Bacteriophages HK97 Structure: Wholesale Covalent Cross-Linking between the Major Head Shell Subunits

MELANIE P. POPA, TERRY A. MCKELVEY, JOHN HEMPEL, AND ROGER W. HENDRIX

Department of Biological Sciences and Department of Molecular Genetics and Biochemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received 19 December 1990/Accepted 15 March 1991

We describe initial genetic and structural characterizations of HK97, a temperate bacteriophage of Escherichia coli. We isolated 28 amber mutants, characterized them with respect to what phage-related structures they make, and mapped many of them to restriction fragments of genomic DNA. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of HK97 virions revealed nine different protein species plus a substantial amount of material that failed to enter the gel, apparently because it is too large. Five proteins are tail components and are assigned functions as tail fiber subunit, tail length template, and major shaft subunit (two and possibly three species). The four remaining proteins and the material that did not enter the gel are head components. One of these proteins is assigned as the portal subunit, and the remaining three head proteins in the gel and the material that did not enter the gel are components of the head shell. All of the head shell protein species have apparent molecular masses well in excess of 100 kDa; they share amino acid sequence with each other and also with a 42-kDa protein that is found in infected lysates and as the major component of prehead structures that accumulate in infections by one of the amber mutants. We propose that all of the head shell species found in mature heads are covalently cross-linked oligomers derived from the 42-kDa precursor during head shell maturation.

Studies of how bacteriophage virions are assembled from their component macromolecular parts have revealed a wealth of information about how complex structures such as phages arise, how their assembly is regulated, and how the requirements for efficient assembly are integrated with the requirements for function in the assembled structure (8). The details of the assembly and structure of double-stranded DNA (dsDNA) phages, though far from completely understood, are probably better known than those for any other biological structures of comparable complexity, and phages stand as a model for asking questions about assembly and structure in other biological systems. Progress in understanding phage assembly has benefitted from the fact that investigators have concentrated their efforts on a relatively small number of phages and also from the fact that this number—about 10 different phages—was not too small. This moderate diversity of experimental systems has made possible a comparison of the assembly strategies used by different phages, and this in turn has made it possible to learn which are the fundamental features of phage assembly and what variations on the basic theme are possible.

We have begun recently to develop a relatively unstudied bacteriophage, HK97, as an experimental system for investigating assembly. HK97 is a temperate phage of Escherichia coli which was isolated in Hong Kong by Dhallon et al. (13). It shares immunity and host range with phage λ and has been reported to recombine with λ to make hybrid phages at a low frequency (13). It belongs to the same morphological class as λ, having an isometric head and a long, flexible, noncontractile tail (13). An understanding of HK97 assembly and structure will no doubt illuminate a number of features of phage assembly already under study in other phages, but our primary interest in it derives from an aspect of its assembly and structure that is unique among well-studied viruses. As we document here, all ~400 copies of the major head protein of HK97 participate in covalent cross-linking reactions among themselves during the process of head maturation.

Although this sort of wholesale cross-linking has not been seen before, posttranslational modification of structural proteins is a rather common feature of phage assembly. In T4 assembly, for example, four head proteins undergo proteolysis catalyzed by a phage-coded protease (4, 31). For λ, gpB (the portal or head-tail connector subunit) loses 12 amino acids from its amino terminus and gpH (the tail length tape measure) loses ~100 amino acids from its carboxy terminus (24, 25, 42). In a more complex reaction, minor λ head protein gpC and a few subunits of major head protein gpE are both cleaved and covalently attached to each other into two differentially cleaved alternative products, X1 and X2 (23). In all of these phage examples, the processing occurs after the protein is assembled into a head or tail precursor structure. Similar examples of posttranslational, postassembly processing have also been reported for other classes of viruses (11, 38). In eukaryotic organisms there are several well-studied examples of protein cross-linking, including covalent attachment of ubiquitin to other proteins and the formation of intermolecular cross-links between structural proteins such as collagen, fibrin, and elastin (33, 43). The plant lectin concanavalin A undergoes a posttranslational cyclic rearrangement of its amino acid sequence as the result of the cleavage of one peptide bond and the formation of another (6). In yeast cells, there is evidence of posttranslational splicing of a 119-kDa protein to yield a 50-kDa protein from the middle of the original sequence and a 69-kDa protein formed by joining the N-terminal and C-terminal thirds of the original sequence (27). In only a few of these cases is there any real understanding of what biological function the posttranslational processing reaction serves.

We report here the isolation of a set of conditional lethal
mutants of HK97 and their initial genetic and phenotypic characterization, followed by identification of the protein components of the virion and their characterization, with a primary emphasis on the posttranslational, postassembly processing of the major subunit of the head shell.

**MATERIALS AND METHODS**

**Media and cell and phage growth.** Cells were grown in tryptone broth (10 g of Bacto-Tryptone per liter, 5 g of NaCl per liter) or, for radioactive labeling, in RG maltose medium (18), a minimal medium supplemented with the 18 sulfur-containing amino acids. Methods for plating phage and carrying out genetic manipulations and tests were identical to those in use for phage λ (for example, see reference 2).

Phage stocks were grown in liquid culture from single plaques by the “pick, dilute, shake” protocol of Blattner et al. (5).

**Strains and mutant isolation.** *E. coli* 594 (Sm Sup) (40) and Ymel (supF) (44) were used as wild-type-nonsuppressing and amber-suppressing hosts, respectively. *E. coli* 159 (Uv+ Gal+ Sup) (34) was the UV-sensitive strain used in labeling infected lysates, as described below. *E. coli* groEAs9 is described by Georgopoulos et al. (18).

Wild-type HK97 was obtained by UV induction of strain K-12 (HK97) (13), provided by T. Dhillon. Amber mutants were isolated from a spontaneous clear plaque mutant of the wild-type phage. The amber mutants were obtained by treating phage with hydroxylamine (12), by treating infected cells with N-methyl-N'-nitro-N-nitrosoguanidine (32), or by growing phage on *E. coli* mutD5 (16) and screening plaques on strains Ymel and 594.

**Restriction mapping, cloning, and marker rescue.** DNA for restriction mapping and subcloning was obtained by phenol extraction of CsCl-banded HK97 virions. Single- and double-enzyme restriction digestions were carried out according to the recommendations of the enzyme suppliers, and fragments were analyzed by agarose gel electrophoresis.

The EcoRI restriction fragments of the phage DNA were cloned into plasmid pBR322 by mixing EcoRI-digested HK97 DNA and EcoRI-digested plasmid DNA, ligating the DNA, transforming *E. coli* 594, and screening the transformants for plasmids carrying the various HK97 restriction fragments.

The HK97 amber mutants were mapped by spotting a series of dilutions of each mutant on each of a series of strains carrying one of the plasmids generated as described above. The results were scored positive and the mutant was considered to have been mapped when the number of plaques on the strain carrying a particular restriction fragment in its plasmid was at least 10-fold higher than the number on a control strain carrying unmodified pBR322.

**Protein analysis.** Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis of virion proteins was carried out as described by Laemmli (31), with an acrylamide concentration of 10%.

Comparison of proteins by partial proteolysis was carried out as described by Cleveland et al. (10). 35S-labeled protein bands were located on a preparative gel by autoradiography, excised, and loaded into the wells of a 15% polyacrylamide gel. Soaking buffer was added to fill the wells and allowed to rehydrate the gel piece for 30 min, and then to each well were added 30 μl of loading buffer and 10 μl of enzyme buffer containing 5 μg of *Staphylococcus aureus* V8 protease. Electrophoresis was interrupted for 15 min just before the dye entered the resolving gel. After electrophoresis, the gel was fixed with 10% acetic acid-30% methanol, impregnated with En3Hance (New England Nuclear), and exposed to X-ray film at −70°C.

Proteins to be reelectrophoresed without digestion were electroeluted from a rehydrated gel piece, extracted with acetone-triethylamine-acetic acid-water (85:5:5.5 by volume), reextracted with acetone, and dried (21) before being dissolved in gel sample buffer for the second electrophoresis.

**Radioactive labeling of phage, heads, tails, proheads, and infected lysates.** *E. coli* 594 was grown in RG maltose medium to 3 × 10^8 cells per ml and infected with wild-type HK97 at a multiplicity of 5 PFU per cell. Growth was continued at 37°C with aeration, and at 15, 30, and 45 min postinfection, 5 μl of L-[35S]methionine (NEG-009H; New England Nuclear) was added per ml of culture. At 90 min, at which time lysis was evident, a few drops of CHCl₃ were added and incubation was continued for an additional 5 min. Debris was removed by centrifugation (5,000 g, 10 min), and the phage were banded in a CsCl step gradient. The phage were dialyzed against TMSP buffer (6 mM Tris-Cl, pH 7.5, 1 mM MgCl₂, 70 mM NaCl, 10 mM putrescine) prior to gel electrophoresis.

Heads, tails, and proheads were labeled in a similar fashion.

For heads, *E. coli* 594 was infected with HK97amE1, and heads were purified on a CsCl step gradient as described above.

For proheads, *E. coli* 594 was infected with HK97amC2 or HK97amU4 and labeled and lysed as described above. After centrifugation to remove debris, the proheads were concentrated by precipitation with 70% saturated (NH₄)₂SO₄, re suspended in a small volume of TMSP buffer, dialyzed against TMSP buffer, and layered onto a 10 to 30% glycerol gradient in TMSP buffer. The gradient was centrifuged in a Beckman SW41 rotor at 200,000 × g (35,000 rpm) for 50 min, fractions were collected, and the peak of prohead radioactivity was pooled.

For tails, *E. coli* groEAs9 was infected with HK97 and labeled and lysed as described above. (The groEAs9 mutation blocks HK97 head assembly and results in the production of free, biologically active tails [our unpublished results].) After clarification, the lysate was passed over a glass bead column (1 by 8 cm; Sigma Chemical Co.) with a nominal pore size of 24.0 nm by using 10 mM Tris-Cl (pH 7.5)-0.1 M NaCl, and the tail peak was collected and subjected to gel electrophoresis.

To label HK97-encoded proteins made in an infected cell, *E. coli* 159 was irradiated with UV light (3,600 ergs/mm²) prior to infection by HK97, in order to reduce incorporation into cellular proteins. The protocol was as described by Hendrix (22) for phage λ except that RG maltose medium and [35S]methionine were used and the multiplicity of infection was 50. The radioactive label was added at 35 min postinfection, and the cells were harvested by centrifugation 10 min later, resuspended in SDS gel sample buffer, and heated at 100°C for 1 min prior to gel electrophoresis.

**Large-scale preparation of proheads.** Large-scale preparation of proheads was as described above for radioactive proheads, but cells were grown in tryptone broth, typically in 1- to 2-liter batches. A small amount of the corresponding radioactive proheads was usually added to the lysate to aid in locating the proheads on the glycerol gradients.

**Electron microscopy.** Virions and heads were adhered to carbon-coated Parlodion films on copper grids, stained with...
0.5% uranyl acetate, and examined in a Philips EM300 electron microscope at an accelerating voltage of 60 kV.

**Densitometry.** Band intensities for the $^{35}$S-labeled virion proteins were measured from an autoradiograph of an SDS gel of virions by using a Biomed Instruments scanning densitometer. Care was taken to ensure that the film densities were within the linear response range of the film. The number of copies of the different proteins per virion was calculated by assuming equal proportions of sulfur atoms in all proteins, correcting for apparent molecular weights, and setting at 12 per virion the number of copies of the protein represented by each band.

**Peptide preparation, separation, and analysis.** Lyophilized proheads (C2 or U4 proheads) were performic acid oxidized by the method of Allet et al. (1). Following oxidation, the protein was dissolved in 70% formic acid at a concentration of 0.5 mg/ml, and CNBr was added at a 20-fold weight excess. The mixture was incubated at room temperature for 24 h, and after reaction, the mixture was then diluted twofold with water and dried under vacuum (19). The CNBr-cleaved peptides were cleaved with trypsin according to the procedure of Fullmer and Wasserman (17). The peptides were dissolved in 0.2 M N-ethylmorpholine acetate (pH 8.0) at a concentration of 1 mg/ml. Tosylsulfonyl phenylalanin chloromethyl ketone-treated trypsin (Sigma Chemical Co.) was added to an enzyme to peptide ratio of 1:100 (wt/wt). Incubation was at 37°C for 15 h and was followed by a second addition of trypsin and an additional incubation of 9 h. The samples were then lyophilized. Prior to chromatography, they were suspended in 88% formic acid and diluted to 30% (vol/vol) formic acid in water with 0.05% trifluoroacetic acid.

Peptides were separated on a Vydac C4 reverse phase column (dimensions, 0.46 by 25 cm; particle size, 5 μm) by using a Waters high-performance liquid chromatography (HPLC) chromatograph. The peptides were eluted over 2 h with a 0 to 35% linear acetonitrile gradient in water and 0.05% trifluoroacetic acid. The eluant was monitored by measuring its $A_{214}$ and $A_{280}$, and peaks were collected, lyophilized, and saved for further analysis.

Amino acid sequences of peptides were determined with a Beckman 890M protein sequenator by using a 0.1 M quadrol program in the presence of precyced Polybrene. The phenylthiohydantoin derivatives were identified after reverse-phase HPLC (Altex Ultrasphere C18 column) by the program of Hawke et al. (20).

<table>
<thead>
<tr>
<th>HK97</th>
<th>λ</th>
<th>HK97</th>
<th>λmm434</th>
</tr>
</thead>
<tbody>
<tr>
<td>HK022</td>
<td>λ</td>
<td>HK022</td>
<td>λ</td>
</tr>
<tr>
<td>I</td>
<td>B</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>1</td>
<td>80</td>
<td>177</td>
<td>68</td>
</tr>
<tr>
<td>95</td>
<td>83</td>
<td>1814</td>
<td>42</td>
</tr>
<tr>
<td>195</td>
<td>60</td>
<td>26</td>
<td>37</td>
</tr>
<tr>
<td>77</td>
<td>22</td>
<td>132</td>
<td>1613</td>
</tr>
<tr>
<td>187</td>
<td>1</td>
<td>218</td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>84</td>
<td>320</td>
<td>5</td>
</tr>
</tbody>
</table>

**RESULTS**

**Isolation of mutants and mapping.** As an initial step in characterizing HK97, we isolated 28 independent amber mutants by screening mutagenized phage on Sup" and SupF strains of E. coli. We examined lysates of cells infected with each mutant by electron microscopy and classified them with respect to what identifiable structures, if any, were produced (Table 1). By this criterion, three mutants made free tails and structures that appeared to be proheads, and we classified these as having defects in genes involved in head production. One mutant made morphologically normal but unjoined heads and tails and is probably defective in one of the final maturation steps of either heads or tails. Two mutants made no detectable phage-related structures and are therefore probably defective in early gene functions. The remaining 22 mutants made no detectable tails but did make what appeared to be mature heads, and they are therefore classified as defective in tail production. We currently have no clear explanation for the surprisingly large fraction of the mutants that fall into this last category.

Figure 1 shows restriction endonuclease maps of HK97 DNA for six enzymes. The map is linear, and for each enzyme, the end fragments usually appeared in the gel in nonstoichiometric amounts and there was an additional nonstoichiometric band with a size equal to the sum of the end fragments. This suggests that HK97, like its relative λ, has a nonpermuted genome with cohesive DNA ends. Based on the restriction fragment sizes, the size of the HK97 genome is about 40.5 kb.

**TABLE 1. HK97 amber mutants**

<table>
<thead>
<tr>
<th>Mutant(s)*</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2, C2, E4</td>
<td>Proheads, unjoined tails</td>
</tr>
<tr>
<td>C4, U1, U3</td>
<td>No structures</td>
</tr>
<tr>
<td>C6</td>
<td>Full unjoined heads, unjoined tails</td>
</tr>
<tr>
<td>C1, C3, C4, C5, E1, E3, E4, E5, F1, F2, F3, F4, U2, U5, U6, U7, U8, U9</td>
<td>Free unjoined heads, no tails</td>
</tr>
</tbody>
</table>

*The mutant names indicate their mapping to the EcoRI restriction fragments of HK97 DNA, as determined by marker rescue experiments. The "U" series of mutants could not be rescued by any of the available EcoRI fragments and are therefore unmapped.
We attempted to clone all of the HK97 EcoRI restriction fragments into plasmid pBR322. This was successful for all but the two terminal fragments, which we were unable to clone even after joining and ligating the two cohesive ends. We presume that the presence of one or both of these end fragments is deleterious to the cell.

We wished to use the cloned EcoRI fragments, which together represent about 86% of the genome, to map the amber mutants by marker rescue. This could be done initially for 10 of the mutants, which were rescued by plasmids carrying EcoRI fragment C or F (Fig. 1; Table 1). None of the mutants mapped to the small EcoRI G (1.7-kbp) or H (0.9-kbp) fragment.

For fragments A, B, and E, mapping by marker rescue could not be done initially because wild-type HK97 does not form plaques on cells carrying these fragments in plasmids. Exclusion by these plasmids extends to other phages as well, as shown in Fig. 1. Thus, fragment A excludes HK97 and its homimmune relative λ but not the heterimmune phages λimm434 (26) and HK022 (14). It thus seems likely that fragment A carries the HK97 immunity region and excludes infecting phages by making the cell immune. Plasmids carrying fragment B or E, in contrast, exclude HK97 and HK022 but not λ or λimm434. We know from experiments to be published elsewhere that fragment B carries the coding region for the major head protein subunit of HK97. Furthermore, HK97 and HK022 have very similar head proteins, as judged by the identical SDS-polyacrylamide gel patterns produced by their heads (our unpublished observations). We hypothesize that plasmids carrying fragment B may disrupt head assembly of infecting HK97 or HK022 by overproducing some but not all of the proteins required for this process and thereby disrupting the correct ratios of assembly components. A similar mechanism could be operating in the case of fragment E, since the amber mutant phenotypes suggest that this region also includes genes for structural proteins. Alternatively, this exclusion could be caused by a gene analogous to the sieB genes of phages λ and P22 (39), which inhibit growth of susceptible phages at an early stage in the life cycle.

Exclusion could be circumvented in the case of fragment E, because we found that mutants of HK97 which are able to overcome exclusion by fragment E can easily be selected. We selected such mutations in all of the phages carrying unmapped amber mutations by platting these amber mutants on an amber-suppressing strain carrying the plasmid with fragment E. Phages that grew in these conditions had presumably acquired a second mutation that overcomes fragment E exclusion; they were tested to confirm that they still retained the amber phenotype as well. These double mutants were then tested for the ability to rescue the am" allele from the various plasmids in amber-suppressing strains. Five of the amber mutations were mapped in this way to fragment E.

**Morphology of HK97 virions.** Figure 2a is an electron micrograph of HK97 virions, negatively stained with uranyl acetate. As shown earlier by Dhillon et al. (13), HK97 has an isometric head and a long, flexible, noncontractile tail. Fine tail fiber material is visible at the tip of the tail. Measurements taken from a micrograph of a mixture of HK97 and λ (not shown) indicate that HK97 heads are 10% smaller in diameter than λ heads and HK97 tails are 18% longer than λ tails. On the basis of the known dimensions of λ (15, 28), this means that HK97 heads are ~54 nm in diameter and HK97 tails are ~177 nm long. HK97 heads, purified from cells infected with tail mutant HK97amEl, are shown in Fig. 2b.

On many of the heads a small structure protrudes slightly from one of the corners, and in others there is a difference in staining intensity at one corner. Since these structures were never seen in virions with tails, they presumably occupy the unique corner of the head where a tail would attach. We believe that this protruding structure is part of the portal (head-tail connector), an oligomeric structure of 12 subunits that is probably universally present in dsDNA phages (3). As with other phages, the portal is more clearly visible in the heads in Fig. 2b that have lost their DNA.

**Protein composition of virions.** Figure 3a shows the autoradiograph of an SDS-polyacrylamide gel of purified HK97 virions. Nine bands are visible, ranging from a triplet around 25 kDa up to a band with an apparent molecular mass above 300 kDa. In addition, a substantial amount of material failed to enter the gel and remained in the well. Panels b and c show gel patterns for purified heads and tails, respectively. Five of the nine bands can be assigned to the tail, and the other four can be assigned to the head. The material that failed to enter the gel appears to be associated with the head.

The relative intensities of the nine 35S-labeled species in Fig. 3a are indicated in Table 2. On the basis of the apparent molecular weights of the proteins indicated by their gel mobilities, it is possible to calculate the relative numbers of subunits per virion by assuming that all of the proteins have similar proportions of sulfur-containing amino acids. We have scaled these values to give 12 subunits of the protein represented by band 6 per virion; for the reasons indicated below, we believe this to be the correct number of subunits for this protein, so the other values should approximately represent the actual numbers of subunits per virion in each case.

**Properties of tail proteins.** Of the five tail proteins identified above, we propose that the second largest (band 5) is the tail length tape measure protein, arguing by analogy with the head-tail protein, gpH, of λ (24, 29, 30). In λ, tail length is roughly proportional to the molecular weight of the tape measure protein (29, 30), and the putative HK97 tape measure protein is the appropriate size to account for the length of the HK97 tail. Thus, the length of the HK97 tail is 118% of the length of the λ tail, and the HK97 band 5 protein has an apparent molecular weight 120% of that of λ gpH*. The λ tape measure protein is thought to be present in the virion in 6 copies, and this agrees reasonably well with the 4.5 copies of the HK97 protein per virion calculated in Table 2. In addition, HK97 band 5 protein, like λ gpH*, appears to be a cleaved form of a larger precursor. Figure 4 shows a comparison of proteins made in HK97-infected cells with the proteins of HK97 virions. The infected lysate does not have a band in the position of band 5, but there is a band at a position corresponding to a protein of ~12 kDa greater molecular mass. Comparison of these two proteins by partial proteolysis of bands cut out of the gel (Fig. 5, lanes a and b) shows that they are closely related in amino acid sequence. This supports the hypothesis that the band 5 protein is derived by proteolysis from the protein in the infected lysate.

We propose that the highest-molecular-weight tail protein, represented by band 4, is the tail fiber protein, on the basis of its similarity in size to the λ tail fiber protein, gpH, and the similar amount present in the virion (2.3 copies per virion, as calculated in Table 2, as compared with a probable value of 3 copies per virion for λ). Figure 4 lane a, shows that a band of the same mobility as band 4 is present in an infected lysate.

If these assignments are correct, then some or all of the
band 7, 8, and 9 proteins, which together account for \(~65\%\) of the mass of the tail, are left to constitute the tail shaft. The similar apparent sizes of these three proteins suggested that they might be related to each other by differential posttranslational modification. Partial proteolysis patterns of the band 7 and 8 proteins are indistinguishable (Fig. 5, lanes c and d), arguing that they are in fact derived from the same primary translation product. We have not examined the band 9 protein in this regard, so its assignment as a component of the shaft is speculative.

Properties of head proteins. The SDS gel pattern of HK97 heads, shown in Fig. 3b, is strikingly different from the patterns of most previously characterized phage heads, which typically have a major band in the 40- to 60-kDa range and a few minor bands. HK97 has one minor band at 45 kDa (band 6), but the remaining three head protein bands have apparent molecular masses well in excess of 100 kDa, and a large portion of the head protein failed to enter the gel matrix.

In order to test whether the material that did not enter the gel could be made to do so, we treated HK97 virions prior to electrophoresis in a variety of ways designed to solubilize insoluble protein. These included (i) extended heating at 100°C (for up to 10 min) in SDS gel sample buffers containing a variety of concentrations of sodium or lithium dodecylsulfate, (ii) heating at 80 to 100°C in 6 to 8 M guanidinium Cl followed by dialysis into 6 M urea and then SDS gel sample buffer, (iii) treatment with 1 N formic acid or 1 N acetic acid at room temperature or 100°C, (iv) reaction with citraconic anhydride, (v) reduction with β-mercaptoethanol or dithiothreitol followed by carboxymethylation, and (vi) electrophoresis in noncontinuous phosphatase-SDS gels. None of these treatments had any discernable qualitative or quantitative effects on the HK97 gel pattern. The relative distribution of material among the nine gel bands and the top of the gel was also insensitive to the amount of material loaded onto the gel, as determined by comparing a heavily loaded Coomassie blue-stained gel with an autoradiograph of a gel which had been loaded with 1,000-fold less physical material. Finally, the presence of DNA in the head is not a factor in producing the unusual gel pattern, since as we show below, an appropriate DNA-free prohead gives the same gel pattern as heads.

Figure 6 shows a comparison of the head proteins by partial proteolysis. The three slowest-migrating bands in the gel, as well as the material recovered from the top of the gel, all have the same partial proteolysis pattern. On the basis of these results, we hypothesize that the three high-molecular-weight proteins and the material on the top of the gel are derived from a common precursor protein and that the
used those purified from E. coli and tails. The apparent molecular masses shown are based on calibrations against phage λ virion proteins and other markers of known molecular mass. (b) Heads and tails. Heads were purified from an infection by tail mutant HK97amE1. (c) Tails and virions. Tails were purified from E. coli groEA59 cells infected with wild-type HK97.

various species resolved by the gel, as well as the material that does not enter the gel, are different covalently linked oligomers of the common precursor. In similar experiments, the 45-kDa protein has a partial proteolysis pattern different from that of the other head proteins (data not shown).

Since the 45-kDa protein, by this test, appears to be distinctly different from the other head protein species, it is presumably encoded by a different gene. We tentatively assign the 45-kDa protein as the subunit of the portal or head-tail connector structure. The portal is a highly conserved structure built of 12 protein subunits which has been found in all dsDNA phages in which it has been sought (3). As pointed out above, HK97 heads have a small structure at one of the icosahedral corners that we interpret as the portal. The 45-kDa protein is the only plausible candidate to be the portal subunit.

The various forms of the other head proteins must make up the main shell of the head, presumably as some sort of mosaic of the different oligomeric forms. To further investigate whether these are truly discrete forms rather than forms that might interconvert with each other through aggregation and disaggregation, we cut each of the high-molecular-weight bands out of a gel and electrophoresed them a second time. Figure 7 shows that each of the bands ran true. The same is true for the 45-kDa proposed portal subunit and a 42-kDa protein from phage-infected lysates described in the next paragraph.
VOL. 65, 1991

COVALENT CAPSID CROSS-LINKING 3233

The three high-molecular-weight forms of the head shell protein did not appear until a few minutes after the beginning of a pulse-chase labeling of phage-infected cells (data not shown), in accord with the hypothesis that mature forms of the head shell protein are processed forms of a precursor protein. It might therefore be anticipated that the proposed common precursor protein could be found in a pulse-labeled lysate of phage-infected cells. The gel pattern shown in Fig. 4, lane a, of proteins made in HK97-infected cells is in fact dominated by an intense 42-kDa band (distinct from the 45-kDa band found in virions). We propose that this 42-kDa protein is the precursor to the higher-molecular-mass forms of the head shell protein found in mature virions; we provide direct evidence for this proposal in the following paragraphs.

Assembly-arrested proheads. In our characterization of the phenotypes of the amber mutants described above, three mutants, am U4, am E2, and am C2, appeared to accumulate large amounts of proheadlike structures when they infected cells. We investigated two of these mutants (am U4 and am C2) further to see whether they would provide easily purified sources of the precursor and cross-linked forms of the head shell protein. Figure 8 shows SDS-polyacrylamide gel patterns of these U4 and C2 proheads after their purification from infected cells. The pattern for the C2 proheads is virtually identical to the HK97 head gel pattern, including the presence of a large amount of material at the top of the gel. The U4 prohead pattern, on the other hand, consists primarily of a 42-kDa subunit. In addition, it has the 45-kDa portal protein band and a small amount of unidentified high-molecular-mass forms. There is no appreciable amount of material that fails to enter the gel. The partial proteolysis patterns of the 42-kDa protein from U4 proheads and the 42-kDa protein from infected cells are indistinguishable (Fig. 9), arguing that they are the same protein.

Peptide analysis of precursor and cross-linked forms of the shell protein. The fact that the 42-kDa major protein of

FIG. 4. Autoradiograph of SDS-polyacrylamide gel showing extract of HK97-infected cells (lane a) and purified virions (lane b). For lane a, the cells were irradiated with UV light prior to infection to suppress host protein synthesis (see Materials and Methods). The positions of band 5 (in virions) and its proposed precursor (in the infected extract) are indicated. The band in the infected extract indicated as 42kD is the proposed head protein precursor.

FIG. 5. Comparison of partial proteolysis patterns for tail proteins. Radioactive bands were excised from an SDS gel, treated with staphylococcal V8 protease, and electrophoresed in a second gel. Lane a, band 5 from virions; lane b, proposed band 5 precursor from infected cells as in Fig. 4, lane a; lane c, band 7 from virions; lane d, band 8 from virions.

FIG. 6. Partial proteolysis patterns of head proteins. Lane a, material from the well of the SDS gel; lane b, band 3 protein; lane c, band 1 protein; lane d, band 2 protein.
infected cells is the major component of U4 proheads says at a minimum that it is capable of assembling into structures with the appearance and dimensions of proheads. If the 42-kDa protein is in fact the precursor of the higher-molecular-mass forms found in heads and in C2 proheads, the 42-kDa protein and the high-molecular-mass forms found in heads and in C2 proheads, the 42-kDa protein and the high-molecular-mass proteins would share amino acid sequences. To test this possibility, we digested the U4 and C2 proheads into peptides, using CNBr treatment followed by trypsin digestion, and separated the peptides by HPLC. Figure 10 shows the chromatograms for the two digests. Although there are some differences between the two patterns, many parts of the chromatograms appear identical for the two digests. From three pairs of identically eluting peaks, we recovered peptides and determined their amino-terminal sequences by Edman degradation. Table 3 shows these sequences. It is apparent that the identically eluting peptides do in fact have identical sequences, implying that the 42-kDa protein of U4 proheads and the high-molecular-weight forms of the major shell protein found in C2 proheads and virions are derived from the same precursor.

A possible cross-link-containing peptide. Among the peaks in Fig. 10 that do not match between the two chromatograms, we determined the amino acid compositions of several and identified two, U4 15 and C2 15, that have similar compositions. We subjected these to Edman degradation. The peptide from the un-cross-linked 42-kDa protein (U4 15) gave the sequence Ala-Leu-Lys-Pro-Glu-Ser-Asp-Ile... (Table 3). The peptide from the cross-linked C2 proheads (C2 15) gave the identical sequence, except there was no signal at position 3. The match between these two sequences implies that they almost certainly have a common genetic origin, while their different elution times and the difference in the results of the Edman degradation at position 3 show that they are not identical. This result could occur by the addition of something to lysine 3. We interpret these observations to mean that peptide U4 15 is one of two peptides, one on each of two adjacent protein subunits in the head, through which these adjacent subunits are linked together. In the cross-linked proheads, this peptide is linked through the lysine at position 3 to a different peptide in the adjacent protein, and peptide C2 15 is actually the cross-linked pair of peptides derived from the two protein subunits. The fact that we see no signal in the Edman degradation that is not attributable to the first peptide could be explained if the side chain of the amino-terminal residue of the second peptide is cross-linked to lysine 3 so that the first residue of peptide 2 remains bound to peptide 1 after the first cycle of Edman degradation and if, in addition, peptide 2 is so short that its remainder washes out of the reaction cell, precluding identification of those residues.
phages—primarily λ—we are able to assign tentative functional roles to all of these protein species. The similarities between the ~100-kDa tail protein represented by band 5 and its proposed λ homolog, gpH*, are particularly striking. Both are present in similar small numbers of copies per tail, and both are derived from a larger precursor by the proteolytic removal of about 100 amino acids. In keeping with its proposed function as a tail length tape measure, the HK97 band 5 protein is larger than λ gpJ in proportion to the difference in tail length between the two phages. The ~130-kDa tail protein (band 4) is assigned as the tail fiber subunit in part on the basis of its similarity to λ gpJ in molecular mass and number of copies per tail and in part because we find no other protein in the tail that could reasonably be assigned as the subunit for the tail fibers that we see by electron microscopy. Similar logic is used to assign the ~45-kDa head protein (band 6) as the portal subunit. The major head shell subunits (bands 1, 2, and 3 and the material at the top of the gel) and the major tail shaft subunits (bands 7, 8, and 9) are assigned as such because they are the proteins that are present in sufficient quantity to plausibly account for the masses of these major components of the structure.

Except for the tail fiber subunit (band 4) and the portal subunit (band 6), all of the protein species in the phage appear to have experienced some form of posttranslational processing. Thus, the tail tape measure protein (band 5) is cleaved from a larger precursor, the major tail protein is present as at least two and possibly three different related forms, and the major head protein is extensively cross-linked into three discrete related forms that enter the gel and an unknown number that fail to do so. (In addition to the cross-linking, we have also found recently that the 42-kDa major head protein loses material from its amino terminus through a proteolytic cleavage that precedes cross-linking [14a].)

The cross-linking of the major head subunits appears to affect all of the subunits, since there is none of this protein found in the virion in any but the high-molecular-weight forms. This is, to our knowledge, the first report of such extensive cross-linking in any virus. The chemical nature of the cross-link is not yet known. However, we present evidence above for a tryptic peptide that, at a minimum, is modified in the cross-linked (C2) proheads as compared with the un-cross-linked (U4) proheads, and we propose that the modification is the addition of a peptide from the adjacent subunit in the shell by virtue of forming the cross-link. If this interpretation is correct, then one half of the cross-link is almost certainly formed from the side chain of the internal lysine residue in this peptide, as discussed above. We expect that the other half of the bond would probably come from a glutamine or asparagine residue, which could form (with a loss of ammonia) an amide linkage to the ε-amino group of the lysine, but other possibilities cannot be ruled out at this time.

Another question that will require further investigation concerns the topology of these cross-links in the head. Evidence from experiments to be published elsewhere suggests that the two most intense bands of those that enter the gel represent pentamers and hexamers (bands 3 and 2, respectively). Viruses built according to the model proposed by Caspar and Klug (9), which are thought to include phages such as HK97, have their head protein subunits arranged around fivefold and quasi-sixfold axes of rotational symmetry. Thus, it is attractive to speculate that these species are groups of five or six subunits cross-linked into rings around

**DISCUSSION**

We report here a preliminary genetic characterization of HK97 in which we have isolated a set of amber mutants, determined some features of their mutant phenotypes, and mapped many of them to specific regions of the physical map. We expect the availability of the amber mutants to be very useful for future studies of the structure and assembly of HK97; in fact, three of them have been crucial in the studies on HK97 composition and structure we report here.

The protein composition of HK97 virions, as revealed by SDS gel electrophoresis, includes nine distinct species that form bands on the gel and a substantial amount of material that fails to enter the gel, most likely because it is too large. Five of these species are components of the tail, and four plus the material at the top of the gel are components of the head. Largely on the basis of analogy with other, similar

---

**TABLE 3. Amino-terminal sequences determined from selected peptides from Fig. 10**

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2 11</td>
<td>...VATAYDTSLN</td>
</tr>
<tr>
<td>U4 12</td>
<td>...VATAYDTSL</td>
</tr>
<tr>
<td>C2 16</td>
<td>...SLGSDADSAGQ</td>
</tr>
<tr>
<td>U4 20</td>
<td>...SLGSDADSAG</td>
</tr>
<tr>
<td>C2 33</td>
<td>...AIYQVTE</td>
</tr>
<tr>
<td>U4 34</td>
<td>...AIYQVTEsEFsAaGIVLNP</td>
</tr>
<tr>
<td>C2 15</td>
<td>...AL/3PSDDit</td>
</tr>
<tr>
<td>U4 15</td>
<td>...ALKPEsDitf</td>
</tr>
</tbody>
</table>

* The sequences are indicated in the one-letter amino acid code. Lowercase letters indicate uncertain determinations. The parentheses at position 3 of peptide C2 15 indicate that no readable signal was obtained in that cycle.
each of these preexisting symmetry axes. The amount of material in the pentamer band can plausibly account for all 11 of the pentameric groupings predicted from the Caspar-Klug model (Table 2). However, there is not enough material in the hexamer band to account for all of the predicted hexameric groupings in the head. (For example, if HK97 is built according to an icosahedral Caspar-Klug model (triangulation number of 7), as λ, P22, and T7 are thought to be [7, 37, 41], there should be 6.5 times as much mass in hexamers as there is in pentamers.) Our current working model is that all of the subunits in the head are cross-linked into hexamers or pentamers but that most of the hexamers are further cross-linked to each other to form the very large material that does not enter the gel. On the basis of this model, we can make a rough estimate of how much of the head protein is in the material that does not enter the gel. There should be 11 pentamers, according to the Caspar-Klug model—1 for each of the 12 icosahedral corners except the corner with the tail. If we assume that all of the expected pentamers are represented in band 5 and that HK97 has a triangulation number of 7, we can calculate that ~77% of the major shell protein is in the material that does not enter the gel.

The biological function of the cross-link is not addressed by our experiments, but we suppose that at least one function may be to make the head of the mature virion more stable against disruption by environmental insults. In all dsDNA phages whose assembly has been studied to date, the head is assembled initially as a prohead that, at a subsequent step in head maturation, expands slightly to the diameter of a mature head. Expansion entails a conformational change in the shell protein subunits which results in a substantial increase in the strength of the noncovalent interactions that hold the subunits together (36). One proposal for the function of the expansion (8) is that it is a transition from a form of the protein subunits that is optimized for efficient assembly into a form that provides for a maximally stable head structure. We imagine a similar role for the cross-linking in HK97 heads: proheads are assembled from a monomer that is able to assemble well and are then converted by cross-linking to a more stable structure. The un-cross-linked structure of the U4 proheads is consistent with this view in that it suggests that shell assembly may be complete before any cross-linking begins. HK97 proheads, like those of other dsDNA phages, undergo expansion during head maturation (our unpublished observations), so it may be that the cross-linking, like the noncovalent stabilization in the other phages, takes place at the time of expansion. Further experiments will be required to clarify these events.

Although this is the first report of a virus that cross-links its capsid as extensively as HK97 appears to do, we do not believe that HK97 is unique in this regard. As mentioned above, bacteriophage HK022, which was isolated from the same source as HK97 (14), gives a head protein pattern on SDS gels that is indistinguishable from the HK97 pattern. More surprisingly, a similar pattern of high-molecular-weight bands, including a large amount of material that fails to enter the gel and similar relative intensities for the bands that do enter, is seen in bacteriophage L5 of Mycobacterium smegmatis (39a) and its relative, mycobacteriophage D29 (35). We predict that these phages and probably others will also be found to cross-link their head protein subunits.

ACKNOWLEDGMENTS

We thank Sherwood Casjens and Robert Duda for useful discussions and comments on the manuscript.
This work was supported by NIH research grant AI12227 to R.W.H.

REFERENCES