Proc. Natl. Acad. Sci. USA
Vol. 76, No. 10, pp. 4976–4980, October 1979
Biochemistry

Tryptic peptide analysis and NH_2-terminal amino acid sequences of polyhedrins of two baculoviruses from *Orgyia pseudotsugata*

(baculovirus polyhedrin structure/genetic stability)

G. F. ROHRMANN*, T. J. BAILEY†, B. BRIMHALL‡, R. R. BECKER†, AND G. S. BEAUDREAU*

*Department of Agricultural Chemistry and †Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331; and ‡Department of Biochemistry, University of Oregon Health Sciences Center, Portland, Oregon 97201

Communicated by Harold J. Evans, July 9, 1979

ABSTRACT Comparative analysis of the tryptic peptides and terminal amino acid sequence was made on polyhedrins from two genetically different baculoviruses that are naturally pathogenic for the same insect host. Comparison of the tryptic peptides of the nucleopolyhedrosis bundle virus and nucleopolyhedrosis single-rod virus of *Orgyia pseudotsugata* by means of cation-exchange resins indicated that the proteins have a closely related amino acid sequence. The NH_2-terminal amino acid sequence of polyhedrins from the two viruses differed in only 4 out of 34 amino acids. The nucleopolyhedrosis bundle virus and the nucleopolyhedrosis single-rod virus also differed in 4 and 5 out of 34 terminal amino acids, respectively, from the sequence reported for polyhedrin of a baculovirus of *Bombyx mori* [Serebryani, S. B., Levitina, T. L., Kautsman, M. L., Radavski, Y. L., Gusak, N. M., Ovander, M. N., Sucharenko, N. V. & Kozlov, E. A. (1977) *J. Invertebr. Pathol.* 30, 442–443]. In addition, the nucleopolyhedrosis single-rod virus had two amino acids (Met-Tyr) on the NH_2 terminus that were not present on the terminus of nucleopolyhedrosis bundle virus or *B. mori* baculovirus polyhedrin. Approximately half (six) of the total tyrosine residues are clustered in the terminal 20 amino acids of the polyhedrins. Secondary structures predicted from the primary sequence suggest that the tyrosines are clustered in two areas. This nonrandom distribution and the pK_a of about 10 for tyrosine may be related to the alkali solubility of the polyhedrin.

A distinctive feature of baculovirus subgroups A [nucleopolyhedrosis viruses (NPVs)] and B [granulosis viruses (GVs)] is the occlusion of the virions in a crystalline protein matrix. This protein is termed polyhedrin and granulin in NPVs and GVs, respectively. The crystal stabilizes the virions outside the host, and viable NPVs have been recovered from soil up to 11 years after a virus outbreak (1). The crystalline protein is composed of protein subunits of 25,000–30,000 daltons (2) and is soluble at alkaline pH. After ingestion, the crystals are dissolved by the alkaline pH of the insect midgut, thereby releasing the virions and allowing the infection to proceed. The primary structure of polyhedrin and granulin appears to be highly conserved. Studies using complement fixation indicated that granulins and polyhedrins from baculoviruses that infect Lepidoptera were antigenically related but differed from those produced by viruses that infect Hymenoptera (3). More recently it has been demonstrated that polyhedrins and granulins from a variety of Lepidoptera viruses have related tryptic peptide maps (4–6).

Orgyia pseudotsugata (the Douglas-fir tussock moth) is infected by two NPVs termed the nucleopolyhedrosis bundle virus (NPBV) and the nucleopolyhedrosis single-rod virus (NPSV) based on the number of nucleocapsids per envelope (7). Although the polyhedrin molecules from these two NPVs are identical in molecular weight and have closely related amino acid composition and antigenicity (8), the viruses are genetically distinct. They have been distinguished based on genome size (NPBV, 86×10^6 daltons; NPSV, 103×10^6 daltons) (9), G·C content of their DNA (10, 11), and restriction endonuclease fragment patterns, and have been demonstrated to have 1% or less DNA sequence homology (12).

This report compares the elution profiles of tryptic fragments of polyhedrin and the amino acid sequences of the NH_2-terminal 36 amino acids of polyhedrins from the two NPVs pathogenic for *O. pseudotsugata*. The sequences are also compared to that reported for the polyhedrin of an NPV pathogenic for the silkworm.

MATERIALS AND METHODS

Production and Purification of NPVs. NPVs were produced in *O. pseudotsugata* larvae infected *per os* or by injection with infectious hemolymph. At death or at an advanced stage of infection the insects were homogenized in a Virtis blender, passed through two layers of gauze, made 50% sucrose (wt/wt), and centrifuged at $17,000 \times g$ for 30 min at 20°C. This supernatant was discarded, and the pellet was suspended in TNS buffer (20 mM Tris·HCl/0.9% NaCl/0.00075% sodium dodecyl sulfate, pH 7.2), layered on a step gradient consisting of 12 ml of 58% sucrose (wt/wt) overlaid with 12 ml of 52% sucrose (wt/wt), and centrifuged at $112,000 \times g$ in a Beckman SW 27 rotor for 1 hr at 15°C. The interface between the 52 and 58% sucrose was removed, diluted with TNS buffer, layered on a 53–58% (wt/wt) sucrose gradient, and centrifuged at $112,000 \times g$ in a Beckman SW 27 rotor for 3 hr at 15°C. The polyhedra were removed, washed by sedimenting three times from TNS buffer, and stored in TNS buffer with 0.1% sodium azide at 4°C.

Production of Radioactively Labeled Polyhedrin. [^{3}H]-Arginine- and [^{3}H]lysine-labeled NPSV polyhedrin was produced by the intrahemocoelic injection of 100 μCi (1 Ci = 3.7×10^{10} becquerels) of [^{3}H]arginine (18.1 Ci/mmol) or [^{3}H]lysine (53.5 Ci/mmol) into larvae 5 or 6 days after infection with NPSV. [^{14}C]Arginine- and [^{14}C]lysine-labeled NPBV polyhedrin was produced by the intrahemocoelic injection of 8 μCi of [^{14}C]arginine (270 Ci/mol) or [^{14}C]lysine (290 Ci/mol) into larvae 3 or 4 days after infection. Larvae weighed approxi-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: NPBV, nucleopolyhedrosis bundle virus; NPSV, nucleopolyhedrosis single-rod virus; NPV, nucleopolyhedrosis virus; GV, granulosis virus.

mately 120 mg at the time of infection, and normally 10 larvae were used for the production of each of the labeled proteins. Labeled polyhedra were purified as described above.

Polyhedrin Preparation. For tryptic peptide analysis, ^{14}C-labeled NPBV and ^{3}H-labeled NPSV polyhedra were mixed with about 10 mg of unlabeled carrier polyhedra, and the protease was inactivated by incubating for 2 hr at 75°C and overnight at room temperature in the presence of 10 mM $HgCl_2$. The polyhedra were washed several times in distilled H_2O and then dissolved in 0.1–0.2 vol of 1 M Na_2CO_3/50 mM NaCl. Solubilization of the polyhedra was assisted by heating to 55°C for 10–20 min. The virions and undissolved polyhedra were removed by centrifugation at 40,000 × *g* for 30 min; the supernatant solution was either dialyzed against distilled water and lyophilized or dialyzed against three changes of the carbonate buffer, then applied to a Biogel A-5m column. The peak protein fractions were pooled and saturated with $(NH_4)SO_4$ to precipitate the protein. At times, 5–10 mg of a protein carrier was added before precipitation to increase the yield. The precipitate was washed with or dialyzed against H_2O and concentrated by lyophilization. The tryptic peptide profile from polyhedrin that had not been purified by passage through the column did not differ from the profile from material that had been passed through the column.

For determining the sequence, the supernatant fluid from the 40,000 × *g* centrifugation was dialyzed exhaustively against distilled water at 4°C and lyophilized.

Peptide Analysis. Polyhedrin was denatured in 8 M urea, adjusted to pH 8–9 with Tris, reduced with 2-mercaptoethanol (0.1 ml/15 mg of protein) for 1 hr, and treated with ethyleneimine (0.5 ml/15 mg of protein) for 1 hr. Then the pH was lowered to 3 with concentrated HCl. The protein was passed through a Sephadex G-25 column equilibrated with 0.2 M acetic acid, lyophilized, dissolved in distilled water, and then adjusted to pH 8.9 with trimethylamine. To this solution was added trypsin (TPCK, Worthington) at a substrate-to-enzyme ratio of 50. After digestion for 3 hr, the solution was adjusted to pH 3.0 by addition of HCl and the mixture was applied to a peptide analyzer equipped with a 0.6 × 15 cm Aminex A5 resin column (Bio-Rad). Peptides were separated at 50°C with a linear gradient derived from two buffers—125 ml of 0.2 M pyridine acetate (pH 3.1) and 125 ml of 2.0 M pyridine acetate (pH 5.0) at a flow rate of 15 ml/hr. Fractions of 2.0 ml were collected, evaporated to dryness at 60°C, dissolved in liquid scintillation fluid, and the radioactivity was measured.

Sequencer Chemicals. Sequencer reagents and solvents were purchased from Pierce. Polybrene, iodoacetic acid, and all other chemicals were of the highest purity grade and also were purchased from Pierce.

Reduction and Alkylation. In a typical alkylation, 5.3 mg of polyhedrin was dissolved in 5.0 ml of 8.0 M urea/1.0 mM EDTA and stirred under N_2 for 1 hr. The flask was opened and 2.0 ml of Tris buffer (17.4 g of Trizma base/30 ml of 1 M HCl made up to 100 ml with H_2O), 1.0 ml of 1 M HCl, and 2.0 mg of dithioerythritol were added. After a N_2 flush, the mixture was stirred for 3 hr and then 10.0 mg of iodoacetic acid dissolved in 1.0 ml of 1 M NaOH was added. The flask was flushed with N_2 and allowed to react for 15 min. The pH was reduced to 2.5 with 4 M HCl and the sample was immediately applied to a Sephadex G-25 (fine) column (3.0 × 40 cm) equilibrated with 0.1 M acetic acid. Fractions of 4 ml were collected at a flow rate of approximately 35 ml/hr. The fractions containing protein, determined spectrophotometrically at 280 nm, were pooled and lyophilized, resulting in 4.9 mg of protein recovered.

Amino Acid Sequence Determination. All sequence determinations were performed on a Beckman 890 C automated sequencer, with either a modified 0.1 M Quadrol program (13) or a modified 0.33 M Quadrol program with polybrene (14). In either case, the protein was added dry to the cup and dissolved by the addition of heptafluorobutyric acid. After an *N*-butylchloride wash, the protein was dried and the degradation was begun. In the runs using polybrene, after the initial drying of the protein, 100 μl of a 30 mg/ml solution of polybrene mixture was run through one full cycle without the addition of phenylisothiocyanate to eliminate most of the polybrene-associated artifacts (15). The resulting anilinothiazolinones from the sequential degradation were converted to their corresponding phenylthiohydantoins by the action of 1 M HCl at 80°C and identified either by gas chromatography (16, 17) on a Beckman GC-65 or by thin-layer chromatography (18). Histidine and arginine were identified by their respective tests (19).

RESULTS

Tryptic peptide patterns

Trypsin catalyzes the hydrolysis of proteins at arginine and lysine residues, and therefore most tryptic peptides contain one of these amino acids at the COOH terminus. After labeling the polyhedrin molecules with either radioactive arginine or radioactive lysine, one should obtain equivalent specific activities for each labeled tryptic peptide. Under certain circumstances double-labeled peptides might be generated (20). By using [^{3}H]arginine and [^{3}H]lysine for labeling the NPSV and [^{14}C]-arginine and [^{14}C]lysine for labeling NPBV polyhedrin and eluting from an ion-exchange resin the trypsin digests of both arginine-labeled polyhedrins or both lysine-labeled polyhedrins, one can compare directly the elution positions of the radioactive peptide fractions from the two viruses. Twelve major arginine peaks are evident in the elution profile from the tryptic digest of the mixture of [^{3}H]arginine-labeled NPSV and [^{14}C]arginine-labeled NPBV polyhedrins (Fig. 1 *upper*). Peaks 4A, 5A, and 6A appear to represent peptides unique to NPBV polyhedrin and peak 1A a fragment unique to NPSV. Seven major peaks are evident in the mixtures of [^{3}H]lysine-labeled NPSV and [^{14}C]lysine-labeled NPBV peptides (Fig. 1 *lower*). Two differences are found in the lysine peptide map. Peak 1L contains a peptide unique to NPSV polyhedrin and peak 2L demonstrates a slight difference in position between the two peptides eluted in this region. Therefore, when all the NPBV and NPSV peptides are enumerated, a total of six differences in 19 major peaks are present.

Polyhedrin from several baculoviruses has been reported to be phosphorylated (4). Although we also found ^{32}P associated with polyhedrin, attempts to measure it by phosphate determination indicated there were less than 0.5 phosphates per polyhedrin molecule. This suggests that either a specific amino acid(s) was phosphorylated in submolar amounts similar to that reported for α-tropomyosin (21) or the polyhedrin was randomly phosphorylated at a low level. We attempted to isolate a phosphorylated peptide from a tryptic digest by using the cation-exchange resin. All ^{32}P from a nuclease-treated sample of NPBV polyhedrin was association with an NPBV minor peak that elutes at the peak designated 1A or 1L (Fig. 1). It is likely that phosphorylated peptides would have altered elution properties from the cation-exchange resin. The fact that the ^{32}P-labeled peptide elutes as one peak suggests that phosphorylation is specific to one or a few peptides. The majority of these phosphorylated peptides would be represented elsewhere in the elution profile by their identical unphosphorylated counterparts.

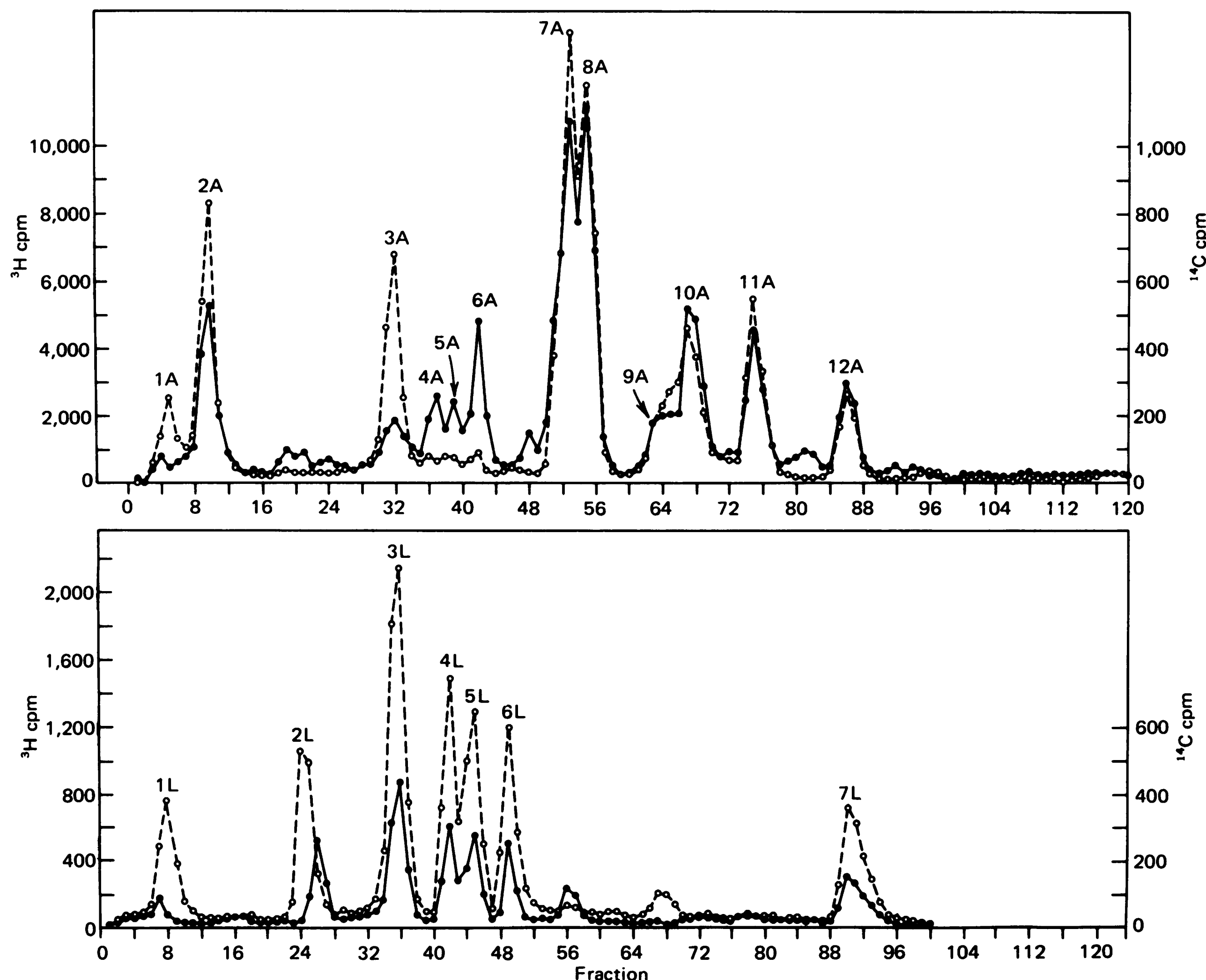

FIG. 1. Elution patterns of tryptic digests of polyhedrin. (*Upper*) Tryptic peptide elution patterns of arginine-labeled NPBV and NPSV polyhedrin. ●—●, [^{14}C]Arginine-labeled NPBV polyhedrin; ○- - -○, [^{3}H]arginine-labeled NPSV polyhedrin. (*Lower*). Tryptic peptide elution patterns of lysine-labeled NPBV and NPSV polyhedrin. ●—●, [^{14}C]Lysine-labeled NPBV polyhedrin; ○- - -○, [^{3}H]lysine-labeled NPSV polyhedrin.

NH_2-terminal sequence

Sequence determinations of the NPSV and NPBV polyhedrin proteins were complicated by their solubility properties. The protein samples were only sparingly soluble in the coupling buffer (0.1 or 0.33 M Quadrol, pH 9.0). NPSV polyhedrin gave nearly quantitative initial yields, whereas the NPBV protein gave less than 25% of the expected yield at the NH_2 termini. This may indicate that most of the NPBV polyhedrin molecules are blocked at the NH_2 terminus. The repetitive yields for both of the proteins were acceptable, giving values of 95.2% and 93.0% for the NPSV and NPBV proteins, respectively, over the full sequence determination.

Sequences for the first 36 amino acids from the NH_2 termini were determined for NPBV and NPSV. Both these sequences were similar to that reported for an NPV of the silkworm (22) (Fig. 2). However, the NPSV had Met-Tyr on the NH_2 terminus, a sequence not found on either the NPBV or the silkworm NPV polyhedrins. With the exception of two amino acids on the NH_2 terminus, the two tussock moth viruses differed in 4 amino acid residues among the subsequent 34 residues studied. Similarly, the NPBV and NPSV differed in four and five residues, respectively, from the silkworm NPV polyhedrin. Only a single base change would be required in either of the first two positions of the DNA codon to bring about these amino acid substitutions, except for an aspartic acid/arginine exchange (position 2).

A striking feature of all three polyhedrin sequences is that, out of a total of about 240 amino acids (22), approximately half (six) of the total tyrosine residues are clustered in the first 20 amino acids. The side chains of normal tyrosine residues have pK_a values varying between 9.5 and 10.5 (23), which are within the pH range at which polyhedrin crystals dissolve. It is possible that this nonrandom distribution of tyrosine is related to the solubility properties of polyhedrin. Another characteristic of the NH_2-terminal regions is the concentration of approximately a third of the total lysines in the first 36 amino acids.

Application of methods to predict protein secondary structure (24, 25) from amino acid sequences resulted in similar structures for all three polyhedrins (Fig. 3). Common to the structures are four β turns with small β sheets on either side side of the β turn beginning at amino acid 15 and an α helix beginning at amino acid 28. Amino acid variations found among the three polyhedrins would not cause major changes in the predicted structure. From the predicted secondary structure, it appears that the tyrosines could be clustered in close proximity in two areas of the NH_2-terminal region.

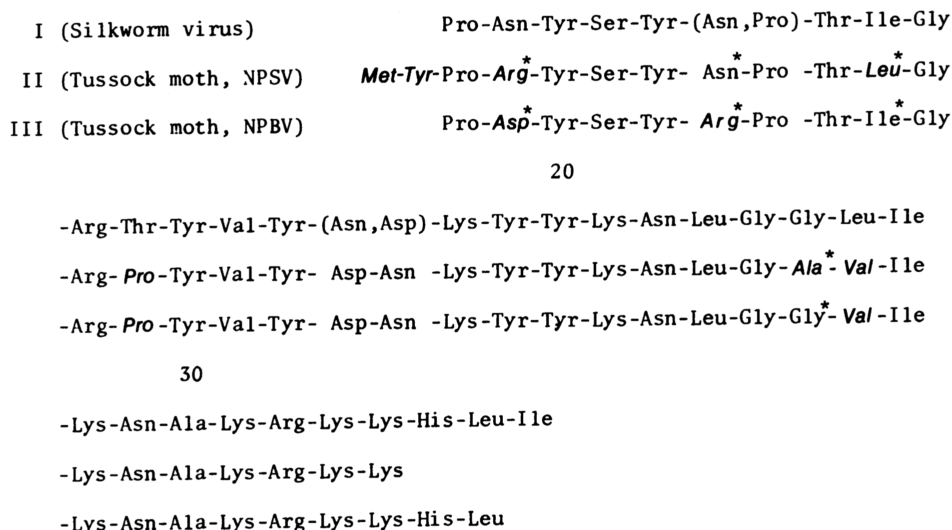

FIG. 2. NH_2-terminal amino acid sequence of two *O. pseudotsugata* baculovirus polyhedrins and that reported by Serebryani *et al.* (22) for the polyhedrin of a silkworm NPV. Sequences are numbered beginning with the first proline. The amino acids in NPSV and NPBV polyhedrins that differ from those of the silkworm virus are italicized. The differences between the amino acid sequence of the two tussock moth viruses are marked with an asterisk.

DISCUSSION

Viruses are capable of rapid evolutionary change because of their vast replicative potential. Although the two NPVs of *O. pseudotsugata* replicate in the same host and are often found in the same geographical region, they have diverged to such an extent that they appear almost completely unrelated genetically. Their genomes differ substantially in G·C content (10, 11) and size (9) and have no recognizable DNA sequence homology (12). In addition, the two viruses show distinct nucleocapsid proteins when analyzed by polyacrylamide gel electrophoresis (unpublished data). In contrast to these marked genetic differences, the polyhedrin molecules from the two viruses are very similar. They have similar amino acid composition, size, and antigenicity (8); in this report we have demonstrated that 68% of the tryptic peptides from both polyhedrins cochromatographed, and both polyhedrins have remarkably similar amino acid sequences near the NH_2 terminus (Fig. 2). Because the two polyhedrins are produced in the same host, the variation observed suggests that polyhedrin is coded for by the virus rather than by the host.

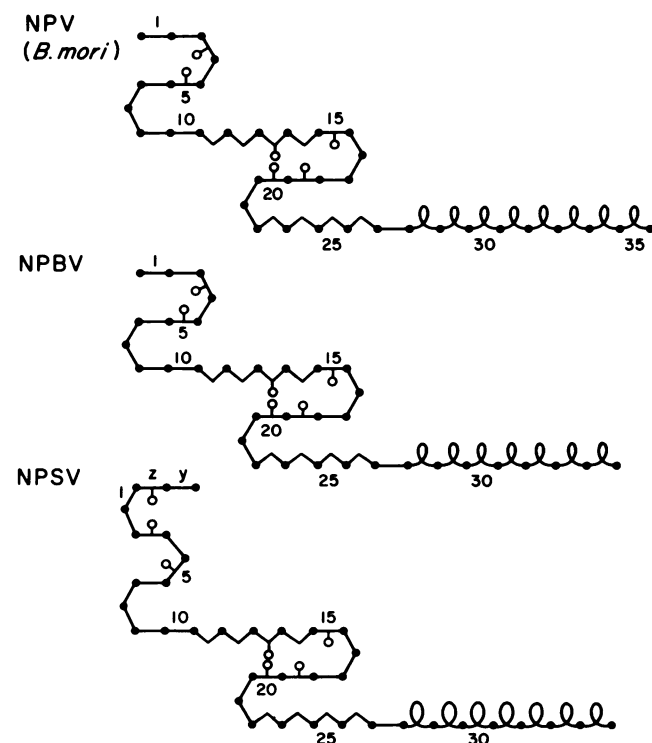

FIG. 3. Predicted secondary structure of the NH_2 termini of three polyhedrins by the method of Chou and Fasman (24, 25). The positions of tyrosine residues are indicated by open circles. Sequences are numbered beginning at the first proline. Amino acids preceding the first prolines are designated by letters.

Polyhedrin has evolved two highly specialized properties that allow it to function in the virus life cycle. First, it has the ability to form a protective crystal around the virions in the nuclear milieu; second, it resists solubilization except under strongly alkaline conditions similar to that found in the insect gut. The stability of the polyhedrin amino acid sequence in the three viruses suggests that the protein structure cannot be extensively modified and still function effectively. In addition, the genetic stability of polyhedrin also reflects its ability to function under a broad spectrum of selective pressures that led to the divergence of the other viral proteins.

In contrast to Lepidoptera baculovirus polyhedrin, a preliminary study on polyhedrin from the reovirus-like cytoplasmic polyhedrosis virus of *O. pseudotsugata* indicates that, although it is of similar size to baculovirus polyhedrin, it differs markedly in amino acid composition, and no relatedness has been detected by means of tryptic peptide analysis, antigenic analysis, or NH_2-terminal amino acid sequence determinations (unpublished data).

We thank Dr. M. Martignoni, U.S. Department of Agriculture, Forestry Sciences Laboratory, Corvallis, OR for supplying the *Orgyia pseudotsugata* larvae used in this study. This work was supported by U.S. Public Health Service Fellowship ES05110 (to G.F.R.) and U.S. Public Health Service Grants AM17850 (to R. T. Jones), ES02129, RR07079, and ES00040. This is Oregon Agricultural Experiment Station Technical Paper No. 5190.

1. Thompson, C. G. & Scott, D. W. (1979) *J. Invertebr. Pathol.* 33, 57–65.

2. Crozier, G. & Crozier, L. (1977) *Arch. Virol.* **55,** 247–250.
3. Krywiencyzk, J. & Bergold, G. H. (1960) *J. Insect Pathol.* **2,** 118–123.
4. Summers, M. D. & Smith, G. D. (1976) *Intervirology* **6,** 168–180.
5. Maruniak, J. E. & Summers, M. D. (1978) *J. Invertebr. Pathol.* **32,** 196–201.
6. Cibulsky, R. J., Harper, J. D. & Gudanskas, R. T. (1977) *J. Invertebr. Pathol.* **29,** 182–191.
7. Hughes, K. M. & Addison, R. B. (1970) *J. Invertebr. Pathol.* **16,** 196–204.
8. Rohrmann, G. F. (1977) *Biochemistry* **16,** 1631–1634.
9. Schafer, M. P., Rohrmann, G., Heine, U. & Beaudreau, G. S. (1979) *Virology* **95,** 176–184.
10. Rohrmann, G. F., Carnegie, J. W., Martignoni, M. E. & Beaudreau, G. S. (1977) *Virology* **80,** 421–425.
11. Rohrmann, G. F. & Beaudreau, G. S. (1977) *Virology* **83,** 474–478.
12. Rohrmann, G. F., McParland, R. H., Martignoni, M. E. & Beaudreau, G. S. (1978) *Virology* **84,** 213–217.
13. Brauer, A. W., Margolies, M. N. & Haber, E. (1975) *Biochemistry* **14,** 3029–3035.
14. Hunkapiller, M. W. & Hood, L. E. (1978) *Biochemistry* **17,** 2124–2133.
15. Klapper, D. G., Wilde, C. E., III & Capra, J. D. (1978) *Anal. Biochem.* **85,** 126–131.
16. Pisano, J. J. & Bronzert, T. J. (1969) *J. Biol. Chem.* **244,** 5597–5607.
17. Pisano, J. J. & Bronzert, T. J. (1972) *Anal. Biochem.* **45,** 43–59.
18. Summers, M. R., Smythes, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53,** 624–628.
19. Yamada, S. & Itano, H. A. (1966) *Biochim. Biophys. Acta* **130,** 538–540.
20. Niece, R. L., Margoliash, E. & Fitch, W. M. (1977) *Biochemistry* **16,** 68–72.
21. Mak, A., Smillie, L. B. & Bárány, M. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3588–3592.
22. Serebryani, S. B., Levitina, T. L., Kautsman, M. L., Radavski, Y. L., Gusak, N. M., Ovander, M. N., Sucharenko, N. V. & Kozlov, E. A. (1977) *J. Invertebr. Pathol.* **30,** 442–443.
23. Tanford, C. (1962) *Adv. Protein Chem.* **17,** 69-165.
24. Chou, P. Y. & Fasman, G. D. (1977) *J. Mol. Biol.* **115,** 135–175.
25. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47,** 45–148.