

# De Novo Molecular Modeling and Biophysical Characterization of *Manduca sexta* Eclosion Hormone<sup>†</sup>

J. Joe Hull,<sup>‡</sup> Kathrin S. Copley, Kathleen M. Schegg, David R. Quilici, David A. Schooley,\* and William H. Welch

Department of Biochemistry, University of Nevada, Reno, Nevada 89557. <sup>‡</sup>Current address: Department of Molecular Biology, University of Wyoming, Laramie, WY 82071.

Received June 25, 2009; Revised Manuscript Received August 10, 2009

**ABSTRACT:** Eclosion hormone (EH) is an integral component in the cascade regulating the behaviors culminating in emergence of an insect from its old exoskeleton. Little is known regarding the EH solution structure; consequently, we utilized a computational approach to generate a hypothetical structure for *Manduca sexta* EH. The de novo algorithm exploited the restricted conformational space of disulfide bonds (Cys14–Cys38, Cys18–Cys34, and Cys21–Cys49) and predicted secondary structure elements to generate a thermodynamically stable structure characterized by 55% helical content, an unstructured N-terminus, a helical C-terminus, and a solvent-exposed loop containing Trp28 and Phe29. Both the strain and pseudo energies of the predicted peptide compare favorably with those of known structures. The 62-amino acid peptide was synthesized, folded, assayed for activity, and structurally characterized to confirm the validity of the model. The helical content is supported by circular dichroism and hydrogen–deuterium exchange mass spectrometry. Fluorescence emission spectra and acrylamide quenching are consistent with the solvent exposure predicted for Trp28, which is shielded by Phe29. Furthermore, thermodynamically stable conformations that deviated only slightly from the predicted *Manduca* EH structure were generated in silico for the *Bombyx mori* and *Drosophila melanogaster* EHs, indicating that the conformation is not species-dependent. In addition, the biological activities of known mutants and deletion peptides were rationalized with the predicted *Manduca* EH structure, and we found that, on the basis of sequence conservation, functionally important residues map to two conserved hydrophobic clusters incorporating the C-terminus and the first loop.

In insects, the physiological processes of growth and development differ fundamentally from our own. Instead of post-embryonic growth occurring gradually, the insect passes through a series of distinct life stages that are punctuated by the physical shedding of (ecdysis), and eventual emergence from (eclosion), the old exoskeleton. The coordinated muscular movements that comprise these stereotypical behaviors are regulated by complex neuroendocrine components (1). Once considered to be the sole component of this system (2, 3), the neuropeptide eclosion hormone (EH)<sup>1</sup> is now considered to be the principle neuro-modulator that regulates the release of the other components (1). Consequently, any disruption of EH function could adversely impact the viability of the insect (4, 5).

EH has been purified to homogeneity from two lepidopteran species, *Manduca sexta* (6, 7) and *Bombyx mori* (8), and has been translated in silico from genomic (9, 10) and cDNA sequences (11, 12) for several insect species from various orders. No structural homologues have been identified in vertebrates. The EH

sequences are highly conserved (59–93% identical with *Manduca* EH) and exhibit two distinguishing features: the presence of three reasonably conserved regions spanning residues 14–24, 28–44, and 46–61 and the invariant position of the six cysteines (Figure 1). The disulfides, biochemically determined to be Cys14–Cys38, Cys18–Cys34, and Cys21–Cys49 pairs (13, 14), are essential for biological activity (16). An intact C-terminus, in contrast to the N-terminus, is also important for activity (14, 17). Point mutations in *Bombyx* EH have identified a number of residues that are important for biological activity, including conserved residues Met24, Phe29, Ile55, Pro47, Phe58, and Leu59, as well as a hydrophobic residue at position 25 and a hydrophilic residue at position 40 (18, 19). Despite these structure–function analyses, little progress has been made in elucidating the EH solution structure. Early attempts to use <sup>1</sup>H NMR proved to be inadequate, because data could be obtained for only a fragment (residues 1–34) of *Bombyx* EH due to reported aggregation of the full-length peptide (20), while a hypothetical model proposed for *Bombyx* EH has not been experimentally confirmed (18).

To determine a plausible solution structure for *Manduca* EH, we used de novo protein structure prediction methodologies that exploited the physicochemical properties intrinsic to the primary sequence to generate a likely three-dimensional structure. To confirm features predicted by the in silico structure, we synthesized and folded *Manduca* EH and used biophysical methods to characterize the biologically active synthetic peptide. *Manduca*

<sup>†</sup>This work was supported by National Institutes of Health Grant GM48172 to D.A.S., with computational support from National Science Foundation Grant MCB9817605 to W.H.W.

\*To whom correspondence should be addressed: Department of Biochemistry, University of Nevada, Reno, NV 89557. E-mail: schooley@unr.edu. Phone: (775) 784-4136. Fax: (775) 784-1419.

<sup>1</sup>Abbreviations: EH, eclosion hormone; CD, circular dichroism; HDX, hydrogen–deuterium exchange; PDB, Protein Data Bank; TFA, trifluoroacetic acid; NAWA, *N*-acetyltryptophanamide; NAYA, *N*-acetyltyrosinamide.

FIGURE 1: Alignment of EH sequences. Alignment performed using CLUSTAL W (<http://npsa-pbil.ibcp.fr>) on 14 of the known EH sequences or fragments with default settings for the gap opening penalty (55). The order of sequences returned by the program approximates the phylogeny. Residues highlighted in black are invariant, whereas conserved residues are highlighted in gray. GenBank accession numbers are listed for sequences obtained from genomic data. Species abbreviations are per NCBI/Swiss Prot: *Tenebrio molitor*, Tenmo (F. M. Horodyski, personal communication, and ref 11); *Tribolium castaneum*, Trica (XP\_969164); *Thermobia domestica*, Thedo (F. M. Horodyski, personal communication, and ref 11); *Pediculus humanus corporis*, Pedhu (XP\_002432310); *Acyrtosiphon pisum*, Acypi (XP\_001943459); *Apis mellifera*, Apime (XP\_001122120); *Drosophila pseudoobscura*, Drops (EAL27464); *Drosophila melanogaster*, Drome (CAA51051); *Anopheles gambiae*, Anoga (9) (XP\_0012230805); *Culex quinquefasciatus*, Culqu (XP\_001864428); *M. sexta*, Manse (6, 7); *Helicoverpa armigera*, Helar (12); *Ostrinia furnacalis*, Ostfu (56); and *B. mori*, Bommo (8). Lines below the sequences represent the disulfide bonds.

FIGURE 2: Predicted secondary structure of *Manduca* EH. Results from various secondary structure prediction algorithms implemented in the SYBYL software package [M&S, Maxfield–Scheraga (23); GOR, Garnier–Osguthorpe–Robson (22); and Q&S, Qian–Sejnowski (24)] or algorithms available online (i.e., GOR1, Garnier–Osguthorpe–Robson method 1978; GOR3, Garnier–Osguthorpe–Robson method 1987; PHD, profile-based neural network prediction; and SOPM, self-optimized prediction method) (25). Only those regions predicted to be  $\alpha$ -helical (h) are shown. The consensus helical assignments of the starting structure used to generate our model (shown in bold) are based on designations from three or more prediction algorithms.

**Molecular Modeling of Manduca EH.** All of the computational methods were performed using SYBYL 6.6–7.3 (Tripos Inc., St. Louis, MO). Previous studies (21) indicated that the first two helices (termed helix I and helix II) are linked by disulfide bonds between Cys14 and Cys38 and between Cys18 and Cys34, while helix I and helix III are linked by a disulfide bond between Cys21 and Cys49. Residues 11–24 (helix I), 34–41 (helix II), and 48–61 (helix III) of the *Manduca* EH sequence were initially modeled as  $\alpha$ -helices with the lowest-energy side chain conformation of the first two helices determined prior to disulfide formation by running helix I through one round of minimization (100000 iterations of the Kollman all atom force field) and helix II through a 10 ps round of molecular dynamics with backbone atoms locked in each case. These and all subsequent molecular dynamics simulations were performed at 310 K with starting velocities derived from a Boltzmann distribution of energies and a sampling of conformational space every 0.5 ps. The ending helical domain from the dynamics run which had reached completion or equilibrium (defined by no further change in the average potential energy, total energy, and oscillation of the radius of gyration about a constant value) was minimized as before.

The software module Flexidock was used to align and orient Cys14 with Cys38 and Cys18 with Cys34. These disulfide bonds were then formed, and the resulting structure (composed of helix I and helix II) was minimized with the backbone atoms locked and a distance constraint of 2 Å placed on each of the disulfide bonds. The resulting molecule was then annealed in a sphere (2 Å in diameter) centered around the Cys14–Cys38 S–S bond such that only the atoms that fell within the sphere were moved. This algorithm was repeated multiple times in an ever-expanding sphere (1 Å increase in sphere diameter per annealing cycle) until the Cys18–Cys34 bond was encompassed. The third helix was then added by formation of the third disulfide bond between Cys21 and Cys49. The annealing process was repeated until the sphere about the Cys14–Cys38 bond reached 27 Å in diameter, thereby encompassing the three helical domains and the disulfide bonds linking them. The structure was then energy minimized to optimize the geometries of the side chains. The first 10 residues of the N-terminus and the C-terminal residue, Leu62, were then added. The N-terminal ends of the two loop regions, residues 25–33 (loop I) and 42–47 (loop II), were attached to the C-terminal ends of helix I (Met24) and helix II (Phe41), respectively. No peptide bond was formed between residues 33 and 34 or between residues 47 and 48, leaving gaps in the peptide. The carboxyl ends of the loops (i.e., Leu33 and Pro47) were brought into alignment with the rest of the peptide backbone in incremental steps with 10 ps molecular dynamics simulations incorporating distance constraints and a strong force constant (200 kJ penalty) but without forming the peptide bond. The backbone atoms of the helices were locked and the loops left unconstrained. The structure was then resubmitted to multiple dynamics runs with decreasing distance constraints, until the distance separating Leu33 from Cys34 (the C-terminus of loop I and the N-terminal end of helix II) and Pro47 from Glu48 (the C-terminus of loop II and the N-terminal end of helix III) was 2.5 Å. The connecting peptide bonds were formed, and the resulting structure was subjected to a 10 ps molecular dynamics simulation with torsional constraints on the backbone atoms. Upon reaching equilibrium, the constrained molecule was energy minimized as before. The torsional constraints were removed, and the structure was reminimized (resulting strain energy of –1042 kcal/mol).

The structure was further optimized<sup>2</sup> with a weak force constraint (0.1 kJ penalty) on the  $\phi$  and  $\psi$  angles of the proposed helical regions (i.e., residues 11–24, 34–41, and 48–61) using canonical values for helices:  $\phi = -57^\circ$ , and  $\psi = -47^\circ$ . The newly constrained structure was minimized as before and submitted to molecular dynamics (20 ps). Upon reaching equilibrium, the structure was minimized in the presence of all constraints. The  $\phi$  and  $\psi$  constraints were then removed, and the structure was energy minimized. The ending conformation was analyzed for helical content (55%) and strain energy (–1058 kcal/mol). This structure will be termed the *Manduca* EH model throughout the paper. Alternate disulfide pairings and disulfide bond torsion angles were tested using the AMBER 99 force field. Thus far, the model presented here has the most favorable strain and pseudo energies.

**Peptide Synthesis.** *Manduca* EH was synthesized using Fmoc (*N*α-9-fluorenylmethoxycarbonyl) chemistry with an Applied Biosystems 431A synthesizer using *p*-hydroxymethyl-phenoxo-derivatized resin on a 0.1 mmol scale, utilizing 1-hydroxybenzotriazole in 1-methyl-2-pyrrolidinone in the

presence of dicyclohexylcarbodiimide for Fmoc-amino acid activation. Extended 2 h single coupling cycles with a 10-fold molar excess of acylating species were employed (26). Protecting groups were as described previously (27) but also utilizing Thr(OtBu) and Cys(*S*-trityl). All of the dry resin-bound peptide (1.12 g) was cleaved in three batches using reagent K (28). The crude peptide (480 mg, 70.5 μmol) was recovered (70.5% yield) after precipitation with ether and being washed with ether and dried. Unfortunately, more than half of the product had cleaved at the acid-labile Asp9–Pro10 bond during acidic cleavage from the resin. This peptide was purified by preparative reversed-phase LC using a YMC C<sub>8</sub> column (300 Å, 20 mm × 250 mm) eluted at 15 mL/min using a gradient from 45 to 60% ethanol (95%) in 0.1% aqueous trifluoroacetic acid (TFA). The product had an average mass of 6813.4 (Bruker Proflex Plus MALDI-TOF), versus a calculated value of 6813.1. A portion of the purified, reduced peptide (4.6 mg) was then oxidized and folded when it was stirred overnight in 27 mL of 30% aqueous 1-propanol in 0.1 M potassium phosphate buffer (pH 8.5) containing 0.1 M KCl with 20 μL of DMSO,<sup>3</sup> with a stream of air sufficient to oxidize the peptide and evaporate the solution. The peptide was recovered by absorption to a cartridge column of ~3 mL of DyChrom Polyoprep 300–30 C<sub>18</sub> support, equilibrated with 0.1% aqueous TFA. Salts were eluted with 0.1% TFA; the peptide was eluted with 0.1% TFA in 60% acetonitrile, and the solvent was concentrated in a vacuum centrifuge. Crude oxidized peptide was then purified by reversed-phase LC using a YMC C<sub>4</sub> column (300 Å, 20 mm × 250 mm) eluted at 10 mL/min using a linear gradient from 30 to 60% ethanol (95%) in 0.1% TFA. The peptide contained few impurities and eluted at 13.8 min into the gradient versus 20.9 min for the reduced peptide. The product had an average mass of 6802.3 (Bruker Proflex Plus MALDI-TOF), versus a calculated value of 6807.0.

**Circular Dichroism.** CD measurements were performed with an AVIV 62-DS circular dichroism spectrometer (University of California, Berkeley, CA) in a 1 cm quartz cuvette at room temperature. CD spectra were recorded over a range of 199–300 nm in 1 nm intervals each averaged for 2 s. Synthetic *Manduca* EH (~2 μM) was analyzed in 20 mM sodium phosphate buffer (pH 6.7, a value used for *Manduca* saline that approximates its hemolymph pH). The observed ellipticity,  $\theta$  (millidegrees), was converted to molar mean residue ellipticity,  $[\theta]$  (degrees per square centimeter per decimole), and helical content was estimated using the suite of programs provided with DICROPROT 2000 version 1.0.4 (<http://dicroprot-pbil.ibcp.fr/>) (29). The EH concentration was determined by amino acid analysis of a hydrolyzed aliquot.

**Hydrogen–Deuterium Exchange.** A 4 μL aliquot of 1 μM synthetic *Manduca* EH in 20 mM ammonium phosphate (pH 6.7) was diluted to a final concentration of 40 nM in 96 μL of 99.8% D<sub>2</sub>O (Cambridge Isotope Laboratories, Inc.). The extent of deuterium incorporation was determined by directly injecting aliquots (2 μL) taken at different time points onto a Thermo-Finnigan LCQ Deca XP mass spectrometer.

**Fluorescence Spectroscopy.** Spectra were collected using a SPEX Fluorolog III spectrofluorimeter with monochromator slits set to 0.5 mm and a 3 mL quartz cuvette with a path length of 1 cm. Measurements were collected in 10 mM sodium phosphate buffer (pH 6.7). Fluorescence emission was monitored over the

<sup>2</sup>A. Dodson and W. H. Welch, unpublished data.

<sup>3</sup>D. S. King (Howard Hughes Medical Institute, University of California, Berkeley, CA), personal communication.



range of 300–400 nm in 0.1 nm intervals at 288 nm. Quenching of the intrinsic tryptophan fluorescence was determined by titration with 10  $\mu$ L aliquots of 3 M acrylamide and monitored at 350 nm (excitation at 295 nm). Data were corrected for the buffer contribution, absorbance at 295 nm, and dilution and analyzed according to the Stern–Volmer relationship (30):

$$F_0/F = 1 + K_{sv}[Q]$$

where  $F_0$  and  $F$  are the fluorescence intensities in the absence and presence of quencher,  $[Q]$  is the quencher concentration in molarity, and  $K_{sv}$  is the Stern–Volmer constant. A linear plot is indicative of collisional-based quenching. The fraction of fluorophore accessible to the quenching agent was determined via a Lehrer plot:

$$F_0/\Delta F = 1/(K_{sv}f_a[Q]) + 1/f_a$$

where  $\Delta F$  is  $F_0 - F$  and  $f_a$  is the fraction of fluorophore accessible to quencher. *N*-Acetyltryptophanamide (NAWA) and *N*-acetyltyrosinamide (NAYA), which serve as mimics of maximally accessible Trp and Tyr side chains in peptides, respectively, were used as controls (both purchased from Sigma).

**Molecular Modeling of a *Manduca* EH Analogue.** A three-dimensional conformation of the *Manduca* EH analogue synthesized by Wang et al. (31) was modeled. Since the analogue incorporated norleucine and naphthylalanine, two unnatural amino acids not included in SYBYL, a mathematical description of each was made. Norleucine was constructed using isoleucine as a template, while naphthylalanine was based on a phenylalanine template. Designation of the atom types of both residues was done according to the Kollman all-atom type, while charge was computed by quantum mechanics scaled to Amber values using common side chain atoms. The *Manduca* EH analogue was modeled by substitution of Asn for Gln20, norleucine residues for Met11 and Met24, and naphthylalanine for Trp28. The resulting structure's side chains were minimized with the backbone atoms locked and then submitted to molecular dynamics (20 ps). Upon reaching equilibrium, the ending structure was minimized and then compared to the native structure. The lipophilic, electrostatic, and hydrogen bonding potentials of the two structures were evaluated using MOLCAD.

**Molecular Modeling of *Bombyx* and *Drosophila* EHs.** The *Bombyx* and *Drosophila* EH models were modeled using the *Manduca* EH model as a template. Amino acids were changed to reflect the *Bombyx* sequence, and the  $\phi$  and  $\psi$  of the residues reported by Fujita et al. (18) to be helical (residues 9–24, 34–40, and 49–55) were constrained with a weak force constant (0.1 kJ penalty). This constrained structure was minimized and then submitted to molecular dynamics (20 ps). Upon reaching equilibrium in dynamics, the ending molecule was minimized as before with the  $\phi$  and  $\psi$  constraints present. A second energy minimization was performed in which the torsional constraints were removed. An alternative model was generated incorporating our initial helical assignments (11–24, 34–41, and 48–61). *Drosophila* EH was also modeled by mutating the appropriate residues in *Manduca* EH to the *Drosophila* sequence and incorporating weak torsional constraints on the residues (11–24, 34–41, and 48–61) assumed to be helical in our starting *Manduca* EH model.

**Analysis of *Manduca* EH, *Bombyx* EH, and *Drosophila* EH Structures.** The hypothetical models were analyzed using the SYBYL tool ProTable to evaluate (i) the allowed and

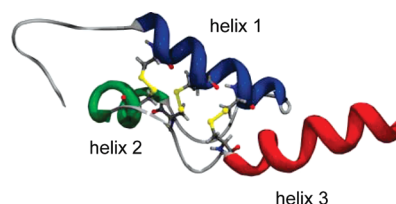


FIGURE 3: Ribbon diagram of the predicted *Manduca* EH solution structure. Lowest-energy conformation of full-length *Manduca* EH following molecular dynamics simulations and unconstrained energy minimization. Disulfide bonds are colored yellow, and helical segments are colored blue (helix I), green (helix II), and red (helix III). This figure was rendered using SwissPdb Viewer (57) and POVRay3.5.

disallowed conformational space of the individual amino acids, (ii) the local geometries of the bond angles and lengths, (iii) the pseudo energies of the residues in relation to their three-dimensional spatial distribution, and (iv) the determination of buried and exposed residues. Using the MOLCAD module, solvent accessible surfaces of the structures were generated using a Connolly surface with a solvent radius of 1.4 Å. The lipophilic, electrostatic, and hydrogen bonding potentials of the three peptides were mapped onto these surfaces using color scales that were normalized to the lowest and highest global values for each property.

## RESULTS

Examination of the aligned EH sequences (Figure 1) indicates (i) six invariant Cys residues, (ii) a high degree of sequence identity at the C-terminus, (iii) a low degree of identity within the N-terminus, (iv) the sequence differences tend to be conservative changes, and (v) the N-terminus is rather variable in both sequence and length, whereas the C-terminus is far more conserved. These observations, in conjunction with the published evidence for cross-reactivity, suggest that EHs exhibit a common structural motif that is necessary for receptor binding and activation. A BLAST search using *Manduca* EH as a query failed to return any peptides or proteins other than known EHs, while a similar search employing the structure prediction tools of ExPASy failed to return any significant threading templates.

**De Novo Modeling of *Manduca* EH.** We proposed to generate de novo a hypothetical three-dimensional model of *Manduca* EH based on three criteria. One, the restrictive nature of the three invariant disulfide bonds would serve as excellent initial constraints in searching the conformational space of the peptide. Two, the conserved nature (Figure 1) of the residues (11–24, 34–41, and 48–61) predicted by multiple secondary structure prediction algorithms to be helical (Figure 2) suggests that they are essential for biological activity, and thus, their presence should be maintained. Placing constraints on the helical  $\phi$  and  $\psi$  angles should greatly facilitate the conformational search process (21). Three, the lowest-energy conformation is expected to approach that of the native conformation.

Using this approach, a model of *Manduca* EH was generated. Initially, the correct disulfide pairings with the appropriate bond lengths, dihedral angles, and torsion angles were established. The full-length peptide was generated in molecular dynamics by slowly splicing in the unstructured regions and further refined by constraining the  $\phi$  and  $\psi$  torsion angles of the predicted helical regions to values typical of  $\alpha$ -helices. The side chains of the peptide backbone were then allowed to find their most

Table 1: Computational Analysis of the Proposed Solution Structures for the *Manduca*, *Bombyx*, and *Drosophila* EHs

	<i>Manduca</i> "preliminary" (21)	<i>Manduca</i>	<i>Bombyx</i>	<i>Drosophila</i>
helical regions <sup>a</sup>	12–16, 34–41, 49–53, 55–57	12–24, 35–41, 49–62	13–24, 34–41, 49–62	4–6, 11–24, 35–41, 49–63
% helix	34	55	55	63
strain energy (kcal/mol)	–218	–1058	–1142	–991
pseudo energy <sup>b</sup>	0.12 kT	–0.04 kT	–0.09 kT	–0.12 kT
no. of solvent-exposed residues	24	22	21	24
disallowed residues <sup>c</sup>	Ala3, Tyr8, Cys34, Ala35, Ala58	Ala27, Glu48	Glu48	Asp27, Ala45
cis proline bond	no	yes	yes	yes
no. of ionic residues	14	14	14	11
no. of salt bridges	6	4	6	2
no. of hydrogen bonds	89	115	122	105
root-mean-square deviation (Å) <sup>d</sup>	9.8	not applicable	3.9	0.9
total surface area (Å <sup>2</sup> ) <sup>e</sup>	3707	3535	3282	3499
polar area (Å <sup>2</sup> ) <sup>f</sup>	2816	1404	1413	1521
hydrophobic area (Å <sup>2</sup> ) <sup>g</sup>	89	2024	1870	1979
Trp/Tyr28 area (Å <sup>2</sup> ) <sup>h</sup>	82	95	98	84
% Trp/Tyr28 exposed <sup>i</sup>	49	60	63	50

<sup>a</sup>Residues located within an  $\alpha$ -helix. <sup>b</sup>A pseudo energy is an indication of the type of contacts that individual residues have with their microenvironment (negative values are favorable, and positive values are unfavorable). <sup>c</sup>Residues whose backbone geometries are suboptimal as evidenced by a Ramachandran plot. <sup>d</sup>Root-mean-square distance deviation of atoms from the backbone atoms of the *Manduca* structure. <sup>e</sup>Total surface area. <sup>f</sup>Total surface area represented by polar residues (area calculated on the basis of the default setting for polar residues). <sup>g</sup>Total surface area represented by hydrophobic residues (area calculated on the basis of the default setting for hydrophobic residues). <sup>h</sup>Total surface area of Trp28 (*Manduca* and *Bombyx*) or Tyr28 (*Drosophila*) exposed to solvent. <sup>i</sup>Percentage of Trp28 (*Manduca* and *Bombyx*) or Tyr28 (*Drosophila*) surface area accessible to a water molecule.

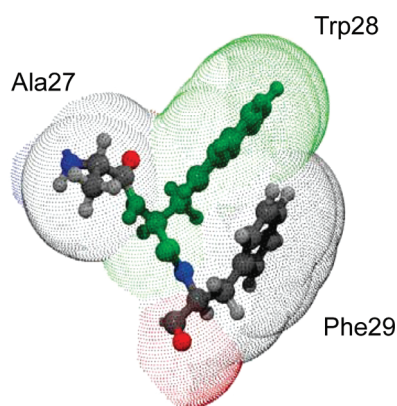


FIGURE 4: Predicted solvent-exposed surface of Trp28. The dots represent the solvent-exposed surfaces of Ala27, Trp28, and Phe29 in *Manduca* EH. The individual residues are depicted in ball-and-stick configuration with Ala27 and Phe29 shown in CPK coloring and Trp28 colored green.

energetically favorable positions by using a combination of molecular dynamics and minimizations. The final EH structure was formed by removing all torsional and distance constraints and minimizing.

Our fully minimized *Manduca* EH model contains disulfide bonds 2 Å in length with torsion angles of approximately +90° or –90°, and helical regions characteristic of  $\alpha$ -helices. Despite the presence of three cis Pro bonds (Pro10, Pro32, and Pro47), the peptide bond geometries are within acceptable ranges. The model is further characterized by a helix–loop–helix–loop–helix motif in which the first and third helices are parallel while the second helix is antiparallel (Figure 3). The structure is 55% helical with  $\alpha$ -helices comprising residues 12–24 (helix I), 35–41 (helix II), and 49–62 (helix III). The discrepancy between the helical domains (11–24, 34–41, and 48–61) initially used to limit the conformational search of the starting structure, and the final helical assignments (12–24, 35–41, and 49–62), is the result of the minimization processes used to identify the lowest-energy conformation. The predicted structure is energetically favorable

(Table 1) with reasonable strain and pseudo energies (i.e., values indicative of the quality of the structure and the types of contacts that individual residues have with their microenvironment). A Ramachandran plot indicates two residues (Ala27 and Glu48) whose backbone geometries are not optimal. Inclusion of water in the simulations was sufficient to nudge Ala27 into an allowed region but did not change the geometry of Glu48. In addition, our model predicts that Trp28 and Phe29, both located in the first loop connecting helices I and II, are located at the peptide surface with the Trp28 indole moiety 60% solvent-exposed and partially occluded by the phenyl group of Phe29 (Figure 4).

Although the solvent-exposed nature of Trp28 and Phe29 was present in the preliminary computational studies of the *Manduca* EH (21), considerable differences (summarized in Table 1) were observed between our model and the ending structure from the pilot study with the preliminary model having a shorter helix I, helix III broken into two shorter helices, and more residues within the disallowed regions of a Ramachandran plot. The differences between the two conformations likely reflect a limitation in the methodology used in the pilot study to identify the lowest-energy conformation following disulfide formation.

The quality of the *Manduca* EH structure was also assessed by comparing the structural features of the model with the solution structures of insulin (PDB entry 1HLS) (32), a two-chain polypeptide composed of two helices joined by three disulfides,  $\alpha$ -cobratoxin (PDB entry 2CTX) from *Naja siamensis* (33), a 62-amino acid peptide with secondary structure elements cross-linked by five internal disulfides (33, 34) and molt-inhibiting hormone (PDB entry 1J0T) (35), a 77-amino acid peptide from the Kuruma prawn (*Marsupenaeus japonicus*) composed of five helices and three disulfides. The average pseudo energy of the *Manduca* EH model (–0.04 kT) was comparable to those of the three known structures (–0.11, –0.09, and –0.08 kT, respectively). Furthermore,  $\alpha$ -cobratoxin and molt-inhibiting hormone both contain sterically hindered residues that fall within disallowed regions of a Ramachandran plot as well as cis-oriented Pro peptide bonds, two features present in our EH model.

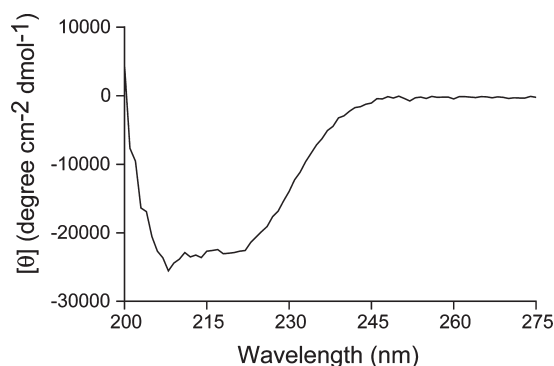


FIGURE 5: CD spectra of synthetic *Manduca* EH. Data were recorded at room temperature in 10 mM sodium phosphate buffer at pH 6.7 with 2  $\mu$ M synthetic *Manduca* EH.

**Peptide Synthesis.** Our model of the putative *Manduca* EH solution structure contains a number of features that are experimentally verifiable, including the predicted secondary structure (55% helix) and the solvent-exposed nature of Trp28 and Phe29. Consequently, to assess the structural validity of these features, we synthesized and folded the 62-amino acid *Manduca* EH. Using Fmoc synthesis protocols with extended coupling and deprotection times (28), we obtained 25 mg of purified synthetic *Manduca* EH. Prior to structural characterization, the biological activity of the synthetic peptide was assayed to verify that it had assumed the correct conformation. Previously, using a different source of synthetic EH, several different folded disulfide isomers were assayed, and only one folded conformer was found to be biologically active.<sup>4</sup>

**Circular Dichroism.** To assess the secondary structure content of the bioactive synthetic *Manduca* EH, CD spectra were recorded in 20 mM sodium phosphate adjusted to pH 6.7 (near hemolymph pH) (36). The CD trace exhibited minima at 208 and 222 nm (Figure 5) characteristic of  $\alpha$ -helices. By averaging the results from the suite of programs available in DICROPROT 2000 (29) for the deconvolution of CD spectra, we obtained a secondary structural content estimation of  $\sim$ 57% helix, which compares quite favorably with that predicted by our model (Table 1).

**Hydrogen–Deuterium Exchange.** An inventory of *Manduca* EH lists 89 total exchangeable hydrogens, of which 56 arise from hydrogens on peptide bonds. The remaining 33 arise from the side chains of Lys, Gln, Asn, Ser, Thr, Trp, Tyr, and the  $\alpha$ -amino group of the peptide. Most side chain hydrogens have undetectably fast exchange rates. To determine the total number of exchangeable hydrogens in *Manduca* EH, we heated non-oxidized EH to 90 °C in 95.8% D<sub>2</sub>O. After 1 min, the mass of the peptide increased 88 Da (corrected for isotopic dilution), indicating that this method is capable of resolving 88 hydrogen–deuterium exchange sites.

Figure 6A shows the progress of incorporation of deuterium into *Manduca* EH at pH 6.7. Attempts to fit the kinetics to three classes of exchange rates using CKS (IBM, Almaden, CA) showed strong systematic deviations from the data at the intermediate times. Resolving the data into two exchanging components was more credible. The class sizes and rate constants are  $54 \pm 6$  hydrogens and  $3 \pm 2 \text{ min}^{-1}$ , and  $34 \pm 16$  hydrogens and  $0.02 \pm 0.01 \text{ min}^{-1}$ , respectively. Although the analysis is

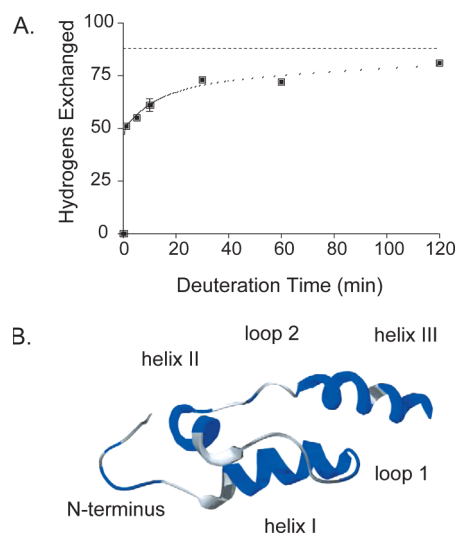


FIGURE 6: Main chain amide protons involved in intramolecular hydrogen bonds. (A) Kinetic profile of hydrogen–deuterium exchange for 40 nM synthetic *Manduca* EH at pH 6.7. Data were obtained following incubation in D<sub>2</sub>O for 1–120 min. The solid line represents the best fit curve for the exchange reaction (goodness of fit of  $R^2 = 0.99$ ), and the dashed line indicates the maximum number of observable hydrogen–deuterium exchange sites. (B) Location of main chain amide protons predicted to be involved in intramolecular hydrogen bonds mapped onto the predicted *Manduca* EH structure and colored blue. Those not involved are colored white.

coarse grained and the uncertainties are large, it is suitable for an overview of *Manduca* EH dynamics. We interpret the slower class as containing  $34 \pm 16$  peptide hydrogens: the faster class contains the remainder of the peptide hydrogens ( $22 \pm 15$ ) and approximately 32 of the 33 side chain hydrogens. Rather than fit all 88 exchangeable hydrogens, one might assume that all side chain hydrogens exchange with a rate sufficiently fast that more than 90% have exchanged prior to the 1 min time point. Calculations based on this assumption do not significantly change the apparent rate constants for direct fits of the data with Origin or use of semilogarithmic plots. Moreover, molecular dynamics simulations suggest several side chains are strongly shielded from water. Therefore, we consider the faster exchanging class to consist of a mixture of peptide and side chain hydrogens. For comparison, a series of insulin derivatives are reported to have rate constants from 0.1 to  $0.005 \text{ min}^{-1}$  at pH 7.5 (37).

Molecular dynamics simulations of *Manduca* EH were conducted in explicit solvent (water, allowing observation of hydrogen bond formation) at constant temperature and pressure at a nominal pH of 7. All peptide hydrogens form internal hydrogen bonds. Formation of hydrogen bonds with water is a comparatively rare event. Hydrogen–deuterium exchange is thought to occur through conformational changes in the peptide which produce alternately intrapeptide hydrogen bonds and peptide–water hydrogen bonds. Gross conformational changes in the peptide were monitored by calculation of the radius of gyration of the backbone atoms of the entire peptide. The radius of gyration oscillated between 14.5 and 15.8 Å with a period of 2.5 ps. Typically, the number of residues in the  $\alpha$ -helical conformation varied from 31 to 34 residues: some transitory turns were observed but no  $\beta$ -structure. It appears that EH has a relatively fast breathing mode that involves little change in its dominant stable secondary structure,  $\alpha$ -helix.

Gross conformational changes can also be monitored by following the energy of interaction (van der Waals, charge–dipole,

<sup>4</sup>D. Zitnan (Institute of Zoology, Slovak Academy of Sciences, Bratislava, Slovak Republic), unpublished data.



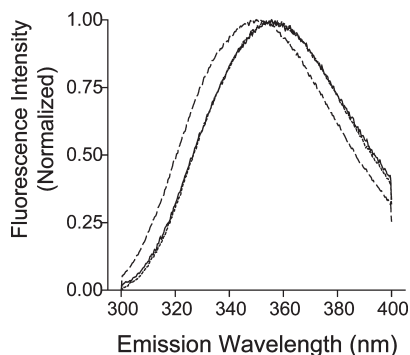


FIGURE 7: Intrinsic fluorescence spectra for 1  $\mu$ M synthetic *Manduca* EH. The fluorescence spectrum was recorded from 300 to 400 nm with excitation at 288 nm. EH (dashed line) exhibited an emission maximum at 351 nm. Both controls, NAWA (solid line) and NAWA/NAYA mixture (1:1, dotted line), exhibited emission maxima at 356 nm.

and dipole–dipole) between the peptide and the solvent water. These oscillated with a period of 20 ps. No correlation was found between the interaction energy and the radius of gyration.

Analyzing the molecular dynamics simulations, we inventoried all exchangeable peptide and side chain hydrogens. Hydrogens were divided into two broad categories: (i) those where hydrogen bonds to water were detected and (ii) those for which no hydrogen bonds to water were detected. Obviously, all must have interacted with water at some point. However, it is reasonable to consider those not detected to represent those hydrogens strongly protected from water and members of the experimentally observed slower class. The simulations show a wide variation in the fraction of time that a particular residue is hydrogen bonded to water. Therefore, the faster class is strongly heterogeneous but unresolvable with the data set presented here. Using this binary classification, 28 hydrogens are assigned to the slower class of exchangeable hydrogens. This is in good agreement with the analysis of the experimental data estimate of  $34 \pm 15$  hydrogens in the slower kinetic class located as shown in Figure 6B.

**Fluorescence Spectroscopy.** Because the intrinsic fluorescence of the tryptophan indole moiety is dependent on the polarity of its local environment, tryptophan fluorescence is commonly used as a means of assessing protein structure. On the basis of our *Manduca* EH model, the Trp28 side chain is predicted to be solvent-exposed (Figure 4). To determine the validity of this prediction, we compared the fluorescence emission profile of synthetic *Manduca* EH (1  $\mu$ M) with the emission profile of NAWA, a soluble Trp analogue generally considered to be a model for a fully solvent-exposed indole side chain (Figure 7). The emission maximum of *Manduca* EH (351 nm) is consistent with Trp in a polar environment (30). The 5 nm blue shift when compared to the emission maximum of NAWA (356 nm) indicates that while Trp28 is solvent-exposed, the local environment surrounding the indole group is less polar than that of NAWA. To eliminate the possibility of instrument artifact as an explanation of the blue shift, we also measured an equimolar ratio of NAWA and NAYA (the same ratio of Trp and Tyr in EH). The near superimposition of the NAWA and NAWA/NAYA spectra indicates the blue shift represents an environmental perturbation of Trp28 (Figure 7).

To further examine the spatial orientation of Trp28, samples were titrated with the neutral quenching reagent acrylamide. The fluorescence intensity of *Manduca* EH decreased as a function of

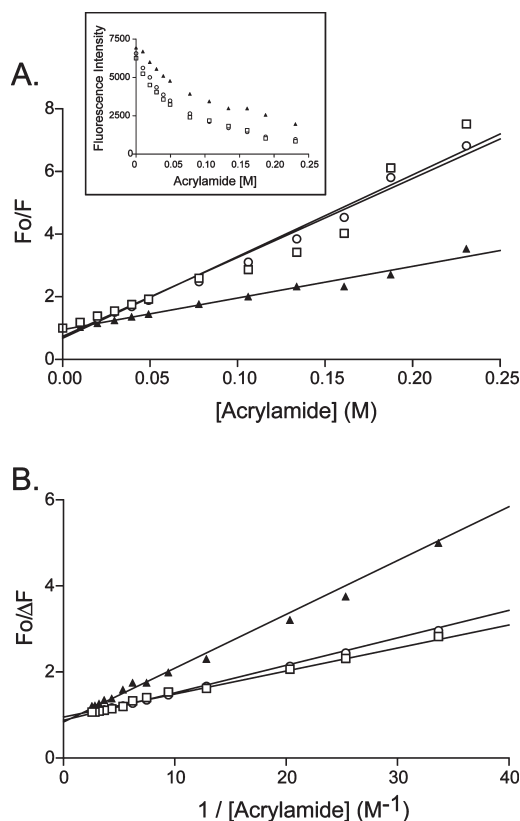


FIGURE 8: Fluorescence quenching by acrylamide. (A) Stern–Volmer plot. The calculated  $K_{sv}$  constants were  $10.1 \pm 0.5 \text{ M}^{-1}$  for *Manduca* EH,  $26.1 \pm 2.1 \text{ M}^{-1}$  for NAWA, and  $25.2 \pm 1.0 \text{ M}^{-1}$  for the NAWA/NAYA mixture. The inset shows the decrease in fluorescence intensity of each as a function of increasing acrylamide concentration. (B) Lehrer plot. All three samples exhibited similar  $y$ -intercepts of near 1, indicating that the Trp28 indole moiety does not oscillate between buried and exposed states. The fluorescence spectrum was recorded at 350 nm with excitation at 295 nm. Data were corrected for buffer effect, acrylamide  $\text{Abs}_{295}$ , and the dilution. Results shown are means  $\pm$  the standard error of the mean of triplicate values and are from a single experiment representative of three independent experiments. Error bars are not visible as they are smaller than the size of the symbols: NAWA ( $\square$ ), NAWA/NAYA mixture ( $\circ$ ), and *Manduca* EH ( $\blacktriangle$ ).

increasing acrylamide concentration (Figure 8A, inset), and as indicated by the linear Stern–Volmer plot, this quenching process was strictly collisional (Figure 8A). At high acrylamide concentrations, NAWA exhibited upward deviations from linearity, demonstrating that the quenching shifted from collisional to static. The absence of this deviation in *Manduca* EH is consistent with partial occlusion of Trp28 by the phenyl group of Phe29 (Figure 4). The calculated Stern–Volmer constant ( $K_{sv}$ ) for *Manduca* EH ( $10.1 \pm 0.5 \text{ M}^{-1}$ ) was  $\sim 2.5$ -fold lower than that of NAWA ( $26.1 \pm 0.5 \text{ M}^{-1}$ ) or the NAWA/NAYA mixture ( $25.2 \pm 1.0 \text{ M}^{-1}$ ), indicating that Trp28 is partially shielded from solvent. Assuming NAWA to be a fully exposed Trp side chain, the solvent exposure of Trp28 was calculated as  $\sim 48\%$ . Computational analysis of the topology of our model predicted 60% exposure, but this value was based on the area exposed to water, not acrylamide. Because of their small diameter, water molecules are able to contact a greater percentage of the peptide surface than the much larger acrylamide molecules which cannot contact the same area.

To determine if the Trp28 functional group oscillates between buried and exposed states, the fluorescence data were analyzed

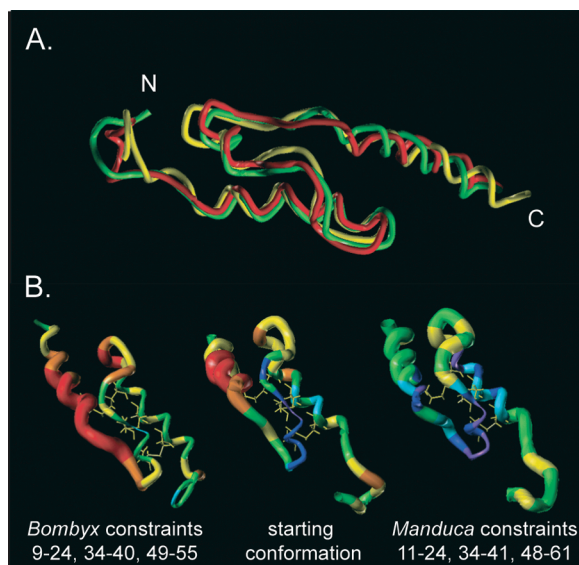


FIGURE 9: Evaluation of differing helical assignments. (A) Alignment of *Manduca* EH conformations after minimization in the presence of differing helical constraints. Snapshot of the ending *Manduca* EH molecules generated from a 20 ps molecular dynamics simulation performed in the presence of weak  $\phi$  and  $\psi$  constraints on residues 11–24, 34–41, and 48–61 (our predicted helical assignments; green) or residues 9–24, 34–40, and 49–55 (the reported *Bombyx* helical assignments; yellow). An unconstrained *Manduca* EH molecule is colored red. (B) Pseudo energies mapped onto tube representations of *Manduca* EH conformations after minimization in the presence of differing helical constraints: (middle) unconstrained peptide, (left) peptide in the presence of strong  $\phi$  and  $\psi$  constraints (residues 9–24, 34–40 and 49–55 from the reported *Bombyx* helices), and (right) peptide with strong  $\phi$  and  $\psi$  constraints on residues 11–24, 34–41, and 48–61 from our initial predicted helical assignments. The coloring and size of the tube representations have been scaled to indicate the location of favorable (violet > blue, small diameter) and unfavorable (red > orange, large diameter) pseudo energies. The cysteines are shown in stick format.

according to the double-reciprocal Lehrer relationship (38), which gives information about the fraction of fluorophore accessible to the quenching agent. As shown in the Lehrer plot (Figure 8B), all three samples exhibited a  $y$ -intercept of  $\sim 1$ , indicating that virtually all of the Trp28 fluorescence was accessible to quencher and that the Trp28 side chain likely does not oscillate between two different states.

**Modeling of a *Manduca* EH Analogue.** The quality of our proposed *Manduca* EH model was further assessed by examining its ability to rationalize known structure–activity relationships. Using our EH conformation as a template, we modeled the EH analogue synthesized by Wang et al. (31). The analogue, which is 20-fold less active than native *Manduca* EH, contains a Q20N amino acid substitution, as well as the unnatural substitutions of norleucine for Met11 and Met24 and a naphthylalanine substituted for Trp28. Structural analyses indicate that the substitutions had minimal conformational effects; both the strain energy and helical content of the analogue were comparable to those of the original EH structure (data not shown). Because of the latter substitutions, the EH analogue exhibits greater overall hydrophobic character.

**Comparison with *Bombyx* and *Drosophila* EH Structures.** Nervous system extracts of insects from a variety of orders contain a factor that stimulates *Manduca* pupal ecdysis activity (2). Given the high degree of sequence identity (81%) between *Manduca* EH and *Bombyx* EH, the cross-reactivity observed in

bioassays suggests that a common conformation exists. A hypothetical *Bombyx* EH structure reported by Fujita et al. (18) differs from our *Manduca* EH model in terms of backbone trajectory, intramolecular contacts, and helical assignments (*Manduca* residues 12–24, 35–41, and 49–62 vs *Bombyx* residues 9–24, 34–40, and 49–55), with our model containing a significantly longer helix at the C-terminus. If the dissimilarities were simply an artifact of the different algorithms used to generate the respective models, then *Manduca* EH should be able to assume the conformation reported by Fujita et al. (18). To test this, we performed a 20 ps dynamics simulation in the presence of weak  $\phi$  and  $\psi$  constraints and compared the backbone atom trajectories of three *Manduca* EH conformations incorporating different helical settings: (i) a *Manduca* EH conformation with no constraints, (ii) a *Manduca* EH conformation restrained with our initial helical assignments (residues 11–24, 34–41, and 48–61), and (iii) a *Manduca* EH conformation constrained with the reported *Bombyx* helical assignments (residues 9–24, 34–40, and 49–55). Alignment of the structures from the simulation (Figure 9A) revealed that the backbones of the conformation constrained with the *Bombyx* helical assignments and with the initial assignments were not significantly different. To determine whether this failure to assume a conformation similar to the reported *Bombyx* structure was the result of a local energy well or barrier, a more rigorous approach was used that incorporated stronger constraints on the  $\phi$  and  $\psi$  angles. Using the three *Manduca* EH conformations incorporating the helical settings described above (i.e., no constraints, our initial helical assignments, or the reported *Bombyx* helical assignments), successive rounds of molecular dynamics and minimizations were performed with strong  $\phi$  and  $\psi$  constraints on two of the three conformations, which maintained the assigned helical segments throughout the simulations. A comparison of the pseudo energies from these simulations, depicted by the tube representations in Figure 9B (colored to reflect the distribution of the pseudo energies along the length of the peptide such that violet > indigo > blue > green > yellow > orange > red), reveals that *Manduca* EH constrained with the reported *Bombyx* helical assignments (Figure 9B, left panel) is the least energetically favorable of the three. These simulations indicate that while *Manduca* EH can be forced to assume the structural elements of the *Bombyx* model (Figure 9B, left panel), the helical assignments of our model clearly represent a more energetically favorable conformation. The published *Bombyx* model also predicts that residues Phe58 and Phe25 (Leu25 in *Manduca*) are separated by an intramolecular distance of 4 Å, a feature which Fujita et al. considered to be evidence of a hydrophobic interaction that stabilizes the peptide conformation (18). A search of residues within 4 Å of Phe58 in our *Manduca* EH model indicates that Leu25 is more distant ( $\sim 14$  Å).

The differences between the published *Bombyx* conformation and the *Manduca* conformation presented here may be (i) the result of the 12-amino acid difference in the two primary structures or (ii) the result of the two different prediction algorithms used. To test the effect of primary structure, we forced the *Manduca* EH sequence to adopt the published *Bombyx* conformation (see paragraph above). Next we coerced the *Bombyx* sequence to adopt the *Manduca* conformation. The amino acid sequence of the *Manduca* model was mutated to that of *Bombyx*. The resulting three-dimensional structure was minimized by two different routes. In one route, the mutated structure was minimized directly without any



constraints. Thus, the conformation fell into the nearest local conformational minimum. In the other route, torsional constraints were applied to the  $\phi$  and  $\psi$  angles according to the published *Bombyx* assignments (residues 9–24, 34–40, and 49–55). Weak force constraints (0.1 kJ/mol) were used to induce the *Bombyx* conformation without removing the possibility of finding other conformational minima. After convergence of the minimizer, the torsion constraints were removed and the structure was reminimized. The molecule constrained with the Fujita et al. helical assignments more closely resembled their structure than our *Manduca* structure. The reminimized structure lacking the helical constraints was closer to our *Manduca* structure. This structure had the lowest-energy conformation produced by these two paths (strain energy of  $-1142$  kcal/mol) and was characterized by  $\alpha$ -helices encompassing residues 13–24, 34–41 and 49–62. These results suggest that computational differences are responsible for the conformational differences between the *Manduca* conformation presented here and the *Bombyx* conformation presented by Fujita et al. (18). Further, we suggest that both *Manduca* EH and *Bombyx* EH have similar conformations.

We next tested if *Drosophila* EH could likewise adopt an energetically favorable conformation similar to that of *Manduca* EH. The *Drosophila* EH gene is predicted to encode a precursor which, if cleaved by the signal peptidase, would generate a peptide with a 10-residue N-terminal extension and one extra residue at the C-terminus compared with *Manduca* and *Bombyx* EH (39). However, this extended N-terminus has a Lys-Arg endoproteolytic processing site that yields a peptide more similar (i.e., 62 residues) to the lepidopteran EHs. Consequently, we modeled the 62-amino acid *Drosophila* EH sequence using our *Manduca* EH conformation as a template and weak constraints on the  $\phi$  and  $\psi$  angles corresponding to our initial helical assignments (i.e., residues 11–24, 34–41, and 48–61). After optimizing the side chain geometries and minimizing in the absence of the helical constraints, we found that *Drosophila* EH adopted a conformation similar to our model of *Manduca* EH, indicating that despite dissimilar sequences, EHs are capable of assuming a shared conformation.

The structures derived for *Manduca* EH, *Bombyx* EH, and *Drosophila* EH have a number of similar characteristics, including a helix–loop–helix–loop–helix motif with the first and third helices parallel, the presence of a smaller, antiparallel second helix, and similar loop regions. Alignment of the backbone atoms of the *Bombyx* and *Drosophila* structures with the *Manduca* structure results in root-mean-square deviations of 3.9 and 0.9 Å, respectively (Table 1). Analysis of the strain energies indicates that the *Bombyx* structure is 84 kcal/mol lower in energy than the *Manduca* model, whereas the calculated energy for the *Drosophila* structure is moderately higher. Further structural analyses of the three models (Table 1) indicated that the pseudo energies among the three models vary only slightly, whereas the degree of helicity is identical in *Manduca* and *Bombyx*, but somewhat higher in *Drosophila*. The EH structures are stabilized by as many as 122 hydrogen bonds (*Bombyx*) or as few as 105 hydrogen bonds (*Drosophila*). The conformations of the three EHs (i) have cis Pro bonds, (ii) contain residues whose backbone geometries are less than optimal, (iii) have comparable solvent-exposed surfaces, and (iv) are similar in size and shape.

Furthermore, the solvent-exposed nature of Trp28 is conserved with the residue estimated to be 63% exposed in *Bombyx*,

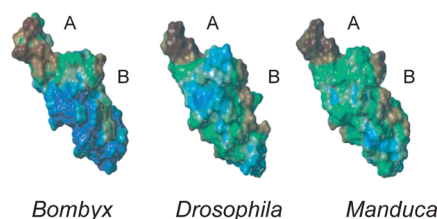


FIGURE 10: Distribution of hydrophobic potentials. View of the hydrophobic potentials at the solvent accessible Connolly surfaces (radius of 1.4 Å) of *Manduca* EH (right), *Drosophila* EH (middle), and *Bombyx* EH (left). Hydrophobicity scaling has been normalized to the lowest and highest global values and decreases from brown (hydrophobic) to blue (polar), with green being indicative of intermediate potentials. Regions that exhibit similarity among the three peptides have been labeled: (A) the hydrophobic residues clustered at the C-terminus and (B) the loop region containing the solvent-exposed Trp28 (Tyr in *Drosophila*) and Phe29.

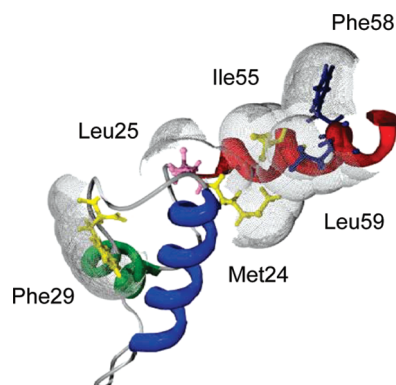


FIGURE 11: Location of some residues essential for biological activity. The predicted solution structure of *Manduca* EH with helices colored blue (helix I), dark green (helix II), and red (helix III). Residues (shown in ball-and-stick format) previously purported to interact with the receptor are colored yellow (Met24, Phe29, and Ile55). Phe58 and Leu59 (both colored blue) are solvent-exposed and crucial for receptor interaction (our data). Leu25 (colored pink) is  $> 14$  Å from Phe58. The solvent accessible surfaces of the individual residues are represented by the light gray dots. This figure was rendered using SwissPdb Viewer (57) and POVray3.5.

whereas in *Drosophila*, which has the conservative W28Y substitution, the smaller Tyr28 residue is predicted to be 50% solvent-exposed. Taken together, these results indicate that, despite differences in primary sequence, all three peptides can assume a similar conformation.

**Regions of Similarity.** Given the reported interspecies cross-reactivity, the models were examined for regions of similarity that could be involved in receptor binding and/or activation. Coloring the solvent accessible surface area of the EHs according to their lipophilic potential reveals (i) the *Bombyx* structure exhibits greater polar character (indicated by blue coloring; see Figure 10) compared to the *Manduca* or *Drosophila* structures and (ii) there are two hydrophobic regions corresponding to the C-terminus (labeled A in Figure 10) and the first loop (labeled B in Figure 10) present in each. No appreciable differences in either electrostatic or hydrogen bonding potentials were observed among the three peptides (data not shown). Interestingly, the third helix in all three EHs encompasses the terminal residues of the peptide, a region that is essential for biological activity (14, 17).

**Structure–Function Analyses of Point Mutations.** Previous studies (18, 19) reported the effects of 29 Gly point mutations on the potency of *Bombyx* EH, which demonstrated

Table 2: Computational Analysis of I55G, F58G, and L59G Point Mutations in the EH Solution Structure<sup>a</sup>

	wild type	I55G	F58G	L59G
side chain pseudo energy (kJ)				
Ile55	0.76	—	0.76	0.81
Phe58	1.43	1.43	—	1.43
Leu59	1.01	1.06	1.01	—
fractional area side chain solvent exposed				
Ile55	0.42	—	0.37	0.47
Phe58	0.76	0.78	—	0.81
Leu59	0.71	0.77	0.74	—
interaction energy (side chain) <sup>b</sup> (kcal/mol)				
Ile55	0.86	—	0.76	0.96
Phe58	1.59	1.59	—	1.7
Leu59	1.25	1.36	1.3	—
radius of gyration (global) (Å)	14.2	13.8	13.6	14
residues with $\alpha$ -helix $\phi$ and $\psi$ angles before and after being heated to 300 K <sup>c</sup>	39/35	40/26	40/25	40/28
interaction energy (global) <sup>d</sup> (kcal/mol)	−1640	−1357	−1293	−1530

<sup>a</sup>Minimizations performed in the presence of explicit solvent (i.e., H<sub>2</sub>O). <sup>b</sup>Energy of side chain interaction with water. <sup>c</sup>The solvated peptide system was heated to equilibrium at 300 K and then returned to the starting temperature of 0 K. Helical residues were defined as those with  $\phi$  and  $\psi$  angles within 15° ( $\pm 15^\circ$ ) of canonical  $\alpha$ -helix values (−57° for  $\phi$  and −47° for  $\psi$ ). <sup>d</sup>Energy of interaction between the entire EH molecule and water.

the particular importance of Met24, Phe25, Phe29, Ile55, Phe58, and Leu59. On the basis of those studies, Phe25 (Leu in *Manduca*), Phe58, and Leu59 were thought to play important roles in stabilizing the EH globular structure, Phe29 and Ile55 were thought to be necessary for receptor interactions, and Met24 was thought to play a role in both functions. Inspection of the solvent accessible surface of our *Manduca* model indicates that the residues previously hypothesized (18) to interact with the EH receptor (Met24, Phe29, and Ile55) are indeed solvent-exposed (colored yellow in Figure 11). However, the completely conserved residues Phe58 and Leu59 are also solvent-exposed in our *Manduca* EH model (colored blue in Figure 11). Because the intermolecular distance between residues Leu25–Phe58 and Phe29–Leu59 is large (> 14 Å) in our model, the hydrophobic interaction between Leu25 and Phe58 and the proximity of Phe29 to Leu59 predicted by Fujita et al. (18) are impossible. This interaction requires a flexible C-terminus with residues 56–62 lacking secondary structure. In our model, these residues constitute the terminal portion of the third helix; as a result, the C-terminus does not have the flexibility required to permit the interaction. The extended third helix is also present in our *Bombyx* and *Drosophila* models, suggesting that it is essential for biological activity. Of the *Bombyx* EH point mutants, the most pronounced effects on EH activity were observed with the I55G, F58G, and L59G mutants, with the L59G mutation having the greatest effect (18); these are all in our predicted C-terminal helix. Because helices are often destabilized by Gly, the decreased activity could be an indication Gly weakens the ordered structure of the helical C-terminus. Consequently, we investigated the roles of Ile55, Phe58, and Leu59 in maintaining the C-terminal helix by modeling the Gly substitutions using our *Manduca* EH structure as a starting point (Table 2). The most pronounced structural effects of the Gly substitutions were observed in the I55G and F58G mutations (Table 2). In both mutants, the C-terminal helix partially collapsed, resulting in more compact EH molecules with smaller radii of gyration and fewer favorable solvent interactions, indicated by the weaker interaction energies (−1357 and −1293 kcal/mol, respectively) compared with that of the nonsubstituted *Manduca* EH structure (−1640 kcal/mol). The L59G mutation likewise destabilized the C-terminal helix, although both the radius of gyration (14 Å) and the

potential interaction energy (−1530 kcal/mol) were closer to those of our nonsubstituted *Manduca* EH structure (14.2 Å and −1640 kcal/mol, respectively). Further clues could be provided by data for I55A, F58A, and L59A mutants, which have minimal side chains but promote rather than break helices. All were reportedly synthesized (18), but biological activity data were reported only for an I55A mutant, which is somewhat more potent than the I55G mutant, but still much less active than the wild-type peptide. This highlights the particular significance of the side chain of Ile55 for activity. Therefore, because they appear to be of key importance, we focused on the solvent-exposed side chains of Ile55 (42%), Phe58 (76%), and Leu59 (71%). In the C-terminal helix encompassing residues 49–62, Ile55 and Leu59 project from the same side of the helix while Phe58 is rotated ~90° from them (see Figure 11). Interestingly, the Ile55 and Leu59 side chains lean toward one another such that the average distance between the two residues decreases from 6.6 Å at the  $\alpha$ -carbons to 4.7 Å at the  $\gamma$ -carbons. This distance, while slightly large for a van der Waals contact, does permit a hydrophobic interaction, which affects the  $\chi_1$  angles (the dihedral axis defined by the  $\alpha$ - and  $\beta$ -carbons) of Ile55 and Leu59, which are  $197 \pm 12^\circ$  and  $186 \pm 12^\circ$ , respectively (molecular dynamics simulation at 300 K). The similarity of these angles reflects the hydrophobic interaction between these side chains. Gly substitution of either side chain disrupts this interaction and results in rotation of the side chain of the nonmutated residue: in the I55G mutant the Leu59  $\chi_1$  angle changes to  $303 \pm 13^\circ$ , while in the L59G mutant the Ile55  $\chi_1$  angle becomes  $289 \pm 12^\circ$ . In addition, the L59G mutant exhibits decreased hydrophobic character (Figure 12A) and a local displacement of steric bulk (Figure 12B). Given these results, an explanation for the pronounced biological effects of the I55G, F58G, and L59G mutants could be that the side chains of the Ile55, Phe58, Leu59 triad are all required to optimally bind the receptor site. Consequently, any disruptions to the shape and surface characteristics of the triad following Gly substitution would be expected to impact the physiological response. The particularly low activity introduced with the L59G mutation, with its steric bulk protruding in a different locus and decreased hydrophobic character, is likely important because repulsion may interfere more with binding than the loss of bulk alone.

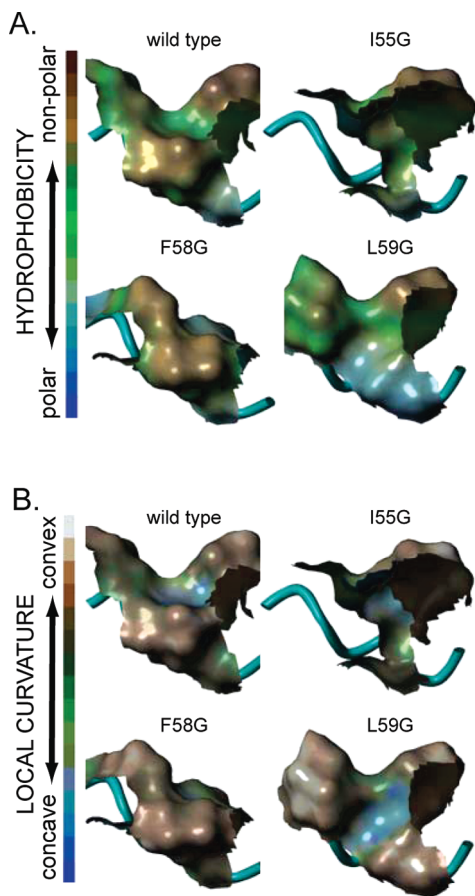


FIGURE 12: Effects of Gly mutations on the solvent accessible surface properties of Ile55, Phe58, and Leu59. (A) Hydrophobic potentials. Hydrophobicity scaling has been normalized to the lowest and highest global values with regions of high hydrophobicity colored brown, polar regions blue, and areas of intermediate potentials colored green. The hydrophobic potential reflects the contribution of the surrounding residues whose surfaces have not been shown for the sake of clarity. (B) Local surface topography. The degree of surface curvature is shown projected onto the solvent accessible surfaces of Ile55, Phe58, and Leu59 with concave regions colored blue, convex regions brown to white, and intermediate regions green. Differences in the surface topography reflect changes in the  $\chi_1$  torsion angles following Gly substitution. Peptides have been aligned to the backbone atoms of residues 55, 58, and 59.

## DISCUSSION

Use of *in silico* methods for predicting the optimal conformations of proteins and peptides is becoming widely accepted and has been applied with greater frequency to develop rational structure-based agonists, antagonists, and inhibitors (40–42). Using a *de novo* modeling method, we generated a plausible solution structure for *Manduca* EH that exhibited a helix–loop–helix–loop–helix motif stabilized by three cystines and characterized by 55% helical content, an unstructured N-terminus, a helical C-terminus, and a solvent-exposed loop containing Trp28 and Phe29 (Figure 3). Two key assumptions were crucial in generating this model: knowledge of the disulfide connectivity in the peptide places considerable constraints on the number of conformations which must be searched, and the approximate location of helical regions predicted by secondary structure algorithms provides additional constraints. In our final model, the predicted location of the three helices (residues 12–24, 35–41, and 49–62) differed somewhat (residues 11–24, 34–41, and 48–61) from the consensus of the algorithms (Figure 2).

We present data to validate the predictions of our model using computational and biophysical approaches. Comparative analysis of the biophysical characteristics of the disulfide and peptide bonds present in our model shows they are comparable with those of known structures. CD analysis (Figure 5) of our bioactive synthetic *Manduca* EH was consistent with the secondary structural features predicted by our model. The predicted solvent-exposed nature of Trp28 and the potential shielding effect arising from the adjacent Phe29 (Figure 4) are supported by the emission spectrum of *Manduca* EH (Figure 7) and a linear Stern–Volmer plot (Figure 8A). Furthermore, the presence of Trp28 at the peptide surface is not the result of natural protein motions (Figure 8B), which suggests that the location of these loop residues is functionally important. Despite the potential destabilizing effect that unburied hydrophobic residues can have on global structure, there are examples where they are necessary for receptor interaction (43, 44). The conserved aromatic nature of positions 28 and 29 in the known EH sequences and the 28-fold decreased biological activity of the F29G *Bombyx* EH point mutation (18) confirm the importance of these residues.

HDX data further support our model with the number of experimentally determined slowly exchanging hydrogens ( $34 \pm 15$ ) in good agreement with our theoretically determined number. Because the hydrogen–deuterium exchange rate of an unhindered peptide bond is influenced by adjacent amino acid side chains, we grouped EH exchange rates into two broad classes and used  $600 \text{ min}^{-1}$  as an average rate constant for an unhindered peptide. Experimentally, for EH the faster class has a protection factor of 200 and the slower class has a protection factor of 30000. In a study of protection factors conferred by conformation changes, Welch and Fasman (45) found protection factors of 50 for simple helix–coil transitions and  $8 \times 10^5$  for self-associated helices. Eclosion hormone falls comfortably within this range and is consistent with the observation that hydrogen exchange rates do not correlate with size but do correlate with the stability of intrapeptide hydrogen bonds.

On the basis of calculated interatomic distances (i.e., charged residues separated by  $\leq 3 \text{ \AA}$ ), *Manduca* EH is predicted to be stabilized by four salt bridges (Asp9–Lys42, Glu30–Lys22, Glu36–Lys40, and Glu48–Lys22). In our *Bombyx* structure, the predicted ionic interactions are Asp9–Arg42, Glu30–Lys22, Glu36–Lys40, Glu48–Lys22, Glu48–Lys44, and Glu50–Lys44 interactions, while in our *Drosophila* EH model, only two salt bridges, Asp9–Lys42 and Asp48–Lys22, are present. Interestingly, these two salt bridges are conserved in all three species, suggesting that these interactions are likewise critical for maintaining the EH structure. Three of the salt bridges (Glu30–Lys22, Glu48–Lys22, and Asp9–Lys42) are situated within five residues of the disulfide bonds, suggesting that they may act in the early stages of peptide folding to aid the spatial alignment and pairing of the Cys residues. The Glu30–Lys22 interaction may assist in the alignment of the Cys18–Cys34 disulfide, the Glu48–Lys22 bond in aligning the Cys21–Cys49 disulfide, and the Asp9–Lys42 bond in aligning the Cys14–Cys38 disulfide. The fourth salt bridge, Glu36–Lys40, likely stabilizes helix II (residues 35–41). The suboptimal orientation of Glu48 in *Manduca* EH, as evaluated by a Ramachandran plot, is likely a result of the predicted salt bridge with Lys22. It is known that proteins tolerate the nonoptimal stereochemical orientation of side chains in exchange for stabilizing the tertiary structure (46). Inspection of the EH sequence alignment (Figure 1) indicates that Glu48 is highly conserved



and thus is likely a key residue involved in conformational stabilization.

Our model of *Manduca* EH contains cis Pro peptide bonds. While side chain steric hindrance usually precludes peptide bonds from assuming the cis position, the cyclic nature of Pro allows for this bond configuration. The activities of a number of proteins and peptides are dependent on the presence of cis Pro bonds (47, 48), which can be found in 10% of proteins and peptides (49). Of the five Pro residues in *Manduca* EH, Pro47 and Pro57 are invariant in all EHs (Figure 1). Gly substitution of Pro47 resulted in a 34-fold decrease in activity in *Bombyx* EH which was attributed to Gly-promoted entropic destabilization (18), suggesting that the increased number of possible conformations in the Gly analogue decreases the effective concentration of peptide with a C-terminus oriented correctly for receptor interaction. This contrasts with the native peptide in which the limited conformational space of the Pro residue results in fewer possible conformations. Substitution of Pro47 with the helix promoter Ala gives an analogue somewhat less active than Gly47 EH (18); this mutation might induce extension of the adjacent C-terminal helix (residues 49–62), concomitantly distorting the conformation. Thus, Pro47 may function as a major conformational determinant in EHs.

A theoretical model of a less potent substituted *Manduca* EH analogue (31) was generated to further test the quality of our model. Wang et al. (31) hypothesized that incorporation of the four residue changes (Met11, Gln20, Met24, and Trp28) into the EH analogue would enhance overall solution stability. While the amino acid substitutions had little effect on the overall conformation, they did promote a net increase in the hydrophobic character of one face of the conformation. The authors attribute the approximate 20-fold reduced activity of the synthetic analogue in a functional assay to this increased hydrophobicity, or “to heterogeneity in conformation” (31), perhaps implying multiple disulfide isomers in their preparation. Alternatively, the naphthylalanine residue, larger and more hydrophobic than Trp, may not fit well in the binding site. Aside from the increased hydrophobicity, the regions thought to be important for activity (i.e., the C-terminus and the first loop) are very similar to those of our *Manduca* EH model.

In silico generation of the *Bombyx* and *Drosophila* EH structures indicated that our predicted EH conformation is not species-dependent, thus providing a structural basis for the reported biological cross-reactivity of EHs (2). The effects of the differing primary sequences are apparent when the physicochemical properties of the three structures are evaluated (Table 1). The surface lipophilic character of the three structures (Figure 10) supports the idea that the C-terminus and the first loop containing Trp28 and Phe29 are specific regions that may interact with the receptor through hydrophobic interactions (18); both positions are expected to be at the solvent interface in all three structures, but with varying degrees of accessibility.

The conformational differences between our *Bombyx* model and that of Fujita et al. (18) likely reflect one of their three central hypotheses, that Ile55 is exposed to solvent. In our model, Ile55 is less solvent exposed than Phe58 and Leu59, which are in a helix in our model but unstructured in theirs. Their third helix is limited to residues 49–55, whereas in our model, the third helix includes residues 49–62 in all three species. The conserved nature of the helix suggests that this structural element is necessary for biological activity. Kono et al. (14) reported that removal of the first six N-terminal residues (an unordered region in our

model) had little effect on activity, while cleavage of residues from the hydrophobic C-terminus dramatically reduced activity. Similar results were seen with Gly substitution of Met24, Phe25, Phe29, Lys40, Pro47, Glu50, Ile55, Phe58, and Leu59 (18, 19). The conservatively substituted hydrophobic residue at position 25 (Phe or Tyr in more than 60% of identified sequences) was postulated to have hydrophobic interactions with Phe58 (18), but in our model, the side chains of these two residues fail to meet the distance criteria for such an interaction (Figure 11). Met24, Phe29, Ile55, Phe58, and Leu59 are all solvent-exposed (Figure 11), which suggests that they are involved in receptor binding and/or activation. While it is currently unknown which residue(s) contacts the binding domain, our data (Figures 11 and 12 and Table 2) coupled with the extreme reduction in biological activity (>1000-fold less potent) of the L59G mutant suggest that Leu59 is the principle receptor interacting residue of a triad including Ile55 and Phe58. Removal of any side chain from this triad decreases the hydrophobicity and perturbs the steric bulk (Figure 12). Also, molecular dynamics simulation of a Gly mutation within the triad produced general, rather than local, perturbations in the structure, with many residues moving away from their  $\alpha$ -helix conformation (Table 2). Peptide–water interactions exhibited large changes compared with that expected from a single side chain mutation. Consequently, not modeling the EH C-terminus as a helix would lead to drastically different orientations of the critical C-terminal residues (i.e., Ile55, Phe58, and Leu59) and different conclusions regarding the geometry of the receptor binding domain. It should be noted that these three residues are highly conserved in all sequences and thus likely important for receptor activation in other species as well.

A number of peptide toxins (reviewed in refs 50 and 51) and growth factors (52, 53) also feature secondary structural elements linked by multiple disulfide bridges. A BLAST query using the *Manduca* EH sequence, however, indicates that little homology exists. Moreover, protein threading algorithms working from the solution structures of a number of the toxins and growth factors fail to generate promising structural templates. Indeed, early attempts at using  $\gamma$ 1-hordothionin (a plant toxin derived from barley endosperm) as a template for *Manduca* EH were unsuccessful (21). Despite some similarities, the peptide toxins and growth factors failed as structural templates for a number of reasons. (i) The essential Cys residues have incorrect disulfide pairings. By labeling the six Cys residues involved in disulfide formation (C1–C6) in their order of appearance in the primary sequence from the N- to C-terminus, we found the toxins and growth factors exhibit a C1–C4, C2–C5, C3–C6 arrangement. EH exhibits a C1–C5, C2–C4, C3–C6 arrangement. (ii) The toxins are primarily composed of  $\beta$ -sheets, whereas EH is essentially  $\alpha$ -helical. (iii) A characteristic feature of many toxins that is noticeably absent in EH is the presence of the cystine knot motif in which a ring formed from two disulfide bonds joining parallel backbone segments is penetrated by a third disulfide bond (54). Despite these differences, the peptides are 40–70 amino acids in length, containing multiple disulfides that function to stabilize secondary structural elements. From a teleological perspective, it is possible that these structures, which must all pass through the hemolymph or blood, evolved as stable, compact structures to avoid enzymatic degradation prior to reaching their target site.

In summary, using an energy-based de novo approach that exploits the restricted conformational space of disulfide-linked helices, we generated an energetically stable three-dimensional

structure for *Manduca* EH characterized by two conserved hydrophobic regions, one encompassing the C-terminus and the other the loop linking the first two helical segments that exhibits a number of features, including a solvent-exposed Trp28, which were verified by biophysical methods. Taken together, the evidence suggests that our model for *Manduca* EH approximates the solution structure because it is energetically favorable, can be assumed by other EHs, and is consistent with known structure–activity relationships. By using this structure as a reference point, a computational search of compounds that mimic the geometry of the conserved functional groups can be used to facilitate the design of peptide antagonists, or possibly development of small molecule agonists or antagonists. Before these processes can begin in earnest, further structure–function studies would be useful for clarifying the role of specific residues responsible for receptor binding and activation of the secondary message cascade.

## ACKNOWLEDGMENT

We are grateful to Dr. David S. King (Howard Hughes Medical Institute, University of California) for supplying the peptide folding conditions and for assistance with the CD study. We are also indebted to Dr. Dusan Zitnan (Institute of Zoology, Slovak Academy of Sciences) for confirming the biological activity of our synthetic *Manduca* EH and to Nami Oguchi and Alaine Garrett for technical assistance in purifying the synthetic peptide.

## REFERENCES

- Truman, J. W. (2005) Hormonal control of insect ecdysis: Endocrine cascades for coordinating behavior with physiology. *Vitam. Horm.* 73, 1–30.
- Truman, J. W., Taghert, P. H., Copenhaver, P. F., Tublitz, N. J., and Schwartz, L. M. (1981) Ecdysis hormone may control all ecdysis in insects. *Nature* 291, 70–71.
- Truman, J. W. (1992) The ecdysis hormone system of insects. *Prog. Brain Res.* 92, 361–374.
- Baker, J. D., McNabb, S. L., and Truman, J. W. (1999) The hormonal coordination of behavior and physiology at adult ecdysis in *Drosophila melanogaster*. *J. Exp. Biol.* 202, 3037–3048.
- McNabb, S. L., Baker, J. D., Agapite, J., Steller, H., Riddiford, L. M., and Truman, J. W. (1997) Disruption of a behavioral sequence by targeted death of peptidergic neurons in *Drosophila*. *Neuron* 19, 813–823.
- Kataoka, H., Troetschler, R. G., Kramer, S. J., Cesarin, B. J., and Schooley, D. A. (1987) Isolation and primary structure of the ecdysis hormone of the tobacco hornworm, *Manduca sexta*. *Biochem. Biophys. Res. Commun.* 146, 746–750.
- Marti, T., Takio, K., Walsh, K. A., Terzi, G., and Truman, J. W. (1987) Microanalysis of the amino acid sequence of the ecdysis hormone from the tobacco hornworm *Manduca sexta*. *FEBS Lett.* 219, 415–418.
- Kono, T., Nagasawa, H., Isogai, A., Fugo, H., and Suzuki, A. (1987) Amino acid sequence of ecdysis hormone of the silkworm, *Bombyx mori*. *Agric. Biol. Chem.* 51, 2307–2308.
- Riehle, M. A., Garczynski, S. F., Crim, J. W., Hill, C. A., and Brown, M. R. (2002) Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* 298, 172–175.
- Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P., Couronne, O., Hua, S., Smith, M. A., Zhang, P., Liu, J., Bussemaker, H. J., van Batenburg, M. F., Howells, S. L., Scherer, S. E., Sodergren, E., Matthews, B. B., Crosby, M. A., Schroeder, A. J., Ortiz-Barrientos, D., Rives, C. M., Metzker, M. L., Muzny, D. M., Scott, G., Steffen, D., Wheeler, D. A., Worley, K. C., Havlak, P., Durbin, K. J., Egan, A., Gill, R., Hume, J., Morgan, M. B., Miner, G., Hamilton, C., Huang, Y., Waldron, L., Verduzco, D., Clerc-Blankenburg, K. P., Dubchak, I., Noor, M. A., Anderson, W., White, K. P., Clark, A. G., Schaeffer, S. W., Gelbart, W., Weinstock, G. M., and Gibbs, R. A. (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. *Genome Res.* 15, 1–18.
- Horodyski, F. M. (1996) Neuroendocrine control of insect ecdysis by ecdysis hormone. *J. Insect Physiol.* 42, 917–924.
- Zhang, M., and Xu, W. H. (2006) Isolation of an ecdysis hormone gene from the cotton bollworm, *Helicoverpa armigera*: Temporal and spatial distribution of transcripts. *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.* 143, 351–359.
- Kataoka, H., Li, J. P., Lui, A. S., Kramer, S. J., and Schooley, D. A. (1992) Complete structure of ecdysis hormone of *Manduca sexta*. Assignment of disulfide bond location. *Int. J. Pept. Protein Res.* 39, 29–35.
- Kono, T., Nagasawa, H., Kataoka, H., Isogai, A., Fugo, H., and Suzuki, A. (1990) Ecdysis hormone of the silkworm *Bombyx mori*. Expression in *Escherichia coli* and location of disulfide bonds. *FEBS Lett.* 263, 358–360.
- Nagasawa, H., Fugo, H., Takahashi, S., Kamito, T., Isogai, A., and Suzuki, A. (1983) Purification and properties of ecdysis hormone in the silkworm, *Bombyx mori*. *Agric. Biol. Chem.* 47, 1901–1906.
- Terzi, G., Truman, J. W., and Reynolds, S. E. (1988) Purification and characterization of ecdysis hormone from the moth, *Manduca sexta*. *Insect Biochem. Mol. Biol.* 18, 701–707.
- Hayashi, H., Nakano, M., Shibasaki, Y., and Fujita, N. (1990) Expression of a silkworm ecdysis hormone gene in yeast. *Biochem. Biophys. Res. Commun.* 173, 1065–1071.
- Fujita, N., Maekawa, T., Ohta, S., and Kikuchi, T. (1998) The functional residues and their representation by a hypothetical 3D model of silkworm ecdysis hormone. *Protein Eng.* 11, 769–773.
- Kikuchi, T., Okamoto, M., Geiser, M., Schmitz, A., Gohda, K., Takai, M., Morita, T., Horii, K., and Fujita, N. (1997) Prediction of the biologically active sites in ecdysis hormone from the silkworm, *Bombyx mori*. *Protein Eng.* 10, 217–222.
- Nosaka, A. Y., Kanaori, K., Umemura, I., Takai, M., and Fujita, N. (1998) Structural study on silkworm ecdysis hormone fragment (1–34) in solution by proton nuclear magnetic resonance spectroscopy. *Bioorg. Med. Chem.* 6, 465–472.
- Copley, K. (1997) Three dimensional structure of insect diuretic hormones. Ph.D. Thesis, pp 60–89, University of Nevada, Reno, NV.
- Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120, 97–120.
- Maxfield, F. R., and Scheraga, H. A. (1976) Status of empirical methods for the prediction of protein backbone topography. *Biochemistry* 15, 5138–5153.
- Qian, N., and Sejnowski, T. J. (1988) Predicting the secondary structure of globular proteins using neural network models. *J. Mol. Biol.* 202, 865–884.
- Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000) NPS@: Network protein sequence analysis. *Trends Biochem. Sci.* 25, 147–150.
- King, D. S. (1996) Synthesis and deprotection of large peptides. *ABRF News* 7, 26–28.
- Furuya, K., Schegg, K. M., and Schooley, D. A. (1998) Isolation and identification of a second diuretic hormone from *Tenebrio molitor*. *Peptides* 19, 619–626.
- King, D. S., Fields, C. G., and Fields, G. B. (1990) A cleavage method which minimizes side reactions following FMOC solid phase peptide synthesis. *Int. J. Pept. Protein Res.* 36, 255–266.
- Deleage, G., and Geourjon, C. (1993) An interactive graphic program for calculating the secondary structure content of proteins from circular dichroism spectrum. *Comput. Appl. Biosci.* 9, 197–199.
- Eftink, M. R. (1991) Fluorescence quenching reactions: Probing biological macromolecular structures. In *Biophysical and biochemical aspects of fluorescence spectroscopy* (Dewey, T. G., Ed.) pp 1–39, Plenum Press, New York.
- Wang, Y. J., Yurttas, L., Dale, B. E., Russell, D. H., Kinsel, G., Preston-Schaffter, L. M., Johnson, V., and Hayes, T. K. (1997) MALDI-MS as a monitor of the purification and folding of synthetic ecdysis hormone. *Peptides* 18, 337–346.
- Ludvigsen, S., Roy, M., Thøgersen, H., and Kaarsholm, N. C. (1994) High-resolution structure of an engineered biologically potent insulin monomer, B16 Tyr → His, as determined by nuclear magnetic resonance spectroscopy. *Biochemistry* 33, 7998–8006.
- Betz, C., Lange, G., Pal, G. P., Wilson, K. S., Maelicke, A., and Saenger, W. (1991) The refined crystal structure of  $\alpha$ -cobratoxin from *Naja naja siamensis* at 2.4 Å resolution. *J. Biol. Chem.* 266, 21530–21536.

34. Chang, L., Lin, S., and Yang, C. (2001) Refolding of Taiwan cobra neurotoxin: Intramolecular cross-link affects its refolding reaction. *Arch. Biochem. Biophys.* 387, 289–296.
35. Katayama, H., Nagata, K., Ohira, T., Yumoto, F., Tanokura, M., and Nagasawa, H. (2003) The solution structure of molt-inhibiting hormone from the Kuruma prawn *Marsupenaeus japonicus*. *J. Biol. Chem.* 278, 9620–9623.
36. Dow, J. A. (1984) Extremely high pH in biological systems: A model for carbonate transport. *Am. J. Physiol.* 246, R633–R636.
37. Brems, D. N., Brown, P. L., Nakagawa, S. H., and Tager, H. S. (1991) The conformational stability and flexibility of insulin with an additional intramolecular cross-link. *J. Biol. Chem.* 266, 1611–1615.
38. Lehrer, S. S. (1971) Solute perturbation of protein fluorescence. The quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. *Biochemistry* 10, 3254–3263.
39. Horodyski, F. M., Ewer, J., Riddiford, L. M., and Truman, J. W. (1993) Isolation, characterization and expression of the eclosion hormone gene of *Drosophila melanogaster*. *Eur. J. Biochem.* 215, 221–228.
40. Sun, J., Feng, J., Li, Y., and Shen, B. (2006) A novel BLYS antagonist peptide designed based on the 3-D complex structure of BCMA and BLYS. *Biochem. Biophys. Res. Commun.* 346, 1158–1162.
41. Low, C. M., Buck, I. M., Cooke, T., Cushnir, J. R., Kalindjian, S. B., Kotecha, A., Pether, M. J., Shankley, N. P., Vinter, J. G., and Wright, L. (2005) Scaffold hopping with molecular field points: Identification of a cholecystokinin-2 (CCK2) receptor pharmacophore and its use in the design of a prototypical series of pyrrole- and imidazole-based CCK2 antagonists. *J. Med. Chem.* 48, 6790–6802.
42. Lozzi, L., Lelli, B., Runci, Y., Scali, S., Bernini, A., Falciani, C., Pini, A., Niccolai, N., Neri, P., and Bracci, L. (2003) Rational design and molecular diversity for the construction of anti- $\alpha$ -bungarotoxin antidotes with high affinity and *in vivo* efficiency. *Chem. Biol.* 10, 411–417.
43. Nachman, R. J., Moyna, G., Williams, H. J., Zabrocki, J., Zadina, J. E., Coast, G. M., and Vanden Broeck, J. (1999) Comparison of active conformations of the insect tachykinin/tachykinin and insect kinin/Tyr-W-MIF-I neuropeptide family pairs. *Ann. N.Y. Acad. Sci.* 897, 388–400.
44. Flohr, S., Kurz, M., Kostenis, E., Brkovich, A., Fournier, A., and Klabunde, T. (2002) Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure-activity relationships and nuclear magnetic resonance studies on urotensin II. *J. Med. Chem.* 45, 1799–1805.
45. Welch, W. H. Jr., and Fasman, G. D. (1974) Hydrogen-tritium exchange in polypeptides. Models of  $\alpha$ -helical and  $\beta$  conformations. *Biochemistry* 13, 2455–2466.
46. Pal, D., and Chakrabarti, P. (2002) On residues in the disallowed region of the Ramachandran map. *Biopolymers* 63, 195–206.
47. Brauer, A. B. E., Domingo, G. J., Cooke, R. M., Matthews, S. J., and Leatherbarrow, R. J. (2002) A conserved cis peptide bond is necessary for the activity of Bowman-Birk inhibitor protein. *Biochemistry* 41, 10608–10615.
48. Goyal, A., Seth, D., and Batra, J. K. (2002) Role of cis prolines 112 and 126 in the functional activity of ribonucleolytic toxin restrictocin. *Biochem. Biophys. Res. Commun.* 295, 812–817.
49. Pal, D., and Chakrabarti, P. (1999) Cis peptide bonds in proteins: Residues involved, their conformations, interactions and locations. *J. Mol. Biol.* 294, 271–288.
50. Goudet, C., Chi, C., and Tytgat, J. (2002) An overview of toxins and genes from the venom of the Asian scorpion *Buthus martensii* Karsch. *Toxicon* 40, 1239–1258.
51. Arias, H. R., and Blanton, M. P. (2000) Review:  $\alpha$ -Conotoxins. *Int. J. Biochem. Cell Biol.* 32, 1017–1028.
52. Murray-Rust, J., McDonald, N. Q., Blundell, T. L., Hosang, M., Oefner, C., Winkler, F., and Bradshaw, R. A. (1993) Topological similarities in TGF- $\beta$ 2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure* 1, 153–159.
53. McDonald, N. Q., and Hendrickson, W. A. (1993) A structural superfamily of growth factors containing a cystine knot motif. *Cell* 73, 421–424.
54. Craik, D. J., Daly, N. L., Mulvenna, J., Plan, M. R., and Trabi, M. (2004) Discovery, structure and biological activities of the cyclotides. *Curr. Protein Pept. Sci.* 5, 297–315.
55. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
56. Wei, Z. J., Hong, G. Y., Wei, H. Y., Jiang, S. T., and Lu, C. (2008) Molecular characters and expression analysis of the gene encoding eclosion hormone from the Asian corn borer, *Ostrinia furnacalis*. *DNA Sequence* 19, 301–307.
57. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723.