Backbone Structure and Dynamics of a Hemolymph Protein from the Mealworm Beetle *Tenebrio molitor*†

Sven Rothemund,‡ Yih-Cherng Liou,§ Peter L. Davies,§ and Frank D. Sönnichsen*,‡

Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, Ohio 44106-4970, and Department of Biochemistry, Queen's University, Kingston, Ontario, Canada K7L3N6

Received June 25, 1997; Revised Manuscript Received August 15, 1997⊗

ABSTRACT: Pheromones play a vital role in the survival of insects and are used for chemical communication between members of the same species by their olfactory system. The selection and transportation of these lipophilic messengers by carrier proteins through the hydrophilic sensillum lymph in the antennae toward their membrane receptors remains the initial step for the signal transduction pathway. A moderately abundant 12.4 kDa hydrophilic protein present in hemolymph from the mealworm beetle *Tenebrio molitor* is ∼38% identical to a family of insect pheromone-binding proteins. The backbone structure and dynamics of the 108-residue protein have been characterized using three-dimensional $^{1}H-^{15}N$ NMR spectroscopy, combined with ^{15}N relaxation and $^{1}H/D$ exchange measurements. The secondary structure, derived from characteristic patterns of dipolar connectivities between backbone protons, secondary chemical shifts, and homonuclear three-bond $J_{HNH\alpha}$ coupling constants, consists of a predominantly disordered N-terminus from residues 1 to 10 and six α-helices connected by four 4−7 residue loops and one β-hairpin structure. The up-and-down arrangement of α-helices is stabilized by two disulfide bonds and hydrophobic interactions between amphipathic helices. The backbone dynamics were characterized by the overall correlation time, order parameters, and effective correlation times for internal motions. Overall, a good correlation between secondary structure and backbone dynamics was found. The ^{15}N relaxation parameters T_1 and T_2 and steady-state NOE values of the six α-helices could satisfactorily fit the Lipari−Szabo model. In agreement with their generalized order parameters (>0.88), residues in helical regions exhibited restricted motions on a picosecond time scale. The stability of this highly helical protein was confirmed by thermal denaturation studies.

Insect pheromone-binding proteins[1] (PBPs), pheromone-binding protein-related proteins (PBPRPs) and general odorant-binding proteins (GOBPs) are secreted small proteins with 12−20 kDa molecular mass that bind volatile, hydrophobic compounds involved in signaling (*1*). Although the mechanism of this signal transduction is not completely understood, insect PBPs, PBPRPs, and GOBPs have been proposed to solubilize and transport hydrophobic pheromones or odorants across hydrophilic fluids to receptors in the dentritic membranes (*2*−*4*). It should be noted that volatile compounds, such as pheromones, odorants, or even fatty acids, show similarities with respect to structure, size, hydrophobicity, and charge. In most cases they are linear aliphatic, saturated or nonsaturated compounds with different head group functionalities such as acetates, alcohols, aldehydes, or acids.

Several PBPs, PBPRPs, and GOBPs have been isolated and described (*5*−*10*). PBPs are typically found in the sensillum of male insects, but the occurrence of PBPRPs and GOBPs is not sex-specific. In addition to some highly conserved regions, the most common structural motif of these proteins is six conserved Cys residues that form intramolecular disulfide bonds (*4*). Three structures of vertebrate pheromone- or odorant-binding proteins have been determined by X-ray crystallography (*11*, *12*). Their three-dimensional structures are similar to those of other proteins from the lipocalin family, consisting of eight- or 10-stranded β-barrels followed by an α-helix. Pheromone-binding characteristics have been studied by Prestwich and coworkers (*13*, *14*). Despite the structural similarity of pheromones, high ligand affinity (micromolar) and specificity were found. Similar to lipocalins, the proposed binding site in pheromone-binding proteins is the most hydrophobic region, which is predicted to be in the interior of the proteins (*11*−*13*).

Recently, Paesen and Happ (*15*) have identified two low molecular mass proteins secreted by the tubular accessory sex glands of the male mealworm beetle, *Tenebrio molitor*. These B1 and B2 proteins show significant sequence identity to the family of pheromone-binding proteins, suggesting a function as carrier proteins for volatile substances (*16*). In contrast to vertebrate pheromone-binding proteins, B proteins

† This work was supported by NIH grant GM55362 (F.D.S.), a grant from the Deutscher Akademischer Austauschdienst (S.R.), and by funds from the Medical Research Council of Canada (P.L.D.).

* To whom correspondence should be addressed: Department of Physiology and Biophysics, Case Western Reserve University, School of Medicine, 10900 Euclid Ave., Cleveland, OH 44106-4970. E-mail frank@herring.phol.cwru.edu; telephone (216) 368-3850; fax (216) 368-1693.

‡ Case Western Reserve University.

§ Queen's University.

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1997.

[1] Abbreviations: 3D, three dimensional; CD, circular dichroism; DQF COSY, double-quantum-filtered correlation spectroscopy; DSS, 4,4-dimethyl-4-silapentane-1-sulfonate; HPLC, high-performance liquid chromatography; HSQC, heteronuclear single quantum coherence; IPTG, isopropyl D-thiogalactopyranoside; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; PMSF, phenylmethanesulfonyl fluoride; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy.

S0006-2960(97)01529-8 CCC: $14.00 © 1997 American Chemical Society

share four of six conserved Cys residues and are predicted to be mostly α-helical. To date, there is neither a crystal structure nor a high-resolution NMR structure of insect PBPs, PBPRPs, and GOBPs.

This paper represents a comprehensive characterization of an abundant 12.4 kDa hemolymph protein (THP12) from mealworm beetle *Tenebrio molitor* by nuclear magnetic resonance spectroscopy, CD, and molecular biology. THP12 shows homology to B1 and B2 proteins as well as similarity in hydropathy plots (*17*) and an alignment of the four Cys residues with identical spacings. Almost complete ^{1}H and ^{15}N NMR resonance assignments and the secondary structure of the protein are reported. Also, the relaxation parameters of the ^{15}N nuclei have been determined to characterize the backbone dynamics of the protein.

MATERIALS AND METHODS

Protein Expression and Sample Preparation. A cDNA clone of THP12 in plasmid pET20b (W. Tang, unpublished results) was transformed into *Escherichia coli* BL21(DE3). For heteronuclear labeling with ^{15}N, THP12 was overexpressed at 37 °C in minimal medium containing 50 mM Na_2HPO_4, 20 mM KH_2PO_4, 8.5 mM NaCl, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.8% glucose, and 0.05% $(^{15}NH_4)_2SO_4$. Protein production was induced using 0.4 mM IPTG when the cell density reached an OD_{600} of 0.8. After induction for 3 h, cells were collected by centrifugation at 3000*g* for 20 min and were resuspended in buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 0.1 mM PMSF) for sonication at 0 °C. The sonicate was centrifuged at 12000*g* for 25 min, and the supernatant fraction was then loaded onto a Sephadex G-75 column (3.0 × 100 cm). The A_{230} value of the elution profile was measured, and the location of THP12 was determined by Tris−Tricine SDS−PAGE. Pooled fractions containing THP12 were lyophilized and resuspended in 0.1% TFA (buffer A) for further purification by reversed-phase HPLC on a C18 preparative column (2.2 × 25 cm). THP12 was eluted by a linear gradient of acetonitrile in 0.1% TFA at a flow rate of 8 mL/min and came off the column at 32% acetonitrile. The mass of the protein was confirmed by ESI-mass spectrometry.

Unlabeled THP12 and the L47M mutant were produced by the same method from *E. coli* grown on LB (*18*). The L47M mutant of THP12 was prepared by site-directed mutagenesis using the method of Kunkel et al. (*19*). The mutagenic oligonucleotide primer (antisense), 5′-CCG-GCTCTCTTCATGATGCAGAAAGC-3′, converted codon 47 from TTG to ATG while also introducing a *Bsp*HIsite (5′-TCATGA-3′) that facilitated the identification of mutant clones.

CNBr Fragmentation. Purified L47M was treated with a 100-fold molar excess of CNBr in 70% formic acid purged with N_2 gas. The reaction was carried out with shaking at room temperature in the dark for 16 h. A control reaction was also performed in which CNBr was omitted. CNBr-treated samples with or without reduction in 2 mM DTT were fractionated by C18 reversed-phase HPLC and analyzed by MALDI-TOF mass spectrometry. Amino acid compositions of cleaved fragments were determined by the Biotechnology Service Center, University of Toronto.

NMR Experiments. NMR measurements were made on a 1 mM protein sample in 95% H_2O/5% D_2O, adjusted to pH 6.9 (not corrected for deuterium isotope effects) by addition of DCl. To increase the protein concentration, a Shigemi microtube was used. NMR experiments were recorded on Varian Unity Plus 600 MHz and Inova 500 MHz spectrometers equipped with pulsed field gradient units and probes with an actively shielded *z* gradient. ^{15}N-NOESY-HSQC (*20*) spectra with a mixing time of 150 ms, ^{15}N-TOCSY-HSQC (*21*) spectra with various mixing times (35−69 ms), and HNHA (*22*) and HNHB (*23*) spectra were recorded at 20 or 25°C. The ^{15}N-NOESY-HSQC, ^{15}N-TOCSY-HSQC, and HNHB experiments are improved pulsed field gradient, selectively enhanced experiments (*24*) with flip-back pulses to prevent water saturation (*25*). The HNHA experiment uses jump and return water suppression. Typical employed carrier positions were 119.5 ppm for ^{15}N and 4.87 ppm for ^{1}H. ^{15}N decoupling was performed during acquisition with WALTZ-16 (*26*). The data sets were processed using shifted sine-bell apodization and zero-filling in both dimensions. Data were processed using nmrPipe (*27*) and analyzed with the program PIPP (*28*).

Hydrogen exchange rates were determined at 25 °C with a 1.0 mM protein solution freshly dissolved in D_2O, adjusted with 0.1 M DCl to pH 6.1. A series of 2D ^{1}H−^{15}N HSQC spectra with a collection time of 15 min were recorded using 1024 complex data points, 4 transients, and 100 complex increments. Exchange rate constants were derived from least-squares fit of the peak intensities to the equation $\ln(h) = -kt + C$, where h is the peak at time t, k is the first-order exchange rate constant (per minute), t is the elapsed time (minutes) after the protein was dissolved, and C is the logarithm of the peak intensity at dead time. Hydrogen exchange was followed for 2 days.

T_1, T_2, and NOE relaxation data were determined by HSQC-based pulse sequences (*29*, *30*) at 20 °C with a spectral width of 8000 Hz in the *F*2 domain and 1700 Hz in the indirectly detected *F*1 domain. Time domain data were recorded with 1024 complex points in *t*2 and with 256 points in *t*1 dimension. ^{15}N T_1 values were measured from spectra recorded with nine different delays of 0, 110, 220, 330, 445, 555, 720, 890, and 1055 ms, whereas 12 delays of 0, 16, 32, 48, 64, 80, 96, 112, 130, 145, 160, and 177 ms were used during the CPMG period of the ^{15}N T_2 experiments. ^{15}N decoupling was carried out using WALTZ-16 decoupling (*26*). Relaxation delays of 1.2 s were employed in the ^{15}N T_1 and T_2 experiments. ^{1}H−^{15}N steady-state NOE values were obtained by recording spectra in the presence and absence of a ^{1}H saturation period of 3 s. In case of the experiment without presaturation, the net relaxation delay was 5 s. All relaxation spectra were measured in duplicate.

Analysis of Relaxation Data. Intensities of resonance peaks in two-dimensional spectra were determined as peak heights. T_1 and T_2 data were fitted to a two-parameter exponential equation using a nonlinear least-squares fitting algorithm. $\{^1H\}$−^{15}N heteronuclear NOE values were calculated as ratios of peak intensities in the spectra recorded with and without ^{1}H saturation. In order to estimate mean values and standard deviations of relaxation data, experiments were repeated. Values of T_1, T_2, and NOE were fitted by different motional models based on the model-free approach (*31*, *32*) with a minimum of parameters to describe the overall tumbling, and model extensions including exchange effects on T_2 and internal motions on two time scales (*33*, *34*). Assuming an isotropic overall motion, the overall

tumbling of the molecule is described by a single correlation time (τ_m) and the faster internal motion as an order parameter (S^2) and an effective internal correlation time (τ_e). In the extended models, additional parameters for conformational exchange (R_{ex}) and fast (S_f^2) and slow local motions (S_s^2) were included, where the time scale of the fast local motion is assumed to be too fast for detection (*33*, *34*). Nonlinear least-squares minimization methods and Monte Carlo error analysis (*24*) were applied for determination of optimal model parameters. The selection of the appropriate model for each residue was performed on basis of the protocol described by Farrow et al. (*30*).

Circular Dichroism. Circular dichroism spectra were recorded on a Jasco J-720 spectropolarimeter (Jasco, Eaton, MD) equipped with a Jasco DP500N data processor. A water bath was used to control the temperature of the cuvette. The temperature dependence of the molar ellipticity was measured at 222 nm by varying the temperature in the range of 5−85 °C. The protein was dissolved in water at pH 7.0 to a final concentration of 1.53×10^{-5} M as determined by amino acid analysis. Because of the expected high helical content of THP12, the calculation of helical content was performed according to Chen et al. (*35*):

$$f_H = \frac{[\Theta]_{222}}{[\Theta]_{222}^{\infty}} + \frac{ik}{N}$$

where f_H is the helix fraction, $[\Theta]_{222}$ is the observed mean residue ellipticity at wavelength 222 nm, i is the number of helical segments, k is a wavelength-dependent constant, N is the total number of residues and $[\Theta]_{222}^{\infty}$ is the reference mean ellipticity for 100% helix.

RESULTS

Expression Yield. THP12 was readily purified to homogeneity by a two-column chromatography procedure in which the Sephadex G-75 column removed the vast majority of soluble *E. coli* proteins and the HPLC profile showed only one major product. Analysis of the HPLC product by ESI-mass spectrometry revealed a single peak of 12.44 kDa, which is consistent with the predicted mass of 12.44 kDa (108 residues plus N-terminal methionine). In initial trials, the yield of THP12 from ^{15}N-labeling in minimal medium was less than 1 mg/L. This was improved to 4−6 mg/L by doubling the concentration of glucose from 0.4% to 0.8% but was still too low to make double-labeling with ^{13}C and ^{15}N feasible. The yields of unlabeled THP12 and L47M mutant on LB medium were 10 times greater.

Resonance Assignments. The uniformly ^{15}N-labeled protein gives good quality NMR spectra with a sufficient dispersion for most of the HN−^{15}N correlations.

Assignments were accomplished using standard methods (*20*, *36*). The spin systems of the various amino acids were identified by their characteristic patterns of cross peaks in the 3D ^{1}H−^{15}N-TOCSY HSQC experiment, which correlates backbone ^{15}N and NH resonances with intraresidue side-chain protons. Generally, complete coherence transfer from the side chain to the backbone NH were observed for most of the amino acids (except for outermost protons of the lysines) at a spin lock time of 69 ms. Unambiguous C_αH and C_βH assignments were obtained from HNHA (*22*) and HNHB (*23*) experiments. With the exception of the N-terminus from Glu 1 to Lys 8 and the three prolines (Pro 37, 75, 92), the 97 remaining residues could be assigned in 3D TOCSY spectra (Table 1). The residues found can be classified as follows: 42 residues that have unique chemical shift patterns (eight Ala, three Gln, 10 Glu, three Gly, nine Ser, two Thr, and seven Val), 29 residues with an AMX spin system (nine Asp, four Asn, four Cys, six His, and six Phe), and 26 residues with longer side chains (four Arg, three Ile, five Leu, and 14 Lys). Additionally, asparagines and glutamines were confirmed by their characteristic intraresidue NOEs between side-chain NH and C_βH and C_γH, respectively. For assignment of prolines 37, 75, and 92, the sequential connectivities between their C_δH and the C_αH of the preceding residue in the 2D spectrum were used. The observation of these cross peaks indicated that the respective X−Pro peptide bonds are in *trans* conformation.

The sequence-specific assignment was carried out by standard procedures (*36*) using complementary experiments (3D ^{1}H−^{15}N NOESY-HSQC and 3D ^{1}H−^{15}N TOCSY-HSQC). The sequential backbone NOESY connectivities, such as NH($i-1$)−NH(i), C_αH($i-1$)−NH(i), and C_βH($i-1$)−NH(i) thus identified and established sequential relationship. Overlap of sequential NH($i-1$)−NH(i) and cross peaks in the fingerprint region of the NOESY spectrum was resolved by varying the temperature and pH. As a representative example, the sequential connectivities between Leu 39 and Phe 52 are shown in Figure 1. Sequential C_αH($i-1$)−NH(i) connectivities were often observed in the HNHB experiment, confirming several of the determined sequential assignments.

Secondary Structure. The patterns of short- and medium-range NOE cross peaks, $^3J_{HNH\alpha}$ coupling constants, and rates of exchanging amide protons, which all give indications of the secondary structure, are summarized in Figure 2. With exception of the regions Glu 32−Lys 38 and Ser 72−Glu 76, short-range NH(i)−NH($i+1$) NOEs, characteristic of α-helical conformation, predominated throughout the protein. Furthermore, medium-range C_αH(i)−NH($i+3$), C_αH(i)−NH($i+4$), and NH(i)−NH($i+2$) NOEs indicate the presence of six α-helices (A−F: A, Ser 11−Gly 19; B, Ser 21−Arg 31; C, Pro 37−Gly 51; D, Leu 61−Glu 70; E, Lys 77−Val 87; and F, Gln 93−Val 103). However, less well defined were the segments between the helices, especially the links between helices B−C and E−F. The links between helices D−E and E−F were confirmed by the presence of long-range NOEs between residues Lys 69 and Val 78, Phe 68 and Asp 79, and Lys 89 and Ser 95. For the segment between the helices C and D, a complete set of sequential, medium-, and long-range NOEs clearly indicated that this region folds into a β-hairpin conformation from residue Phe 52 to Gln 60, shown schematically in Figure 3. Intense dNN(i, $i+1$) and weak dαN(i, $i+1$) NOEs were found for residues Ala 55−Gly 57 in the turn region, whereas for residues Phe 52−Asp 54 and Glu 58−Gln 60 in the β-sheet weak dNN(i, $i+1$) and intense dαN(i, $i+1$) were observed (Figure 2). In addition, the presence of long-range NOEs, typical of β-sheets, between Ile 53 and Gln 60, Ile 53 and Glu 58, Asp 54 and Glu 58, and Asp 54 and Phe 59 confirms the β-hairpin structure.

Chemical shift deviations from random coil values are largely used for secondary structure assignments (*37*−*40*). In particular, chemical shifts of C_α protons within α-helices or β-sheets tend to experience an upfield or downfield shift,

Table 1: ^{15}N and ^{1}H Chemical Shifts for THP12 at pH 6.9 and 25 °C[a]

residue	N	HN	$C_\alpha H$	$C_\beta H$	other
Gln 9	118.12	7.844	4.222	2.052, 2.146	γ 2.410, 2.568; 7.014, 7.559
His 10	118.50	8.108	4.748	3.047, 3.271	
Ser 11	113.74	8.134	4.439	4.025, 3.917	
Asp 12	119.38	8.971	4.470	2.830*	
Ala 13	122.16	8.080	4.350	1.610	
Cys 14	115.69	8.078	4.685	2.964*	
Lys 15	123.10	8.233	4.247	2.096, 1.998	1.345; ϵ2.996
Ala 16	120.78	8.165	4.300	1.620	
Glu 17	116.62	7.808	4.367	2.323*	
Ser 18	114.77	8.950	4.410	4.308, 4.013	
Gly 19	110.15	8.293	4.170, 3.985		
Val 20	120.25	7.750	4.003	1.765	γ 0.961, 0.174
Ser 21	122.18	8.980	4.571	4.445, 4.171	
Glu 22	123.03	8.927	4.086	2.050*	γ 2.432
Glu 23	118.00	8.692	4.155	2.228, 2.034	γ 2.406
Ser 24	116.65	8.096	4.263	4.063, 3.788	
Leu 25	120.32	7.465	4.010	1.960, 1.359	γ 1.598; δ 0.689, 0.660
Asn 26	118.46	8.777	4.497	2.979, 2.879	δ 7.149, 7.743; Nδ 113.74
Lys 27	120.32	7.860	4.128	2.086, 1.730	γ 1.962, 1.376; δ 1.697; ϵ 3.044
Val 28	118.47	7.640	3.935	2.213	γ 1.224, 0.796
Arg 29	120.78	8.086	4.040	2.018*	γ 1.843, 1.680; δ 3.250
Asn 30	116.61	7.870	4.914	3.137, 2.829	δ 6.855, 7.810; Nδ 113.73
Arg 31	115.22	8.042	4.101	2.146*	γ 1.723, 1.673; δ 3.358
Glu 32	119.39	8.379	4.517	2.077, 1.805	γ 2.301
Glu 33	121.71	8.514	4.283	2.023*	γ 2.370, 2.312
Val 34	122.17	8.319	4.398	2.156	γ 1.060*
Asp 35	128.18	8.795	4.882	2.836, 2.628	
Asp 36	120.83	7.660	5.122	2.777*	
Pro 37			4.420	2.222, 2.609	δ 4.210, 4.081
Lys 38	117.53	8.660	4.325	2.128, 1.810	γ 1.710, 1.610; δ 1.870; ϵ 3.180
Leu 39	119.85	7.715	4.167	2.110, 1.775	γ 1.782; δ 1.049, 1.001
Lys 40	118.44	7.706	4.218	2.173, 1.925	γ 1.492, 1.338; δ 1.670; ϵ 3.019
Glu 41	118.00	7.457	4.439	2.325*	γ 2.587, 2.552
His 42	121.70	8.640	4.277	3.358, 3.404	6.500
Ala 43	118.93	8.568	3.762	1.499	
Phe 44	118.93	7.982	5.061	3.457, 3.306	7.430, 7.340
Cys 45	118.93	8.366	3.925	3.514, 3.270	
Ile 46	122.17	8.414	3.287	2.035	γ 1.560, 0.482, 1.111; δ 0.683
Leu 47	116.61	8.400	3.946	2.191, 1.455	γ 1.303; δ 1.110, 2.407
Lys 48	120.80	8.516	4.324	1.655, 1.535	γ 1.095, 1.060; δ 1.215, 1.085; ϵ 3.094
Arg 49	121.71	8.160	3.861	1.801, 1.691	γ 1.597, 1.503; δ 3.270
Ala 50	117.53	7.821	3.829	0.498	
Gly 51	105.06	7.590	4.078, 3.926		
Phe 52	117.07	8.622	4.825	3.216, 3.034	7.001, 7.159
Ile 53	120.77	7.686	4.922	1.193	γ 1.127, 0.799, 0.634; δ 0.232
Asp 54	124.94	7.860	4.973	3.383, 2.622	
Ala 55	118.50	8.476	4.227	1.567	
Ser 56	112.00	8.114	4.603	4.118, 4.061	
Gly 57	109.21	8.392	4.249, 3.279		
Glu 58	118.93	8.201	4.510	1.969, 1.915	γ 2.385, 2.213
Phe 59	126.33	9.213	4.844	2.921, 2.702	6.890, 7.120
Gln 60	123.09	8.318	4.575	2.447, 2.364	γ 2.323, 1.940; 6.796; ϵ 7.118; Nϵ 115.93
Leu 61	119.85	7.715	3.691	1.949, 1.387	γ 1.724; δ 1.060, 0.810
Asp 62	116.64	8.610	4.475	2.769*	
His 63	122.09	7.979	4.562	3.335*	7.180, 7.030
Ile 64	117.08	7.978	2.541	1.469	γ 1.733, −0.090, −0.343; δ 0.718
Lys 65	116.16	8.114	3.645	2.069, 1.839	γ 1.316; δ 1.765; ϵ 2.907
Thr 66	114.21	8.200	3.922	4.339	γ 1.360
Lys 67	119.39	8.267	4.310	2.151, 2.100	γ 1.570; δ 1.834; ϵ 3.176, 3.040
Phe 68	117.44	9.025	4.683	3.520, 3.125	7.092
Lys 69	117.54	8.176	4.130	2.135*	γ 1.628; δ 1.900, ϵ 3.114
Glu 70	120.78	7.398	4.178	2.277, 2.164	γ 2.688, 2.359
Asn 71	126.34	9.213	4.628	3.136, 2.859	
Ser 72	111.53	7.539	4.668	4.094, 3.649	
Glu 73	125.86	9.040	4.246	1.911*	γ 2.280, 2.106
His 74	114.3	8.392	5.401	3.131*	7.240
Pro 75			4.385	2.465, 2.289	δ 3.889, 3.552; δ 2.285
Glu 76	120.36	10.239	4.489	2.295, 2.233	γ 2.702, 2.294
Lys 77	119.85	7.795	4.672	2.192, 1.938	γ 1.570; δ 1.768, 1.678; ϵ 2.947
Val 78	120.76	7.623	3.699	2.306	γ 1.261, 1.161
Asp 79	121.71	8.868	4.208	2.828, 2.710	
Asp 80	119.47	8.460	4.471	2.702*	
Leu 81	122.63	7.525	4.170	2.146, 1.761	γ 1.473; δ 0.980, 0.924
Val 82	118.46	8.665	3.297	2.104	γ 1.004, 0.912

Table 1 (Continued)

residue	N	HN	$C_\alpha H$	$C_\beta H$	other
Ala 83	120.31	7.982	4.122	1.571	
Lys 84	115.60	7.806	4.070	1.976*	γ 1.493, 1.335; δ 1.665, ϵ 3.017
Cys 85	110.61	8.010	5.202	3.004, 2.477	
Ala 86	127.25	8.901	3.751	0.679	
Val 87	124.48	7.255	4.333	1.980	γ 1.002, 0.933
Lys 88	130.03	9.152	4.393	2.060,2.006	γ 1.749; δ 1.870; ϵ 3.001
Lys 89	127.73	9.336	4.822	2.088,1.807	γ 1.686; δ 1.550, 1.500; ϵ 3.180
Asp 90	118.01	8.826	4.407	2.955,2.853	
Thr 91	108.29	7.510	5.012	4.584	γ 1.247
Pro 92			4.549	1.642,0.695	γ 1.920, 1.638; δ 3.920, 3.780
Gln 93	119.86	8.909	3.660	2.074,1.750	γ 2.879, 1.962; δ 7.741, 7.151
His 94	120.68	8.146	4.725	3.514,3.372	7.240, 7.130
Ser 95	116.17	9.238	4.567	4.111,3.983	
Ser 96	118.09	8.430	4.435	4.255,3.800	
Ala 97	125.87	8.716	4.184	1.798	
Asp 98	119.40	8.664	4.730	3.185,2.912	
Phe 99	123.09	8.357	4.504	3.692*	7.157, 6.768
Phe 100	119.90	8.153	4.023	3.420,3.465	
Lys 101	119.87	8.125	4.258	2.145*	γ 1.612; δ 1.687; ϵ 3.000
Cys 102	120.00	8.470	4.372	3.386,3.289	
Val 103	116.61	8.398	3.692	1.918	γ 0.532
His 104	118.00	7.633	4.272	3.377,3.069	7.305, 6.800
Asp 105	119.85	8.297	4.603	2.850*	
Asn 106	118.44	7.840	4.778	2.826,2.714	7.899, 7.325
Arg 107	122.16	8.250	4.465	2.019,1.906	γ 1.746, δ 3.265
Ser 108	122.63	8.020	4.327	3.912*	

[a] Chemical shifts are given in parts per million. Degenerate $C_\beta H$ frequencies are marked with an asterisk.

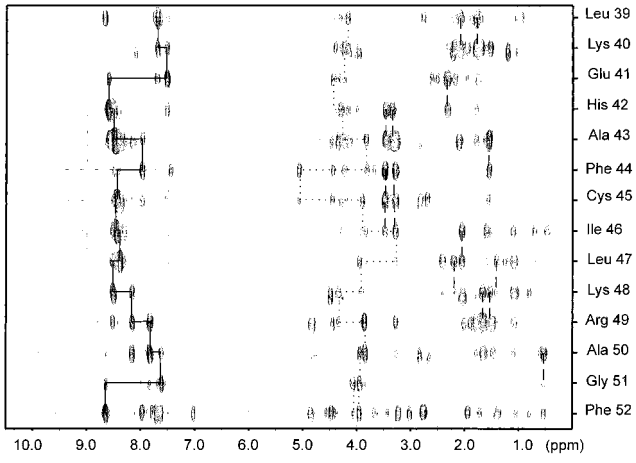

FIGURE 1: 2D strip plot from a 3D $^1H-^{15}N$ NOESY-HSQC spectrum of uniformly ^{15}N-labeled THP12 showing sequential NH(i)−NH(i + 1), $C_\alpha H(i)$−NH(i + 1), and $C_\beta H(i)$−NH(i + 1) NOE connectivities for residues Leu 39−Phe 52. Lines highlight sequential connectivities. The spectrum was obtained at 20 °C with a mixing time of 150 ms.

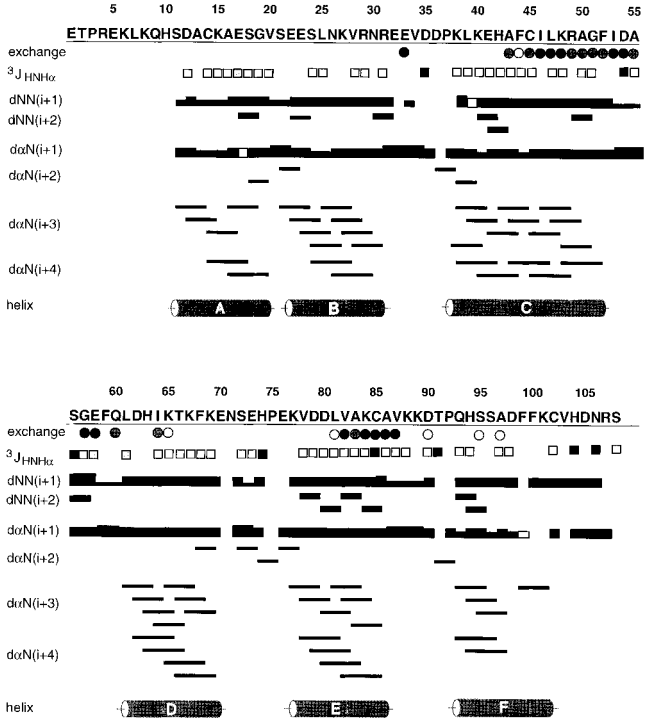

FIGURE 2: Amino acid sequence and summary of sequential connectivities, $^3J_{HNH\alpha}$ coupling constants, and amide exchange data. NOEs involving NH and $C_\alpha H$ are represented as bars; the intensity of NOE cross peaks is indicated by the heights of the bars. Exchanging backbone amide protons are indicated as circles (black, gray, and white circles indicate signals still present after 12 h, 1 h, and 10 min, respectively). Black, gray, and white rectangles identify residues with $^3J_{HNH\alpha} > 7$ Hz, $^3J_{HNH\alpha}$ 6−7 Hz, and $^3J_{HNH\alpha} < 6$ Hz, respectively. The locations of helices are identified at the bottom.

respectively, relative to chemical shift values typical for random coil conformation. The plot of $C_\alpha H$, NH, and ^{15}N chemical shift deviations vs the sequence number is shown in Figure 4. Negative $C_\alpha H$ shift differences up to 1 ppm correlate with the six proposed α-helical regions. Comparing the average $C_\alpha H$ chemical shift differences per helix, the values for helices A, B, and F are somewhat smaller than values for helices C, D, and E, suggesting an overall higher stability for helices C, D, and E. Except for Phe 44 and Cys 85, large positive $C_\alpha H$ chemical shift differences were exclusively found in regions between the helices. The downfield shifts at residues Phe 44 and Cys 85 can be explained by backbone disturbances resulting from disulfide bond formation at residue Cys 45 and Cys 85, respectively. Interestingly, the $C_\alpha H$ chemical shift pattern (↑↑↑↓↑↓↑↑↑) of the loop region Phe 52−Gln 60 is in excellent agreement with a β-hairpin structure (*41, 42*). Downfield shifts with respect to random coil values are observed for residues Phe 52−Asp 54 and Glu 58−Gln 60 in the sheet conformation, whereas upfield shifts were found for two of the three residues in the turn region.

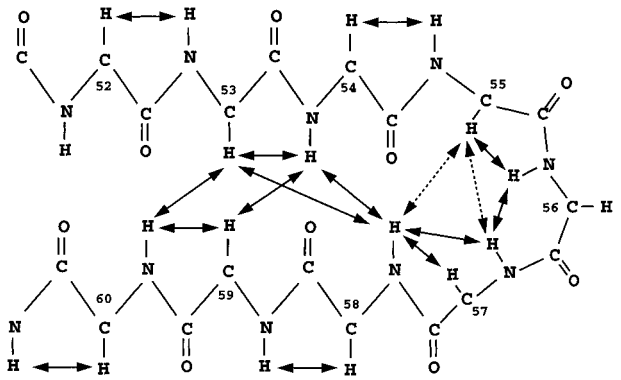

FIGURE 3: Schematic representation of the β-hairpin structure in THP12 as deduced from qualitative analysis of 3D NOESY-HSQC spectra. The observed connectivities between backbone NH and $C_\alpha H$ are indicated by solid arrows. Dashed arrows represent NOE connectivities that cannot be assigned unambiguously due to overlap.

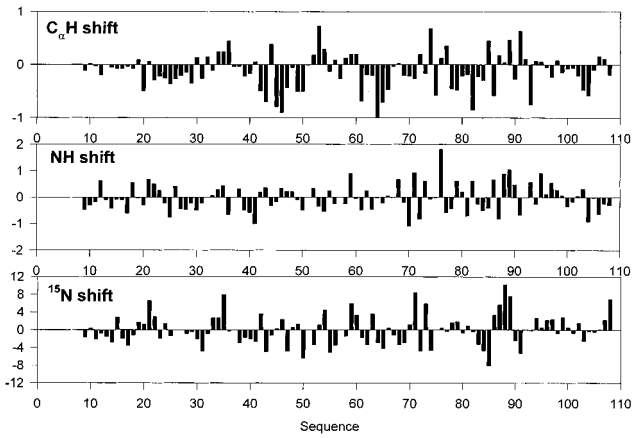

FIGURE 4: Plot of chemical shift differences of $C_\alpha H$, NH, and ^{15}N nuclei between observed and random coil values vs the residue number. The reference values are taken from Wishart et al. (*40*). All shifts were recorded at pH 6.9 and 20 °C.

Due to a sufficient dispersion of $^1H-^{15}N$ correlations, 67 $^3J_{HNH\alpha}$ coupling constants could be estimated by the measurement of the diagonal-peak to cross-peak intensity ratio in the 3D ^{15}N-separated HNHA spectrum (*22*). Coupling constants <6 Hz are consistent with an α-helical conformation in the regions Val 20−Arg 29, Lys 40−Ala 50, Ile 64−Glu 70, Glu 78−Lys 88, and Ser 95−Cys 102, supporting assignment of helices B−E (except for Cys 85) (Figure 2). The distortion at Cys 85 seems to ensure that Cys 85 and Cys 102 are properly oriented to make a disulfide bond between helices E and F. Values of coupling constants for helices A and F were determined to be in the range 3−7 Hz. In contrast, coupling constants of 7−10 Hz were measured for Asp 35, Asp 54, Ser 56, Glu 58, His 74, and Thr 91, all residues in regions between helices. In the β-hairpin structure large coupling constants ≈8 Hz were found for residues Asp 54 and Ser 56.

Amide Exchange. Hydrogen exchange rates were obtained by directly monitoring the disappearance of $^1H-^{15}N$ cross peaks in 2D HSQC spectra as a function of time after dissolution in D_2O at 25 °C and pH 6.1. While the exchange for most of backbone amide protons in the N-terminal region (residues 1−40) was too rapid to be measured, the C-terminal part of THP12 has a distinctly different exchange behavior. Initial data points taken after 10 min show completely observable amide proton resonances in the regions His 43−Gly 51, Asp 62−Lys 65, Leu 81−Val 87, and Ser 95−Ala 97, spanning the helices C−F. In addition, several amide proton resonances corresponding to loop regions between the helices C and D as well as E and F were still present. Of these amide proton resonances the signals of helices D and F with the exception of Ile 64 disappeared after 1 h, while amide protons of the helices C and E and the β-hairpin structure exhibited protection against exchange. The data collected after 12 h showed stable amide protons for residues Ile 46−Lys 48 (helix C), Val 82−Val 87 (helix E), and the β-hairpin residues Ile 53, Asp 54, Gly 57, and Glu 58. After 24 h, amide proton signals were still observed for five residues (Leu 47, Ile 53, Val 82, Lys 84, and Val 87).

Backbone Dynamics. Values of spin−lattice and spin−spin relaxation rate constants, and steady-state $\{^1H\}-^{15}N$ NOEs were determined as described in Materials and Methods. The $^1H-^{15}N$ NOE correlation spectrum without ^1H saturation as well as some representative plots of ^{15}N T_1 and T_2 decays are displayed in Figure 5 as examples of the resolution and precision of the data. The summary of the relaxation data plotted as a function of residue number is shown in Figure 6. Uncertainties of ^{15}N T_1 and T_2 values were calculated as standard deviations from the nonlinear fitting procedure using two complete data sets, respectively. Calculated standard deviations for ^{15}N T_1 and T_2 values were always smaller (average standard deviation T_1 = 1.5%, T_2 = 3.5%) than those calculated from the root-mean-square (rms) base plane noise (*30*). Uncertainties of NOE values were estimated as standard deviations from duplicate measurements and averaged to 6%. A total of 75 sufficiently resolved resonances were characterized. For most residues, the values are rather uniform (average $T_1 = 592 \pm 28$ ms, $T_2 = 103 \pm 20$ ms, and NOE = 0.78 ± 0.08), with small deviations in T_2 and NOE for residues Glu 17, Glu 32−Asp 35, Ala 55, Glu 70−Glu 73, and Thr 91. The only major deviation was observed for the C-terminal residue Ser 108. Assuming isotropic tumbling of THP12 and a T_2 unaffected by chemical exchange, the overall rotational correlation time (τ_m) was obtained from T_1/T_2 ratios of all residues within 1 standard deviation of the average value (*20*, *33*). Excluding 12 residues, the calculated average τ_m was 7.2 ± 0.6 ns. The small change in the average calculated correlation time using all 75 residues (7.3 ± 0.9 ns) suggests that there is no anisotropy in the rotational behavior of THP12. As described by Farrow et al. (*30*), the relaxation parameters of each residue were analyzed by the model-free Lipari-Szabo formalism (*31*, *32*) and its extended forms. Extensions by an effective correlation time (τ_e), conformational exchange parameter (R_{ex}), and parameters for motions on the fast and slow time scales (S_f^2, S_s^2, τ_s) were only included if the simplest model failed to fit the experimental data. In Tables 2 and 3, the model-free parameters are summarized. For 45 residues, the three relaxation parameters T_1, T_2, and NOE could fit the simplest models using the order parameter alone or the order parameter and the effective correlation time within 95% confidence limits (Figure 6D). A conformational exchange parameter was required for 15 residues, and 15 residues were best fit by the two time scale spectral density function. In general, the more extended models were necessary to fit the relaxation data for residues connecting the proposed helices, such as Glu 32−Asp 35, Ile 53−Gln 60, Ser 72−His 74, and Lys 89−Thr 91. Except for Cys 14 and Ser 108, all other residues exhibiting a two time scale

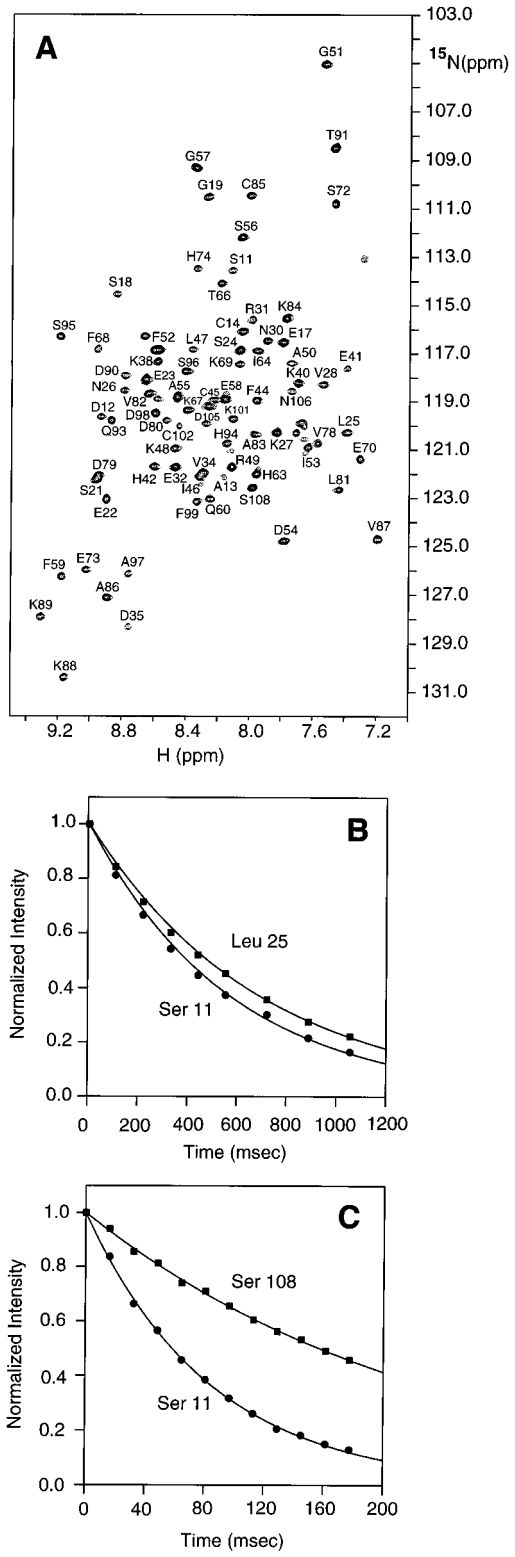

FIGURE 5: (A) Contour plot of a region of the ^{1}H−^{15}N heteronuclear NOE spectrum without presaturation. The cross peak of Glu 76 with ^{1}H and ^{15}N frequencies of 10.2 and 120.8 ppm is not shown. Experimental ^{15}N T_1 (B) and T_2 (C) decay curves of selected residues with substantially different relaxation data. Drawn curves represent calculated least-squares fits to a single-exponential function.

motion with $\tau_s > 1$ ns were located in the loop regions between helices B and C, E and F, and the β-hairpin.

Mapping the Disulfide Linkages by NMR Spectroscopy and CNBr Fragmentation. There are three possible arrangements for disulfide bridges between the four Cys in THP12. Long-range NOEs in the 2D NOESY spectrum between β-protons of Cys 14 and the α-protons of Cys 45 indicated close spatial proximity and suggested the presence of a disulfide bond between these two residues. The second disulfide bond could only be inferred by the presence of NOEs between β_1 and β_2 protons of Cys 85 and the γ protons of Val 103, the residue adjacent to Cys 102.

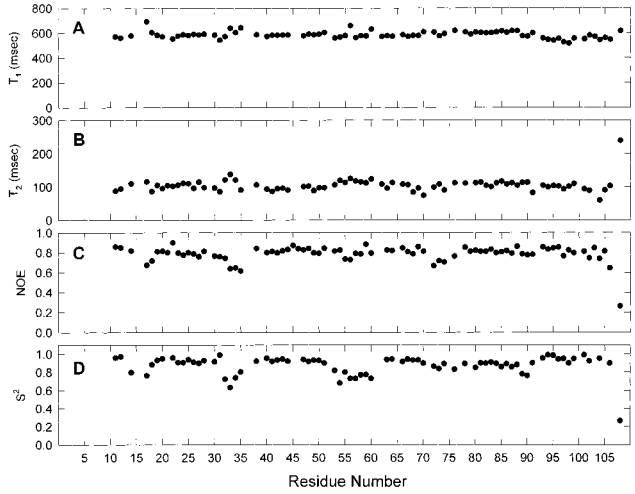

FIGURE 6: Plots of relaxation data as a function of residue number: the measured (A) T_1 and (B) T_2 values, (C) {^{1}H}−^{15}N NOE ratios, and (D) order parameter S^2. The only residues used were those for which the ^{15}N−^{1}H correlation is sufficiently resolved. All data were determined at 600 MHz and 20 °C.

Table 2: Spectral Density Models Used To Fit T_1, T_2, and NOE Data for THP12

model	parameter	THP12
1	S^2	31
2	S^2, τ_e	12
3	S^2, R_{ex}	8
4	S^2, τ_e, R_{ex}	7
5	S_s^2, S_f^2, τ_s	15

The L47M CNBr cleavage experiment was designed to test the prediction made by NMR analysis that Cys 14 is linked to Cys 45, and Cys 85 is linked to Cys 102. The Met for internal cleavage was introduced at Leu 47 because sequence comparisons of THP12 homologs indicated that Leu at this position is not conserved. The L47M mutant eluted from HPLC at approximately the same acetonitrile concentration as wild-type THP, suggesting that the fold of THP12 was not affected by the Leu to Met replacement. This result was confirmed by ^{1}H NMR analysis, which gave virtually identical spectra for both THP12 and L47M (data not shown).

CNBr cleavage of L47M produced peptides of mass 5.52 and 6.92 kDa, consistant with cleavage at the internal Met position. The fragments were separated by HPLC, and their identities as the N- and C-terminal cleavage products, respectively, were confirmed by amino acid analysis. For example, the smaller fragment contained 2% Phe and the larger fragment contained 7% Phe.

CNBr cleavage was done under highly acidic experimental conditions that could potentially cleave the peptide bond between Asp 36 and Pro 37. When the L47M control sample (incubated in 70% formic acid in the absence of CNBr) was analyzed by HPLC it coeluted with THP12. However, on reduction with DTT a fragment of mass 4.23 kDa was

Table 3: Values of Model-Free Parameters for THP12

residue	model	S^2	S_f^2	τ_e	R_{ex}
Ser 11	3	0.959 ± 0.011			1.177 ± 0.403
Asp 12	1	0.972 ± 0.010			
Cys 14	5	0.798 ± 0.048	0.892 ± 0.020	7.220 ± 2.614	
Glu 17	2	0.763 ± 0.008		0.037 ± 0.006	
Ser 18	4	0.884 ± 0.009		0.065 ± 0.011	2.223 ± 0.374
Gly 19	1	0.932 ± 0.007			
Val 20	1	0.950 ± 0.007			
Glu 22	1	0.960 ± 0.015			
Glu 23	2	0.908 ± 0.017		0.090 ± 0.062	
Ser 24	2	0.909 ± 0.009		0.038 ± 0.011	
Leu 25	1	0.937 ± 0.007			
Asn 26	3	0.898 ± 0.010			0.621 ± 0.219
Lys 27	2	0.907 ± 0.009		0.054 ± 0.023	
Val 28	1	0.929 ± 0.007			
Asn 30	2	0.918 ± 0.009		0.046 ± 0.006	
Arg 31	3	0.998 ± 0.005			0.975 ± 0.306
Glu 32	5	0.725 ± 0.035	0.876 ± 0.024	1.877 ± 1.964	
Glu 33	5	0.633 ± 0.030	0.805 ± 0.015	1.349 ± 0.157	
Val 34	5	0.841 ± 0.013		0.097 ± 0.015	
Asp 35	4	0.805 ± 0.023		0.079 ± 0.027	2.221 ± 0.400
Lys 38	1	0.922 ± 0.009			
Lys 40	2	0.951 ± 0.010		0.028 ± 0.008	
Glu 41	3	0.919 ± 0.009			1.780 ± 0.517
His 42	1	0.933 ± 0.007			
Ala 43	1	0.945 ± 0.012			
Phe 44	3	0.923 ± 0.008			1.220 ± 0.377
Leu 47	1	0.942 ± 0.010			
Lys 48	1	0.914 ± 0.010			
Arg 49	3	0.934 ± 0.007			1.366 ± 0.430
Ala 50	1	0.930 ± 0.007			
Gly 51	1	0.902 ± 0.006			
Ile 53	5	0.820 ± 0.043	0.942 ± 0.015	7.220 ± 1.275	
Asp 54	5	0.682 ± 0.053	0.867 ± 0.019	7.220 ± 1.792	
Ala 55	5	0.802 ± 0.031	0.904 ± 0.015	1.398 ± 0.314	
Ser 56	5	0.731 ± 0.030	0.811 ± 0.026	1.243 ± 0.315	
Gly 57	5	0.731 ± 0.047	0.879 ± 0.019	3.408 ± 2.246	
Glu 58	5	0.771 ± 0.042	0.880 ± 0.017	3.239 ± 0.618	
Phe 59	5	0.773 ± 0.050	0.882 ± 0.026	7.220 ± 2.046	
Gln 60	5	0.732 ± 0.051	0.816 ± 0.021	3.608 ± 1.081	
His 63	1	0.943 ± 0.008			
Ile 64	1	0.946 ± 0.010			
Thr 66	1	0.917 ± 0.008			
Lys 67	1	0.944 ± 0.006			
Phe 68	3	0.934 ± 0.009			2.026 ± 0.358
Lys 69	1	0.935 ± 0.008			
Glu 70	3	0.900 ± 0.011			4.126 ± 0.493
Ser 72	4	0.865 ± 0.006		0.082 ± 0.007	0.826 ± 0.330
Glu 73	1	0.932 ± 0.033			
His 74	4	0.893 ± 0.022		0.090 ± 0.041	1.402 ± 0.354
Glu 76	2	0.832 ± 0.018		0.023 ± 0.008	
Val 78	1	0.895 ± 0.005			
Asp 80	2	0.851 ± 0.019		0.076 ± 0.037	
Leu 81	1	0.903 ± 0.007		0.003 ± 0.001	
Val 82	1	0.899 ± 0.007			
Ala 83	1	0.911 ± 0.005			
Lys 84	1	0.896 ± 0.006			
Cys 85	2	0.859 ± 0.013		0.030 ± 0.020	
Ala 86	1	0.892 ± 0.010			
Val 87	2	0.857 ± 0.010		0.029 ± 0.016	
Lys 88	1	0.883 ± 0.008			
Lys 89	5	0.782 ± 0.045	0.890 ± 0.019	2.896 ± 1.681	
Asp 90	5	0.762 ± 0.030	0.879 ± 0.015	4.263 ± 2.308	
Thr 91	4	0.902 ± 0.011		0.024 ± 0.003	2.451 ± 0.178
Gln 93	1	0.956 ± 0.010			
His 94	1	0.990 ± 0.004			
Ser 95	1	0.987 ± 0.009			
Ser 96	1	0.944 ± 0.029			
Ala 97	2	0.950 ± 0.029		13.873 ± 2.253	
Asp 98	1	0.901 ± 0.030			
Phe 99	1	0.962 ± 0.009			
Lys 101	1	0.986 ± 0.008			
Cys 102	4	0.923 ± 0.008		0.079 ± 0.019	1.305 ± 0.494
His 104	4	0.952 ± 0.020		0.154 ± 0.038	6.869 ± 0.806
Asn 106	2	0.899 ± 0.036		0.776 ± 0.225	
Ser 108	5	0.266 ± 0.028	0.782 ± 0.016	1.185 ± 0.028	

[a] $\tau_e = \tau_f$ for models 1−4 and $\tau_e = \tau_s$ for model 5.

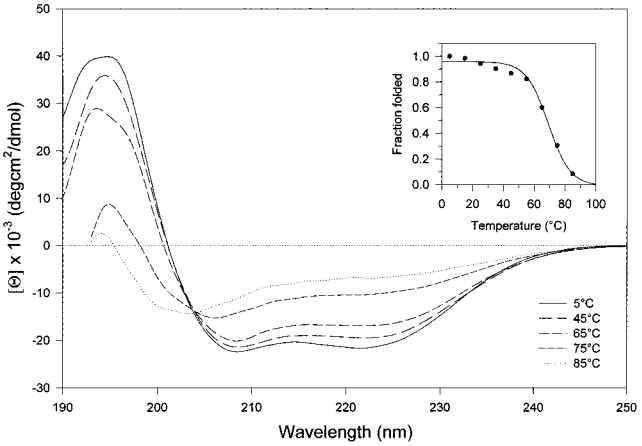

FIGURE 7: Temperature dependence of the circular dichroism spectra as well as the calculated fraction folded (inset) for THP12. The fraction folded protein was calculated as $([\Theta] - [\Theta]_U)/([\Theta]_N - [\Theta]_U)$, where $[\Theta]$ is the observed mean residue ellipticity at 222 nm and $[\Theta]_U$ and $[\Theta]_N$ are the mean residue ellipticities at 222 nm of the unfolded and folded states, respectively.

released in ~5% yield consistent with limited cleavage of the acid-labile Asp 36−Pro 37 linkage. The fact that this peptide was only released on reduction shows that the disulfide bridges were kept intact through the cleavage conditions. Taken together, these findings demonstrated that the disulfide bonds in THP12 are formed between Cys 14 and Cys 45 and between Cys 85 and Cys 102.

CD Spectroscopy. All spectra were measured at a protein concentration of 1.53×10^{-5} M under benign conditions (pH 7.0). The far-ultraviolet CD spectra of THP12 are shown in Figure 7. The spectra in the temperature range of 5−65 °C are characteristic for an α-helical conformation and show two minima at 222 and 206 nm as well as a maximum at 192 nm. Using the equation for chain-length dependency of the expected molar ellipticity (*35*), a helical content of 64% was calculated for 25 °C.

In order to determine the stability of this highly helical protein, the equilibrium between folded and unfolded state as a function of temperature were investigated (*43*). As the temperature was increased, the mean molar ellipticity at 222 nm decreased, which is consistent with a loss of helicity (Figure 7). The isodichroic point at 202 nm is characteristic for a two-state transition from the α-helix to random coil conformation (*44*), indicating thermal unfolding with increasing temperature. A classical evaluation of the unfolding curve of THP12 (Figure 7, inset) yielded a transition midpoint (T_M) of 67 °C.

DISCUSSION

THP12 present in hemolymph from the mealworm beetle *Tenebrio molitor* is the first member of the insect pheromone- and odorant-binding proteins for which the solution structure and backbone dynamics have been characterized. Based on the matching Cys and Pro residues with identical spacings (Figure 8), and similar hydropathy plots (data not shown), the likelihood of structural and/or functional similarity between THP12 and the B proteins secreted by tubular accessory glands of the male mealworm beetle *Tenebrio molitor* is very high.

The secondary structure of THP12 was determined by interpretation of distinctive sequential and medium-range NOE patterns, chemical shifts of C_α protons, and three-bond

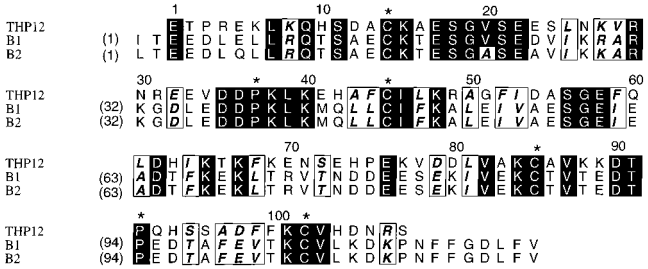

FIGURE 8: Multiple alignment of amino acid sequences of THP12 and B1 and B2 proteins (*15*) from *Tenebrio molitor* using Alscript (*58*). Signal sequences are not shown. Conserved Cys and Pro residues are labeled with an asterisk. Identical residues are highlighted by a black background, while conservatively conserved residues (bold-italic) are boxed.

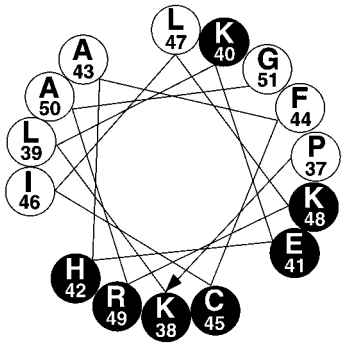

FIGURE 9: Amino acid sequence of the amphipathic helix C represented in a helical wheel projection. The distribution of polar residues is indicated by black circles.

$J_{HNH\alpha}$ coupling constants. While the N-terminal residues Glu 1−Lys 8 do not appear to form any secondary structure, the majority of the remaining residues in THP12 are involved in α-helical structure. NMR data suggest that THP12 consists of six α-helices interspaced by four loops and one *β*-hairpin. In particular, the first turn of helix A is capped by Ser 11, which can form a side-chain hydrogen bond to the backbone NH of Cys 14 at position $i + 3$ (*45*). On the basis of medium-range NOE connectivities, upfield $C_\alpha H$ chemical shifts, and small $^3J_{HNH\alpha}$ coupling constants, helix A extends to residue Gly 19. Applying identical criteria, helix B was identified from residue Ser 21 to Arg 31. The lack of $d\alpha N(i, i + 3)$ NOEs, typical for α-helical structure, between residues Ala 16 and Gly 19 and residues Glu 17 and Val 20, as well as the positive $C_\alpha H$ chemical shift difference at residue 21, suggests a short loop or a hinge between helices A and B. Due to the lack of an amide proton and prevention of α-helix-typical hydrogen bonds between the carbonyl of residue $(i - 3)$ and the NH $(i + 1)$, proline frequently found in first turns of α-helices (*46*, *47*). For helices C and D we found continuous pathways of unambiguous $dNN(i, i + 1)$, $d\alpha N(i, i + 1)$, $d\alpha N(i, i + 3)$, and $d\alpha N(i, i + 4)$ NOE connectivities starting with residue Pro 37 and extending up to residue Gly 51 in the case of helix C and from residue Leu 61 to residue Glu 70 for helix D. This result is also supported by negative $C_\alpha H$ chemical shift differences and small $^3J_{HNH\alpha}$ coupling constants. With the exception of Lys 40, the distribution of hydrophobic and hydrophilic residues in helix C is characteristic of an amphipathic helix with residues Leu 39, Ala 43, Phe 44, Ile 46, Leu 47, and Ala 50 facing the hydrophobic side (Figure 9). The patterns of sequential, medium-, and long-range NOEs, $C_\alpha H$ chemical shift differences, and $^3J_{HNH\alpha}$ coupling

FIGURE 10: Schematic representation of the 3D model of THP12. First or last residues of the helices are labeled. The position of Lys 40 is indicated by the NH_3^+ label.

constants for the region Phe 52−Gln 60 all give indications for a β-hairpin structure connecting helices C and D. The C-terminal helices D and E are again clearly identifiable from the patterns of short- and medium-range NOEs and are further supported by $C_\alpha H$ chemical shifts and coupling constants. Overall, 67 residues of THP12 are involved in helical arrangement and specify an α-helical content of 62%. This observation correlates well with the estimated α-helical content of 64% by CD.

Assuming isotropic tumbling for the protein, the overall correlation time should correlate with its molecular weight. The measured value of 7.2 ns for a 108-residue protein is consistent with those obtained for ribonuclease H (9.7 ns, 155 residues), SH2 domain (6.5 ns, 105 residues), calbindin (4.1 ns, 75 residues), and IgG-binding domain (3.3 ns, 56 residues) (*30, 48*−*50*).

The averaged value of the generalized order parameter for the 75 analyzed residues of THP12 was determined to be 0.87 ± 0.10, similar in size to values observed in stable helices of soluble proteins (*30, 48, 49, 51*). Averaged values of the generalized order parameters were also calculated for individual secondary structure elements in THP12. The six helices displayed values between 0.88 and 0.95 (helix A, 0.87; B, 0.93; C, 0.93, D, 0.93; E, 0.88; and F, 0.95), suggesting fairly rigid conformations in these regions. In contrast, significantly lower values of 0.72 and 0.75 indicate a considerable degree of internal motion in the loop region between helices B and C and the β-hairpin between helices C and D. Overall, a good correlation between secondary structure elements and the magnitude of the generalized order parameter was found.

Based on the determined secondary structure and the results from hydrogen exchange and backbone dynamic studies, a three-dimensional model of THP12 is proposed (Figure 10). The hydrophobic core of the protein is formed by an up-and-down arrangement of helices C, D, and E, where the central amphipathic helix C makes hydrophobic contacts to helices D and E. For example, helix E contributes residues Val 82, Ala 83, Ala 86, and Val 87 to the hydrophobic core. The amino acid composition of helix D with a high content of hydrophilic amino acids leads to a somewhat weaker amphipathicity with residues Leu 61, Ile 64, and Phe 68 corresponding to a small hydrophobic face. This hypothesis of folding in THP12 is strongly supported by slowly exchanging amide protons of residues Ile 46, Leu 47, and Lys 48 (helix C) and Val 82, Lys 84, Cys 85, Ala 86, and Val 87 (helix E). Slowly exchanging amide protons were also observed for residues Ile 53, Asp 54, Gly 57, and Glu 58, confirming solvent shielded residues in the β-hairpin structure. Due to the lack of long-range NOE connectivities, the arrangement of helices A, B, and F with respect to the central three-helix bundle is ambiguous. However, close proximity between helices A and C as well as E and F can be expected by the formation of disulfide bonds connecting helix A and C (Cys 14−Cys 45) and helix E and F (Cys 85−Cys 102). Further evidence is given by interhelical contacts between helices D and E (Phe 68−Asp 79 and Lys 69−Val 78).

The binding of hydrophobic pheromones and odorants in aqueous media is probably based on a specific arrangement of hydrophobic and hydrophilic subdomains inside and outside the binding protein. However, little structural information is available on binding of hydrophobic insect pheromones to their binding proteins. As discussed by Du and Prestwich (*13*), the preferred binding site to pheromones is the most hydrophobic domain in all PBPs and is predicted to be a helix−loop/turn−helix motif. Moreover, a favorable orientation of hydrophilic residues in the binding interface and the resulting electrostatic interactions to the polar head group of pheromones is suggested to contribute to the high binding specificity.

Given that the suggested pheromone-binding site in PBPs is a helix−loop/turn−helix motif, it is tempting to speculate about the role of the helix C−β-hairpin−helix D region in THP12. First, the most hydrophobic domain in THP12 is identified as the protein core created by an up-and-down arrangement of the three amphipathic helices C, D, and E. Second, considering the results from backbone dynamic studies, the residues Asp 54, Gly 57, and Glu 58 of the β-hairpin region between helices C and D show increased mobility in a nanosecond time scale. On the other hand, hydrogen exchange studies indicate protection against solvent exchange even for residues Asp 54, Gly 57, and Glu 58, suggesting a solvent-shielded location of residues in the β-hairpin. As discussed by many authors, conformational flexibility is important for ligand binding, particularly for binding of ligands in a central cavity of proteins (*52, 53*). This is also supported by backbone dynamic studies of protein−ligand complexes, where a stiffening of the protein structure upon ligand binding was observed (*30, 54*). Third, despite the hydrophilic character, the conserved residue Lys 48 is highly protected against amide proton exchange, which could suggest a role as charge donor in the hydrophobic core for interactions with negatively charged ligands. Alternatively, Lys 40 could fulfill this role since it is also located in the hydrophobic face of the amphipathic helix C. Fourth, together with the expected sensitivity of this highly charged protein against intermolecular electrostatic interactions resulting in a conformational transfer upon binding to biological membranes, the concept of a ligand entry portal similar to fatty acid binding proteins (*55*−*57*) involving the β-hairpin region seems reasonable.

In conclusion, ^{15}N T_1, T_2, and NOE experiments indicate that this highly helical protein is fairly rigid on a picosecond time scale except for the loop regions and the β-hairpin. It has been suggested that the combination of restricted α-helices as well the mobility of this β-hairpin in a

nanosecond time scale may be important for binding of small hydrophobic ligands to THP12.

ACKNOWLEDGMENT

We thank W. Tang and J. G. Baust for the gift of the THP12 DNA clone, F. Delaglio and D. Garret for the computer programs NMRpipe and PIPP, and Alexander Wu for assistance with mapping the disufide bonds. We also thank L. Kay and N. Farrow for providing the source code for the pulse sequences and software for relaxation analysis.

REFERENCES

1. Pelosi, P. (1994) *Crit. Rev. Biochem. Mol. Biol. 29*, 199−228.
2. Vogt, R. G., and Riddiford, L. M. (1981) *Nature 293*, 161−163.
3. Vogt, R. G. (1987) in *Pheromone Biochemistry* (Prestwich, G. D., and Blomquist, G., Eds.) pp 385−431, Academic Press, Orlando, FL.
4. Prestwich, G. D. (1993) *Arch. Insect Biochem. Physiol. 22*, 75−83.
5. Pikielny, C. W., Hasan, G., Rouyer, F., and Rosbash, M. (1994) *Neuron 12*, 35−49.
6. McKenna, M. P., Hekmat-Scafe, D. S., Gaines, P., and Carlson, J. R. (1994) *J. Biol. Chem. 269*, 16340−16347.
7. Vogt, R. G., Prestwich, G. D., and Lerner, M. R. (1991) *J. Neurobiol. 22*, 74−84.
8. Vogt, R. G., Rybczynski, R., and Lerner, M. R. (1991) *J. Neurosci. 11*, 2072−2984.
9. Gyorgi, T. K., Roby-Shemlovich, A. J., and Lerner, M. R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9851−9855.
10. Raming, K., Krieger, J., and Breer, H. (1989) *FEBS Lett. 256*, 215−218.
11. Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Philips, E. V., Cavaggioni, A., Findlay, J. B. C., and North, A. C. T. (1992) *Nature 360*, 186−188.
12. Bianchet, M. A., Bains, G., Pelosi, P., Pevsner, J., Snyder, S. H., Monaco, H. G., and Amzel, L. M. (1996) *Nat. Struct. Biol.* 934−939.
13. Du, G., and Prestwich, G. D. (1995) *Biochemistry 34*, 8726−8732.
14. Feixas, J., Prestwich, G. D., and Guerro, A. (1995) *Eur. J. Biochem. 234*, 521−526.
15. Paesen, G. C., and Happ, G. M. (1995) *Insect Biochem. Mol. Biol. 25*, 401−408.
16. Paesen, G. C., Feng, X., and Happ, G. M. (1996) *Biochim. Biophys. Acta 1293*, 171−176.
17. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105−132.
18. Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Kunkel, T. A., Roberts, J. D., and Zarkour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
20. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry 28*, 6150−6156.
21. Muhandiram, D. R., Farrow, N. A., Xu, G.-Y., Smallcombe, S. H., and Kay, L. E. (1993) *J. Magn. Reson. B 102*, 317−321.
22. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
23. Archer, S. H., Ikura, M., Torchia, D. A., and Bax, A. (1991) *J. Magn. Reson. 95*, 636−641.
24. Palmer, A. G., Rance, M., and Wright. P. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.
25. Kay, L. E. (1995) *Curr. Opin. Struct. Biol. 5*, 674−681.
26. Shaka, A. J., Keeler, J., Frenkiel, T., and Freeman, R. (1983) *J. Magn. Reson. 52*, 335−338.
27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
28. Garret, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) *J. Magn. Reson. 95*, 214−220.
29. Bodenhausen, G., and Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−189.
30. Farrow, N. A., Muhandiram, D. R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
31. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559−4570.
32. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4558.
33. Clore, G. M., Driscoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990) *Biochemistry 29*, 7387−7401.
34. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) *J. Am.Chem. Soc. 112*, 4989−4991.
35. Chen, Y., Yang, J., and Chau, K. (1974) *Biochemistry 13*, 3350−3359.
36. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, J. Wiley and Sons, New York.
37. Clayden, N. J., and Williams, R. J. P. (1982) *J. Magn. Reson. 49*, 383−396.
38. Dalgarno, D. C., Levine, B. A., and Williams, R. J. P. (1983) *Biosci. Rep. 3*, 443−452.
39. Jimenez, M. A., Nieto, J. L., Herranz, R. J., and Santoro, J. (1987) *FEBS Lett. 221*, 320−324.
40. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
41. Ramirez-Alvarado, M., Blanco, F. J., and Serrano, L. (1996) *Nat. Struct. Biol. 3*, 604−612.
42. Williamson, M. P. (1990) *Biopolymers 29*, 1423−1431.
43. Shortle, D. (1989) *J. Biol. Chem. 264*, 5315−5318.
44. Padmanaban, S., Marqsee, S., Ridgway, T., Laue, T. M., and Baldwin, R. L. (1990) *Nature 344*, 268−270.
45. Harper, E. T., and Rose, G. D. (1993) *Biochemistry 32*, 7605−7609.
46. MacArthur, M. W., and Thornton, J. M. (1991) *J. Mol. Biol. 218*, 397−412.
47. Richardson, J. S., and Richardson, D. C. (1988) *Science 240*, 1548−1652.
48. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) *J. Mol. Biol. 246*, 144−163.
49. Kördel, J., Skelton, N. J., Akke, M., Palmer, A. G., and Chazin, W. J. (1992) *Biochemistry 31*, 4856−4866.
50. Barchi, J. J., Grasberger, B., Gronenborn, A. M., and Clore, G. M. (1994) *Protein Sci. 3*, 15−21.
51. Williams, K. A., Farrow, N. A., Deber, C. M., and Kay, L. E. (1996) *Biochemistry 35*, 5145−5157.
52. Heinemann, B., Andersen, K. V., Nielsen, P. R., Bech, L. M., and Poulsen, F. M. (1996) *Protein Sci. 5*, 13−23.
53. Markus, M. A., Hinck, A. P., Huang, S., Draper, D. E., and Torchia, D. A. (1997) *Nat. Struct. Biol. 4*, 70−77.
54. Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993) *Nature 362*, 87−89.
55. Kim, H.-K., and Storch, J. (1992) *J. Biol. Chem. 267*, 77−82.
56. Kim, H.-K., and Storch, J. (1992) *J. Biol. Chem. 267*, 20051−20056.
57. Herr, F. M., Aronson, J., and Storch, J. (1996) *Biochemistry 35*, 1296−1303.
58. Barton, G. J. (1993) *Protein Eng. 6*, 37−40.

BI971529K