Proc. Nat. Acad. Sci. USA
Vol. 71, No. 4, pp. 1451–1455, April 1974

Protein Fusion: A Novel Reaction in Bacteriophage λ Head Assembly

(protein processing/viral morphogenesis/protein splicing)

ROGER W. HENDRIX* AND SHERWOOD R. CASJENS†

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

Communicated by Dale Kaiser, November 26, 1973

ABSTRACT **Parts of two phage-coded head proteins, pE and pC, become fused during bacteriophage λ head assembly. pE is the main structural component of λ heads and pC is a minor head protein that is not found as such in mature heads. The bond joining the two proteins appears to be covalent and is not a disulfide bond. Only a specific subset of the sequences of each protein is found in the fusion products, and these sequences are found in the products in equimolar amounts. Two nearly identical fusion products, X1 and X2, are detected; X2 is slightly smaller than X1 and appears to be a proteolytic cleavage product of X1. The fusion reaction probably takes place on a nascent head structure.**

It is likely that all of the genes of bacteriophage λ which control head assembly have been identified: nine head genes are known from genetic data, and the proteins corresponding to seven of them have been identified (1–3). Fig. 1 is a diagram of the head region of the physical and genetic map, with the lengths of the bars below the map indicating the length of DNA required to code for each known protein. These proteins account for about 95% of the coding capacity of the head region. Consequently, it has been puzzling to us that phage heads always contain several additional minor proteins; two of these, which we have called X1 and X2, have molecular weights of 31,000 and 29,000, respectively, and calculations indicate that the DNA required to code for them is equivalent to about 20% of the head region. X1 and X2 are phage-specific late proteins by the criterion that they can be labeled radioactively at late times after phage infection but cannot be if label is present only at early times or prior to infection (4).

Protein cleavage is a common feature of bacteriophage head assembly (5–9, 27), and we have recently shown that a minor protein component of λ heads, called h3, is produced by proteolytic cleavage of the B protein, pB (10). (The protein products of the *B*, *C*, and *E* genes are referred to as pB, pC, and pE, respectively.) By analogy, it seemed possible that X1 and X2 could be other chemically processed forms of pB or of pC. We have examined this question by making fingerprints of tryptic digests of the various λ head proteins. This paper describes the unusual interrelationship we have discovered among four phage proteins, X1, X2, pC, and pE, the major structural components of heads.

Abbreviations: pB, pC, pE, the protein products of the *B*, *C*, and *E* genes, respectively; SDS, sodium dodecyl sulfate.

* Current address: Department of Biophysics and Microbiology, University of Pittsburgh, Pittsburgh, Pa. 15260.
† Current address: Department of Biology, Massachusetts Institute of Technology, Cambridge, Mass. 01239.

MATERIALS AND METHODS

Radioactive Labeling. ^{35}S-labeling of induced cells or infected UV-irradiated cells was carried out as described by Georgopoulos *et al.* (11). Proteins for fingerprinting were obtained from lysates, petit λ, or phage. Fingerprints of a given protein from these different sources are indistinguishable. ^{35}S-Labeled heads were prepared as described previously (12) from *Escherichia coli* strain 594 *mal*$^-$ (λ *Jam*27 *Sam*7 *c*I857), using two successive CsCl step gradients for purification.

Electrophoresis and Tryptic Digestion. ^{35}S-Labeled material (purified phage, purified petit λ, or λ-infected lysates) was fractionated on sodium dodecyl sulfate (SDS)-polyacrylamide gels, eluted, treated with trypsin, and prepared for fingerprinting as described by Crawford and Gesteland (13). Samples of the proteins were occasionally taken prior to trypsin treatment and found to be intact upon rerunning on an analytical SDS gel.

Fingerprinting. The tryptic digest was taken up in water and spotted onto an Eastman no. 6061 silica gel 20 × 20-cm plastic thin-layer plate; two samples were spotted symmetrically on each plate, 4 cm from one edge and 1.5 cm from the center line of the plate. Electrophoresis was carried out at 400 V for 3 hr at 10°, in water:acetic acid:pyridine (500:50:5). Plates were then air dried, cut down the center line, and the two halves chromatographed in water:acetic acid:pyridine:1-butanol (143:50:143:204). The plates were autoradiographed in contact with Kodak SB-54 single-coated x-ray film.

Quantitation of Spots. Spots on the autoradiogram were traced through the center in a horizontal direction with a Joyce-Lobel microdensitometer. Care was taken to assure that the optical densities were within the linear response range of the film. The tracings were integrated using a Hewlett-Packard 9864A computer with a 9810A digitizer tablet.

Strains. Y mel *sup F* (λ *Eam* 4 *Sam* 7 *c*I857) was constructed by infection. Other strains were as described elsewhere (10, 11).

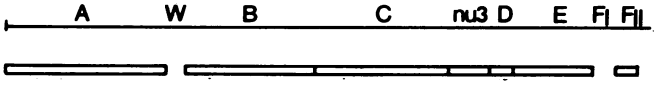

FIG. 1. Physical and genetic map of the head region of the λ genome. The lengths of the bars, representing known head proteins, were calculated from the protein molecular weights assuming an average molecular weight of 110 for an amino acid.

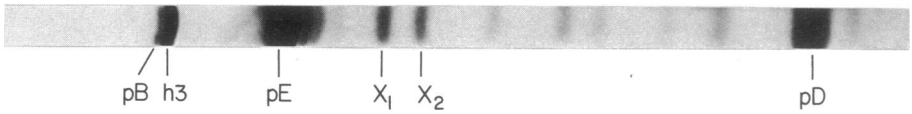

FIG. 2. Autoradiogram of SDS-polyacrylamide gel of ^{35}S-labeled λ heads. The apparent molecular weights of the indicated bands are: pB, 62,000; h3, 56,000; pE, 37,500; X1, 31,000; X2, 29,000; pD, 11,000. The genetic origins of the unnamed minor bands have not been determined.

RESULTS

The Minor Proteins of the Phage λ Head. Fig. 2 shows an autoradiogram of the proteins present in an SDS-polyacrylamide gel of highly purified ^{35}S-labeled phage λ heads. It can be seen that in addition to the products of genes *B*, *E*, and *D* there are several minor bands which cannot be purified away from the particles. These proteins are also present in whole phage, as well as in incomplete heads made during infection by phage defective in genes *W* or *F*, but are completely absent from purified tails (4, 14). h3 has been shown to be the product of a proteolytic cleavage of pB (10), whereas the genetic origins of the proteins X_n have been unknown. The products of genes *A*, *C*, and *nu3* are not components of phage heads (1, 4).

X1 and X2 are Fusion Proteins. Fig. 3*a–d* shows tryptic fingerprints of ^{35}S-labeled pE, pC, X1, and X2. Notice first the similarity between the fingerprints of X1 and X2. The two patterns are identical with the exceptions of spots 17 and 18, which are present only in the X1 fingerprint, and spot 19, present only in the X2 fingerprint. On the basis of these fingerprints and the molecular weights of X1 (31,000) and X2 (29,000), we postulate that X2 is derived from X1 by proteolytic cleavage. Spot 17 (or 18) presumably represents a peptide that is lost in the cleavage process and spot 18 (or 17) one that is converted by cleavage to spot 19. The most surprising feature of these fingerprints, however, is that X1 and X2 both have several spots in common with pE and several in common with pC. From this and other evidence presented below we conclude that X1 and X2 are new proteins produced by the covalent fusion of parts of pE and pC.

Fig. 3*e–h* is a schematic drawing of the fingerprints. According to this scheme, the X1 fingerprint shares at least five spots in common with pE and at least five in common with pC. The large number of spots involved makes it unlikely that the correspondence is fortuitous. Although the mobilities of tryptic spots for a given protein may vary slightly from one experiment to another (compare fingerprints of pE in Figs. 3*a* and 4*a*), mobilities are highly reproducible within each experiment. The positions of the pE or pC spots and of the corresponding X1 or X2 spots are always perfectly superposable within one experiment.

Fig. 3 also shows that X1 does not contain all of the sequences in pC or in pE. Rather, there are at least two spots in the pE fingerprint and four in the pC fingerprint that are absent from the X1 fingerprint. This observation rationalizes the otherwise confusing fact that the fusion products have lower apparent molecular weights than either pC or pE.

X1 and X2 Contain Sequences from pE. In order to prove that spots in the X1 and X2 fingerprints actually do correspond to those in the pE fingerprint, we sought evidence that a mutation in the *E* gene changes the amino-acid sequence in X1 and X2. We grew phage carrying an amber mutation in *E*, λ*E*am4, in cells carrying the *sup*F suppressor. *Sup*F inserts tyrosine at the site of amber mutations (15), so the pE formed in these cells, designated $pE_4 \cdot supF$, contains a tyrosine substituted for the wild-type amino acid at the *E*am4 site. Fig. 4*a* and *b* show fingerprints of pE and $pE_4 \cdot supF$, which are identical with one exception: spot 5 in the $pE_4 \cdot supF$ fingerprint is missing from its usual position and has apparently moved to a position under the edge of spot 2_E. Thus we presume that spot 5 contains the peptide which covers the *E*am4 locus. The corresponding fingerprints for X2 are shown in Fig. 4*c* and *d*. Fig. 4*c* is a fingerprint of X2 from the wild-type infection and Fig. 4*d* is a similar fingerprint of X2 from the λ *E*am4 infection of *sup*F cells. The fingerprints are again identical with the exception of spot 5, which is missing from its normal position in the mutant fingerprint. This shows directly that the fusion products contain sequences coded by the *E* gene and therefore derived from pE.

Stoichiometry of Fusion is 1:1. Most proteins have a substantial excess of basic amino acids over sulfur-containing amino acids. Since each tryptic peptide has only one basic amino acid, its is expected that most tryptic peptides will contain either 0 or 1 sulfur atom. Therefore, most of the radioactive spots on fingerprints from ^{35}S-labeled proteins should represent peptides containing a single ^{35}S atom. In this case, the intensities of the various spots within the fingerprint of one of the fusion proteins (X1 or X2) should indicate the molar ratio of pE and pC sequences in the fusion protein.

Fig. 5 displays the intensities of spots from fingerprints of pE, pC, and X2. As expected, the intensities of spots in the pE and pC fingerprints fall into a single cluster, with the exception of one spot in each fingerprint which is 2–3 times more intense than the others. We assume that the spots in the main clusters contain a single ^{35}S atom and that the others have 2 or perhaps 3. (In fact, the pE spot which appears to have two ^{35}S atoms can be resolved into two distinct spots by longer electrophoresis.) The intensities in the X2 fingerprint show the same clustering pattern: all of the spots which were found to have a single sulfur atom in the pE and pC fingerprints have equal intensities, and the two spots which appear to have more than one sulfur in pE and pC behave similarly in X2. We conclude, therefore, that the pC and pE sequences are present in the fusion proteins in equimolar amounts.

The simplest model consistent with this stoichiometry is that one molecule of pC reacts with one molecule of pE to yield a fusion protein. Since the number of pE molecules synthesized is 30–40 times greater than the number of pC molecules (4), it appears that the pC molecules react with some small fraction of the pE molecules to produce the fusion proteins.

DISCUSSION

Phage λ heads contain two major protein components and about 12 minor proteins, only two of which, pB and h3, have a known genetic origin. Since the DNA comprising the head genes is too short to code for all the minor proteins, we asked

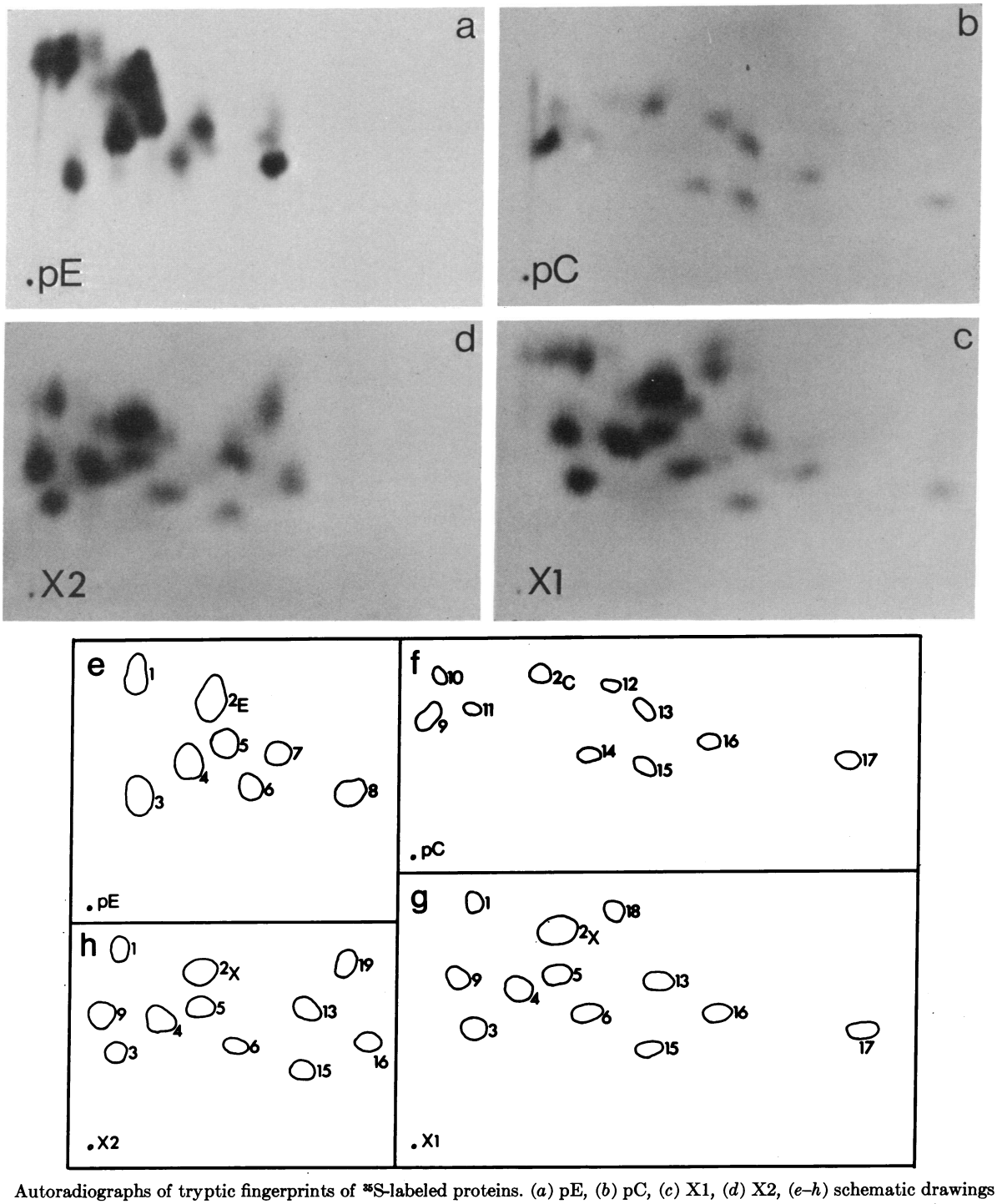

FIG. 3. Autoradiographs of tryptic fingerprints of ^{35}S-labeled proteins. (*a*) pE, (*b*) pC, (*c*) X1, (*d*) X2, (*e–h*) schematic drawings of the fingerprints. The numbering of the spots is meant to indicate the correspondence between spots in the different patterns. Since spots 2_E and 2_C have approximately the same mobility, it is not clear whether spot 2_X is derived from pE or pC or both. Electrophoresis is from left to right and chromatography is from bottom to top.

whether these might be chemically processed forms of proteins of known genetic origin. Tryptic fingerprints of the minor proteins suggested that a novel protein processing reaction occurs during head assembly. Of the minor protein bands seen in SDS gels of highly purified λ heads, two, X1 and X2, were found to contain some of the amino acid sequences of pC and some of those from pE. The simplest explanation of this observation is that there has been a covalent joining, or fusion of parts of pC and pE to form the new proteins X1 and X2.

Since X1 and X2 are defined by bands on SDS-polyacrylamide gels, it might be argued that these results could also be explained by unlinked fragments of pC and pE which had identical molecular weights and consequent comigration on the SDS gel. However, this is very unlikely. Both X1 and X2 contain equimolar amounts of peptides from pC and from pE and, moreover, the bandwidths of X1 and X2 are the same as those of other proteins of similar molecular weights. Thus for this hypothesis to be correct, the phage would have to carry a

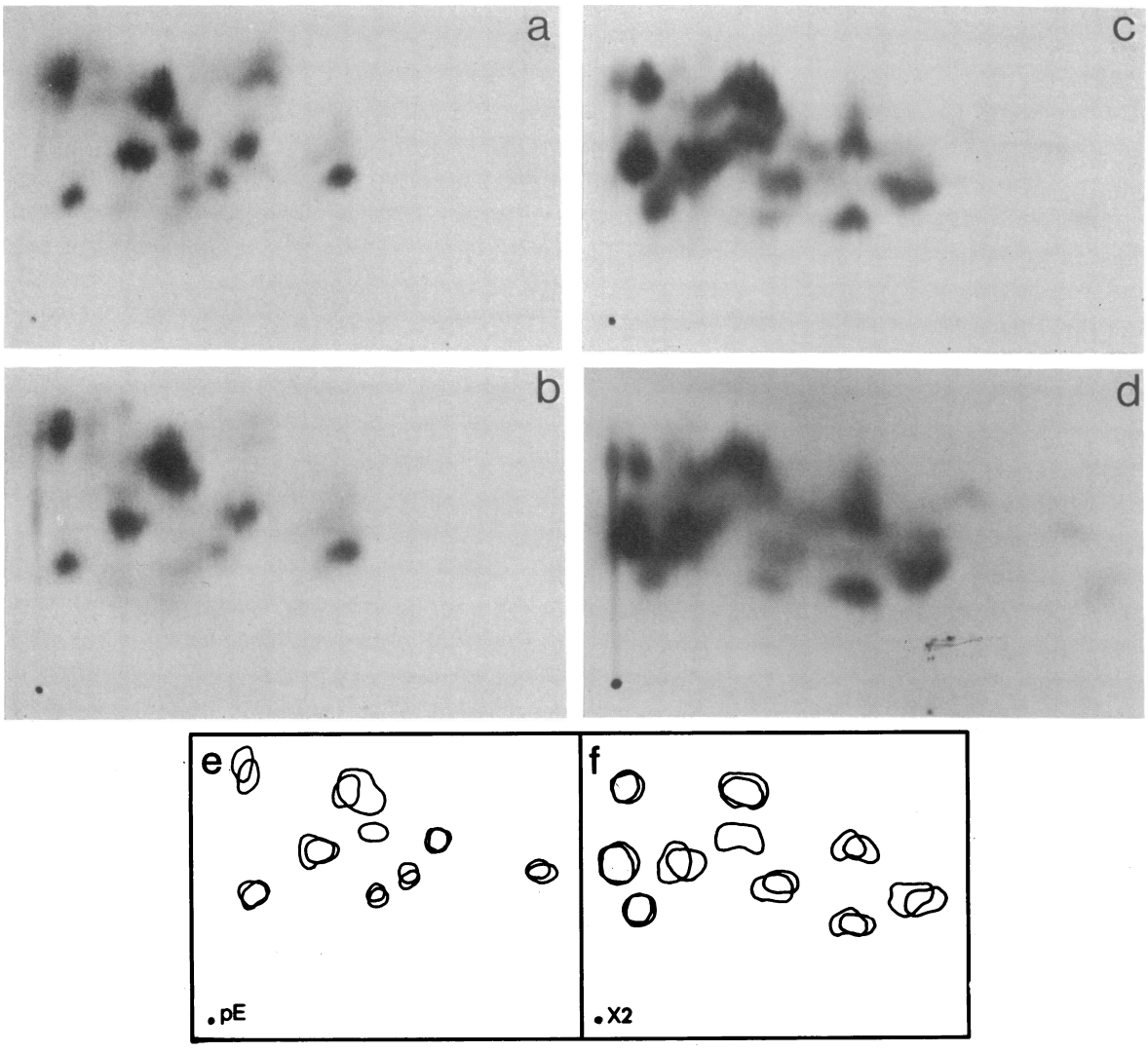

FIG. 4. Autoradiographs of tryptic fingerprints of ^{35}S-labeled proteins. (*a*) pE, (*b*) pE_4·supF, (*c*) X2, (*d*) X2 (E_4·supF), (*e*) superposition of tracings of 4*a* and 4*b*, (*f*) superposition of 4*c* and 4*d*.

fragment of pC and a fragment of pE, both of 31,000 daltons and both present in the same amounts, and also a fragment of pC and one of pE, both of 29,000 daltons and also present in equal amounts. Consequently, we conclude that the pC and pE sequences found in the X1 and X2 bands are actually linked to each other and not simply comigrating. Our preliminary fingerprint evidence (unpublished), that a third minor protein, X4, may be related to X2 by proteolytic cleavage further supports this contention.

The linkage between pC and pE appears, furthermore, to be covalent by the following criteria. Phage and other structures containing X1 and X2 are routinely boiled in a buffer containing 1% SDS and 1% β-mercaptoethanol prior to electrophoresis on an SDS gel. This treatment breaks known noncovalent bonds between polypeptide chains efficiently (16). The fact that X1 and X2 survive this treatment intact argues that the linkage between the pC and pE sequences is most likely covalent and almost certainly not a disulfide bond. In addition, if phage are heated with 6 M guanidine hydrochloride and 1% β-mercaptoethanol, or with 66% acetic acid before SDS treatment, no change in the band pattern results. It remains to be determined what type of covalent bond joins the pC and pE sequences in the fused protein.

There are other ways to explain the production of a fused protein than by the joining of two separate proteins. These include various hypotheses of gene fusion at the DNA or mRNA level, or of translational read-through from *C* into *E*. Both classes of hypotheses seem unlikely, since they are difficult to reconcile with the apparance of a discrete pC protein in lysates and in *groE*-defective head structures (10, 11). The read-through hypothesis also seems improbable on the grounds that a read-through protein would be expected to contain the tryptic peptides of the nu3 and pD proteins (see Fig. 1). This is not the case.

The fusion of pC with pE is consistent with several observations that are reported separately: unfused pC is not found in mature heads (1, 4). Unfused pC can be detected in lysates of λ-infected cells, but it disappears with time in a pulse-chase experiment (1, 11). This disappearance depends on the presence of a functional *E* gene (1, 4). Furthermore, when head-related structures (petit λ) from various mutant infections are examined, X1 and X2 are missing from them in two specific cases (10). The first is the case of infections by a phage carrying an amber mutation in the *C* gene. The second is the case of a wild-type λ infection of a *groE* host. *GroE* mutants of *E. coli* specifically block λ head assembly (11, 17), and the

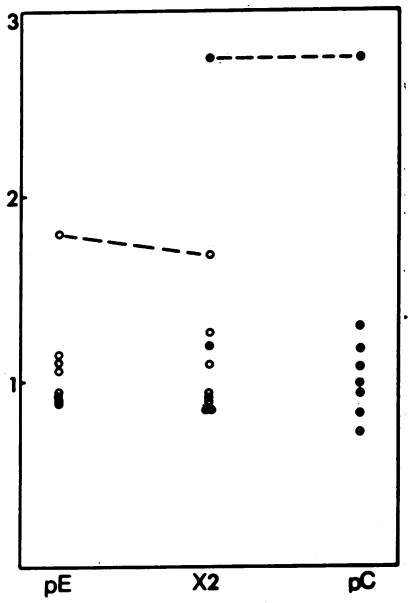

Fig. 5. Intensities of tryptic spots. Values for the intensities of spots were normalized for each pattern by averaging the values in the main cluster and setting the average equal to 1. The *type of symbol* and the *dashed lines* indicate the source of each spot in X2. Spots 2_E, 2_C, and 2_X are not included in the diagram.

petit λ structures isolated from these cells, besides missing X1 and X2, contain pC in a molar amount approximately equal to the wild-type complement of X1 plus X2. The fact that pC sediments with head-related structures from *groE* cells implies that nascent head structures have a binding site for unprocessed pC, and therefore, that the fusion reaction probably takes place after pC and pE join in a nascent head structure.

The assembly pathways of large bacteriophages, unlike those of many simple viruses, require information which is not contained in the structural components of the virion. The host cell plays a direct role in head assembly for λ, T4, and probably other similar phages (10, 11, 17–20). Cases are known in which a phage-coded protein participates in assembly but is not included in the finished virus structure (21–23). And some phage structural proteins undergo chemical modification during assembly (1, 5–10). These processes presumably allow a degree of control over assembly or a level of structural complexity that are not available to the simpler self-assembling viruses.

There is no clear indication of the biological function of the fusion reaction described here, or, for that matter, of any of the protein processing reactions reported to occur during the assembly of phages. It may be that some of the processing reactions render the assembly process irreversible, or are concerned with the ordering of assembly steps. Likewise it is possible that processing reactions may be involved in DNA packaging. In the case of the fusion reaction, it is interesting that presence or absence of the fusion proteins correlates with a difference in sedimentation properties of head-related structures (10). This difference may reflect a change in the spatial relationships between protein subunits.

We believe that the protein fusion described here represents a new type of biological reaction. The closest known analogies in nature are the crosslinking reactions reported for fibrin, elastin, and collagen (24–26). The fusion of pE and pC differs from these in at least two important respects. First, the crosslinking reactions join several molecules to produce a matrix of crosslinked proteins. The reaction of pE and pC most likely involves one molecule of each protein, and results in a small reaction product. Second, the crosslinking reactions produce links between essentially identical molecules. The fusion described here is the first example we know of in which two different proteins from two different genes are covalently joined *in vivo*. We have recently learned of experiments by Apte and Zipser (28) showing that fragments of the enzyme β-galactosidase can join *in vivo* to produce a molecule that is indistinguishable from the wild-type enzyme. This reaction may well involve the same mechanisms or enzyme systems as the reaction described here.

We wish to thank Drs. R. W. Davis, A. D. Kaiser, P. Schedl, and J. M. Syvannen for useful discussions and Drs. R. F. Gesteland and P. Greene for experimental techniques. This work was supported by NIH Grant AI-04509 to A. D. Kaiser and was carried out during the tenures of a Damon Runyon Postdoctoral Fellowship to R.H. and an NIH predoctoral training grant to S.C.

1. Murialdo, H. & Siminovitch, L. (1972) *Virology* **48**, 785–823.
2. Murialdo, H. & Siminovitch, L. (1972) *Virology* **48**, 824–835.
3. Boklage, C. E., Wong, E. C. & Bode, V. (1973) *Genetics* **75**, 221–230.
4. Casjens, S. (1972) Ph.D. Thesis, Stanford University, Stanford, Calif.
5. Dickson, R., Barnes, S. & Eiserling, F. (1970) *J. Mol. Biol.* **53**, 461–474.
6. Hosoda, J. & Cone, R. (1970) *Proc. Nat. Acad. Sci. USA* **66**, 1275–1281.
7. Kellenberger, E. & Kellenberger-Van der Kamp, C. (1970) *FEBS Lett.* **8**, 140–144.
8. Laemmli, U. (1970) *Nature* **227**, 680–685.
9. Lengyel, J., Goldstein, R., Marsh, M., Sunshine, M. & Calendar, R. (1973) *Virology* **53**, 1–23.
10. Hendrix, R. & Casjens, S., manuscript in preparation.
11. Georgopoulos, C., Hendrix, R., Casjens, S. & Kaiser, D. (1973) *J. Mol. Biol.* **76**, 45–60.
12. Casjens, S., Hohn, T. & Kaiser, D. (1972) *J. Mol. Biol.* **64**, 551–563.
13. Crawford, L. V. & Gesteland, R. F. (1972) *J. Mol. Biol.* **74**, 627–634.
14. Murialdo, H. & Siminovitch, L. (1971) in *The Bacteriophage Lambda*, Hershey, A. D. ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), pp. 711–723.
15. Smith, J., Abelson, J., Clark, B., Goodman, H. & Brenner, S. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 479–485.
16. Maizel, J. (1971) *Methods Virol.* **5**, 179–246.
17. Sternberg, N. (1973) *J. Mol. Biol.* **76**, 25–44.
18. Georgopoulos, C., Hendrix, R., Kaiser, D. & Wood, W. (1972) *Nature New Biol.* **239**, 38–41.
19. Takano, T. & Kakefuda, T. (1972) *Nature New Biol.* **239**, 34–37.
20. Coppo, A., Manzi, A., Pulitzer, J. & Takahashi, H. (1973) *J. Mol. Biol.* **76**, 61–87.
21. Wood, W. & Henninger, M. (1969) *J. Mol. Biol.* **39**, 602–618.
22. King, J. & Laemmli, U. (1971) *J. Mol. Biol.* **62**, 465–477.
23. King, J., Lenk, E. & Bottstein, D. (1974) *J. Mol. Biol.* **80**, 697–731.
24. Traub, W. & Piez, K. (1971) *Advan. Protein Chem.* **25**, 243–352.
25. Piez, K. (1968) *Annu. Rev. Biochem.* **37**, 547–570.
26. Chen, R. & Doolittle, R. (1971) *Biochemistry* **10**, 4486–4491.
27. Zweig, M. & Cummings, D. (1973) *J. Mol. Biol* **80**, 505–518.
28. Apte, B. & Zipser, D. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2969–2973.