Molec. gen. Genet. 180, 489–494 (1980)

MGG

Recognition of Two Initiation Codons for the Synthesis of Phage fd Gene 2 Protein

Thomas F. Meyer,[1] Konrad Beyreuther,[2] and Klaus Geider[1]

[1] Max-Planck-Institut für Medizinische Forschung, Abteilung Molekulare Biologie, Jahnstraße 29, D-6900 Heidelberg, Federal Republic of Germany

[2] Institut für Genetik der Universität Köln, Weyertal 121, D-5000 Köln 41, Federal Republic of Germany

Summary. Bacteriophage fd gene 2 protein was specifically labeled with radioactive amino acids and was isolated from membranous cell structures as an apparently homogenous protein. Amino acid sequence analysis revealed that the protein was initiated at two distinct AUG codons close to the ribosome binding site. The two resulting translation products were found to begin with a deformylated methionine residue. Initiation at the first signal was used for 90% of the chains and at the second signal for 10% of the sequenced molecules. The use of one or the other chain start may influence functions of gene 2 protein.

Introduction

Replication of double-stranded phage fd DNA depends on the function of phage coded gene 2 protein (Lin and Pratt 1972; Fidanián and Ray 1972). The recently isolated phage protein is a very specific endonuclease of molecular weight 46,000 dalton which can also seal previously cleaved DNA (Meyer and Geider 1979a; Meyer and Geider 1979b). Gene 2 protein specifically reacts with supercoiled replicative form DNA of filamentous bacteriophages (fd, fl, M13) and cleaves the viral strand of the replicative form I in the intergenic region of the phage genome (Meyer et al. 1979). This nick is subsequently used as the origin for synthesis of the viral strand (Meyer and Geider 1979a; Geider and Meyer 1978).

The sequence of the gene 2 protein can be deduced from the DNA sequence (Beck et al. 1978; Schaller et al. 1978). The protein was suggested to start at either one of the two ATG triplets beginning at positions 6007 and 6016, respectively, and to terminate at position 828 of the fd genome. Both ATG triplets belong to the same translational reading frame. A sequence characteristic for a ribosomal binding site is found three to nine nucleotides in front of the first potential initiation codon of gene 2 and 12 to 18 nucleotides in front of the second AUG codon (Schaller et al. 1978). Thus, there are three possibilities for translation of the gene 2 protein: (i) the first methionine codon is used for initiation, or (ii) the second methionine is used or (iii) both methionine codons are used. The distances between the ribosome binding sites and the two initiation signals are extreme values observed in translation of prokaryotic genes (Steitz 1979). Sequence analysis of the radioactively labeled in vivo product revealed that both start codons are used. The major species starts at the first AUG triplet of the mRNA. The C-terminal sequence of the protein was analyzed by digestion with carboxypeptidase C, and the results were also in agreement with the sequence predicted from the DNA sequence.

Offprint requests to: Dr. Klaus Geider

Materials and Methods

Escherichia coli strain Hfr C7 is met$^-$, thi$^-$, phage fd am 51 is mutated in gene 5. Tritium-labeled amino acids were from Amersham Buchler (Braunschweig, FRG) and carboxypeptidase C was from Carl Roth KG (Karlsruhe, FRG).

Preparation of Radioactively Labeled Gene 2 Protein. Strain Hfr C7 was grown in 20 ml M9-medium (supplemented with methionine and thiamine) (Pratt and Erdahl 1968) at 37° C to 2×10^8 cells/ml and irradiated with UV (about 10 kerg mm^{-2}). The culture was supplemented with all 20 amino acids (20 µg/ml, except 5 µg/ml for the tritiated amino acid to be incorporated) and the cells were then infected with 20 phages per cell. After 10 min at 37° C 5 mCi of the [^{3}H]-labeled amino acid (leucine, valine, and proline, respectively) were added and the incubation continued for 60 min at 42° C. The cells were then converted to spheroplasts (Osborn et al. 1972), the spheroplasts were sonicated and membranous cell structures were pelleted in a sucrose gradient (Osborn et al. 1972; Lin and Pratt 1974) in a Beckman 50 Ti rotor for 2 h at 50,000 rpm

0026-8925/80/0180/0489/$01.20

and 10° C. Gene 2 protein was separated from other labeled proteins of the fast sedimenting material by sodium dodecyl sulfate polyacrylamide gel electrophoresis. The slab gel containing 12.5% acrylamide and 0.5% bisacrylamide (Lämmli 1970) was loaded with material of the pellet, which was dissolved in 150 µl of 20% glycerol, 3% sodium dodecyl sulfate, 3% 2-mercaptoethanol and 0.005% bromo-phenol blue. At both sides of the application line on the gel, the marker ovalbumin (M_r 45,000) was added and the gel was run for 10 h at 6 V/cm. The ovalbumin-containing part of the gel (longitudinal strips) was stained with Coomassie brilliant blue, destained and fixed in 10% acetic acid. The tritiated gene 2 protein which coelectrophoretes with the marker protein was eluted electrophoretically from the corresponding unstained gel strips as described under Results.

Amino Acid Sequence Analysis. The radioactive samples were supplemented with 4 mg of oxidised chicken lysozyme and subjected to automated Edman degradation in an updated Beckman model 890B sequencer equipped with a Sequemat P-6 automatic converter. A single-coupling (using 0.1 M Quadrol/trifluoroacetic acid buffer, pH 9.5), single-cleavage program including a combined benzene/ethyl acetate wash (Brauer et al. 1975) was used. Each degradation was started with three nondegradive cycles (without phenylisothiocyanate) to wash out potential radioactive contaminants. The analysis of the phenylthiohydantoin derivatives of the carrier was performed as described (Beyreuther et al. 1977; Beyreuther 1977; Beyreuther et al. 1978). The radioactivity of phenylthiohydantoin derivatives of the gene 2 protein was determined on an aliquot in 10 ml of Aquasol-2 (New England Nuclear). Repetitive degradation yields over the first 25 steps of oxidised lysozyme as carrier ranged between 95% and 96%. Sequencer runs resulting in lower repetitive yields for the carrier were repeated.

Digestion with Carboxypeptidase C. [^{3}H]leucine fd gene 2 protein (62,000 cpm) was supplemented with 0.8 mg of oxidized chicken lysozyme and denatured by heat treatment at 100° C for 20 min. Digestion was performed in 0.125 ml of 0.1 M ammonium acetate buffer, pH 5.5 with 1,000 mU of enzyme and 0.02 ml were withdrawn after 0, 6, 9, 12, 24, and 48 h of digestion. These samples were lyophilized and assayed for radioactivity released (Beyreuther et al. 1980).

Results

Isolation of in vivo Radioactively Labeled Gene 2 Protein

Although biologically active gene 2 protein of filamentous phages is found in the cytoplasm (Meyer and Geider 1979a), it is trapped in membranous cell structures, when the phage infected cells are grown at 42° C (Linn and Pratt 1974). Two other phage proteins, the gene 8 protein (fd coat protein) and the gene 5 protein (fd DNA binding protein) are also associated with such structures (Meyer and Geider 1979a; Webster and Rementer 1980).

Gene 2 protein in membranous cell structures was separated from other proteins by sodium dodecyl sulfate polyacrylamide gel electrophoresis (Fig. 1). To elute gene 2 protein from the gel, the gel slice containing the protein was transferred into a gel electrophoresis tube, the slice was embedded in polyacrylamide and the protein was electrophorized for 12 h at 5 V/cm into a dialysis bag which was sealed at the anodic end of the tube. The protein eluted was then dialysed for 48 h at 37° C against 50 mM Tris/HCl, pH 7.4 in 8 M urea, for 24 h against 0.1 M ammonium bicarbonate and lyophilized. The yields of labeled protein are listed in Table 1. The three tritiated amino acids used had the same specific radioactivity (50 mCi/mmol), but the uptake of an amino acid, measured by its incorporation into gene 2 protein, increased from [^{3}H]leucine and [^{3}H]proline to [^{3}H]valine. Considering the frequency of the amino acid in the protein the specific radioactivity of the [^{3}H]valine-protein is much higher than that of the two other derivatives (Table 1).

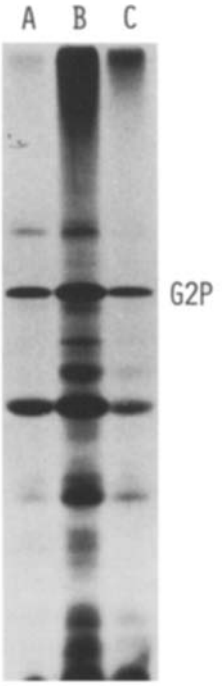

Fig. 1. Polyacrylamide gel electrophoresis of fast sedimenting material (membrane fraction) from *E. coli* cells infected with phage fd. [^{3}H]labeled phage proteins were enriched as described in Materials and Methods. The gene 2 protein band at M_r 45,000 was localized via the marker protein ovalbumin which was stained with Coomassie brilliant blue. The section of the gel containing gene 2 protein was cut out with a razor blade and eluted as described in Results. This figure shows an analytical gel-fluorograph of an aliquot of the labeled membrane fraction. *A.* [^{3}H]leucine; *B.* [^{3}H]valine; *C.* [^{3}H]proline; G2P: fd gene 2 protein. The other bands are host proteins

N-terminal Sequence of the Gene 2 Protein

Gene 2 protein preparations labeled with either leucine, valine or proline were subjected to automated

Table 1. In vivo incorporation of tritiated amino acids into phage fd gene 2 protein

Incorporated tritiated amino acid	Radioactivity (cpm)		Total number of residues/ gene 2 protein	Specific radioactivity of gene 2 protein (cpm/ residue)
	Membrane fraction	Purified gene 2 protein		
Leucine	1.7×10^6	1.2×10^5	46	2,926
Valine	5.6×10^6	6.3×10^5	26	28,636
Proline	1.1×10^6	0.9×10^5	16	6,428

The total number of amino acid residues of gene 2 protein were taken from the nucleotide sequence (Schaller et al. 1978). The radioactive samples were obtained as described in Materials and Methods

Table 2. Carboxypeptidase C digestion of [^{3}]leucine-labelled fd gene 2 protein

Digestion time (h)	Gene 2 protein (input in cpm)	[^{3}H]Leucine released[a] (cpm)	Number of [^{3}H]leucine residues[b] released
0	9,875	0	0
6	9,710	187	0.89
9	10,150	276	1.25
12	9,920	438	2.03
24	9,485	866	4.20
48	10,430	948	4.18

[a] determined by radiolabel amino acid analysis
[b] on the basis of 46 leucine residues/chain (Schaller et al. 1978)

sequence analysis (Fig. 2). The three amino acids appeared in the following positions: [^{3}H]leucine: (2), (4), 5, 7, 9, (12), 15, 20, 25, 31, 37 and 42; [^{3}H]valine: (3), 6, (13), 16 and 29; [^{3}H]proline: (7) and 10. Values in parantheses indicate steps where about 10% of the expected radioactivity per residue were released. In addition automated sequence analysis of [^{3}H]leucine gene 2 protein was carried out after prior chemical deformylation treatment. When the tritiated protein was incubated with 0.1 ml of 1 M HCl in methanol at 25° C for 90 min (Clark and Marcker 1966) in the presence of lysozyme as carrier (4 mg), the recovery of [^{3}H]leucine in the corresponding degradation steps was only five percent higher than compared to the untreated protein. This shows that the protein became deformylated before complexing to membranous cell structures.

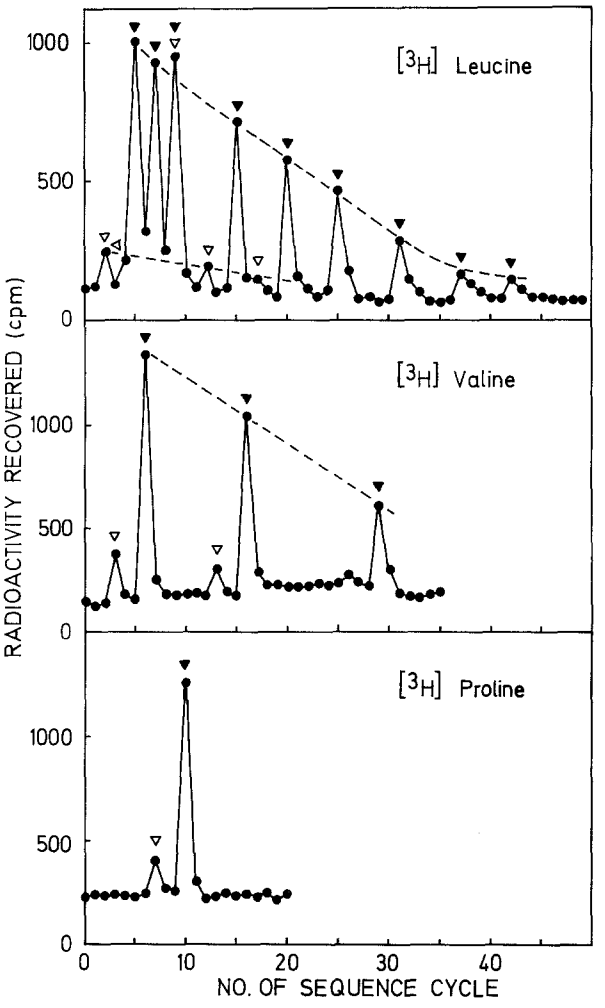

Fig. 2. Sequence analysis of phage fd gene 2 protein labeled with [^{3}H]leucine or [^{3}H]valine and [^{3}H]proline. Phage fd gene 2-protein and oxidized chicken lysozyme as carrier were subjected to 20–50 cycles of Edman degradation and the radioactivity released was analyzed as described in Materials and Methods. The radioactivities in the samples analysed were: 73,800 cpm for [^{3}H]leucine; 79,300 cpm for [^{3}H]valine, and 48,000 cpm for [^{3}H]proline. The [^{3}H]leucine protein was pretreated with 1 M HCl in methanol (deformylation treatment). The untreated [^{3}H]leucine derivative gave the same sequence positions as shown here (with 5% lower initial yield). The filled arrows indicate sequence positions assigned to sequence I, the open arrows to sequence II of Fig. 3. Cycle zero represents a blank cycle (without phenylisothiocyanate)

Release of [^{3}H]Leucine From the C-Terminus

Carboxypeptidase C treatment of [^{3}H]leucine-gene 2 protein resulted in the release of a total of four leucine residues per molecule (Table 2). Virtually all radioactivity coeluted with leucine as shown by analyses of radioactively labeled amino acids in carboxypeptidase digests and hydrolysates of the [^{3}H]leucine-protein. Only traces of radioactivity (less than 1% of the input) coeluted with the amino acids methionine, tyrosine

and phenylalanine. Glycine is the only amino acid residue known to significantly decrease the rate of release of C-terminal amino acids (Tschesche 1977). The release of four leucine residues suggested that the fourth leucine residue from the C-terminus is preceded by a sequence containing a residue which is only slowly released by carboxypeptidase C. Indeed, a glycine residue is found between the fourth and fifth leucine residue from the C-terminus of the protein (Schaller et al. 1978).

Discussion

Two methionine start codons are found close to the ribosome binding site of phage fd gene 2 (Schaller et al. 1978). Amino acid sequence data presented here show that both signals are used for the initiation of the peptide chain. 90% of gene 2 protein molecules start at the first methionine codon after the ribosome-binding site, 10% of the radioactivity of gene 2 protein eluted in positions which do not agree with the assignment of the translational start at the first AUG codon (Fig. 2). These peaks cannot be due to a flow of label from the tritiated amino acid to other amino acids, as a flow of label was not detected by acid hydrolysis of [^{3}H]leucine-containing gene 2 protein. Radioactivity coeluting with amino acids other than leucine was lower than 1% of the input.

Incomplete degradation of precursor protein could also generate two size classes of a peptide. For filamentous phages the products of gene 3 and gene 8 were found larger when synthesized in vitro in a protein synthesizing system compared to the size of these proteins from the virions (Konings et al. 1975). They are apparently processed in vivo. Our data on the sequence of gene 2 protein show that processing of the enzyme in vivo is unlikely, because most of a precursor protein should be converted into the mature product. Furthermore, the protein becomes deformylated in the cell but the N-terminal methionine residues are not released. As already shown for dihydrofolate reductase of *E. coli* (Brown et al. 1967), the enzyme which removes methionine residues seems not to act on Met-Ile sequences apparently, on Met-Leu sequences (sequence II in Fig. 3).

The minor peaks appearing in Fig. 2 most likely suggest a chain start which uses the second start codon leading to a shift of three residues in the protein sequence. Because of interference with the major protein not all expected positions could be detected for the second peptide. Increasing background in the course of degradation made the detection of minor peaks more and more difficult after 15 steps of sequence determination.

Unprecise chain initiation was also detected for gene 0.3-protein of bacteriophage T7 (Dunn et al. 1978 and F.W. Studier, pers. commun.). When the ribosome binding site for the gene 0.3 was mutated, the first methionine was used to initiate half the peptides translated and the second and third methionine were signals to initiate the other peptide chains. The early ribosome binding site is presumably used for the first and second chain start, whereas a second

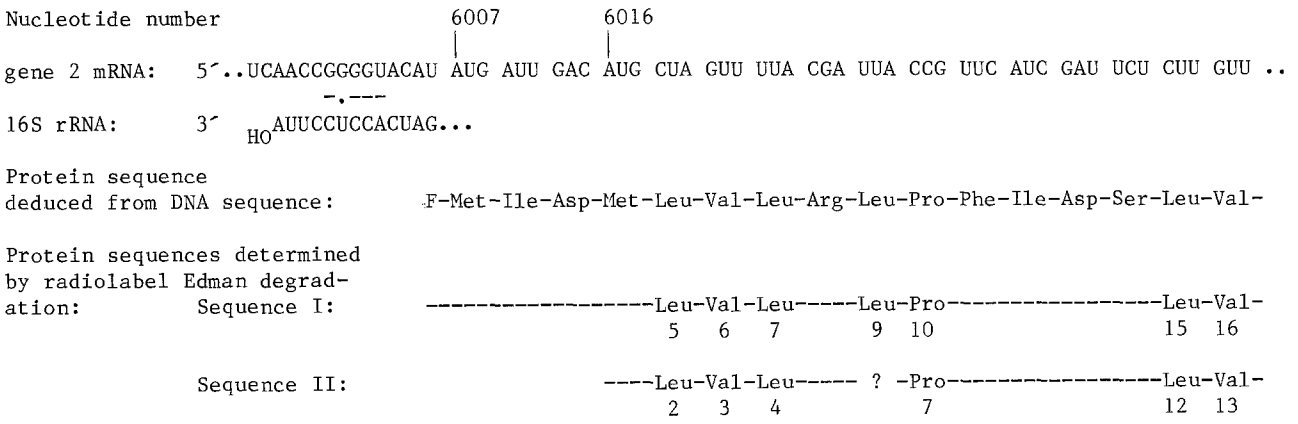

Fig. 3. Comparison of the two possible start positions of gene 2 protein deduced from the DNA sequence (Schaller et al. 1978) with the partial N-terminal protein sequence obtained by radiolabel Edman degradation. The ribosome binding site (underlined) and the protein sequence deduced from the DNA sequence were taken from Schaller et al. (1978), codon assignments of gene 2 mRNA are from Beck et al. (1978). Sequence I represents 87%–90% and sequence II 13%–10% of total gene 2 protein

ribosome binding site may account for the third chain start (Dunn et al. 1978). In wild-type cells the second and third initiator AUG are used in less than 1% of the chain starts. – Thiogalactoside transacetylase, the *lacA* gene product of *E. coli* might also be initiated at two codons (Büchel et al. 1980; Bennet 1974).

The data on fd gene 2 protein support the idea that more than one initiation codon may be available for chain start in vivo. A single ribosome binding site might be used for the two products of gene 2, since both initiator AUG codons are preceded in a proper distance by a 'Shine-Dalgarno sequence' (Schaller et al. 1978; Steitz 1979). The two resulting peptide chains may not have the same activities attributed to gene 2 protein. The choice for one or the other codon may regulate the enzyme functions required for phage replication.

Aside from the two sequences discussed for initiation of gene 2 protein a third peptide can be synthesized starting at a distance of $^3/_4$ within gene 2. This peptide, called X-protein or gene 10 protein, is in phase with gene 2 protein, but initiated by a promoter and a ribosome binding site within gene 2 (Van den Hondel et al. 1975). Phage ΦX174 gene A codes for a protein with a role comparable to fd gene 2 protein in DNA replication of the phage. A peptide which corresponds to fd gene 10 protein is called A* protein and is also synthesized in the reading frame of gene A protein (Linney and Hayashi 1974). The role of A* protein may be a regulatory function for the action of gene A protein (Ikeda et al. 1979; Langeveld et al. 1980).

The small single stranded DNA phages could possibly regulate the activity of genes and gene products required for the replication of their DNA by the synthesis of various peptides, which are coded for by the complete or partial information of the corresponding gene. The use of predominantly a first or a second initiation codon could regulate the amount of gene 2 product synthesized assuming that a chain start in some distance of the ribosome binding site is less efficient. Missing amino acids at the N-terminal end could furthermore change properties of gene 2 protein in the life cycle of phage fd.

Acknowledgement. We wish to thank K. Neifer for excellent technical assistance. Part of this work was supported by a grant from Deutsche Forschungsgemeinschaft through SFB 74 to K.B.

References

Beck E, Sommer R, Auerswald EA, Kurz C, Zink B, Osterburg C, Schaller H, Sugimoto K, Sugisaki H, Okamoto T, Takanami M (1978) Nucleotide sequence of bacteriophage fd DNA. Nucleic Acids Res 5:4495–4503

Bennet CD (1974) Similarity in the sequence of *Escherichia coli* dihydrofolate reductase with other pyridine nucleotide-requiring enzymes. Nature 248:67–68

Beyreuther K (1977) Coupling of phenylthiocarbamyl peptides to glass supports. In: Previero A, Coletti-Previero MA (eds) Solid Phase Methods in Protein Sequence Analysis. Elsevier, Amsterdam, p 107

Beyreuther K, Raufuss H, Schrecker O, Hengstenberg W (1977) The phospho*enol*pyruvate-dependent phosphotransferase system of *Staphylococcus aureus*. Eur J Biochem 75:275–286

Beyreuther K, Böhmer H, Dimroth P (1978) Amino-acid sequence of citrate-lyase acyl-carrier protein from *Klebsiella aerogenes*. Eur J Biochem 87:101–110

Beyreuther K, Ehring, Overath P, Wright JK (1980) Microsequence analysis of lactose permease of *Escherichia coli*. In: Birr Ch (ed) Solid Phase Methods in Protein Sequence Analysis. Elsevier, Amsterdam, p 199

Brauer AW, Margolies MN, Haber E (1975) The application of 0.1 M Quadrol to the microsequence of proteins and the sequence of tryptic peptides. Biochemistry 14:3029–3035

Brown JL, Koorajian S, Zabin I (1967) Thiogalactoside transacetylase. Amino-acid and carboxyl-terminal studies. J Biol Chem 242:4259–4262

Büchel DE, Gronenborn B, Müller-Hill B (1980) Sequence of the lactose permease gene. Nature 283:541–545

Clark BFC, Marcker KA (1966) N-Formyl-methionyl-*s*ribonucleic acid and chain initiation in protein biosynthesis. Polypeptide synthesis directed by a bacteriophage ribonucleic acid in a cell-free system. Nature 211:378–380

Dunn JJ, Buzash-Pollert E, Studier FW (1978) Mutations of bacteriophage T7 that affect initiation of synthesis of the gene 0.3 protein. Proc Natl Acad Sci USA 75:2741–2745

Fidanián HM, Ray DS (1972) Replication of bacteriophage M13. VII. Requirement of the gene 2 protein for the accumulation of a specific RFII species. J Mol Biol 72:51–63

Geider K, Meyer TF (1978) Gene II-protein of bacteriophage fd in enzymatic replication of viral duplex DNA. Cold Spring Harbor Symp Quant Biol 43:69–72

Ikeda J, Yudelevich A, Shimamoto N, Hurwitz J (1979) Role of polymeric forms of the bacteriophage ΦX174 coded gene A protein in ΦX174 cleavage. Proc Natl Acad Sci USA 254:9416–9428

Konings RNH, Hulsebos T, Van den Hondel CA (1975) Identification and characterization of the in vitro synthesised gene products of bacteriophage M13. Virology 15:570–584

Lämmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Langeveld SA, van Arkel GA, Weisbeek PJ (1980) Improved method for the isolation of the A and A* proteins of bacteriophage ΦX174. FEBS Lett 114:269–272

Lin NSC, Pratt D (1972) Role of bacteriophage M13 gene 2 in viral DNA replication. J Mol Biol 72:37–49

Lin NSC, Pratt D (1974) Bacteriophage M13 gene 2 protein: Increasing its yield in infected cells, and identification and localization. Virology 61:334–342

Linney EA, Hayashi M (1974) Intragenic regulation of the synthesis of ΦX174 gene A proteins. Nature 249:345–348

Meyer TF, Geider K (1979a) Bacteriophage fd gene II-protein. I. Purification, involvement in RF replication, and the expression of gene II. J Biol Chem 254:12636–12641

Meyer TF, Geider, K (1979b) Bacteriophage fd gene II-protein. II. Specific cleavage and relaxation of supercoiled RF from filamentous phages. J Biol Chem 254:12642–12646

Meyer TF, Geider K, Kurz C, Schaller H (1979) Cleavage site of bacteriophage fd gene II-protein in the origin of viral strand replication. Nature 278:365–367

Osborn MJ, Gander IW, Parisi E, Carson J (1972) Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. J Biol Chem 247:3962–3972

Pratt D, Erdahl WS (1968) Genetic control of bacteriophage M13 DNA synthesis. J Mol Biol 37:181–200

Schaller H, Beck E, Takanami M (1978) Sequence and regulatory signals of the filamentous phage genome. In: Denhardt DT, Dressler D, Ray DS (eds) The single-stranded DNA phages. Cold Spring Harbor Laboratory, p 139

Steitz JA (1979) Genetic signals and nucleotide sequences in messenger RNA In: Goldberger RF (ed) Biological Regulation and Development Plenum Press, New York London, p 349

Tschesche H (1977) Carboxypeptidase C. Methods Enzym 47, 73–84 (1977)

Van den Hondel CA, Konings RNH, Schoenmakers JGG (1975) Regulation of gene activity in bacteriophage M13 DNA: Coupled transcription and translation of purified genes and gene-fragments. Virology 67:487–497

Webster RE, Rementer M (1980) Replication of bacteriophage f1: A complex containing gene II protein in gene V mutant-infected bacteria. J Mol Biol 139:393–405

Communicated by E. Bautz

Received September 10, 1980