THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 259, No. 21, Issue of November 10, pp. 13159-13165, 1984
Printed in U.S.A.

The Covalent Protein Structure of Insecticyanin, a Blue Biliprotein from the Hemolymph of the Tobacco Hornworm, *Manduca sexta* L*

(Received for publication, June 21, 1984)

Clark T. Riley‡, Barbara K. Barbeau, Pamela S. Keim§, Ferenc J. Kézdy, Robert L. Heinrikson, and John H. Law§

From the Departments of Biochemistry and Chemistry, The University of Chicago, Chicago, Illinois 60637

The amino acid sequence has been determined for the insecticyanin from the hemolymph of the fifth instar larvae of the tobacco hornworm, *Manduca sexta*. The apoprotein is a single polypeptide chain of 189 amino acids, molecular weight 21,378, containing two disulfide bridges, 9-119 and 42-176. The sequence analysis was performed by automated Edman degradation of reduced and carboxymethylated insecticyanin and fragments generated therefrom by cyanogen bromide, trypsin, chymotrypsin, and *Staphylococcus aureus* proteinase. Most of the peptides were purified by reverse-phase high-performance liquid chromatography. A purification procedure for the isolation of insecticyanin in high yields and a simple method of determining disulfide linkages are also reported.

Camouflage is an essential ingredient in the survival strategy of an insect, and an understanding of the basic biology and biochemistry of this system may point the way to new avenues for insect control. The remarkable matching of insect color to plant pigments was first attributed to the accumulation and use of chlorophyll as a mimetic agent by the insect (Meldola, 1873). However, later investigation of the chemical nature of insect coloration (Przibran and Lederer, 1933) revealed that the green pigmentation of many insects is due to two components, the yellow contributed by carotenoids and the blue due to bile pigments, most often biliverdins.

Biliverdins have been isolated from several insect species (McDonagh, 1979), and isomers of biliverdin IX have been identified in the hemolymph and integument of several insects. Chino *et al.* (1983) have recently reported the isolation and characterization of a blue biliprotein from the hemolymph of the locust (*Locusta migratoria*); its properties, however, differ substantially from those of the insecticyanin from *Manduca sexta*. Dahlman (1969) showed that in the integument of the larva of the tobacco hornworm, *M. sexta* L, the blue pigment is intimately associated with a protein. Cherbas (1973) developed a purification procedure for the blue biliprotein in the pupal hemolymph of the same species. He crystallized and characterized this biliprotein, which he named insecticyanin. He also presented data suggesting that insecticyanin had a molecular weight of approximately 70,000 and consisted of 23,000-dalton monomers. The chromophore could not be removed from the holoprotein by extensive dialysis, and only by treatment with formamide could the prosthetic group and the apoprotein be dissociated. Cherbas (1973) proposed that the chromophore was biliverdin IXγ, citing as evidence the spectrum of the free chromophore and the chromatographic behavior of the dimethyl ester of the chromophore. Riddiford (1982) has shown that insecticyanin is only produced during the larval life stage. Yet, the protein persists throughout the pupal stage and into the adult female hemolymph from which it is ultimately sequestered into the egg (Cherbas, 1973).

The strong attachment of the chromophore to the protein and the remarkable stability of the holoprotein during the radical restructuring that accompanies insect metamorphosis underscore the importance of an investigation of the protein chemistry of insecticyanin. Furthermore, the determination of the covalent structural properties of insecticyanin is essential before one can attempt to manipulate camouflage as an insect control strategy. In this paper, we report an improved purification for insecticyanin from larval hemolymph of *M. sexta*, the complete covalent structure of the apoprotein, and the determination of some properties of the apoprotein and holoprotein.

EXPERIMENTAL PROCEDURES AND RESULTS[1]

DISCUSSION

This paper reports the amino acid sequence of insecticyanin from *M. sexta* (Fig. 1), as determined by automated Edman degradation of the intact biliprotein and derivative fragments generated by several selective cleavage procedures. The strategy normally employed in sequencing proteins containing around 200 amino acids is cleavage of the polypeptide into large segments either with cyanogen bromide or with trypsin under conditions where cleavage is restricted to arginine residues. Indeed, this strategy seemed promising since the amino acid composition of insecticyanin (Table I) indicated the presence of only 2 residues each of methionine and arginine. Thus, our initial efforts centered on the determination of the amino acid sequence of the fragments generated by cyanogen bromide treatment of the reduced and carboxymethylated insecticyanin. Initially, only one cleavage was ob-

* This work was supported by National Science Foundation Grants PCM 74-21379, 75-23506, and 83-02670 and by National Institute of General Medical Sciences Grant 29238. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address, Department of Pathology, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637.

§ Present address, Department of Biochemistry, Biosciences West, University of Arizona, Tucson, AZ 85721.

[1] Portions of this paper (including "Experimental Procedures," "Results," Figs. 1–8, and Tables II–IV) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 84M-1889, cite the authors, and include a check or money order for $9.20 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

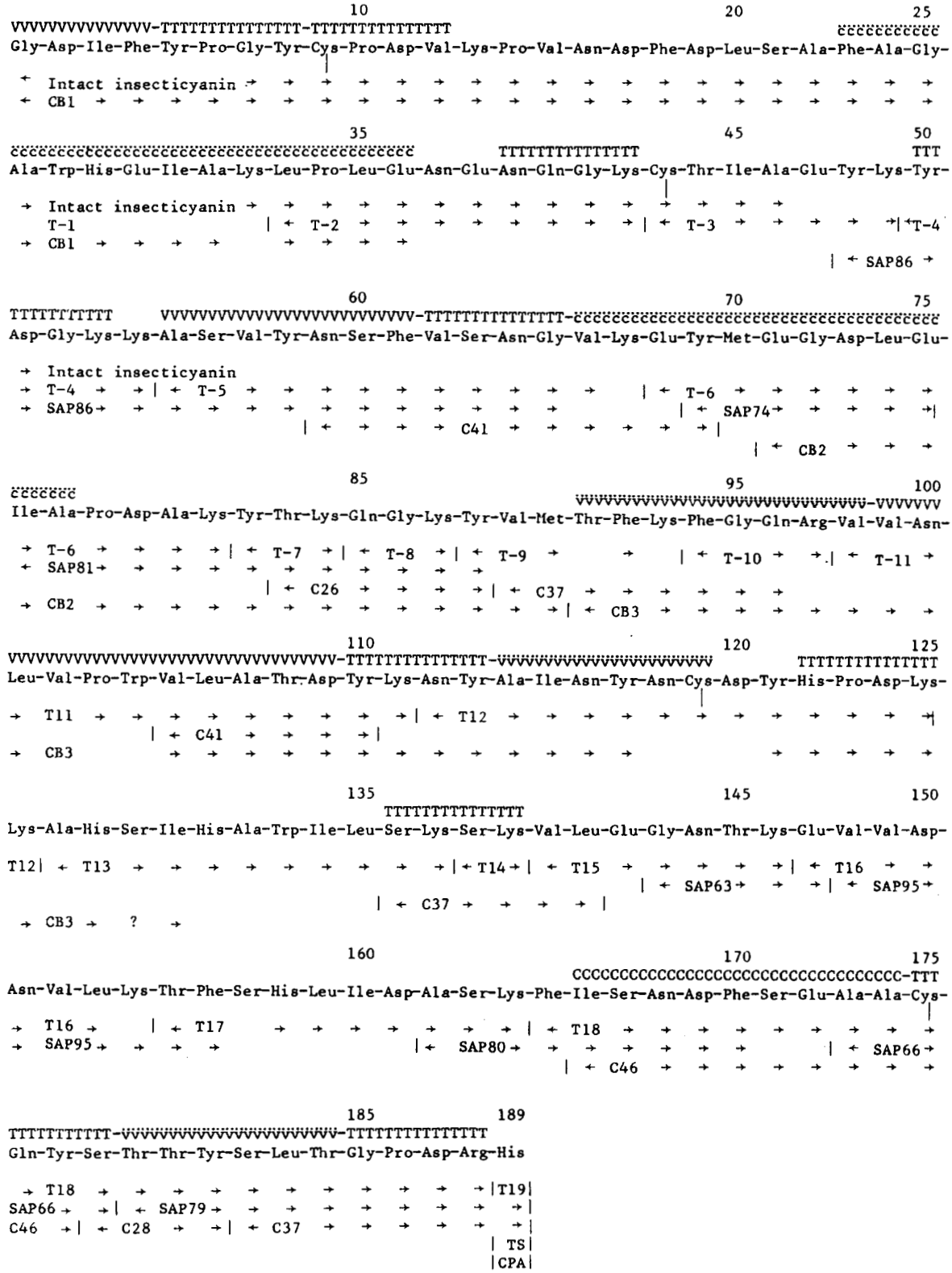

FIG. 1. **The amino acid sequence of insecticyanin and a summary of its peptide fragments used in establishing the sequence.** The sites of cleavage are indicated by vertical bars (|). For the methods of generating the fragments, see the accompanying Miniprint Section. The symbols identifying the individual peptides begin with a letter code reflecting their origin: tryptic digestion (T), cyanogen bromide cleavage (*CB*), *Staphylococcus aureus* protease digestion (*SAP*), chymotryptic digestion (*C*), tryptic digestion of the succinylated insecticyanin (*TS*), and carboxypeptidase A (*CPA*). *Horizontal arrows* indicate residues identified by Edman degradation. The probability of occurrence of secondary structural domains as predicted by the methods of Chou and Fasman (1978) and Garnier *et al.* (1978) are indicated as follows: right-handed helical regions are indicated by (*CCCCCCCC*). β strands are indicated by (*VVVVVV*). Predicted turns are indicated by (*TTTTTT*). Amphiphilicity is indicated by one-sided shading (*ÿÿÿÿÿÿ* or *ĉĉĉĉĉĉ*).

TABLE I
The amino acid composition of insecticyanin

Residue	Analysis[a]	Expected from sequence
Cys[b]	3.4	4
Asp[c]	26.9	15
Asn		11
Thr	7.9	9
Ser	8.5	13
Glu[c]	14.1	10
Gln		4
Pro	7.4	8
Gly	12.2	11
Ala	15.6	15
Val	15.4	14
Met	1.6	2
Ile	8.4	9
Leu	11.5	11
Tyr	13.9	14
Phe	8.7	9
His	6.2	6
Lys	18.8	19
Trp[d]		3
Arg	2	2
Total		189

[a] Numbers are expressed as residues per 2 arginine residues.
[b] Cystine was measured as the *S*-carboxymethylcysteine derivative.
[c] Asparagine and glutamine were measured as aspartic acid and glutamic acid, respectively.
[d] Tryptophan was not determined by amino acid analysis.

served. Once the sequence had been determined, it became clear why the other methionine was refractory to CNBr treatment; Met 90 is followed by a threonine, and such bonds have been observed to be resistant to CNBr cleavage. These early experiments with CNBr yielded sequence information for only 19 residues in the center of the molecule in addition to 32 residues at the NH_2 terminus. Our efforts toward producing large fragments were further thwarted by the fact that one arginine residue is located in the immediate vicinity of Met 90 and the other is the penultimate residue of the protein. Thus, limited cleavage of the polypeptide provided very little sequence information.

Deprived of the full benefit of these powerful procedures, we turned our attention to the complete set of tryptic peptides, which we were able to separate by reversed-phase HPLC[2] (Hermodson and Mahoney, 1981). Each of the tryptic peptides generated from reduced and carboxymethylated insecticyanin was thus purified and sequenced. The sequential order of these peptides in the parent protein was established with the aid of overlapping peptides generated by digestion of the apoprotein with chymotrypsin or with the proteinase from *Staphylococcus aureus*. These peptides were also separable by reversed-phase HPLC.

Fig. 1 shows the amino acid sequence of the apoinsecticyanin from *M. sexta*, together with the structure of the peptides used in the establishment of the sequence and of the overlaps. The anhydrous molecular weight of this apoprotein calculated from the sequence is 21,382. Details concerning peptide nomenclature, peptide purification, and determination of the sequence overlaps are reported in the Miniprint Sections. Insecticyanin contains four half-cystines as shown by the lack of reactivity toward 5,5′-dithiobis-2-nitrobenzoic acid. The disulfide bonds link positions 9-119 and 42-176, as shown by amino acid analysis of peptides from the tryptic digest of unreduced insecticyanin. Because of the remarkable resolution and reproducibility of the reversed-phase HPLC techniques of peptide separation, the determination of the positions of disulfide linkages can now be reduced to a simple comparison of HPLC peptide maps combined with amino acid analysis. Further details of this new procedure will be reported elsewhere.

The composition of the protein reveals an amino acid distribution typical of a globular protein, perhaps with the exception of tyrosine, which is twice as abundant as would be expected for an average protein (Dayhoff *et al.*, 1976). Application of the semiempirical predictive methods of Chou and Fasman (1978) and Garnier *et al.* (1978) suggests that the secondary structure of the protein may be very similar to that commonly seen in globular proteins, namely a mixture of α helical regions and β strands punctuated with β turns (Fig. 1). Most notable is the region from 90 to 120 which appears to exist as an array of β segments separated by turns, perhaps forming an antiparallel domain. It appears that the central strand is composed of strongly hydrophobic residues whereas the two flanking strands are clearly amphiphilic β structures (Kaiser and Kézdy, 1984). The other salient feature of the predicted secondary structures is the clear amphiphilicity of the two α helical segments 23-36 and 66-77.

This sequence is the first to be determined for a protein involved in insect camouflage. We have no functionally similar protein with which to compare sequences. However, since the functional role of insecticyanin is to maintain the level of biliverdin in the hemolymph of the insect, one might predict some functional similarity with other proteins capable of interacting with bile pigments or perhaps even with porphyrins. Accordingly, we have searched for possible sequence homologies between insecticyanin and the following proteins: cytochrome *c* from *M. sexta* (Chan, 1970), human serum albumin (Behrens *et al.*, 1975), allophycocyanin from *Anabaena variablis* (DeLange *et al.*, 1981), phytochrome from oats (Lagarias *et al.*, 1979; Lagarias and Rapoport, 1980), and the phycocyanins from *Mastigocladus laminosus* (Frank *et al.*, 1978), *Synachococcus* 6301 (Williams and Glazer, 1978), and *Cyanidium caldarium* (Offner *et al.*, 1981; Troxler *et al.*, 1981). In the cases of human serum albumin and allophycocyanin, we observed a few homologous tetra- to octapeptides (Fig. 2), suggesting possible functional similarity. The fact that the bile pigment binding regions of both albumin and allophycocyanin are included in these regions of homology leads us to speculate that the segments of insecticyanin from 60 to 90 and from 125 to 135 might participate in the ligand binding interaction.

During the course of our study of insecticyanin we have observed a number of properties which we believe would be worth investigating. The first property is the unusually tight binding of the chromophore to the apoprotein. In light of the rather mild conditions under which apophytochromes separate from their covalently bound chromophores (Rüdiger, 1970), we cannot rule out a labile covalent linkage between apoinsecticyanin and its chromophores. The visible spectra of insecticyanin and of the chromophores isolated from it suggest that the chromophores may consist of three or more distinct species. It is essential that the identities and stoichiometries of these chromophores be established unambiguously before the nature of the binding can be determined. Also, upon thin-layer isoelectric focusing (pH 3–9) insecticyanin segregates into several closely spaced bands despite its appar-

[2] The abbreviations used are: HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid; PTH, phenylthiohydantoin; TPCK, *S*-1-chloro-3-tolylsulfonamido-4-phenylbutan-2-one; TLCK, *S*-1-chloro-3-tolylsulfonamido-7-aminoheptan-2-one; SDS, sodium dodecyl sulfate.

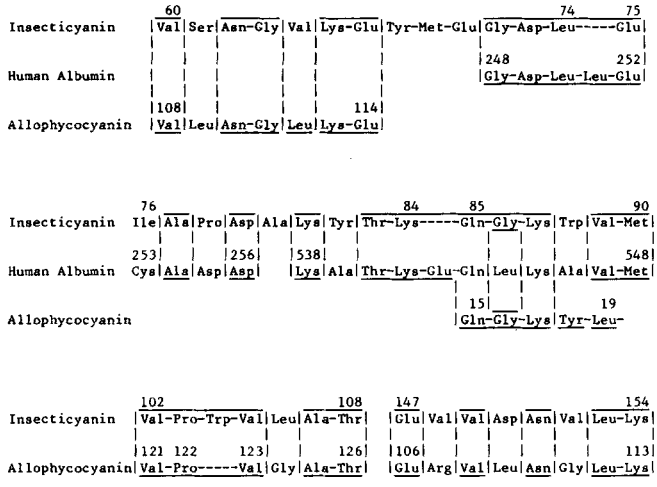

FIG. 2. **Sequence homologies between insecticyanin, human serum albumin (Behrens *et al.*, 1975), and allophycocyanin from *Anabaena variabilis* (DeLánge *et al.*, 1981).** *Numbers above* the amino acid residues indicate their positions in the respective proteins. Gaps introduced to maximize homologies are indicated by dashes (- - -). Amino acids which are identical with those of insecticyanin are *enclosed* while conservative replacements are *underlined.*

ent homogeneity by SDS-polyacrylamide gel electrophoresis and sequence determination. This charge microheterogeneity then might also reflect a multiplicity of chromophores. Compounding the complexities of microheterogeneity of insecticyanin due to possible difference in the chromophores is the fact that insecticyanin is an oligomeric protein in the native state. Chemical cross-linking by dimethyl suberimidate, gel permeation chromatography, and sedimentation studies indicate that the oligomer is, most likely, a tetramer. The physical chemistry of the self-association and of chromophore binding must be investigated before one can ascertain whether chromophore heterogeneity is a necessary structural feature of the tetrameric insecticyanin or whether the various chromophore species are distributed randomly among the subunits in the tetramer. Like the holoprotein, *apoinsecticyanin* appears tetrameric in structure based on our experiments with dimethyl suberimidate (data not shown). It is still possible that the various chromophores are generated as artifacts during dissociation, though the observation of the spectra of fresh hemolymph and the individual spectra of the separate chromophores strongly suggest the presence of several different species in the native state.

With the determination of the amino acid sequence, it is now possible to design experiments for the location of the "active" site or sites of the molecule. The modification of specific residues such as histidine or methionine should confirm whether the histidine-rich segment from insecticyanin is involved in the binding of the chromophore and whether both halves of the molecule are essential for the association of the chromophore. Knowledge of the sequence would then permit the immediate identification of specific amino acid side chains which interact with chromophores bearing affinity-labeling groups. Finally, the comparison of the sequence of the insecticyanin from *M. sexta* with those from insects of other genera and orders would be useful in probing the principles underlying the strategies and regulation of color mimicry among the insects. In this regard it is of interest to compare a blue hemolymph protein from *L. migratoria* reported by Chino *et al.* (1983). This biliprotein has a much higher monomer molecular weight, M_r = 83,000, than that of *M. sexta* insecticyanin and seems to be a rather unrelated molecule. Thus the means by which pigmented proteins are made and deployed in insects may vary widely over the class.

Acknowledgments—We wish to thank Dr. J. P. Reinecke, United States Department of Agriculture, Fargo, ND for generously providing eggs of *M. sexta*, Ilene Reardon for her kind assistance and for sequencing several peptides, and Dr. Roger Poorman for many helpful discussions.

REFERENCES

Beckman Instruments, Inc. (1971) *In Sequence*, p. 31-5, Palo Alto, CA
Behrens, P. Q., Spiekerman, A. M., and Brown, J. R. (1975) *Fed. Proc.* **34,** 591
Chan, S. K. (1970) *Biochim. Biophys. Acta* **221,** 497–501
Cherbas, P. K. (1973) Ph.D. thesis, Harvard University
Chino, H., Abe, Y., and Takahashi, K. (1983) *Biochim. Biophys. Acta* **748,** 109–115
Chou, P. Y., and Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251–276
Crestfield, A. M., Moore, S., and Stein, W. H. (1963) *J. Biol. Chem.* **238,** 622–627
Dahlman, D. L. (1969) *J. Insect Physiol.* **15,** 807–814
Dayhoff, M. O., Hunt, L. T., and Hurst-Calderone, S. (1976) *Atlas of Protein Sequence and Structure* Vol. 5, Suppl. 2, p. 301, National Biomedical Research Foundation, Washington, D. C.
DeLange, R. J., Williams, L. C., and Glazer, A. N. (1981) *J. Biol. Chem.* **256,** 9558–9566
Edman, P., and Begg, G. (1967) *Eur. J. Biochem.* **1,** 80–91
Frank, G., Sidler, W., Widmer, H., and Zuber, H. (1978) *Hoppe-Seyler's Z. Physiol. Chem.* **359,** 1491–1507
Garnier, J., Osguthrope, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120
Heinrikson, R. L., and Meredith, S. C. (1984) *Anal. Biochem.* **136,** 65–74
Hermodson, M., and Mahoney, W. C. (1981) in *Chemical Synthesis and Sequencing of Peptides and Proteins* (Liu, T.-Y., Schecter, A. N., Heinrikson, R. L., and Condliffe, P. G., eds) pp. 119–130, Elsevier/North-Holland, New York
Kaiser, E. T., and Kézdy, F. J. (1984) *Science (Wash. D. C.)* **223,** 249–255
Laemmli, U. K. (1970) *Nature (Lond.)* **227,** 680–685
Lagarias, J. C., Glazer, A. N., and Rapoport, H. (1979) *J. Am. Chem. Soc.* **101,** 5030–5037
Lagarias, J. C., and Rapoport, H. (1980) *J. Am. Chem. Soc.* **102,** 4821–4828
McDonagh, A. (1979) in *The Porphyrins* (Dolphin D., ed) Vol. 6, pp. 293–491, Academic Press, New York
Meldola, R. (1873) *Proc. Zool. Soc. London*, 153–162
Offner, G. D., Brown-Mason, A. S., Ehrhardt, M. M., and Troxler, R. F. (1981) *J. Biol. Chem.* **256,** 12167–12175
Porter, W. H., Cunningham, L. W., and Mitchell, W. M. (1971) *J. Biol. Chem.* **246,** 7675–7682
Przibram, H., and Lederer, E. (1933) *Akad. Wiss. Wien* **70,** 163–165
Riddiford, L. M. (1982) *Dev. Biol.* **92,** 330–342
Riddiford, L. M., and Law, J. H. (1983) in *The Larval Serum Proteins* (Scheller, K., ed) pp. 75–85, Georg Thiene Verlag, New York
Rose, S. M., and Schwartz, B. D. (1980) *Anal. Biochem.* **107,** 206–213
Rüdiger, W. (1970) *Angew. Chem. Int. Ed. Engl.* **9,** 473–480
Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971) *Biochemistry* **10,** 4912–4921
Spackman, D. H., Stein, W. H., and Moore, S. (1958) *Anal. Chem.* **30,** 1190–1206
Tarr, G. E., Beecher, J. F., Bell, M., and McKean, D. J. (1978) *Anal. Biochem.* **84,** 622–627
Troxler, R. F., Ehrhardt, M. M., Brown-Mason, A. S., and Offner, G. D. (1981) *J. Biol. Chem.* **256,** 12176–12184
Tsugita, A., and Scheffler, J.-J. (1982) *Eur. J. Biochem.* **124,** 585–588
Wang, D., Wilson, G., and Moore, S. (1976) *Biochemistry* **15,** 660–665
Williams, V. P., and Glazer, A. N. (1978) *J. Biol. Chem.* **253,** 202–211
Yamamoto, R. T. (1969) *J. Econ. Entomol.* **62,** 1427–1431

SUPPLEMENTAL MATERIAL TO
THE COVALENT PROTEIN STRUCTURE OF INSECTICYANIN, A BLUE HILIPROTEIN
FROM THE HEMOLYMPH OF THE TOBACCO HORNWORM, MANDUCA SEXTA L

Clark T. Riley, Barbara K. Barbeau, Pamela B. Keim,
Ferenc J. Kézdy, Robert L. Heinrikson and John H. Law

EXPERIMENTAL PROCEDURES

Materials.--Eggs of Manduca sexta were a gift from Dr. J. P. Reinecke, United States Department of Agriculture, Fargo, North Dakota. The larvae were reared at 27°C with a 15-hour-light-9-hour-dark photoperiod using standard diet (Yamamoto 1969), as modified by R. A. Bell.

Media for gel permeation chromatography and for ion exchange chromatography were from Pharmacia. Sodium azide, 1-phenyl-2-thiourea, and EDTA (disodium salt) were from Eastman Kodak Company. Azeotropic hydrochloric acid was prepared by distillation in an all-glass apparatus. N-ethylmorpholine, formamide, cyanogen bromide, 2-iodoacetic acid, and 2-mercaptoethanol were obtained from Aldrich Chemical Company. Guanidinium chloride, ultra pure grade, was obtained from Schwarz/Mann, Orangeburg, New York. Ethyl acetate, benzene, butyl chloride, acetonitrile, and 1-propanol were obtained from Burdick and Jackson Laboratories, Muskegon, Michigan. TFA (Sequenal grade), PTH standards, Polybrene (1,5-dimethyl-1,5- diazundecamethylene polymethobromide), Quadrol (N,N,N',N'-tetrakis(2-hydroxypropyl) ethylenediamine), dimethyl suberimidate, and heptaflouro butyric acid were purchased from Pierce Chemical Company, Rockford, Illinois. Formic acid (88% w/v in water) was obtained from Mallinckrodt Chemical Co. Ammonium bicarbonate was from Fisher Chemical Company, Itasca, Illinois. Bovine trypsin, treated with TPCK, bovine carboxypeptidase A, and protease from Staphylococcus aureus were purchased from Worthington, Freehold, New Jersey. Bovine chymotrypsin and TLCK were purchased from Sigma Chemical Company. TLCK-treated chymotrypsin was prepared by the method of Porter et al. (1971). Acrylamide, bis-acrylamide, ammonium persulfate, and the polyacrylamide-based separation media (Bio-Gels) are products of Bio-Rad, Richmond, California. Reverse phase HPLC columns (C8 or C18), 250 X 4.5 mm (5 um bead size), were obtained from Altex, Beckman Instruments, Berkeley, California or from DuPont. Glass scintillation vials and foil lined caps were from Research Products, Inc., Mount Prospect, Illinois.

Preparation of hemolymph.--Hemolymph was collected from fifth instar Manduca sexta larvae exceeding 6.5 grams in weight. For a batch of larvae, routinely 10 to 30 animals, 20 mM pH 8.3 Tris HCl buffer was prepared containing 100 mM NaCl, 0.02% NaN_3, and saturated with 1-phenyl-2-thiourea. The buffer, 0.4 ml per larvae, was cooled in an ice bath and stirred continuously. The larvae were first cooled in crushed ice, then a first proleg was cut and haemolymph was collected by applying gentle pressure to the larvae. The hemolymph was expressed directly into the stirred buffer. After collection, the diluted hemolymph was frozen immediately and stored at -20°C until needed. We obtained a yield of approximately 0.75 ml of hemolymph per larva.

Analytical procedures.--To aid in the identification and quantitation of cysteine, insecticyanin was reduced and carboxymethylated using a modification of the procedure of Crestfield, et al. (1963). Amino acids were quantitated in acid hydrolysates by automated ion exchange chromatography following the general procedures of Spackman et al. (1958), using a Durrum D-502 amino acid analyzer, and more recently by the method of Heinrikson and Meredith (1984). Select phenylthiohydantoins were converted to amino acids by hydrolysis in 57% HI as described by Smithies, et al. (1971) and analysed by the amino acid analyzer. Some of our peptide samples were hydrolysed by the procedure of Tsugita and Scheffler (1982) for the rapid hydrolysis of proteins. Phenylthiohydantoin amino acids were identified by reversed phase HPLC by the procedure of Rose and Schwartz (1980) using an IBM Instruments model 9533 liquid chromatograph and a Du Pont Zorbax C18 reversed phase column (4.6 mm x 25 cm; 5 micron). The effluent was monitored spectrophotometrically at 254 nm. No corrections were made for degradative losses of serine, threonine or histidine. Chemical crosslinking, using dimethylsuberimidate and polyacrylamide gel electrophoresis was performed by the method of Wang, et al. (1976).

Sequence analysis.--Native insecticyanin or peptides generated from the reduced and carboxymethylated protein were sequenced by automated Edman degradation (Edman and Begg, 1967), using either a Beckman 890B sequencer and the manufacturer's Quadrol program #011576, or a Beckman 890C sequencer modified with a cold trap and the Quadrol program #121078. One hundred ul of a 2% aqueous solution of Polybrene was added to the peptide solutions to aid in the retention of the peptide in the cup (Tarr, et al., 1978). The anilinothiazolinone product of the automated degradation procedure was converted to the more stable phenylthiohydantoin by treatment with a 1 N HCl solution containing 0.1% butane dithiol according to the general procedure of Beckman Instruments (1971).

Gel permeation chromatography.--The dilute hemolymph from as many as 40 larvae was thawed and centrifuged at 30,000 x g for 20 minutes at 4°C to remove a small amount of insoluble debris. The resulting clear supernatant was applied to a column (2.0 x 150 cm) of Sephadex G100 (medium) equilibrated at 4°C with a 20 mM Tris HCl buffer, pH 8.3, containing 100 mM NaCl, 0.2 mM 1-phenyl-2-thiourea, and 0.02% NaN_3. The column was eluted with the same buffer and the eluent was collected in fractions of 6.5 ml. The insecticyanin, with its typical blue color, eluted as a single peak at approximately 1.3 x void volume (Fig. 1*). The blue fractions were pooled, frozen, and stored at -20°C.

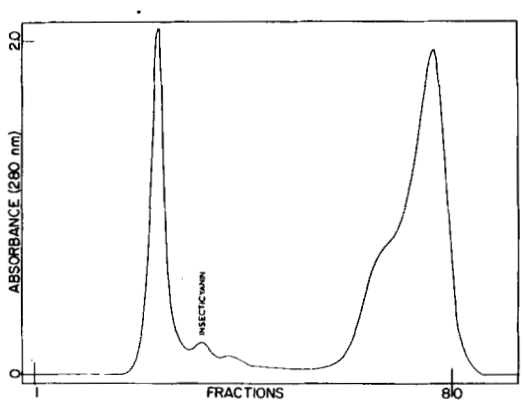

Fig. 1*. Elution profile of G100 gel permeation chromatography of the hemolymph of Manduca sexta. Hemolymph stored at -20°C was thawed and applied to a 2 X 150 cm column of Sephadex G100 (medium) equilibrated at 4° C with a 20 mM Tris HCl buffer, pH 8.3, containing 100 mM NaCl, 0.2 mM 1-phenyl-2- thiourea, and 0.02% NaN_3. The column was eluted with the same buffer and the eluent was collected in fractions of 6.5 ml. The effluent was monitored spectrophotometrically at 280 nm.

Cation exchange chromatography.--The pooled fractions from Sephadex G100 chromatography were concentrated by ultrafiltration using an Amicon cell with a UM10 membrane (Amicon Corporation, Lexington, Mass.). The solubility of insecticyanin was high enough to allow us to concentrate the macromolecules from several liters of pooled fractions into less than 100 ml. The resulting concentrate was centrifuged and repurified by gel permeation chromatography if turbidity was present, as this normally indicated the presence of high amounts of contaminating storage proteins (Riddiford and Law, 1983). The clear concentrate was dialysed against several changes of buffer containing 10 mM acetic acid brought to pH 5.5 with NaOH and 10 mM NaCl. Following centrifugation to remove the slight precipicate formed, the dialysate was applied to a column (2.0 x 50 cm) of SP Sephadex C50 equilibrated with the pH 5.5 buffer containing 10 mM acetate and 10 mM NaCl. The blue chromoprotein adsorbed completely as a narrow blackish-blue band at the top of the cation exchanger. The column was first washed with starting buffer and then eluted using a linear salt gradient from 10 mM sodium chloride, 10 mM acetate, pH 5.5 to 75 mM sodium chloride, 10 mM acetate, pH 5.5. The gradient was constructed using 100 ml of each buffer, and the protein eluted as a single peak about midway through the gradient. The protein was detected at 280 nm and could also be followed by visual observation. The insecticyanin-containing fractions (Fig. 2*) were pooled, concentrated to about five ml by ultrafiltration, and applied to a desalting column (2.0 x 50 cm) of Sephadex G25 (coarse) equilibrated with deionized water. The intensely blue protein that was eluted from the column was found to be homogeneous by SDS polyacrylamide gel electrophoresis. This solution was flash frozen and lyophilized. The resulting light blue powder was stored, under vacuum over $CaSO_4$ in the dark at 4°C.

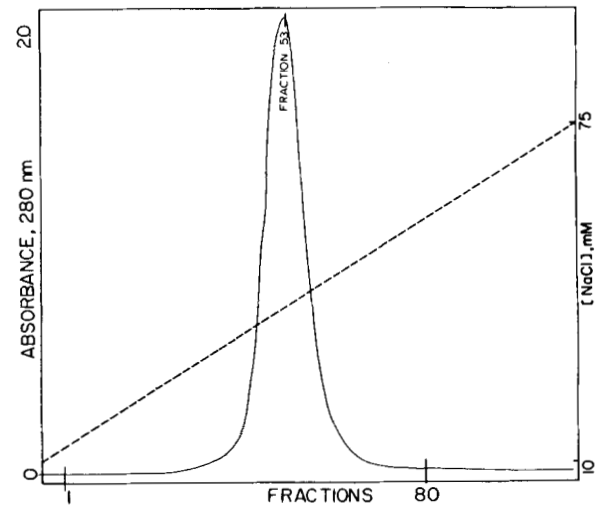

Figure 2*. Elution profile of SP Sephadex C50 cation exchange chromatography of partially purified insecticyanin. The pooled fractions from G100 gel permeation chromatography were applied to a 2.0 x 50 cm column of SP Sephadex C50 equilibrated with a 10 mM acetate buffer, pH 5.5, containing 10 mM NaCl. The protein was eluted using a linear gradient generated from 100 ml each of the starting buffer and the starting buffer containing 75 mM NaCl. Six ml fractions were collected. The effluent was monitored spectrophotometrically at 280 nm.

Peptide nomenclature.--Tryptic peptides from the reduced and carboxymethylated protein are designated by the symbol T followed by an arabic numeral indicating their relative position in the amino acid sequence. Staphylococcus aureus protease peptides and chymotryptic peptides are denoted by SAP and C respectively, followed by Arabic numerals indicating their elution order from high performance reversed phase liquid chromatography. Fragments generated by cyanogen bromide are labelled CBI, II, and III according to their position in the protein sequence.

Fragmentation by cyanogen bromide.--Reduced and carboxymethylated insecticyanin (42 mg) was desalted, lyophilized, and dissolved in 3.0 ml of aqueous 70% formic acid. The solution was treated with 40 mg of CNBr (a 100-fold molar excess with respect to methionine) for 24 hours and then with another 40 mg of CNBr for an additional 24 hours. The resulting peptides were separated on a column (2 X 200 cm) of Sephadex G75 (fine) equilibrated with 10% aqueous formic acid. The effluent was monitored spectrophotometrically at 280 nm and the collected fractions were pooled as shown in Fig. 3*. The pools were diluted five-fold with water, frozen and lyophilized.

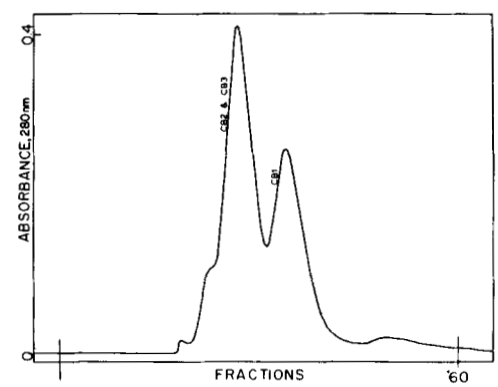

Figure 3*. Gel permeation chromatography of the cyanogen bromide fragments of insecticyanin. Insecticyanin, after treatment with cyanogen bromide, was applied to a 2 x 200 cm column of Sephadex G75 (fine) equilibrated with 10% aqueous formic acid and eluted with the same solvent. Six ml fractions were collected. The effluent was monitored spectrophotometrically at 280 nm.

SDS Polyacrylamide gel electrophoresis.--The apparent molecular weight of insecticyanin, apoinsecticyanin, and the cyanogen bromide fragments thereof was estimated using SDS polyacrylamide gel electrophoresis with a 16% polymer concentration in a vertical slab gel electrophoresis apparatus, according to the method of Laemmli (1970). The molecular weight standards were purchased from Bio Rad.

Tryptic peptides.--Reduced and carboxymethylated insecticyanin (14.5 mg) was dissolved in 1.0 ml of 0.1 M N-ethylmorpholine adjusted to pH 8.0 with acetic acid. To this solution was added 0.22 mg of TPCK-treated bovine trypsin and the mixture was stirred briefly. The vial containing the solution was capped and the mixture was incubated at 37°C for 24 hours. The reaction product was flash-frozen, lyophilized, and then redissolved in 1.0 ml of 0.1% (v/v) aqueous TFA. The tryptic peptides were separated by HPLC using a C18 reversed phase column eluted at a flow rate of 0.8 ml/minute with a solvent gradient constructed as follows: 10 minutes of 0.1% (v/v) aqueous TFA followed by a linear segment of 0.1% TFA in 1-propanol increasing to 50% 1-propanol in 180 minutes. The effluent was monitored spectrophotometrically at 230 nm (Fig. 4*). Each peak was collected in a separate scintillation vial and an aliquot of 10% of each sample was taken for hydrolysis and amino acid analysis. The remaining sample was flash frozen, lyophilized, and stored in a vacuum desiccator until needed. Those samples containing pure peptides - as judged by amino acid analysis - were sequenced directly and the others were purified further by gel permeation chromatography on columns (0.5 x 200 cm) of Bio-Gels P2 or P4 (200 - 400 mesh). For this purpose, the lyophilized peptide mixtures were dissolved in minimum volumes of aqueous 25% (v/v) 1-propanol, 0.1% (v/v) TFA, and applied to a column (0.5 X 200 cm) of the appropriate Bio-Gel equilibrated with the same solvent. The column effluent, collected in three ml fractions, was monitored spectrophotometrically at 230 nm. The fractions corresponding to the absorbance peaks were pooled, and a 10% aliquot was removed for amino acid analysis. The remaining sample was flash frozen, lyophilized and stored in a vacuum desiccator until needed.

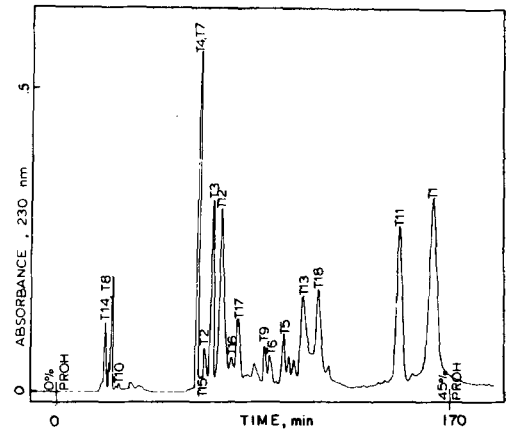

Figure 4*. Reverse phase HPLC separation of the tryptic fragments of reduced and carboxymethylated insecticyanin. Insecticyanin, after digestion with trypsin, was dissolved in 0.1% v/v aqueous TFA. Two hundred microliters of the 14.5 mg/ml solution were injected into a C18 reversed phase column equilibrated with 0.1% v/v aqueous TFA. The peptides were eluted at a flow rate of 0.8 ml/min at room temperature using a gradient constructed of 10 minutes of the equilibration solvent followed by a linear segment of 0.1% v/v TFA in 1-propanol increasing to 50% 1-propanol in 180 minutes. The effluent was monitored spectrophotometrically at 230 nm and each eluting peak was collected in a separate scintillation vial.

Digestion with Staphylococcus aureus protease.--Reduced and carboxymethylated insecticyanin (12.0 mg) was suspended in 5.0 ml of a 1% (w/v) ammonium bicarbonate solution in a scintillation vial. To this suspension was added 0.77 mg of Staphylococcus aureus protease and the mixture was stirred briefly. The vial was capped and the mixture was incubated at 37°C for 24 hours. The reaction product was flash frozen, lyophilized, and then redissolved in 0.2 ml of 10% aqueous acetic acid. The peptides were separated by HPLC using a C8 reversed phase column eluted at a flow rate of 1.5 ml/minute with a solvent gradient as follows: 10 minutes of 0.1% (v/v) aqueous TFA followed by a linear segment of 0.1% TFA in acetonitrile increasing to 38% acetonitrile in 140 minutes. The effluent was monitored spectrophotometrically at 230 nm (Fig. 5*). Each eluted peak was collected in a separate vial and an aliquot of 10% of each collected peak was taken for hydrolysis and amino acid analysis. The remaining sample was frozen and stored at -20°C until needed. Those samples that contained pure peptides whose amino acid compositions indicated potentially overlapping sequences were sequenced as previously described.

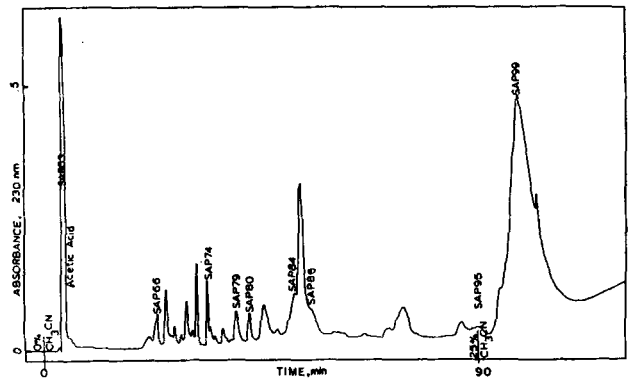

Figure 5*. Reverse phase HPLC separation of the Staphylococcus aureus protease fragments of reduced and carboxymethylated insecticyanin. Insecticyanin, after digestion with Staph. aureus protease, was dissolved in 10% aqueous acetic acid. Two hundred microliters of the 48 mg/ml solution were injected into a C8 reversed phase column equilibrated with 0.1% aqueous TFA. The peptides were eluted at a flow rate of 1.5 ml/min at room temperature using a gradient constructed of 10 minutes of 0.1% aqueous TFA followed by a linear segment of 0.1% v/v TFA in acetonitrile increasing to 38% acetonitrile in 140 minutes. The effluent was monitored spectrophotometrically at 230 nm and the eluting peaks were collected in individual scintillation vials.

Chymotryptic peptides.--Reduced and carboxymethylated insecticyanin (11.2 mg) was suspended in 1.2 ml of one percent (w/v) ammonium bicarbonate in a scintillation vial. To this solution was added 0.7 mg of TLCK-treated bovine chymotrypsin, the solution was stirred briefly, and maintained for 2 hours at 37°C. The reaction product was flash frozen, lyophilized, and then redissolved in 50% aqueous acetic acid. The chymotryptic fragments were separated by HPLC using a C18 column eluted at 0.8 ml/minute with a solvent gradient constructed of 10 minutes of 0.1% (v/v) aqueous TFA followed by a linear segment of 0.1% TFA in 1-propanol increasing to 30% 1-propanol in 160 minutes. The effluent was monitored spectrophotometrically at 230 nm (Fig. 6*) and each eluting peak was collected in a separate scintillation vial. A 10% aliquot was removed for amino acid analysis and the remainder was stored frozen at -20°C until needed.

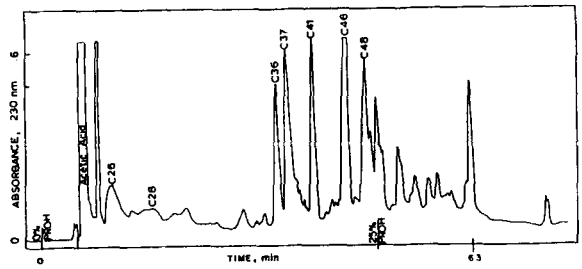

Figure 6*. Reverse phase HPLC separation of the chymotryptic fragments of reduced and carboxymethylated insecticyanin. Insecticyanin, after digestion with chymotrypsin was dissolved in 50% aqueous acetic acid. Two hundred microliters of the 11.2 mg/ml solution was injected into a C18 reversed phase column equilibrated with 0.1% v/v aqueous TFA. The peptides were eluted at a flow rate of 0.8 ml/min with a gradient constructed of 10 minutes of the equilibrating solvent followed by a linear segment of 0.1% v/v TFA in 1-propanol in 160 minutes. The effluent was monitored spectrophotometrically at 280 nm and the eluting peaks were collected in individual scintillation vials.

C-terminal analysis.--To 2.0 mg of native insecticyanin dissolved in 100 ul of 0.2 M N-ethylmorpholine-acetate buffer, pH 8.0, was added 20 ul of a solution consisting of carboxypeptidase A (5 mg/ml) in 0.04 N sodium hydroxide, 0.4% (w/v) aqueous sodium bicarbonate. The sample was incubated at 37°C and 20 ul aliquots were removed at 5, 10, 20, 40, and 80 minutes. These aliquots were added to equal volumes of 88% (v/v) aqueous formic acid and the samples were dried under nitrogen and analyzed for free amino acids.

Only one amino acid, histidine, was detected from the reaction of carboxypeptidase A with reduced and carboxymethylated insecticyanin.

Separation of individual chromophores of insecticyanin.--Native insecticyanin (2 mg) was dissolved in 1 ml of 6 M guanidinium chloride and the solution was applied to a column (0.5 x 30 cm) of Sephadex G25 (superfine) equilibrated with 6 M guanidinium chloride. The sample was eluted with the same solution and the effluent was monitored spectrophotometrically at 280 nm (Fig. 7*). The colorless protein eluted as the first peak, followed by at least four absorbing species which were violet to blue-green in color. The UV-Visible spectrum of each of these species was obtained and compared and it was found that the composite of these spectra approximated the spectrum of pure, native insecticyanin (Fig. 8*).

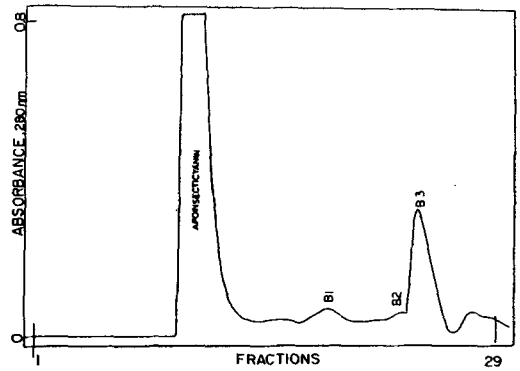

Figure 7*. Separation of individual chromophores of insecticyanin. Native insecticyanin (2.0 mg) was dissolved in 1 ml of 6 M guanidinium chloride and the solution was applied to a column (0.5 X 30 cm) of Sephadex G25 (superfine). The sample was eluted with 6 M guanidinium chloride and the effluent was monitored spectrophotometrically at 280 nm.

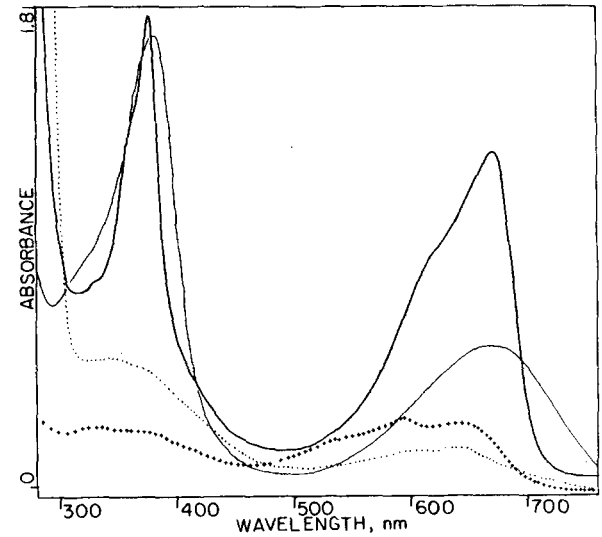

Figure 8*. Composite chromatogram of insecticyanin and its components. The heavy line (——) represents the chromatogram of fresh insecticyanin. The light line (——) represents the spectrum of peak B3 of figure 7*, which is indistinguishable from authentic biliverdin. The dotted line (.....) represents the spectrum of the first peak of figure 7*, which was shown to be apoinsecticyanin. The crossed line (+++++) represents the spectra of peaks B1 and B2 of figure 7*. The identity of these species has not been determined.

RESULTS

The complete amino acid sequence of the insecticyanin from Manduca sexta, shown in Fig. 1 was determined by Edman degradation of the N-terminal portion of the intact protein (Table II), the cyanogen bromide fragments and tryptic, chymotryptic, and S. aureus protease peptides all generated from reduced and carboxymethylated insecticyanin. This section will be devoted to details regarding the isolation and analysis of the peptides crucial to the sequence determination.

TABLE II.

IDENTIFICATION OF INDIVIDUAL AMINO ACID RESIDUES IN THE SEQUENCE OF INSECTICYANIN FROM MANDUCA SEXTA. (1)
(All quantities are nanomoles.)

Residue	Native			Residue	T10 (cont.)		CBIII(cont.)
Gly 1	+ (g,t)			Gln 96	57.92		+ (g,t)
Asp 2	+ (g,t)			Arg 97	169.83	TS	
Ile 3	+ (g,t)			Val 98	T11:106 b	20.17 (2)	+ (g,t)
Phe 4	+ (g,t)			Val 99	106.41 b	6.32	+ (g,t)
Tyr 5	+ (g,t)			Asn 100	134.98 b	4.54	+ (g,t)
Pro 6	+ (g,t)			Leu 101	241.41 b	7.20	+ (g,t)
Gly 7	+ (g,t)			Val 102	168.73 b	11.10	8.78 + (g,t)
Tyr 8	+ (g,t)			Pro 103	26.46 b	8.86	
Cys 9	89.03 + (g,t)			Trp 104	35.51 b	6.64	
Pro 10	+ (g,t)			Val 105	68.73 b	17.93	+ (g,t)
Asp 11	+ (g,t)			Leu 106	74.77 b	15.67	+ (g,t)
Val 12	+ (g,t)			Ala 107	11.59 b	28.36	10.87 + (g,t)
Lys 13	+ (g,t)			Thr 108	2.96 b	3.60	1.01
Pro 14	+ (g,t)			Asp 109	159.81 b	18.33	5.57
Val 15	+ (g,t)			Tyr 110	20.95 b		6.55
Asn 16	+ (g,t)			Lys 111	10.88 b		4.47
Asp 17	+ (g,t)			Asn 112	T12: 8.71		5.39
Phe 18	+ (g,t)			Tyr 113	11.83		6.13
Asp 19	+ (g,t)			Ala 114	40.82		6.91
Leu 20	55.91 + (g,t)			Ile 115	23.59 b		4.90
Ser 21	+ (g,t)			Asn 116	++ b		5.75
Ala 22	+ (g,t)			Tyr 117	16.42 b		
Phe 23	+ (g,t)			Asn 118	++ b		
Ala 24	+ (g,t)			Cys 119	22.47		
Gly 25	+ (g,t)			Asp 120	++		
Ala 26	+ (g,t)			Tyr 121	20.67		3.17
Trp 27	+ (g,t)			His 122	17.75		0.93
His 28	+ (g,t)			Pro 123	6.73		1.42
Glu 29	22.02 + (g,t)			Asp 124	++		1.70
Ile 30	+ (g,t)			Lys 125	1.58		1.60
Ala 31	+ (g,t)			Lys 126			1.68
Lys 32	+ (g,t)	T2		Ala 127	T13:13.06 b		3.35
Leu 33	+ (g,t)	93.48 b		His 128	3.23 b		0.78
Pro 34	+ (g,t)	16.99 b		Ser 129	4.14 b		
Leu 35	31.36	64.59 b		Ile 130	9.01 b		+ (3)
Glu 36		59.04 b		His 131	3.23 b		0.56
Asn 37	+ (g,t)	71.52 b		Ala 132	10.35 b		
Glu 38		43.38 b		Trp 133	5.72 b		
Asn 39	+ (g,t)	80.81 b		Ile 134	5.98 b		
Gln 40		62.00 b		Leu 135	5.80	C37-1c	
Gly 41	11.00,g,t	13.36 b		Ser 136	1.75	2131.3 (2)	
Lys 42	6.80,g,t	7.68 b		Lys 137	3.68 b	116.9	
Cys 43	4.87	T3:84.96 b		Ser 138	T14:53.05	32.9	
Thr 44	+ (g,t)	9.29 b		Lys 139	33.57	174.4	
Ile 45	+ (g,t)	242.75 b		Val 140	T15:97.02	69.6	
Ala 46	2.50	194.73 b		Leu 141	85.79	20.5	
Glu 47		118.32 b	SAP86	Glu 142	147.39	SAP63	
Tyr 48		44.13 b	2.30	Gly 143	42.64	72.9	
Lys 49	T4	7.23 b	1.68	Asn 144	35.27	37.4	
Tyr 50	23.03		1.77	Thr 145	4.43	7.9	
Asp 51	20.17		1.59	Lys 146	11.33	26.9	
Gly 52	6.46		1.17	Glu 147	T16:103.6	13.9	
Lys 53	11.04		1.71	Val 148	234.61	SAP95:26.42	
Lys 54	7.97		2.70	Val 149	249.31	27.82	
Ala 55	T5:36.54 b		.97	Asp 150	161.12	15.50	
Ser 56	0.08 b		.65	Asn 151	204.64	17.20	
Val 57	36.64		.55	Val 152	90.77	13.40	
Tyr 58	34.36		.57	Leu 153	80.61	11.43	
Asn 59	27.64 b		.38	Lys 154		4.56	
Ser 60	4.20 b		.43	Thr 155	T17:	2.57	
Phe 61	17.93	C4l	.58	Phe 156	42.94	9.41	
Val 62	14.37 b	37.6 (2)	.32	Ser 157			
Ser 63	b	37.2	+	His 158	24.53		
Asn 64	9.04 b	8.7	.72	Leu 159	16.22		
Gly 65	1.68 b	5.6	.12	Ile 160	6.68		
Val 66	3.09	6.4	.42	Asp 161	2.70		
Lys 67		2.7		Ala 162	1.60		SAP80:10.29
Glu 68	T6:6.35 b	1.9		Ser 163	+ (3)		+ (3)
Tyr 69	6.05 b	7.4		Lys 164			3.81
Met 70	3.52 b	CBII		Phe 165	T18:117.12 b		5.30
Glu 71	1.52 b	3.20 + (g,t)		Ile 166	101.31 b		3.85
Gly 72	1.60	8.60 + (g,t)		Ser 167	+ b (3)		+ (3)
Asp 73	6.02 b	+ (g,t)		Asn 168	24.40 b		3.75
Leu 74	3.08	+ (g,t)		Asp 169	58.39 b		2.26
Glu 75	4.34 b	+ (g,t)		Phe 170	61.32 b		2.16
Ile 76	6.30 b	+ (g,t)		Ser 171	+ b (3)		
Ala 77	1.91 b	+ (g,t)		Glu 172	23.07 b		
Pro 78	b	+ (g,t)		Ala 173	19.11		
Asp 79	b	+ (g,t)		Ala 174	17.58		
Ala 80	b	+ (g,t)		Cys 175	30.91		
Lys 81	b	+ (g,t)		Gln 176	27.26		
Tyr 82	T7:225.6a	6.50 + (g,t)		Tyr 177	29.47		
Thr 83	204.4 (a)	+ (g,t)		Ser 178	3.01		SAP79
Lys 84	236.4 (a)	+ (g,t)		Thr 179	2.31		.64
Glu 85	T8:51.06	+ (g,t)		Thr 180	+ (3)		2.18
Gly 86	53.79	+ (g,t)		Tyr 181	21.65	C37-1a	15.49
Lys 87	163.90	7.29 + (g,t)		Ser 182	1.79	2131.3 (2)	+ (3)
Tyr 88	T9:20.68	6.80 + (g,t)		Leu 183	10.02	245.4	15.49
Val 89	12.91	5.59 + (g,t)		Thr 184	4.72 b	50.2	1.07
Met 90	9.80	0.17		Gly 185	11.66	130.2	4.75
Thr 91		CBIII:1.72 + (g,t)		Pro 186	6.43	68.8	2.89
Phe 92	1.20	13.40 + (g,t)		Asp 187	6.54	72.8	5.13
Lys 93		+ (g,t)		Arg 188		357.0	2.93
Phe 94	T10:117.46	+ (g,t)		His 189	TS:92.43	+ (3)	+ (3)
Gly 95	44.38	+ (g,t)					

(1) Peptides shown are those which most clearly demonstrated the overlaps. All peptide sequences are consistent with this protein sequence. All quantitations shown were from HPLC. Where gas chromatography, thin-layer chromatography, amino acid analysis, or back-conversions were employed for identification, they are indicated by g, t, a, and b respectively.

(2) In some cases, the amino acids determined were shared by more than one peptide.

(3) In some cases, amino acids were present, but not integrable due to peak shape.

Cleavage by cyanogen bromide .--We observed that the reaction of CNBr with insecticyanin in the presence of 50% formic acid results in the cleavage of only the Met 70 bond. In order to cleave - even partially - the Met 90 bond, it was necessary to use 70% formic acid as the solvent. Thus the gel permeation chromatography of the reaction product in the presence of 50% formic acid yielded two peptides which proved to be CBI and CBII-III. The latter was sequenced through 22 residues. Figure 3* shows the gel permeation chromatography of the CNBr fragments generated in the presence of 70% formic acid. The unequal yield of peptides in pool I allowed the simultaneous sequencing of both peptides.

SDS polyacrylamide gel electrophoresis.--Molecular weights were estimated by plotting first the electrophoretic mobility of the protein standards versus the log of the molecular weight and then interpolating the mobilities of the unknowns on the straight line defined by the standards. The results of these measurements are reported in table III. It should be noted that the samples of insecticyanin and of apoinsecticyanin showed multiple faint bands corresponding to the apparent molecular weights of a series of oligimers of the protein, indicating extreme oligimerization even in the presence of SDS.

Table III. Polyacrylamide gel electrophoresis of Insecticyanin and of its cyanogen bromide fragments.

Peptide	Experimental M_r	Expected M_r
Native Insecticyanin	19,500	21,382
CB2-3 uncleaved	14,000	13,555
CB3	11,100	11,316
CB1	7,700	7,828
CB1/CB2-3 *	20,900	21,382
CB1/CB3 *	17,700	19,134

* These peptides were unreduced and hence linked by their disulfide bridges.

Tryptic digest.--It should be noted that 10 out of 19 tryptic peptides shown in Fig. 4* were eluted in a state pure enough for immediate sequencing. Peptides T9 and T6, T4 and T7, and T16 and T17 were obtained by subsequent gel permeation chromatography using Biogel P2, P4, or P6 as the stationary phase. Theoretically, one would expect the generation of 22 tryptic peptides, since the molecule contains 19 lysines and 2 arginines. In fact, we only obtained 18 peptides, and histidine and lysine as free amino acids due to the fact that the molecule contains two lysyl lysine bonds, one lysyl proline bond and the C-terminal arginyl histidine linkage.

Hydrolysis by Staphylococcus aureus protease.--After amino acid analysis, those peptides deemed crucial to the proof of the overlaps of cyanogen bromide or tryptic peptides were sequenced. All of the remaining peptides could be accounted for on the basis of their amino acid composition, the specificity of the enzyme, and by the known sequence of the protein.

Chymotryptic peptides.--We found that many of the peaks shown in figure 6* consisted of mixtures of peptides. Those peptides which appeared by virtue of their amino acid compositions to be able to complete the proof of the structure of the protein were sequenced. All of the remaining peptides could be accounted for on the basis of their amino acid composition, the specificity of chymotrypsin, and by the known sequence of the protein.

Separation of individual chromophores of insecticyanin.--Fig. 7* represents the gel permeation chromatography elution profile monitored at 280 nm of the products of the solution of native insecticyanin in 6M guanidinium chloride. Fig. 8* is an overlay of the spectra of the three major absorbance peaks shown in Fig. 7*. Peak 1 is the apoprotein. The spectrum of peak 3 is similar to the spectrum of authentic biliverdin and the spectrum of peak 2 approximates the difference between the sum of peaks 1 and 2 and that of the native insecticyanin. We have not as yet established the identity of any of these compounds. The absorbancies of insecticyanin in water are given in table IV.

Table IV. Spectroscopic constants for Insecticyanin from Manduca sexta.

Wavelength (nm)	$E_{1\%}$	Molar absorptivity
668.7	11.73	25760
374.4	16.78	36851
273.5	22.80	50071

Absorbance spectra of pure insecticyanin, 1.07 mg/ml in water were measured using a Perkin-Elmer Lambda 5 spectrophotometer with a 2 mm pathlength quartz cuvette. Molecular weight is the weight of apoinsecticyanin based on the sequence of the protein plus one molecule of biliverdin IX gamma.

Peptide Alignments

Residues 1 to 51 (T1 through T4)--Automated Edman degradation of 8.3 mg (400 nmol) of native insecticyanin provided the sequence of the first 51 residues (Table II). The amino acid composition of T1 agreed with the expected composition of residues 1 through 32. The sequence derived from 108 nmol of T2 proved identical to the sequence of residues 33 through 42 of native insecticyanin (Table II). The sequence derived from 148 nmol of T3 was identical to the sequence of residues 43 through 69 of the native insecticyanin. The first two residues of T4 (120 nmol) were the same as the last two residues which we were able to identify in the sequence analysis of native insecticyanin. The sequence determination for the first 42 residues of CBI was identical to that found for native insecticyanin. The overlap of T3 and T4 was again confirmed by the determination of the sequence of 18 nmol of SAP86.

Residues 52 through 92 (T4 through T9)--The amino acid sequence joining T4 and T5 (172 nmol) was found in SAP 86. The region from Lys-67 to Met-70 proved particularly rich in overlap information as each residue serves as a cleavage site for the reagents employed. T5 contains Lys-67, SAP-86 contains Glu-68, C4la (25.6 nmol) contains Tyr-69, while CBI contains Met-70. The fragment CBI must be linked to CBII since in the presence of 50% formic acid, only 2 fragments are generated. The CBI-CBII overlap was also confirmed by the sequence analysis of 80 nmol of T6 and 31 nmol of SAP 74. Peptides T7 (236 nmol), T8 (208 nmol), and T9 (60 nmol) were all contained in SAP81 and the overlaps were also confirmed by C26. The CBII-CBIII junction was shown to be contained in peptides T9, SAP81, and C37.

Residues 91 through 137 (T9 through T13)--The location from the analysis of the automated Edman degradation of CBIII resulted in the placement of T10 (143 nmol), T11 (100 nmol), T12 (160 nmol), and T13 (164 nmol) in sequence. The junction between T12 and T13 was also contained in fragment C37-1b (8.8 nmol).

Residues 138 through 189 (T14 through T19)--The sequence assignment of T13, T14(102 nmol), and T15 (160 nmol) was based on the sequence of C37-1c (29 nmol). The alignment of T15 and T16 (188 nmol) was determined by sequencing peptides SAP63 (89 nmol) and C48 (12.7 nmol), while the T16-T17 (176 nmol) junction was contained in SAP95 (36 nmol). SAP80 (35 nmol) represented residues 162 through 172 which were contained in T17 and T18 (165 nmol). The sequence of T18 also included the sequeace of C46, SAP66, and C28. Histidine was the sole amino acid product of the digestion of insecticyanin with carboxypeptidase A and was one of the products of the trypsin digestion of succinylated, reduced and carboxymethylated insecticyanin. Since, in addition, histidine was observed as the terminal amino acid of SAP79 (29.2 nmol) and C37-1a (15.6 nmol), we conclude that His-189 is the carboxyl terminal moiety.

Disulfide Bonds

SDS polyacrylamide gel electrophoresis.--The bands generated by the cyanogen bromide fragments of insecticyanin on SDS gels in the absence of mercaptoethanol are fully consistent with the fragmentation pattern expected from sequence analysis. Vigorous conditions of cleavage by cyanogen bromide generate three fragments corresponding to CBI/CBIII, CBI/CBII-III, and CBII. SDS polyacrylamide gel electrophoresis of CBI/CBIII in the presence of mercaptoethanol yielded two fragments corresponding to CBI and CBIII thereby showing that the two major cyanogen bromide fragments containing all four half-cysteines of the molecule are linked together by disulfides. Since both peptides contain two half-cysteines, they must be joined by two disulfide bonds either by a Cys 9 - Cys 119 and Cys 42 - Cys 176 combination or by a Cys 9 - Cys 176 and Cys 42 - Cys 119 combination.

Location of the disulfide linkages by HPLC.--For the location of the disulfide bonds, we took advantage of the fact that all of the tryptic peptides could be seperated in a single HPLC experiment. Accordingly, we first digested with TPCK-treated trypsin 1.4 mg of unreduced insecticyanin which had been denatured by a one hour exposure to 6 M guanidinium chloride and then desalted on a column of Sephadex G25 superfine. The tryptic digest was lyophilized, dissolved in 0.1% (v/v) TFA and subjected to HPLC under the same conditions as used in figure 4*. The elution pattern coincided with that of figure 4* except for the absence of peaks T3, T12, T18, and T1 and the appearance of two new peaks, one located between T5 and T13, and the other immediately following T11. The two new peaks were collected and subjected to amino acid analysis. The amino acid composition of the first peak agreed with the composition expected for T3 - T18 and not with any other combination expected of the disulfide peptides. In confirmation of this result, the composition of the second new peak clearly showed that it consists of T1 - T12. Thus, the two disulfide bonds of insecticyanin are Cys 9 - Cys 119 and Cys 42 - Cys 176.

We found that the combination of reversed phase HPLC with a shallow propanol gradient, and gel permeation chromatography on Biogel was sufficient to isolate all the major peptides of insecticyanin. From molecular weight determination by SDS polyacrylamide gel electrophoresis and from the amino acid analysis of the intact protein at the outset of this project, we knew that apoinsecticyanin should be a single peptide chain containing some 200 amino acids. By the sequence analysis described above, we found that the protein indeed consists of a single polypeptide chain of 189 amino acids the composition of which agrees, within experimental error, with the composition measured experimentally. In addition, the fragments generated by four different procedures have been separated chromatographically and the sequence presented in Fig. 1 accounts for all experimentally observed peptides. All peptides predicted from this sequence were found in every cleavage procedure.