembly assays, this mutant tailspike trimer can be shown to assemble with capsids at very high tailspike concentrations. From these assays, we estimate that this single missense mutation decreases by 100–500-fold the affinity of the tailspike for capsids. Furthermore, hmH3034 tailspike protein has a structural defect which makes the mature tailspike trimers sensitive to SDS at room temperature and causes the trimers to "partially unfold." Spontaneously arising intragenic suppressors of the capsid assembly defect have been isolated. All of these suppressors are changes at amino acid 13 of the tailspike protein, which substitute His, Leu or Ser for the wild type amino acid Arg. These hmH3034/sup3034 mutants and the separated sup3034 mutants form fully functional tailspike proteins with assembly activities indistinguishable from wild type while retaining the SDS-sensitive structural defect. From the analysis of the hmH3034 mutant and its suppressors, we propose that in the wild-type tailspike protein, the Asp residue at position 100 and the Arg residue at position 13 form an intrachain or interchain salt bridge which stabilizes the amino terminus of the tailspike protein and that the unneutralized positive charge at amino acid 13 in the hmH3034 protein is the cause of the assembly defect of this protein. To test this hypothesis we have generated suppressors of the hmH3034 mutation by site-directed, random mutagenesis of codon 13. From the broad spectrum of amino acids at position 13 which function as suppressors of hmH3034 we have concluded that elimination of Arg at position 13 is sufficient in most cases to restore capsid assembly activity to the hmH3034 protein.

THE tailspike protein of bacteriophage P22 is a multifunctional trimeric protein (GOLDENBERG, BERGET and KING 1982) which functions in the adsorption of the phage to susceptible Salmonella strains. The tailspike protein, coded for by P22 gene *9* (BOTSTEIN, WADDELL and KING 1973; BERGET and POTEETE 1980) has proven to be a rich model system in which to study protein structure/function relationships at the protein sequence level. Each monomer in the mature tailspike trimer contains 666 amino acids after removal of its N-terminal Met residue (SAUER *et al.* 1982). The mature wild type trimer is extremely thermostable with a melting temperature of 88° (STURTEVANT *et al.* 1989). It is also resistant to proteolysis by a wide range of proteases and maintains its trimeric structure in the presence of SDS at room temperature. The pathway by which newly synthesized monomers fold and assemble into trimers is, in contrast, partially sensitive to high physiological temperatures of 39°–42° (GOLDENBERG, BERGET and KING 1982). All mutations in gene *9* which result in a

temperature sensitive phenotype produce polypeptide chains which seem to be blocked in *in vivo* protein folding and trimer assembly (SMITH, BERGET and KING 1980; KING *et al.* 1986; VILLAFANE and KING 1989) and aggregate *in vivo* at high physiological temperatures (HAASE-PETTINGELL and KING 1988). These ts mutations cluster in the central third of the gene *9* coding region. Most absolute lethal mutations isolated in gene *9* generate polypeptides which are blocked in trimer assembly; however these mutant polypeptides do not aggregate but accumulate as SDS soluble monomers or are degraded by the cell (BERGET and CHIDAMBARAM 1989; SCHWARZ and BERGET 1989a). Most of these mutations map in the last 20% of the gene *9* coding region. Thus the vast majority of missense mutations in gene *9* (over 100) define steps in protein folding and trimerization in the *in vivo* maturation of nascent tailspike polypeptides into functional trimers and thus provide an entré into the genetic analysis of these processes.

The mature tailspike trimer functions in P22 phage adsorption to Salmonella through its endo-rhamnosidase enzyme activity by binding to and hy-

[1] Current address: Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York 14853.

drolyzing the O-antigen on the surface of the cell (IWASHITA and KANEGASAKI 1973). Through this hydrolysis reaction the phage, attached to the tailspike protein, is brought to the cell surface where DNA ejection occurs resulting in the transfer of the P22 genome into the cell. Phage capsids which lack the tailspike protein neither adsorb to nor infect Salmonella. In contrast to the number of mutations which generate folding and trimerization defects, only two mutations have been identified which specifically compromise this endorhamnosidase activity (BERGET and POTEETE 1980; SCHWARZ and BERGET 1989b). Substitution of Gly for Arg at position 505 or Tyr for Gln at position 489 reduces the endorhamnosidase activity of these mutant tailspike proteins to 1–3% of the wild-type level. These mutant tailspike proteins trimerize and assemble onto phage capsids normally; and although phage carrying these mutant tailspike trimers adsorb to Salmonella, they fail to infect these cells presumably because they cannot hydrolyze O-antigen and bring the phage to the cell surface (BERGET and POTEETE 1980).

The assembly of the tailspike protein onto the capsid occurs *in vivo* as the last step in P22 morphogenesis. This assembly reaction was one of the first phage assembly reactions demonstrated *in vitro* (ISRAEL, ANDERSON and LEVINE 1967) and was shown to proceed under a wide variety of solution conditions. The noncovalent interaction between phage capsids and the tailspike protein is quite remarkable in that no evidence of cooperativity during the assembly reaction has been observed, yet the tailspike protein attachment is so strong that no release or exchange of tailspike protein molecules can be detected *in vitro* (BERGET and POTEETE 1980). Until recently, no genetic evidence has been available which suggests what portion or domain of the tailspike is involved in this assembly reaction and what types of intermolecular interactions are involved. In addition no evidence is available to indicate which of the capsid proteins participate in this assembly reaction. The mutant hmH3034 was isolated in a mutagenesis protocol where the tailspike protein gene cloned into a high copy number plasmid was mutagenized with hydroxylamine (SCHWARZ and BERGET 1989a). This mutant gene 9 produces tailspike protein trimers with <1% the assembly activity of wild type protein (SCHWARZ and BERGET 1989b). To identify the molecular nature of this assembly defect we have isolated both spontaneous and directed intragenic pseudorevertants which restore the assembly activity of this altered tailspike protein and determine the molecular nature of these "suppressing" mutations. Through this analysis we have refined the location of the domain of the tailspike protein which is responsible for capsid-tailspike assembly and identified two amino acid residues in the

tailspike polypeptide which must be in contact with each other in the mature trimer.

## MATERIALS AND METHODS

**Bacterial strains:** *Salmonella typhimurium* DB7000 was used as the standard strain for propagating P22 phage. MS1868 (RENNELL and POTEETE 1985) was used as the primary recipient for plasmid transformations because of its increased transformation frequency compared to DB7000. *Escherichia coli* KK2186 (ZAGURSKY and BERMAN 1984) was used in transformation experiments and as a host for preparing single stranded pJS28 DNA. The $dut^-$, $ung^-$ *E. coli* strain BW313 was used for the preparation of uracil-substituted single stranded DNA for site directed mutagenesis.

**Phage strains:** P22 phage strains deficient in the production of active tailspike protein were propagated by the addition of $10^{11}$ phage equivalents of wild-type tailspike protein in soft agar overlays in plating experiments. For liquid cultures, LB broth (LEVINE 1957) was supplemented with $10^{11}$ phage equivalents of tailspike protein per ml. M13-IR1 was used as the helper phage to produce single stranded plasmid DNA as described in SCHWARZ and BERGET (1989a).

**Phage/plasmid crosses:** Crosses to move gene 9 alleles between P22 phage and plasmid replicons were performed as described in SCHWARZ and BERGET (1989a). pJS28 is a pBR322-derived plasmid which carries the tailspike gene under the control of the lacUV5 promoter. It also contains the origin of replication of the filamentous phage f1 so that single-stranded plasmid DNA can be obtained in filamentous phage particles after infection of $F^+$ strains carrying this plasmid with helper filamentous phage. Because of this latter property, we refer to this plasmid as a filamid. P22 $9^-$hmH3034 was generated by crossing the hmH3034 mutation from pJS28 hmH3034 onto P22. The hmH3034 mutation at codon 100 and suppressors of this mutation at codon 13 were separated from each other by plasmid heteroduplex crosses (SHORTLE 1983). In these crosses, heteroduplexes are formed *in vitro* between pJS28-hmH3034/sup3034 plasmid DNA and pJS28-D1 (SCHWARZ and BERGET 1989a) DNA which carries a deletion removing roughly the first 23 codons of gene 9. This *in vitro* manipulation effectively generates a recombinational intermediate between the two plasmid molecules which is resolved by mismatch repair into recombination products after transformation into competent cells. Gene 9 DNA from nucleotide −39 to 70 (where nucleotide 1 represents the first nucleotide of the codon representing the first amino acid in the mature protein) is missing in pJS28-D1 and is replaced by a HindIII octamer linker. pJS28-D1 and pJS28-hmH3034/sup3034 plasmid DNAs were linearized with HindIII and PstI respectively, phenol extracted, precipitated and resuspended at a concentration of 625 μg/ml in TE (10 mM Tris HCl, 1 mM EDTA, pH 8). To denature the plasmids, 1 μl of each plasmid DNA was mixed together with 6 μl separating buffer (0.13 N NaOH, 0.13 mM EDTA) and incubated for 15 min at room temperature. To anneal the separated strands the reactions were incubated at 50° for 1 hr after adding 10 μl reannealing buffer (0.08 N HCl, 0.05 M Tris HCl, pH 8). This mixture of reannealed linear homoduplexes and circular heteroduplexes was used to transform *E. coli* KK2186 to ampicillin resistance. Plasmid DNA was isolated from several transformants and analyzed with HpaI and HindIII restriction endonucleases. Those which had lost the HpaI site at codon 100 and did not contain a HindIII site were presumed to be the desired recombinants and were verified by DNA sequencing.

**Isolation of revertants of hmH3034:** P22 9⁻hmH3034 phage can only form plaques on plates in soft agar overlays containing purified P22 tailspike protein. Independent "stocks" of hmH3034 were isolated by plating P22 9⁻hmH3034 phage permissively on plates seeded with DB7000 and added tail protein. Individual permissively grown plaques were picked and resuspended in 1 ml of LB broth. One half of each picked plaque was plated in soft agar overlays containing only DB7000 to isolate spontaneously arising revertants. Plaques formed on these plates at a frequency of $10^{-5}$ to $10^{-7}$. One revertant was saved from each original permissively isolated plaque to guarantee independence.

The mutation present in hmH3034 is a G to A transition at nucleotide 298 in the structural gene for the P22 tailspike protein (SCHWARZ and BERGET 1989a). This mutation changes the DNA sequence at codons 99 and 100 from GTT-GAC to GTT-AAC. This transition mutation changes the wild type sequence spanning these codons from a HincII site to a HpaI site and thus represents a restriction fragment length polymorphism (RFLP) mutation. To distinguish between true revertants and pseudorevertants of hmH3034, DNA was prepared from the independent revertant phage and its HpaI digestion pattern compared to DNA isolated from wild type phage and the hmH3034 mutant. DNA was isolated from 30 ml lysates of P22 by the method used for phage λ (MANIATIS, FRITSCH and SAMBROOK 1982).

**DNA sequencing:** DNA sequencing was performed off of single stranded pJS28 DNA templates using the method of SANGER, NICKLEN and COULSON (1977). Single stranded plasmid DNA was produced by infection of E. coli KK2186 containing mutant pJS28 plasmids with phage M13-IR1 according to the method of DENTE, CESARENI and CORTESE (1983). DNA sequencing was done with reagents and procedures provided in the Sequenase kit using the P22 gene 9 custom primers described in SCHWARZ and BERGET (1989a). [³⁵S]dATP was supplied by Amersham.

**Oligonucleotide-directed mutagenesis:** Random mutagenesis of pJS28 plasmids at codon 13 of gene 9 was accomplished using the 64-fold degenerate oligonucleotide 5'-TCT AAC CCT NNN CCA ATC TTC where N indicates all four nucleotides were added at that cycle of synthesis. This oligonucleotide was synthesized on an Applied BioSystems synthesizer at the University of Pittsburgh DNA synthesis facility. Uracil-substituted single stranded pJS28 DNA was prepared from E. coli BW313 carrying the appropriate pJS28 derivative by M13-IR1 infection and was used as a template for the in vitro extension of the above primer as described by KUNKEL (1985). The products of this in vitro reaction were used directly to transform KK2186 to ampicillin resistance. These transformants were screened for active tailspike protein production by the "zone of lysis" test previously described (SCHWARZ and BERGET 1989a).

**Tailspike protein purification:** Tailspike protein was purified from either Salmonella or E. coli strains carrying the high level expression plasmid pJS28 onto which various combinations of mutant alleles of gene 9 had been crossed. Four one liter cultures of these strains were grown in 2.8-liter Fernbach flasks with maximal aeration at 30° in Super Broth (3.2% Bacto Tryptone, 2% Bacto Yeast Extract, 0.5% NaCl, 5 mM NaOH) until the cells had just reached stationary phase. The cells were harvested by centrifugation and the cell paste was resuspended in 10 ml of B buffer (BERGET and POTEETE 1980) containing 25 mM NaCl (B25 buffer) per liter of starting culture and frozen at −80°. The frozen cell paste was thawed, adjusted to 15 mM EDTA, 1 mM dithiothreitol and pH 8 by the addition of 2 M Tris base and the cells lysed by the addition of 15 mg/ml lysozyme

and incubation at 4° for 60 min. Cell debris was removed by centrifugation at 110,000 × g for 90 min. The clarified supernatant fraction was brought to 40% saturation with ammonium sulfate and stirred for 24 hr. The ammonium sulfate precipitate was collected by centrifugation and resuspended in 10 ml of B25 and dialyzed against 4 liters of the same buffer for 24 hr. The dialyzed sample was clarified by centrifugation to remove a small amount of precipitate which formed during dialysis. The tail protein was purified from this dialyzed sample by chromatography on either DEAE Sephadex or on a Waters DEAE 5 PW, 21.5 mm × 15 cm high performance liquid chromatography column. For DEAE Sephadex chromatography 10 ml of the dialyzed sample were loaded onto a 50-ml column equilibrated in B25. The sample was eluted with a 200-ml linear gradient of NaCl from 25 to 200 mM in B buffer. Fractions of 8 ml were collected. For DEAE high performance liquid chromatography, 12 ml of the dialyzed sample were loaded onto the column equilibrated with B25 buffer. The column was washed for 5 min with B25 buffer and then eluted with a 60 min linear gradient from B25 to B215 (B buffer containing 215 mM NaCl) buffer. The flow rate for this chromatography was 7 ml/min and 14-ml fractions were collected. For either chromatographic separation, the P22 tailspike protein was located by SDS polyacrylamide gel electrophoresis (PAGE). Fractions containing pure tailspike protein were pooled and dialyzed against B25 and frozen at −80°. Protein concentrations were determined using the Bio-Rad protein assay using bovine serum albumin as a standard.

**Assembly activity measurements:** Assembly reactions were performed using P22 9⁻D10 capsids as described in SCHWARZ and BERGET (1989b). P22 9⁻D10 carries a deletion in gene 9 which removes codons 473 to 504. In these capsid-tailspike assembly reactions, the tailspike protein attaches to capsids converting them to infectious particles. The deletion phage formed in these in vitro assembly reactions were titered on S. typhimurium DB7000 carrying the plasmid pPB10 (BERGET, POTEETE and SAUER 1983) which contains the wild-type gene 9. Deletion phage which infect these cells recombine with the wild-type gene 9 on pPB10 to generate wild-type phage which form a plaque.

**PAGE:** SDS-PAGE was performed essentially as previously described (GOLDENBERG, BERGET and KING 1982). Wild-type P22 tailspike protein is completely denatured by heating to 100° in SDS cracking buffer and migrates in SDS polyacrylamide gels as a monomer of 72,000 D. However, if the gel samples are not heated above room temperature, the wild-type tailspike protein remains as a trimer and migrates with a mobility characteristic of a much higher molecular weight species. This difference in mobility depending on incubation conditions allows one to determine the quaternary structure attained by mutant tailspike proteins.

## RESULTS

**Properties of the hmH3034 mutant:** The P22 gene 9 mutation hmH3034 is a hydroxylamine induced transition mutation at codon 100 of the P22 tailspike gene which results in the substitution of Asn for Asp in the tailspike protein at this position (SCHWARZ and BERGET 1989a). This amino acid substitution has two major effects on the tailspike protein. It lowers the affinity of the tailspike protein for capsids to <1% that of wild type as measured in an in vitro capsid-tailspike assembly assay (Fig. 1A) (SCHWARZ and BER-

A.



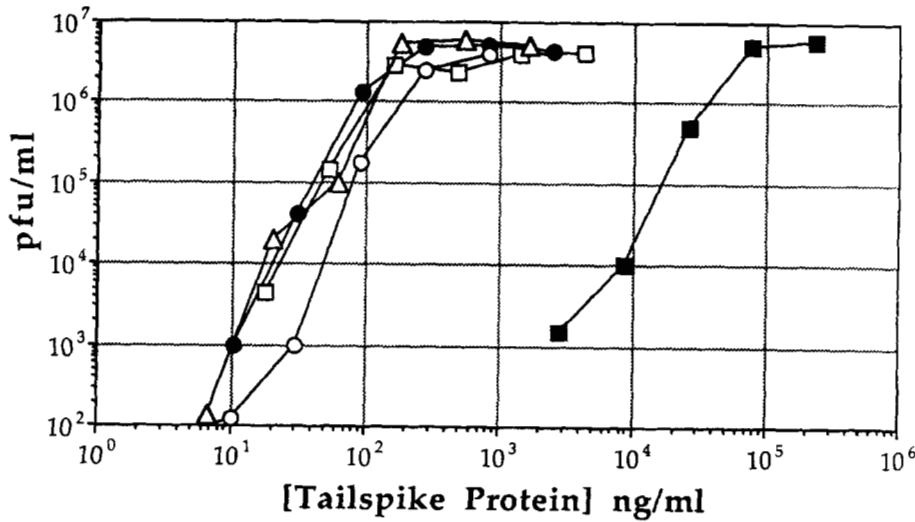[Tailspike Protein] ng/ml
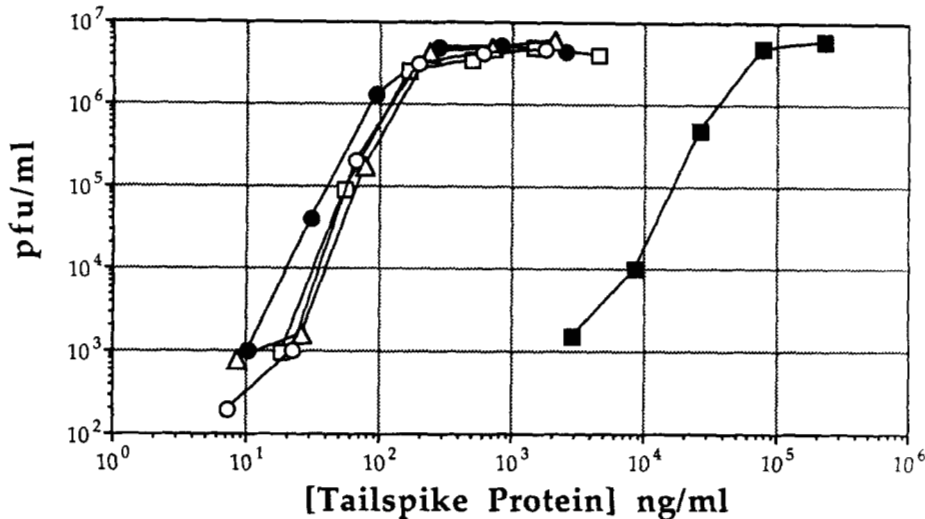
B.



[Tailspike Protein] ng/ml

FIGURE 1.—Quantitative tailspike-capsid assembly assays. A, hmH3034/sup3034 double mutants (●), wild type (■), hmH3034 (○), hmH3034/sup3034-Ser (□), hmH3034/sup3034-His (△), hmH3034/sup3034-Leu. B, sup3034 single mutants (●), wild type (■), hmH3034 (○), sup3034-Ser (□), sup3034-His (△), sup3034-Leu.

GET 1989b). In this assay the plateau which is reached when tailspike proteins is in excess and all capsids are fully occupied by tailspike protein is at the same level as wild type protein indicating that the hmH3034 tailspike protein suffers no loss in endorhamnosidase activity which is required for infectivity. In addition, the hmH3034 tailspike trimer is somewhat unstable in the presence of SDS at room temperature. This can be observed by its unique electrophoretic mobility compared to wild-type protein when run on a SDS polyacrylamide gel without heating the samples in "cracking buffer" above room temperature (Figure 2B). This unique mobility is somewhere between that observed for the wild-type trimer and the wild-type monomer. However, the hmH3034 protein has a mobility indistinguishable from wild-type protein in native gel electrophoresis (SCHWARZ and BERGET

1989b); and, the hmH3034 monomer has the same mobility as the wild-type protein in the SDS gel system after the samples are heated to 100° for 5 min to completely denature the proteins (Figure 2A). This suggests that rather than being due to the charge change associated with the amino acid substitution, the mobility change in the hmH3034 protein is due to partial denaturation in the presence of 1% SDS at room temperature and the binding of more SDS than the wild-type protein.

The hmH3034 mutation also results in the creation of a RFLP in gene 9 and thus in the chromosome of P22. As described in MATERIALS AND METHODS, the hmH3034 mutation changes a HincII site in gene 9 to a HpaI site. This suggested to us that revertants of the P22 hmH3034 mutation could quickly and easily be divided on the basis of their chromosomal HpaI

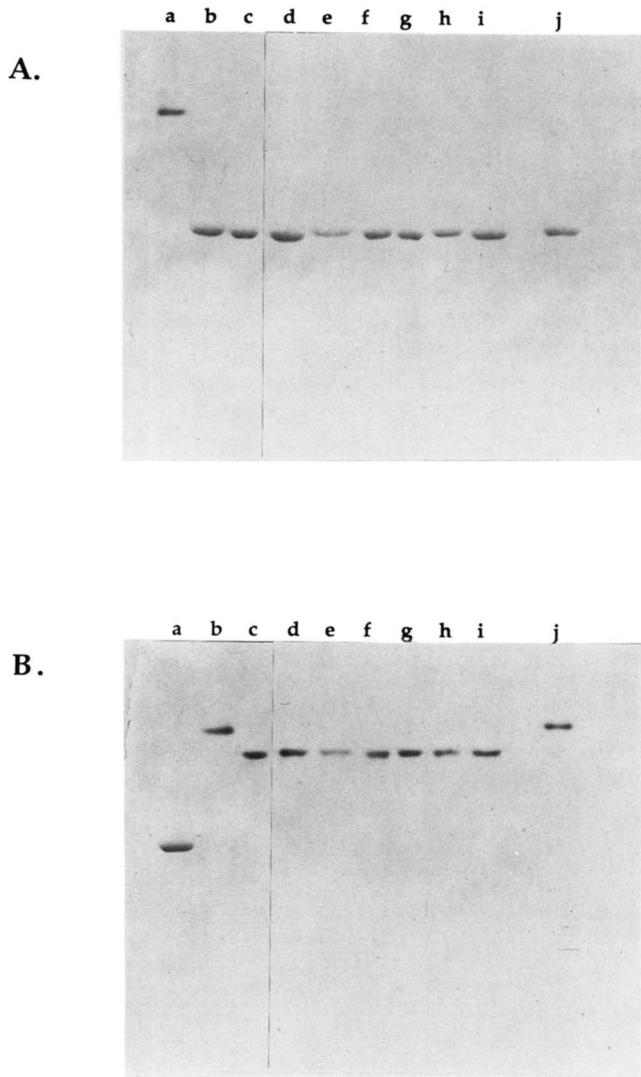a b c d e f g h i j

**A.**



a b c d e f g h i j

**B.**



FIGURE 2.—7.5% SDS-PAGE of wild-type and mutant P22 tailspike proteins. Approximately 2.5 μg of purified tailspike protein were loaded per lane. A, for the analysis of monomer mobilities samples were heated to 100° for 5 min unless otherwise indicated (a), unheated wild type [trimer] (b), wild type (c), hmH3034 (d), sup3034-Ser (e), hmH3034/sup3034-Ser (f), sup3034-His (g), hmH3034/sup3034-His (h), sup3034-Leu (i), hmH3034/sup3034-Leu (j), wild type. B, for the analysis of trimer mobilities samples were not heated above room temperature unless otherwise indicated (a), heated wild type [monomer] (b), wild type (c), hmH3034 (d), sup3034-Ser (e), sup3034/sup3034-Ser (f), sup3034-His (g), hmH3034/sup3034-His (h), sup3034-Leu (i), hmH3034/sup3034-Leu (j), wild type.

digestion pattern into two groups, those which represented true revertants to wild type and those which represented pseudorevertants carrying a suppressor mutation. In that suppressor mutations often provided information as to the nature of the defect that they suppress or information about interacting proteins, we sought to isolate suppressors of the hmH3034 mutation.

**Isolation of spontaneous hmH3034 suppressor mutations:** Twenty-seven independent revertants of P22 9⁻hmH3034 were selected by their ability to

form plaques on DB7000 lawns in the absence of added tailspike protein as described in MATERIALS AND METHODS. Of these, 20 showed a chromosomal *Hpa*I restriction pattern identical to the starting hmH3034 parent (not shown) and thus were presumed to be pseudorevertants, and are indicated as hmH3034/sup3034. To determine whether the pseudorevertants carried either intragenic or extragenic suppressors of the hmH3034 mutation, crosses were performed between each of these phage and the cloned P22 gene *9* carrying the hmH3034 mutation on pJS28 (SCHWARZ and BERGET 1989a). The results from these crosses (not shown) indicated that all of the suppressor mutations were intragenic, *i.e.*, pJS28-hmH3034/sup3034 plasmids were recovered from all pseudorevertant crosses. As a control, one true revertant (as judged by RFLP analysis) was crossed with pJS28-hmH3034 and a plasmid recovered which had lost the associated RFLP.

**Sequence determination of hmH3034/sup3034 mutations:** Ten pJS28-hmH3034/sup3034 (pseudorevertant) plasmids and 1 true revertant plasmid recovered from the above crosses were chosen for subsequent sequence analysis. Because no easy genetic method could be devised to map the locations of the suppressor mutations, the DNA sequence of the region surrounding the hmH3034 mutation was determined using custom oligonucleotide primers for gene *9* as described in MATERIALS AND METHODS. The true revertant plasmid sequence indicated, as suspected from the RFLP analysis, that the hmH3034 mutation at codon 100 had reverted to the wild type sequence. On the other hand, each of the 10 hmH3034/sup3034 plasmids was shown by sequence analysis to carry the hmH3034 mutation at codon 100 and an accompanying change in the DNA sequence at codon 13. Three distinct changes at codon 13 were identified; each of these changed the wild type Arg codon CGT to either CAT (His, 7 isolates), AGT (Ser, 2 isolates) or CTT (Leu, 1 isolate). No other sequence changes were observed within the first 200 codons for each suppressor.

**Characteristics of the sup3034 mutations:** A representative hmH3034/sup3034 double mutant phage from each sequence class described above was plated at 20°, 30° and 40° on lawns of DB7000. Each of these double mutants plated with an efficiency identical to wild type at these different temperatures (not shown). Thus these suppressor mutations in combination with hmH3034 seem to have no distinguishing phenotype other than the suppression of the hmH3034 mutation. In order to assess more carefully the suppression of the assembly defect of hmH3034 by these secondary missense mutations and their phenotypes, the three sup3034 mutations were separated from their accompanying hmH3034 mutations by the

plasmid heteroduplex technique described in MATE-RIALS AND METHODS. pJS28-sup3034 plasmids with a substitution of His, Ser or Leu at position 13 were recovered from these heteroduplex crosses. *E. coli* carrying each of these plasmids was examined by the "zone of lysis" test (SCHWARZ and BERGET 1989a) and shown to produce active tail protein at either 20°, 30° or 40°. Thus the separated sup3034 mutations have no secondary phenotype either by themselves or in combination with the hmH3034 mutation.

**Assembly activity measurements:** To test directly the specific capsid assembly activity of these altered P22 tailspike proteins, each protein was purified as described in MATERIALS AND METHODS and assayed by the quantitative tailing assay. Figure 1A shows the results for the hmH3034/sup3034 double mutant proteins compared to wild-type and the hmH3034 proteins. In these particular assays, the hmH3034 protein has ca. 0.3% the specific activity of wild-type protein, requiring nearly 500 times as much tailspike protein to saturate P22 capsids. In contrast, all of the hmH3034/sup3034 proteins behave nearly as wild type with specific activities in the range of 53–89%. Figure 1B shows the results of quantitative tailing assays performed with the sup3034 proteins. These proteins have assembly activities in the range of 51% to 80% that of wild type and are indistinguishable from the corresponding hmH3034/sup3034 double mutants. Thus within the experimental error of these assays the suppression of the hmH3034 mutation by these three different suppressors is complete and identical.

**Physical characterization of the suppressor proteins:** As described above, the hmH3034 mutation also causes a destabilization of the tailspike trimer as judged by its aberrant mobility on SDS-PAGE when samples are not heated above room temperature. Each of the purified tailspike proteins used in the quantitative tailing assay was examined on such a gel system to determine whether or not the suppression of the assembly defect was accompanied by a suppression of this instability. As can be seen in Figure 2B, each of the hmH3034/sup3034 proteins has the same characteristic trimer instability of the hmH3034 protein. This indicates that the structural defect associated with the hmH3034 mutation is not suppressed by these suppressors. Furthermore, each of the sup3034 proteins has the same characteristic trimer instability of the hmH3034 protein. All of these proteins have the same mobility when the samples are heated to 100° before electrophoresis (Figure 2A) thus the mobility differences apparent in Figure 2A are not due to proteolysis or the charge changes associated with these missense mutations. These data suggest that the instability of the hmH3034, hmH3034/sup3034 and sup3034 proteins is due to a loss of an important

interaction between residues 13 and 100 in the mature polypeptide chain in all of these altered proteins. The most obvious interaction which could be proposed between residues 13 (Arg) and 100 (Asp) in the wild type protein would be an ionic interaction or salt bridge between the negatively charged carboxylate moiety of Asp and the strong positively charged guanido moiety of Arg. This putative salt bridge would be absent in the original hmH3034 mutant protein by the replacement of Asn for Asp at position 100. It would also be absent in the hmH3034/sup3034 proteins again due to the missense mutation at position 100 and, in addition, the loss of the Arg residue at position 13 with its replacement by Ser, Leu or His. Each of the solo sup3034 mutant proteins would have also lost this salt bridge due to the loss of the Arg residue at position 13.

These considerations and the fact that the amino acid residues found at position 13 in the three spontaneously isolated suppressors of hmH3034 bear no structural or functional similarities (Ser, His and Leu) suggest that the capsid assembly defect of the hm-H3034 protein is caused by an indirect effect; that is, the assembly defect is due to the presence of the Arg residue at position 13 which is not "charge-neutralized" by the salt bridge with Asp at position 100 found in the wild-type protein. This free positive charge must somehow compromise the formation of the tail-spike protein-capsid complex.

**Isolation of site-directed hmH3034 suppressor mutations:** To test this hypothesis directly an attempt was made to isolate suppressors of the hmH3034 mutation by introducing "random" amino acids at position 13 in the hmH3034 protein. If the hypothesis were true, one would expect to find a large number of amino acid residues which could suppress the hmH3034 mutation if the amino acid requirement at position 13 were quite flexible. Single-stranded, ura-cil-substituted DNA of the plasmid pJS28-hmH3034 was subjected to a round of site directed mutagenesis using the 64-fold degenerate oligonucleotide spanning codon 13 described in MATERIALS AND METHODS. Plasmids from 30 independent transformants for which tailspike protein activity was restored were sequenced to determine the amino acid replacements at position 13 which suppressed the hmH3034 mutation.

Table 1 catalogs the replacements at position 13 which suppress the hmH3034 mutation generated by site-directed "randomization" of codon 13. Twelve different amino acid changes (including the three isolated as spontaneous suppressors) are capable of suppressing the hmH3034 defect. With few exceptions (noted below) representatives from all structural and functional categories of amino acids can serve as suppressors of the hmH3034 defect. Thus to a first

## TABLE 1

### Codon 13 substitutions in the P22 tailspike gene

| Amino acid | Codon (No. of isolates) |
|---|---|
| Amino acids which suppress the hmH3034 mutation[a] | |
| Ala | GCG (2) |
| Cys | TGC (1), TGT (3) |
| Gln | CAG (1) |
| Glu | GAG (1) |
| His | CAT (2)[b] |
| Leu | CTG (1), CTT (2),[b] CTA (2) |
| Met | ATG (1) |
| Ser | AGT (5),[b] AGC (1), TCG (1) |
| Thr | ACA (1), ACT (1), ACG (1) |
| Trp | TGG (7) |
| Tyr | TAT (1) |
| Val | GTC (3), GTA (1), GTG (1) |
| Amino acids which inactivate tailspike activity[c] | |
| Gly | GGC (1), GGA (1) |
| Pro | CCT (1) |
| Amino acids which produce functional tailspike[d] | |
| Ala | GCT (1), GCC (5) |
| Arg[e] | CGC (1), CGT (1), CGG (1) |
| His | CAT (1)[f] |
| Leu | CTT (1)[f] |
| Ser | AGC (2), AGT (1)[f] |
| Thr | ACC (3) |
| Trp | TGG (1) |
| Val | GTG (1), GCC (1) |

[a] Isolated by site-directed randomization of codon 13 in pJS28-hmH3034 background.

[b] Also isolated as spontaneous suppressor by 9⁻ hmH3034 phage reversion.

[c] Isolated by site-directed randomization of codon 13 in pJS28 (wild type) background.

[d] Isolated by site-directed randomization of codon 13 in pJS28 (Arg 13 Gly [GGC]) background.

[e] Wild-type amino acid.

[f] Isolated as spontaneous suppressor of 9⁻ hmH3034 and separated by an *in vitro* heteroduplex cross as described in MATERIALS AND METHODS.

approximation, removal of the Arg residue at position 13 suppresses the hmH3034 mutation.

The failure of a particular amino acid to appear at position 13 as a suppressor of hmH3034 in the above scheme could be due to several reasons. Besides the fact that a particular amino acid either may not function as a suppressor of the hmH3034 mutation or simply not have been generated in the mutagenesis reaction, some substitutions at position 13 could themselves have a tailspike-negative phenotype. To explore this latter possibility another round of site-directed "randomization" of codon 13 was performed starting with a wild-type pJS28 template. Transformants were isolated which had lost the ability to produce functional tailspike protein and the plasmids from these transformants were sequenced around codon 13. Three different missense mutations resulting in only two amino acid substitutions at codon 13 were identified which caused the loss of tailspike activity, they were Gly (GGC and GGA) and Pro (CCT). Thus by

this analysis Gly and Pro might not be able to serve as suppressors of hmH3034.

**Analysis of silent amino acid substitutions at position 13:** The isolation of missense mutations at codon 13 with a negative tailspike phenotype allows a direct analysis of those amino acids substitutions which have a silent phenotype at position 13. The plasmid pJS28-Arg 13 Gly (GGC) was used as a template in as yet another round of site-directed "randomization" of codon 13, this time screening for mutants which had regained tailspike activity. Nineteen independent tailspike positive transformants were isolated and the sequence at codon 13 determined. Six different amino acid substitutions were recovered which restored tailspike protein activity (see Table 1). These were Ala, Ser, Thr, Trp, and Val which appear as suppressors of hmH3034 as well as the wild-type amino acid Arg.

## DISCUSSION

The P22 9⁻hmH3034 mutation is pleiotropic. It causes a destabilization of the tailspike trimer structure in solutions containing SDS and simultaneously affects the domain of assembly of the tailspike for the P22 capsid. From the above genetic and biochemical analysis it appears that the stability defect caused by the Asp to Asn change at residue 100 in this mutant is direct while the assembly defect is indirect. The exclusive nature of the intragenic suppression of the hmH3034 mutation by alterations at codon 13 leads to the strong prediction that the side chains of amino acid residues 13 and 100 directly interact with each other in the mature tailspike protein structure. While there is no evidence which would suggest that this interaction is either intrapolypeptide or interpolypeptide in the final trimeric structure, the data are consistent with an overall model in which Asp 100 and Arg 13 participate in the stabilization of the amino termini of gp9 in the wild type tailspike trimer by the formation of a salt bridge between these residues. The loss of either of these amino acid residues results in the destabilization of the tailspike trimer in solutions containing SDS. It is not clear whether the loss of this interaction changes the structure of the tailspike trimer in solutions not containing SDS; however, the hmH3034, hmH3034/sup3034 and sup3034 mutant proteins are capable of assembling with capsids *in vitro* and thus retain enough "wild-type" structure in nondenaturing solutions to function in this reaction. The presence of the strongly positive Arg residue at position 13 without an Asp residue at position 100 to neutralize its charge seems to be the cause of the assembly defect found in the hmH3034 mutant. In most cases, changing the amino acid at position 13 to one without such a positive charge is sufficient to

suppress the assembly defect of the hmH3034 mutation.

The hmH3034/sup3034 double mutant in which Arg at codon 13 is replaced by His in the background of the Asp to Asn mutation at codon 100 should provide the starting material to test the charge neutralization hypothesis. Although the *in vitro* capsidtailspike assembly reactions are normally done at pH 7.6, this assembly reaction proceeds efficiently between pH 3 and 10 (ISRAEL, ANDERSON and LEVINE 1967). If the amino acid residue at position 13 is solvent accessible, the His residue at position 13 of this mutant tailspike protein ought to have little if any positive charge characteristics at the pH of the standard *in vitro* assembly reaction. It should be possible by decreasing the pH of the assembly reaction to protonate this His residue and determine if this double mutant protein loses its assembly activity and behaves more like the hmH3034 single mutant.

Twelve of the possible 19 amino acids which could have replaced Arg at position 13 were recovered as suppressors of the hmH3034 mutation. The search by site directed randomization for missense knock-out mutations at codon 13, although not rigorously exhaustive, yielded only two amino acids (Gly and Pro) which result in a nonfunctional tailspike protein in the otherwise wild type background. Thus it is formally possible that the remaining five amino acids (Asn, Asp, Ile, Lys and Phe) may not be able to act as suppressors of the hmH3034 mutation. It is not surprising that Lys did not appear as a suppressor of hmH3034 if the charge neutralization hypothesis is true. However, amino acids structurally similar to the remaining four were recovered as suppressors; thus, the set of isolated suppressors may be incomplete. How these five amino acids function at position 13 with respect to suppression of the hmH3034 mutation or with respect to tailspike function can be determined by performing specific site-directed mutagenesis to create these substitutions.

The list of amino acids which either function as hmH3034 suppressors or are phenotypically silent at position 13 leads one to suspect that the "information content" at that position is low. The wild type amino acid sequence around position 13 is Pro-Arg-Pro. Thus the two substitutions at position 13 which inactivate the tailspike protein (Gly and Pro) generate the sequences Pro-Gly-Pro and Pro-Pro-Pro. These changes might significantly alter the secondary structure in this region of the protein and thus account for their unsuitability. At this time we have not investigated the intracellular fate of these particular mutant proteins. We have, however, examined in a preliminary fashion the SDS stability of the tailspike trimers produced by all the hmH3034/sup3034 double mutants and the silent phenotype mutants at codon 13

listed in Table 1. All of the proteins examined are unstable compared to wild type and have an increased mobility similar to the hmH3034, hmH3034/sup3034 and sup3034 proteins shown in Fig. 2B. A more detailed examination of this set of mutant proteins is in progress.

The three-dimensional structure of the tailspike protein has not been solved. However laser Raman spectroscopy has revealed that the tailspike trimer is at least 50% $\beta$-sheet secondary structure (SARGENT *et al.* 1988). Residues 13 and 100 are most certainly on the surface of the tailspike trimer. Secondary structure prediction algorithms employed by the MacVector 3.0 computer program suggest that residue 100 in the tailspike polypeptide is near the beginning of a short stretch of $\beta$-sheet structure. From inspection and computer analysis of the local sequence immediately surrounding residue 13 it is likely that this region is involved in a polypeptide chain turn. This analysis also suggests that the intervening polypeptide sequence contains at least two regions of $\beta$-sheet secondary structure with an intervening turn near residue 50. It is possible that these two $\beta$ stretches could form an antiparallel sheet punctuated by the turn near residue 50. The N-terminal and C-terminal ends of these secondary structures could be held together by the proposed salt bridge between residues 13 and 100. This could then be the structure destabilized by the loss of this salt bridge in the hmH3034 mutant and its intragenic suppressors. Confirmation of this speculation will have to await the solution of the three-dimensional structure by x-ray crystallography.

If the sole purpose of the Arg residue at position 13 and the Asp residue at position 100 is to provide a stabilizing salt bridge either between subunits of the trimer or within the monomers themselves, it might be possible to engineer this stabilizing effect by some other means. For example, the two amino acid residues could be "switched" in their primary sequence through site directed mutagenesis. In this context Glu and Lys residues could also be tested for their suitability in forming such a salt bridge. And although it is unlikely that disulfide bonds would form *in vivo*, it would be interesting to determine whether two Cys residues placed at positions 13 and 100 could properly interact once the protein is purified to form a disulfide bond which would stabilize the trimer. It seems likely that such substitutions at position 13 would have little effect on the tailspike protein structure or function because of the broad spectrum of position 13 substitutions recovered in this study; however, the effect of each of these substitutions at position 100 would have to be determined independently.

**Assembly site analysis:** It is difficult at this time to locate more precisely the region of the tailspike protein which assembles or makes contact with P22 cap-

sids. It is quite clear, however, that the N termini of the tailspike polypeptides play an important role. In that sense, the tailspike of Salmonella phage P22 is similar to the thin fibers of *E. coli* phage T3 which are homotrimers of gp17 that attach by their N termini to T3 capsids (KATO, FUJISAWA and MINAGAWA 1985a, b; 1986). The amino acid at position 13 in the P22 tailspike most likely does not play a specific role in the assembly of the tailspike with the capsid because at least 13 different amino acid substitutions at this position produce functional tailspike protein. However, an unneutralized, positively charged amino acid at this position, such as that found in the hmH3034 mutant, can interfere with *in vivo* capsid-tailspike assembly. This suggests that it may be possible to isolate an extragenic suppressor which would change the target protein for tailspike assembly to the capsid so that this positive charge could be accommodated. Such a suppressor would identify the as yet unknown target gene and protein for tailspike assembly.

## LITERATURE CITED

BERGET, P. B., and M. CHIDAMBARAM, 1989 Fine structure genetic and physical map of the phage P22 tail protein gene. Genetics **121**: 13–28.

BERGET, P. B., and A. R. POTEETE, 1980 Structure and functions of the bacteriophage P22 tail protein. J. Virol. **34**: 234–243.

BERGET, P. B., A. R. POTEETE and R. T. SAUER, 1983 Control of phage P22 tail protein expression by transcription termination. J. Mol. Biol. **164**: 561–572.

BOTSTEIN, D., C. H. WADDELL and J. KING, 1973 Assembly and DNA encapsulation in *Salmonella* phage P22. I. Genes, proteins, structure and DNA maturation. J. Mol. Biol. **80**: 669–695.

DENTE, L., G. CESARENI and R. CORTESE, 1983 pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11**: 1645–1655.

GOLDENBERG, D., P. B. BERGET and J. KING, 1982 Maturation of the tail spike endorhamnosidase of Salmonella phage P22. J. Biol. Chem. **257**: 7864–7871.

HAASE-PETTINGELL, C. A., and J. KING, 1988 Formation of aggregates from a thermolabile *in vivo* folding intermediate in P22 tailspike maturation. J. Biol. Chem. **263**: 4977–4983.

ISRAEL, J. V., T. F. ANDERSON and M. LEVINE, 1967 *In vitro* morphogenesis of phage P22 from heads and base-plate parts. Proc. Natl. Acad. Sci. USA. **57**: 284–291.

IWASHITA, S., and S. KANEGASAKI, 1973 Smooth specific phage adsorption: endorhamnosidase activity of tail parts of P22. Biochem. Biophys. Res. Commun. **55**: 403–409.

KATO, H., H. FUJISAWA and T. MINAGAWA, 1985a Genetic analy-

sis of subunit assembly of the tail fiber of bacteriophage T3. Virology **146**: 12–21.

KATO, H., H. FUJISAWA and T. MINAGAWA, 1985b Purification and characterization of gene 17 product of bacteriophage T3. Virology **146**: 22–26.

KATO, H., H. FUJISAWA and T. MINAGAWA, 1986 Subunit arrangement of the tail fiber of bacteriophage T3. Virology **153**: 80–86.

KING, J., M-H. YU, J. SIDDIQI and C. HAASE, 1986 Genetic identification of amino acid sequences influencing protein folding, p 275–291 in *Protein Engineering. Applications in Science, Medicine and Industry*, edited by M. INOUE and S. SARMA. Academic Press, New York.

KUNKEL, T. A., 1985 Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82**: 488–492.

LEVINE, M., 1957 Mutations in the temperate phage P22 and lysogeny in *Salmonella*. Virology **3**: 22–41.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

RENNELL, D., and A. R. POTEETE, 1985 Phage P22 lysis genes: nucleotide sequences and functional relationships with T4 and λ genes. Virology **143**: 280–289.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463–5467.

SARGENT, D., J. M. BENEVIDES, M-H. YU, J. KING and G. J. THOMAS, 1988 Secondary structure and thermostability of the P22 tailspike. XX. Analysis by Raman spectroscopy of the wild-type protein and a temperature sensitive folding mutant. J. Mol. Biol. **199**: 491–502.

SAUER, R. T., W. KROVATIN, A. R. POTEETE and P. B. BERGET, 1982 Phage P22 tail protein: gene and amino acid sequence. Biochemistry **21**: 5811–5815.

SCHWARZ, J. J., and P. B. BERGET, 1989a The isolation and sequence of missense and nonsense mutations in the cloned bacteriophage P22 tailspike protein gene. Genetics **121**: 635–649.

SCHWARZ, J. J., and P. B. BERGET, 1989b Characterization of bacteriophage P22 tailspike mutant proteins with altered endorhamnosidase and capsid assembly activities. J. Biol. Chem. **264**: 20112–20119.

SHORTLE, D., 1983 A genetic system for analysis of staphylococcal nuclease. Gene **22**: 181–189.

SMITH, D. H., P. B. BERGET and J. KING, 1980 Temperature-sensitive mutants blocked in the folding and subunit assembly of the bacteriophage P22 tailspike protein. II. Fine-structure mapping. Genetics **96**: 331–352.

STURTEVANT, J., M-H. YU, C. HAASE-PETTINGELL and J. KING, 1989 Thermostability of temperature-sensitive folding mutants of the P22 tailspike protein. J. Biol. Chem. **264**: 10693–10698.

VILLAFANE, R., and J. KING, 1988 Nature and distribution of sites of temperature sensitive folding mutations in the gene for the P22 tailspike polypeptide chain. J. Mol. Biol. **204**: 607–619.

ZAGURSKY, R. J., and M. L. BERMAN, 1984 Cloning vectors that yield high levels of single-stranded DNA for rapid DNA sequencing. Gene **27**: 183–191.

Communicating editor: G. MOSIG