Proc. Natl. Acad. Sci. USA
Vol. 73, No. 11, pp. 4205–4209, November 1976
Microbiology

Protein cleavage during virus assembly: A novel specificity of assembly dependent cleavage in bacteriophage T4

(T4 morphogenesis)

TOSHIAKI ISOBE*, LINDSAY W. BLACK†, AND AKIRA TSUGITA‡

Abt. Mikrobiologie, Biozentrum der Universität Basel, CH-4056 Basel, Switzerland

Communicated by Heinz Fraenkel-Conrat, September 1, 1976

ABSTRACT **Cleavage of precursor proteins occurs during assembly of numerous viruses. Seven bacteriophage T4 head-related proteins are known to be cleaved during morphogenesis. Sequences surrounding the cleavage sites in T4 head precursors P23 and IPIII are reported here. We previously determined the sequences of precursor and processed forms of IPII and IPI. Cleavage occurs at a glutamyl-alanyl bond in each protein. By comparison of sequences around five cleaved and four uncleaved Glu-Ala bonds in head precursors, it appears that cleavage is limited to the primary structure H_1-X_1-Glu-Ala-X_2-H_2, where H_1 and H_2 are hydrophobic residues Leu, Ile, or Val, and X_1 is possibly limited to Thr or Ala, and X_2 to hydrophilic residues. The results suggest the viral-induced assembly protease recognizes and cleaves an extended primary structure in the structurally dissimilar precursors.**

Precursor proteins are known to be cleaved during assembly of numerous animal and bacterial viruses (1, 2). The function of protein cleavage in viral morphogenesis, and the relationship of such cleavages to other types of post-translational modification of proteins remain to be established. One possible function of these reactions is to allow the accumulation of large pools of inactive precursors at times when assembly would lead to the formation of aberrant or inactive virions. Cleavage of seven proteins has been demonstrated upon maturation of the head of bacteriophage T4: two capsid shell proteins P23 and P24 (3–6), four "assembly core" components P22, IPI, IPII, and IPIII (7–9), and B1 (alt) protein (10).

Cleavage of T4 head precursors *in vivo* is dependent on the presence and assembly of all of the head protein precursors, but *in vitro* only on gene *21* function (11). The apparently viral origin of the T4 head assembly proteolytic mechanism, and its dependence on formation of head structures, make T4 an attractive system for study of the genetic control and function of cleavage mechanisms in biological assembly systems.

Cleavage of P23 (P23 indicates the product of gene 23 etc.) was shown by Celis *et al.* (12) to result in the loss of the NH_2-terminal portion of the precursor. In recent publications, the *in vivo* cleavage of proteins IPI, IPII, IPIII, and P23 was shown to occur near the NH_2-terminus which resulted in creation of new NH_2-terminal alanine in each protein (8). One of the precursor proteins, IPII, was sequenced, and its *in vivo* cleavage was shown to occur at a Glu-Ala bond in the NH_2-terminal part of the precursor (residues 15–16) by comparison with IPII* (* indicates the processed form of a protein after maturation) (9). *In vitro* cleavage of purified P22 with a partially purified enzyme preparation from T4 infected cells, resulted in the appearance of new NH_2-terminal alanine and new COOH-terminal glutamic acid (8). These observations strongly suggested that a single protease with specificity for glutamyl-alanyl bonds cleaves all the precursor proteins.

At the same time, we also found two glutamyl-alanyl bonds in the primary structure of IPII which are not cleaved *in vivo*. This suggested a requirement of the proteolytic activity not only for the glutamyl-alanyl dipeptide, but also for primary and/or higher-order structure around the sensitive glutamyl-alanyl bond (9). Recently, we have sequenced another precursor protein IPI and its cleaved product IPI*[§], as well as internal peptide VII[¶], which results from assembly dependent proteolysis of P22 (13).

Here we report the sequences surrounding cleavage sites in two other precursor proteins P23 and IPIII. Thus, we have now determined in head precursor proteins P23, IPI, IPII, IPIII, and P22, the extended primary sequences around five cleaved, and four uncleaved glutamyl-alanyl bonds. By comparison of the sequences around cleaved and uncleaved glutamyl-alanyl bonds, we find evidence for a proteolytic mechanism with a novel degree of specificity for an extended primary sequence.

MATERIALS AND METHODS

Purification of P23*. Cleaved capsid shell protein P23* was isolated from T4 D^+ particles purified by repeated differential centrifugation. The purified phage particles were disrupted by four times-repeated freezing and thawing in 10 mM Na-PO_4 (pH 7.0), 1 mM $MgSO_4$ in the presence of pancreatic DNase (deoxyribonucleate 5′-oligo-nucleotidohydrolase, EC 3.1.4.5) (5 μg/ml). After centrifugation, the precipitate was resuspended in 10 mM Na-PO_4 (pH 7.0) containing 0.5% (wt/vol) 2-mercaptoethanol and 2 mM EDTA, and powdered guanidine hydrochloride (GdnHCl) was added to a final concentration of 6 M. After stirring at room temperature for 30 min, the turbid solution was clarified by centrifugation. The supernatant was diluted 10 times with Na-PO_4 (pH 7.0) and powdered ammonium sulfate was added to 30% saturation. After centrifugation, the precipitate was dissolved in 0.05 M Tris-HCl (pH 7.8) containing 6 M GdnHCl and 1 mM dithiothreitol and the solution was subjected to gel chromatography on Bio-Gel A-5m. The elution buffer was 0.05 M Tris-HCl (pH 7.8) containing 4 M GdnHCl and 1 mM dithiothreitol. This step was repeated twice.

The P23* preparation was 95% pure by sodium dodecyl sulfate ($NaDodSO_4$)/polyacrylamide gel electrophoresis based

Abbreviation: $NaDodSO_4$, sodium dodecyl sulfate.

* Present address: Department of Chemistry, Faculty of Science, Tokyo Metropolitan University.
† Present address: Department of Biological Chemistry, School of Medicine, University of Maryland, Baltimore, Md. 21201.
‡ On leave of absence from Laboratory of Molecular Genetics, Medical School, Osaka University, Osaka, Japan.

§ T. Isobe, L. W. Black, and A. Tsugita (submitted to *J. Mol. Biol.*).
¶ J. P. van Eerd, S. P. Champe, L. Yager, and A. Tsugita (submitted to *J. Mol. Biol.*).

on a densitometric tracing of the stained gel. NH_2-terminal analysis of this preparation by Edman degradation (step one), and then by back hydrolysis of the phenylthiohydantoin amino acid (14) provided only alanine which is known to be the NH_2-terminus of P23* (8).

Purification of an *am* B272 Fragment. The *am* B272 fragment was isolated from *Escherichia coli* infected with an amber mutant in gene 23 (*am* B272) by a procedure described elsewhere[‖]. The isolated fragment appeared homogeneous on 12.5% $NaDodSO_4$/polyacrylamide gels[‖].

Purification of Internal Protein III and III*. Cleaved internal protein III (IPIII*) was isolated from purified T4 D^+ particles. The purification procedure of Black and Ahmad-Zadeh (15) was used. Uncleaved internal protein III (IPIII) was isolated from *E. coli* B^E infected with a triple mutant T4·21^- (*am* N90). 23^- (*am* H11) e^- (*am* H26). The purification procedure of Isobe *et al.* was also used. Both IPIII and IPIII* preparations were homogeneous on $NaDodSO_4$/polyacrylamide gels (9).

Sequence Analysis of Proteins and Peptides. Cleavage of an *am* B272 fragment (12 mg) with CNBr was carried out in 70% formic acid for 24 hr at room temperature. Tryptic digestion was carried out on the *S*-carboxymethyl derivative of the fragment (15 mg) with 0.25 mg of L-1-tosylamido-2-phenethylchloromethyl ketone-treated trypsin (Worthington Biochemical Corp.) at 37° for 19 hr. Chymotryptic digestion was carried out on 6 mg of the fragment with 0.12 mg of α-chymotrypsin (Sigma) at 37° for 9 hr, in the presence of 5 mM $CaCl_2$ and 0.06 mg of soybean trypsin inhibitor (Worthington Biochemical Corp.). Peptides were separated by ion exchange chromatography on SP-Sephadex G-25 with a gradient of linearly increasing pH and ionic strength in pyridine acetate buffer, and then by gel chromatography on Bio-Gel P6 or P10 (minus 400 mesh). Chymotryptic digestions of CNBr peptides were carried out in 0.1 M *N*-ethylmorpholine acetate buffer (pH 8.0) for 5–12 hr at 37°, at an enzyme-substrate ratio of 1/30–1/50. Peptides were separated by column chromatography either on SP-Sephadex G-25 or Bio-Gel P6, or both. Details of the methods will be described elsewhere[‖].

Edman degradation (direct identification method) was carried out on 20–50 nmol of peptides or proteins according to the procedure described by van Eerd and Takahashi (16). The coupling buffer was 0.4 M dimethylallylamine-trifluoroacetic acid (pH 9.5) containing 60% propanol. For proteins, the phenylthiocarbamoyl amino acids which result from the cleavage step were extracted directly with 1-chlorobutane without addition of water. Phenylthiohydantoin amino acids were identified by high-pressure liquid chromatography essentially according to the method of Frank and Strubert (17). When necessary, thin-layer chromatography on polyamide sheets (18) was employed for the identification.

Other sequencing methods including subtractive Edman degradation, hydrazinolysis, or carboxypeptidase digestions were as described in an earlier paper (9).

RESULTS

NH_2-Terminal Amino Acid Sequence of P23*. The NH_2-terminal amino acid sequence of P23* was determined to be ^{1}Ala-Glu-Ile-Gly-^{5}Gly-Asp-His-Gly-Tyr-^{10}Asn by manual Edman degradation on 50 nmol of the protein (Fig. 1). The phenylthiohydantoin amino acids from each step of the degradation, except for step seven, were identified by high pressure liquid chromatography as well as by thin-layer chromatography

[‖] T. Isobe, M. Yanagida, and A. Tsugita (in preparation).

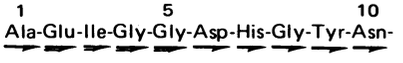

FIG. 1. NH_2-terminal amino acid sequence of P23*. Arrows (→) indicate the results of Edman degradations on P23*, and (⇀) indicate those on P23* modified with Braunitzer reagent. Numbers written above the amino acid residues indicate the residue number from the NH_2-terminus of P23*.

with the same results (*Materials and Methods*). The histidine residue at position 7, the phenylthiohydantoin derivative of which is known to be soluble in water, was identified by analysis of the corresponding HCl layer with thin-layer chromatography on polyamide. The sequence between residues ^{2}Glu and ^{5}Gly was confirmed by repeated degradations performed on a P23* preparation modified with Braunitzer reagent (19). The NH_2-terminal Ala was consistent with the previous results obtained by the dinitrophenylation and dansylation methods (8).

Partial Amino Acid Sequence of an *am* B272 Fragment and the Cleavage Site in P23*. For the determination of the cleavage site in P23, sequence studies were made on the NH_2-terminal regions of the cleaved protein P23* and on a peptide fragment produced by an amber mutation in gene 23 (*am* B272), which was predicted to cover the cleavage site. This prediction was based on the following data (see Fig. 2): molecular weight for P23 and P23* are known to be 57,000–59,000 and 46,000–48,000, respectively, and the difference in molecular weight is calculated to be 9,000–13,000 or about 100 amino acid residues (8). Celis *et al.* (12) showed the cleavage of P23 to occur in the NH_2-terminal portion of the molecule. This was confirmed by terminal analysis of both P23 and P23* (8), where a common amino acid Ile was found at the COOH-termini of both proteins but different amino acids (Met[‖] for P23 and Ala for P23*) at their NH_2-termini. Sarabhei *et al.* (20) studied a series of host conditional nonsense (amber) mutants in gene 23, and demonstrated that the sizes of the peptide fragments produced by these amber mutants were colinear with their genetic map positions. The colinearity of the amber fragments has been shown directly by $NaDodSO_4$/polyacrylamide gel electrophoresis of the various amber fragments by Celis *et al.* (12). According to their results, the smallest amber fragment found was the product of *am* E161, molecular weight 22,000; assuming colinearity, the molecular weight of the product of *am* B272 was calculated to be 11,000. C. Kellenberger and E. Kellenberger independently identified the *am* B272 fragment on $NaDodSO_4$/polyacrylamide gels, where the molecular weight was estimated as 18,000 (personal communication). Therefore, it seemed likely that the amber fragment B272 contains the cleavage site of P23 towards the COOH-terminal part of the molecule.

This amber fragment was identified on 12.5% $NaDodSO_4$/polyacrylamide gels in *E. coli* B^E infected with an amber mutant in gene 23 (*am* B272), and was purified[‖]. The molecular weight of the purified fragment was estimated to be 16,000-18,000 by $NaDodSO_4$/gel electrophoresis as well as by equilibrium sedimentation ultracentrifuge analysis, as expected from the above prediction. For the determination of the cleavage site, we looked for a sequence in the *am* B272 fragment identical to the NH_2-terminal sequence of P23*. The amber fragment was cleaved with CNBr, and four peptides which included about 90% of the residues in the fragment, were isolated. The three large peptides were subjected to chymotryptic digestion for further sequencing.

Fig. 3A shows the amino acid sequence of a chymotryptic peptide which was derived from one of the large CNBr peptides. The NH_2-terminal sequence of this peptide was estab-

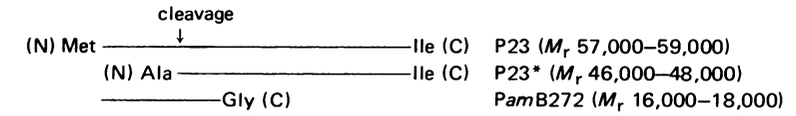

FIG. 2. Location of an *am* B272 fragment (P*am*B272) in P23. Molecular weight and terminal amino acid residues of P23 and P23* are based on Tsugita *et al.* (8), and of *am* B272 on Isobe *et al.*‖ Details are given in the *text*. M_r indicates molecular weight.

lished as ^{1}Leu-Thr-Glu-Ala-^{5}Glu-Ile-Gly-Gly-Asp-^{10}His-Gly-^{12}Tyr. The amino acid composition and the sequence between ^{4}Ala and ^{12}Tyr coincided with those of the first nine residues from the NH_2-terminus of P23*. In addition, in accordance with the sequence of the other CNBr peptides and the results from tryptic and chymotryptic digestions of the amber fragment, about 87 amino acid residues (about 10,000 molecular weight) separate ^{4}Ala and the COOH-terminal Gly of the fragment‖. This distance agrees with that expected for the cleavage site, derived from the difference in molecular weight between P23 and P23*, and the molecular weight of the amber fragment (Fig. 2). From the location and the coincidence of the sequences, we conclude that the cleavage in P23 takes place at the ^{3}Glu-^{4}Ala bond in the peptide described above.

In analyzing chymotryptic peptides of the CNBr fragments of the *am* B272 fragment, we found one peptide BC4-19 containing a Glu-Ala dipeptide sequence which is not cleaved *in vivo*, because this peptide can be shown to be located between the cleavage site and the COOH-terminus of the amber fragment based on overlapping tryptic and chymotryptic peptides from the amber fragment (Fig. 3B)‖. Homoserine (Hse) was found at the COOH-terminus of BC4-19 which shows that it was derived from the COOH-terminus of the CNBr peptide. Because of the low yield of this chymotryptic peptide, the sequence studies were actually carried out on overlapping tryptic and chymotryptic peptides derived from the whole amber fragment (Fig. 3B).

Two tryptic peptides T1-13 and T14-28, bridged by a chymotryptic peptide C4-16, covered amino acid residues in BC4-19. The sequences of these two tryptic peptides were determined by Edman degradation as well as by carboxypeptidase P (penicillum acid peptidyl-L-amino acid hydrolase, EC 3.4.4.17) digestions (Fig. 3B). The Glu residue at position 14 was confirmed to be Glu (not Gln) by carboxypeptidase Y [yeast peptidyl-L-amino acid (-L-proline) hydrolase, EC 3.4.12.1] digestion of C4-16. From the above results, the sequence of BC4-19 was established as ^{4}Gly-Lys-Asp-Pro-Val-Ala-^{10}Ala-Gly-Ala-Lys-Glu-^{15}Ala-Phe-His-Pro-^{19}Hse. The Glu-Ala sequence at residues 14–15 is the same as that of the cleavage site. Because this Glu-Ala bond is located between the cleavage site and the COOH-terminus of the amber fragment, it should be included in the NH_2-terminal region of the cleaved protein P23* (see Fig. 2). Therefore, this sequence is another instance of a Glu-Ala bond in a precursor protein which is not cleaved *in vivo*.

NH_2-Terminal Sequence of IPIII and IPIII*, and the Cleavage Site in IPIII. The sequence of the first 16 amino acid residues from the NH_2-terminus of IPIII (except for residue 15) was determined by direct identification of the phenylthiohydantoins by high-pressure liquid chromatography following Edman degradations (Fig. 4A). The NH_2-terminus was determined to be Met, which is consistent with previous results obtained by dinitrophenylation and dansylation methods (8). Residue 15 was not identified either by high-pressure liquid chromatography nor by thin-layer chromatography.

Because it was previously shown that the cleavage in IPIII takes place near the NH_2-terminus of the molecule (8), the cleaved protein IPIII* was subjected to Edman degradation to find a sequence overlapping with that of IPIII. The sequence of the first six residues from the NH_2-terminus (except for residue 5) was determined to be ^{1}Ala-Thr-Val-Val-X-^{6}Ala by identification of phenylthiohydantoin amino acids with high-pressure liquid chromatography (Fig. 4B). This sequence is identical with that of IPIII from residues 11 to 16, although residues 15 in IPIII and 5 in IPIII* were not determined. This coincidence of their sequences shows the cleavage in IPIII *in vivo* to occur at the Glu-Ala bond at positions 10 and 11.

Based on this result, the molecular weight reduction of IPIII by cleavage is calculated to be 1239. This is in good agreement with the difference in molecular weight between IPIII and IPIII* estimated by NadDodSO_4/polyacrylamide gel electrophoresis (1600).

DISCUSSION

Our intention in investigating the primary structures of precursor and processed forms of the phage T4 head proteins was to determine the specificity of the morphogenetic cleavages. Incidentally in the course of this work, we observed an unusual homology between the NH_2-terminal sequences of IPII (9) and IPIII, namely:

IPII: Glx-Pro-Ala-Leu-Lys-Met-Lys-*Thr*-Tyr-Gln-Glu-Phe-Ile-Ala-Glu-Ala-*Arg*-Val-*Gly*

IPIII: Met-Lys-*Val*-Tyr-Gln-Glu-Phe-Ile-Ala-Glu-Ala-*Thr*-Val-*Val*

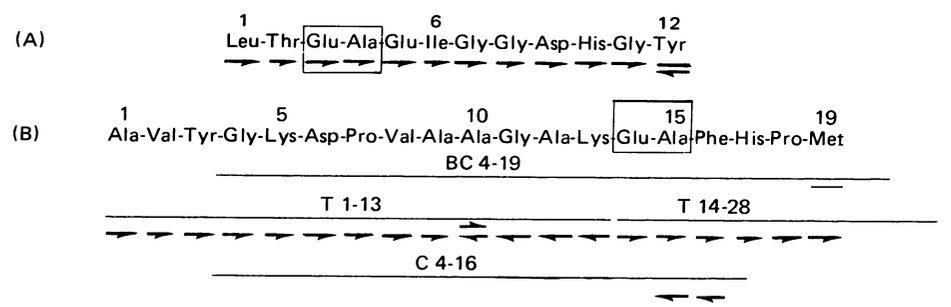

FIG. 3. Amino acid sequences around the dipeptide Glu-Ala, found in an *am* B272 fragment. (A) Sequence around the Glu-Ala dipeptide cleaved *in vivo*. (B) Sequence around another Glu-Ala dipeptide that is not cleaved *in vivo*. (The peptide numbers are written above solid lines. The abbreviations used are as follows: BC, chymotryptic peptide of a CNBr peptide; T, tryptic peptide of the amber fragment; C, chymotryptic peptide of the amber fragment.) Arrows (→) indicate the sequence proved by Edman degradation; (⟶) indicate the sequence proved by carboxy peptidase digestions; (—) indicate COOH-terminal amino acid residue proved by hydrazinolysis.

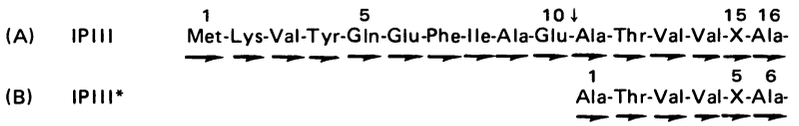

FIG. 4. NH_2-terminal amino acid sequences of IPIII and IPIII*. Arrows (→) indicate the results of Edman degradations. Numbers written above the amino acid residues indicate residue numbers from the NH_2-termini of the proteins. The residue 15 in IPIII and the residue 5 in IPIII* (X) were not determined.

Of the three differences in this run of 14 amino acids, (Thr ↔ Val, Arg ↔ Thr, and Gly ↔ Val), the latter two are changes compatible with single base changes in codons. Such sequence homology is not found between IPII or IPIII and the internal protein IPI, T4 phage lysozyme (21), and other available head protein sequences[§¶‖]. Although the sequence of IPIII available for comparison is so far limited, the striking homology of the NH_2-terminal structures of IPIII and IPII suggests that the IPII and IPIII genes are interrelated by a gene duplication. This hypothesis is compatible with the locations of the internal protein genes, because the genes for IPII and IPIII are apparently adjacent, while IPI is located about 6000 base pairs away (22).

The morphogenetic cleavage of IPII *in vivo* was previously demonstrated to occur at a specific Glu-Ala bond, but not at two other Glu-Ala dipeptides found in the protein (9). A second precursor protein IPI and its cleaved product IPI* have been sequenced, and cleavage of IPI *in vivo* was also shown to occur at a glutamyl-alanyl bond, which resulted in the removal of four amino acid residues from the NH_2-terminus. However, no other Glu-Ala bonds occur in the molecule[§]. In this communication, through sequence studies on P23*, amber fragment B272 of P23, IPIII, and IPIII*, we have determined sequences that include the morphogenetic cleavage sites in P23 and IPIII. For another precursor (P22), it was shown that the protein undergoes fragmentation by cleavage during maturation processes to the extent that its products are not detectable on Na-$DodSO_4$/polyacrylamide gels (4). In the finished capsid, two peptide components (internal peptides, IpepII and VII) were found, and it was suggested that these are the proteolytic cleavage products of precursor protein(s) (23). Goldstein *et al.* (13) recently identified the precursor of one of the internal peptides, IpepVII, to be P22. We can conclude this because the amino acid sequence of IpepVII has been determined[¶] and the corresponding sequence was found in P22 (24). The head precursor proteins which are cleaved during assembly—P23, P22, IPI, IPII, and IPIII—would appear to differ greatly in overall structure by comparison of their molecular weights, net charges, amino acid compositions, solubilities, primary structures, and functions. Nevertheless, cleavage of all these proteins appears to share a common structural specificity.

Table 1. Amino acid sequences around cleaved and uncleaved Glu-Ala dipeptides in T4 head precursors

Experiment	Precursor	Amino acid sequence
A	P23	-Phe-Leu-Thr-Glu↓Ala-Glu- Ile -Gly-
	IPI	NH_2Thr- Ile -Thr-Glu↓Ala-Thr-Leu-Thr-
	IPII	-Phe- Ile -Ala-Glu↓Ala-Arg-Val-Gly-
	IPIII	-Phe- Ile -Ala-Glu↓Ala-Thr-Val-Thr-
	IpepVII	-Lys- Ile -Ala-Glu↓^{COOH}
B	P23	-Gly-Ala-Lys-Glu-Ala-Phe- His-Pro-
	IPII	-Gly-Lys-Leu-Glu-Ala-Ala-Val-Asn-
		-Gln-Gln- Ile -Glu-Ala-Ala-Ala-Ala-
	IpepVII	-Asn-Lys-Glu-Glu-Ala-Glu-Glu-Lys-

Experiment A lists the sequences cleaved *in vivo* and in experiment B, those uncleaved *in vivo*. Glu^{COOH} indicates the COOH-terminus.

In Table 1 (number A), sequences including the morphogenetic cleavage sites determined for five precursor proteins, P23, IPI, IPII, and IPIII, and a part of the sequence of IpepVII (for one of the cleavage sites in P22) are listed. Table 1 (number B) also summarizes sequences that include uncleaved glutamyl-alanyl bonds found in the precursors. The presence of additional uncleaved Glu-Ala bonds in P23 and IPIII is not excluded because sequences of these proteins are not completely determined. Primary (21) and tertiary (25) structures of the phage T4 lysozyme are known. This protein appears not to undergo cleavage and contains two Glu-Ala bonds. However, it is uncertain whether these can be classified as protease resistant Glu-Ala bonds for head assembly, because it is possible that the active protease encounters only proteins present in the T4 prehead. Therefore, these and related sequences produced by lysozyme gene mutants are not included in Table 1. From the primary sequence data, it is obvious (Table 1, number A) that all cleavages *in vivo* take place at Glu-Ala bonds. Taking into account this specificity, the COOH-terminal Glu of IpepVII probably arises as a result of the same specific cleavage of P22. It is also clear that there is an additional requirement for cleavage, because at least four Glu-Ala bonds in the precursors are not cleaved *in vivo* (Table 1, number B).

What is the distinction between the two types of Glu-Ala bonds? The sequences that include cleavage sites have a very characteristic primary structure, as is shown in Fig. 5. In this figure, H1 and H2 represent the aliphatic hydrophobic amino acid residues Ile, Leu, and Val, and X1 and X2 represent amino acid residues between the hydrophobic residues and the cleavage site Glu-Ala. As H1 we found Ile (IPI, II, III, and IpepVII) or Leu (P23) and as H2, Leu (IPI), Val (IPII and IPIII), or Ile (P23). The residue X_1 is limited to either Thr (IPI and P23) or Ala (IPII, IPIII, and IpepVII), while the residue X_2 is quite variable: Thr (IPI and IPIII), Arg (IPII), and Glu (P23), or, in general, hydophilic residues. The common features of the cleaved sequences suggest the importance of hydrophobic residues at H1 and H2 and possibly, of the presence of Ala or Thr at X1 and of a hydrophilic residue at X2. None of the sequences around uncleaved Glu-Ala bonds satisifies this schematic sequence for the cleavage site. Therefore, it seems reasonable to suggest that the head morphogenesis protease requires for activity: (*i*) the presence of a Glu-Ala bond, (*ii*) surrounded by aliphatic hydrophobic residues at the H1 and H2 positions and possibly (*iii*) by threonyl or alanyl residues at the X_1 position and (*iv*) by hydrophilic residues at the X2 position. A valyl residue was found at the H2 position for the uncleaved sequence in IPII (Table 1, B). However, the sequence neither includes hydrophobic residue at the H1 position, nor satisfies

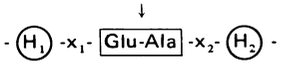

FIG. 5. Schematic representation of the substrate specificity for morphogenesis-dependent cleavage in bacteriophage T4. An arrow indicates the cleavage site. Details are given in the *text*.

conditions (*iii*) and (*iv*). The present model for cleavage specificity which includes a requirement for hydrophobic amino acid residues at positions H1 and H2 is consistent with the previous observation that organic solvents such as chloroform or butanol inhibit the proteolytic activity *in vitro* (8).

This model based on protease specificity for primary structure does not exclude a contribution of the precursor molecule configurations, but indeed suggests the importance of it, by indicating a possible interaction of hydrophobic residues with a primary Glu-Ala sequence through secondary or tertiary conformation. It is evident that the cleavage sites in the head precursors should be located on the surface of the molecules so as to be accessible to the morphogenetic protease. This was in fact suggested in our sequence studies of IPII by the accessibility of peptide bonds in IPII near the cleavage site to trypsin and chymotrypsin (9). Predictive analysis according to the method of Chou and Fasman (26) suggests the cleavage sites of IPI, IPII, IPIII, and P23 are located in α-helical regions. However, it appears unlikely that a protease would recognize a hexapeptide in α-helical conformation. The cleavage sites may rather be located in random coil conformations which locate on the surfaces of the protein molecules, as is believed to be the case for many proteolytic cleavages. Alternatively, distribution of hydrophobic and hydrophilic residues around the cleavage sites could suggest the presence of β-sheet conformation (27, 28). Clearly the question of the contribution of the higher-order structure to cleavage specificity awaits further structural analysis of the precursors, and investigation of cleavage of model peptides with purified morphogenetic protease.

It is known that cleavage of the head precursors *in vitro* is directly dependent only upon the function of gene *21*, which suggests that gene *21* is the structural gene for a protease (17). However, cleavage *in vivo* is considerably more complex than the recognition and cleavage of various head proteins by a soluble protease because: (*i*) cleavage *in vivo* depends upon the function of each of the head genes *20, 21, 22, 23, 24, 31,* and *40* and not only gene *21;* (*ii*) cleavage of all of the head precursors is blocked by temperature sensitive mutations affecting the structures of these precursors, even when the mutation allows defective head structures to be formed, as for example gene *23* temperature sensitive mutants; and (*iii*) amber mutant fragments of the precursors are not cleaved (3–7). These requirements for cleavage could be due either to a mechanism which only activates the morphogenetic protease upon correct prehead structure formation, or to formation of higher order structure in the precursors around the cleavage sites essential for protease activity. In addition, we must emphasize that it is not known whether all of the head assembly precursor processing is caused by a single viral-induced protease. Indeed, the finding of an NH_2-terminal lysine in IpepVII[§], and the observation that the bulk of the cleavage fragments cannot be recovered (7), due presumbly to further degradation in the cell, suggests participation by another proteolytic enzyme(s). It should be noted that both termini of I pep II, Gly-Glu-Glu-Pro ..... Asp-Asp-Glu[‖], does not fit the present enzyme cleavage specificity though the cleavage was shown to be due to the gene *21* function. This observation further suggests the participation of other enzymes or perhaps relaxed specificity of the enzyme.

Thus, the striking degree of extended primary structure similarity in the *in vivo* cleavage sites of the dissimilar head precursors we have sequenced, supports the notion that one enzyme is responsible for all of these cleavages. Specificity towards such an extended sequence further suggests that the assembly protease and the precursors are part of one viral coded system, and that cleavage of the precursors is important for their functions. It appears that sophisticated proteases with specialized biological roles may have a high capacity to recognize extended structures in proteins.

We thank Drs. E. Kellenberger, S. P. Champe, and M. K. Showe for advice, and Dr. T. Bickle for a critical reading of the manuscript. This work has been supported by the Swiss National Foundation for Scientific Research (3.639.75) and by a grant from the National Institutes of Health (USA) Institute of Allergy and Infectious Diseases (for L.W.B.).

1. Casjens, S. & King, J. (1975) *Annu. Rev. Biochem.* **44,** 555–615.
2. Hershko, A. & Fry, M. (1975) *Annu. Rev. Biochem.* **44,** 775–797.
3. Hosoda, J. & Cone, R. (1970) *Proc. Natl. Acad. Sci. USA* **66,** 1275–1281.
4. Laemmli, U. K. (1970) *Nature* **227,** 680–685.
5. Kellenberger, E. & Kellenberger-van der Kamp, C. (1970) *FEBS Lett.* **8,** 140–144.
6. Dickson, R. C., Barnes, S. L. & Eiserling, F. A. (1970) *J. Mol. Biol.* **53,** 461–473.
7. Showe, M. K. & Black, L. W. (1973) *Nature New Biol.* **242,** 70–75.
8. Tsugita, A., Black, L. W. & Shwe, M. K. (1975) *J. Mol. Biol.* **98,** 271–275.
9. Isobe, T., Black, L. W. & Tsugita, A. (1976) *J. Mol. Biol.* **102,** 349–365.
10. Horvitz, H. R. (1974) *J. Mol. Biol.* **90,** 739–750.
11. Onorato, L. & Showe, M. K. (1975) *J. Mol. Biol.* **92,** 395–412.
12. Celis, J. E., Smith, J. D. & Brenner, S. (1973) *Nature New Biol.* **241,** 130–132.
13. Goldstein, J. G., McCullough, J. E. & Champe, S. P. (1976) *J. Virol.,* **18,** 894–903.
14. Mendez, E. & Lai, C. Y. (1975) *Anal. Biochem.* **68,** 47–53.
15. Black, L. W. & Ahmad-Zadeh, C. (1971) *J. Mol. Biol.* **57,** 71–92.
16. van Eerd, J. P. & Takahashi, K. (1976) *Biochemistry* **15,** 1171–1180.
17. Frank, G. & Strubert, W. (1973) *Chromatographia* **6,** 522–524.
18. Summers, M. R., Smythies, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53,** 624–628.
19. Brauitzer, G., Schrank, B. & Ruhfus, A. (1970) *Hoppe-Seylers Z. Physiol. Chem.* **351,** 1589–1590.
20. Sarabhei, A. S., Stretton, A. D., Brenner, S. & Bolle, A. (1964) *Nature* **201,** 13–17.
21. Tsugita, A. & Inouye, M. (1968) *J. Mol. Biol.* **37,** 201–212.
22. Black, L. W. (1974) *Virology* **60,** 166–179.
23. Goldstein, J. & Champe, S. P. (1974) *J. Virol.* **13,** 419–427.
24. Showe, M. K. & Tsugita, A. (1974) *Experientia.* **30,** 709.
25. Matthews, B. W. & Remington, S. J. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 4178–4182.
26. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 222–245.
27. Lim, V. I. (1974) *J. Mol. Biol.* **88,** 857–872.
28. Lim, V. I. (1974) *J. Mol. Biol.* **88,** 873–894.