A Complex Family of Highly Heterogeneous and Internally Repetitive Hyperactive Antifreeze Proteins from the Beetle *Tenebrio molitor*[†,‡]

Yih-Cherng Liou,[§] Pierre Thibault,[∥] Virginia K. Walker,[⊥] Peter L. Davies,[§,⊥] and Laurie A. Graham*[,§]

Departments of Biochemistry and Biology, Queen's University, Kingston, Ontario, Canada K7L 1N6, and the Institute for Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario, Canada K1A 0R6

Received March 17, 1999; Revised Manuscript Received June 18, 1999

ABSTRACT: We have previously identified a Thr- and Cys-rich thermal hysteresis (antifreeze) protein (THP) in the beetle *Tenebrio molitor* that has 10−100 times the freezing point depression activity of fish antifreeze proteins. Because this 8.4 kDa protein is significantly different in its properties from THP preparations previously reported from this insect, a thorough search was undertaken for other antifreeze types. Many active proteins were observed, but all appeared to be isoforms of the THP that differed in their number of 12-amino acid repeats (consensus sequence CTxSxxCxxAxT), amino acid substitutions, and N-linked glycosylation. Mass spectral analysis has matched most of these isoforms with cDNA sequences of 17 different clones from a larval fat body library that encode eight different mature THPs containing 84, 96, or 120 amino acids. Genomic Southern blots suggest there may be 30−50 tightly linked copies of the gene, which is a signature consistently seen with unrelated fish antifreeze protein genes, and one that has been associated with the need to rapidly increase gene product in response to climate change. A three-dimensional model is proposed for the fully disulfide-bonded structure of *T. molitor* THP, which can accommodate addition or deletion of 12-amino acid repeats. The structure is a β-helix that places most of the Thr in a regular array on one side of the protein to form a putative ice-binding surface.

The common yellow mealworm beetle, *Tenebrio molitor*, is a pest of stored grain products in cold temperate regions (*1*). Larvae are freeze-intolerant but are able to survive subzero temperatures by supercooling to an average of −12 °C (*2*). A second line of defense against freezing is the thermal hysteresis (TH)[1] activity of their hemolymph and other extracellular fluids (*3*). This activity allows the insects to depress their freezing points in the presence of ice or ice nucleators. It is defined as the temperature difference (degrees Celsius) between the freezing (nonequilibrium) and melting (colligative) temperatures of a solution containing ice. The reported values for TH in *T. molitor* hemolymph are quite varied, ranging from 0.68 to 10 °C according to the method of measurement and the conditions under which the insects were reared (*4*−*6*). By visual inspection of ice growth and melting with a microscope, we have measured 5.5 °C of TH in the hemolymph of *T. molitor* larvae (*7*).

In the three decades since TH was first described in *T. molitor* (*3*), a number of groups have attempted to purify *T. molitor* THPs (*4*, *8*−*14*). These putative THPs differed markedly in their amino acid compositions, and the TH activity was low. Recently, an 8.4 kDa protein with a specific activity approximately 10−100 times that of fish antifreeze proteins (AFPs) was isolated in microgram quantities from the hemolymph of *T. molitor* larvae (*7*). Internal sequencing allowed the design of degenerate oligonucleotide primers, which led to the isolation of four THP cDNA clones. The overriding structural feature of these Cys-, Thr-, and Ser-rich proteins was a tandemly repeated 12-amino acid sequence, CTxSxxCxxAxT. Bacterial expression of one of these cDNAs produced a protein with high TH activity, thereby unequivocally identifying it as an insect AFP.

Subsequently, sequences of several THP isoforms from another beetle species, *Dendroides canadensis*, have been reported (*15*, *16*). These THPs are clearly homologous to the *T. molitor* THPs, sharing 40−66% amino acid identity and being composed of a tandemly arranged 12-amino acid repeat with the same consensus sequence but with occasional insertion of an additional amino acid. In contrast, the other recently discovered insect THP, isolated from the spruce budworm, *Choristoneura fumiferana*, does not appear to be homologous and does not have a well-defined repeating sequence (*17*). However, it too is hyperactive relative to fish AFPs, Thr-, Ser-, and Cys-rich, fully disulfide-bonded, and of a similar size (9 kDa).

In this paper we describe the first comprehensive characterization of a large family of THPs, and their corresponding cDNA sequences, from an insect. Together, these THPs

[†] This work was supported by grants from the Medical Research Council of Canada to P.L.D. and the National Sciences and Engineering Research Council of Canada to V.K.W.

[‡] The sequences of cDNAs encoding each protein isoform have been deposited in GenBank under Accession Numbers AF010329 through AF010332 and AF159114 through AF159118.

* To whom correspondence should be addressed: Phone 613-533-2984; Fax 613-533-2497; E-mail 3lag6@biology.queensu.ca.

[§] Department of Biochemistry, Queen's University.

[∥] National Research Council of Canada.

[⊥] Department of Biology, Queen's University.

[1] Abbreviations: THP, thermal hysteresis protein; TH, thermal hysteresis; AFP, antifreeze protein; HPLC, high-performance liquid chromatography; ESMS, electrospray mass spectrometry; H, high apparent molecular weight gel-exclusion peak; L, low apparent molecular weight gel-exclusion peak.

10.1021/bi990613s CCC: $18.00
Published on Web 08/12/1999

can account for the extremely high TH activity observed in *T. molitor* hemolymph. Both sequence polymorphism and posttranslational modification contribute to isoform diversity. In addition, a model for the structure of this THP has been developed that can accommodate variation in the number of 12-amino acid repeats.

EXPERIMENTAL PROCEDURES

Insect Rearing and Hemolymph Collection. *T. molitor* was reared in the laboratory at 22 °C in darkness on wheat bran and provided with water-soaked paper towels 2−3 times a week as their only source of moisture. Hemolymph was obtained from >100 large active larvae, most of which were in the final larval instar. A leg was severed and 2−20 μL of hemolymph was collected from each insect with a microcapillary tube. Hemolymph was diluted 1:1 with ice-cold hemolymph buffer [50 mM Tris-HCl (pH 8.0), 150 mM NaCl and 1 mM phenylthiocarbamide], briefly centrifuged, and quick-frozen in liquid N_2.

THP Purification. Pooled, diluted hemolymph (1.2−2.5 mL) was applied to an S-100 Sephacryl HR column (92 cm × 1.6 cm; Pharmacia), equilibrated with hemolymph buffer without phenylthiocarbamide, and eluted with the same buffer at a flow rate of 0.25 mL/min. The absorbance of each 1.5 mL fraction was measured at 230 nm, and selected fractions were assayed for TH activity in the column buffer. Combined fractions from size-exclusion chromatography of a 2.5 mL sample were rechromatographed by reversed-phase HPLC on a C18 semipreparative column (Vydac), using a linear gradient of 10−50% B (80% acetonitrile in 0.05% trifluoroacetic acid) in 20 min, with a flow rate of 3 mL/min. Selected fractions (3 mL) containing TH activity were pooled and rechromatographed as above but on an analytical column (Vydac) with a linear gradient of 10−30% B in 80 min at a flow rate of 1 mL/min (0.5 mL/fraction). Fractions (0.5 mL) were dried and then resuspended in 20 μL of 0.1 M NH_4HCO_3 (pH 8.0) for TH measurements.

Antifreeze Activity Measurements. TH was measured with a nanoliter osmometer (Clifton Technical Physics, Hartford, NY) following standard procedures (*18*) with the following exception. For each fraction, a trial reading (not reported) was taken at a cooling rate of 0.1 °C every 15 s. Subsequent readings were done at this cooling rate to within 0.4 °C of the trial freezing point before reverting to the standard cooling rate of 0.02 °C every 15 s.

THP Characterization. Electrospray mass spectrometry (ESMS) experiments were conducted on a Q-TOF (Micromass, Manchester, U.K.) hybrid quadrupole/time-of-flight instrument (*19*). An HPLC HP1100 (Hewlett-Packard, Palo Alto, CA) was used for both flow injection and on-line liquid chromatography−ESMS experiments. Chromatographic separations were done on a 15 cm × 0.32 mm PepMap capillary column (LC Packings, San Francisco, CA) with a linear gradient of 5−95% acetonitrile containing 0.2% formic acid over 40 min. Conventional mass spectra were obtained by operating the quadrupole in a radio frequency-only mode while a pusher electrode was pulsed (∼16 kHz) to transfer all ions to the time-of-flight analyzer. For tandem mass spectrometry experiments, precursor ions identified in a preliminary survey scan were selected by the first quadrupole while a pusher electrode was pulsed (∼16 kHz) to transfer fragment ions formed in the radio frequency-only hexapole cell to the time-of-flight analyzer. Detection was done with a dual-stage microchannel plate and a time-to-digital converter operating at 1 GHz (Precision Instruments, Knoxville, TN). Mass spectral resolution was typically 4000−5000. Scan durations of 1 and 2 s were set for conventional and MS−MS mass spectral acquisition, respectively. Collisional activation was performed with argon collision gas with a 25 V offset between the DC voltage of the entrance quadrupole and the radio frequency-only hexapole cell. Data were acquired and processed in the Mass Lynx Windows NT-based data system. Amino acid analysis was done by the Harvard Microchemistry Facility, Cambridge, MA, for both protein composition and quantitation.

PCR Amplification of THP cDNA. Degenerate sense oligonucleotides were designed based on the amino acid sequence of an internal THP fragment (*7*). PCR was done on an aliquot (20 000 plaque-forming units) of a λ-Zap cDNA library, derived from the fat body mRNA of *T. molitor* larvae reared at 22 °C (*20*), with 100 pmol of the degenerate primer and 5 pmol of the vector primer, T7. Following an initial 5 min denaturation at 95 °C, 30 cycles of PCR were done under the following conditions: 95 °C for 1 min, 56 °C for 1 min, and 72 °C for 5 min, ending with a 20 min extension at 72 °C. Selected DNA fragments were purified and reamplified with a nested degenerate primer and T7 under similar PCR conditions. Purified bands were sequenced using an ABI Prism dye terminator cycle sequencing kit and AmpliTaq DNA polymerase FS, on an ABI 377 DNA sequencer system (Applied Biosystems, Foster City, CA).

cDNA Library Screening. Approximately 2×10^5 plaques, derived from 10 individually amplified aliquots from the primary larval fat body cDNA library, were screened following standard methodologies. Plaques were lifted onto colony/plaque screen hybridization transfer membranes (NEN Research Products), treated according to the manufacturer's instructions, and hybridized with a portion of the THP PCR fragment that was ^{32}P-labeled by using the random primers DNA labeling system (BRL Life Technologies). The membranes were prehybridized in 6× SSPE, 10× Denhardt's, 0.5% SDS, and 200 mg/mL salmon sperm DNA at 60 °C for 6 h and then hybridized in the same solution plus probe for 18 h at 60 °C. The membranes were washed 4 times for 20 min in 2× SSC and 0.1% SDS at 60 °C and twice in 0.5× SSC and 0.1% SDS at 60 °C. Exposures were taken at −70 °C with Kodak X-OMAT AR film and a Dupont Cronex Lightning Plus intensifying screen. Secondary screening of selected plaques was performed as above, and phage from isolated positive plaques were subjected to in vivo excision by R408 helper phage (Stratagene) as per manufacturer's instructions. Double-stranded pBluescript DNA was purified from cultures derived from single colonies by the QIAprep spin plasmid miniprep kit (Qiagen) and sequenced as above by using the T7 and T3 vector primers and, when necessary, internal primers.

Southern Blot Analysis. Genomic DNA was isolated as described by Ashburner (*21*) with the addition of an RNase digestion step of 100 μg/mL RNase A for 1 h at 37 °C. DNA from individual larvae was digested with *Hin*dIII. Digests (10 μg of DNA/lane) were electrophoresed in 0.7% agarose, and the gel was blotted onto a Hybond-N^+ membrane (Amersham). The membrane was hybridized and washed

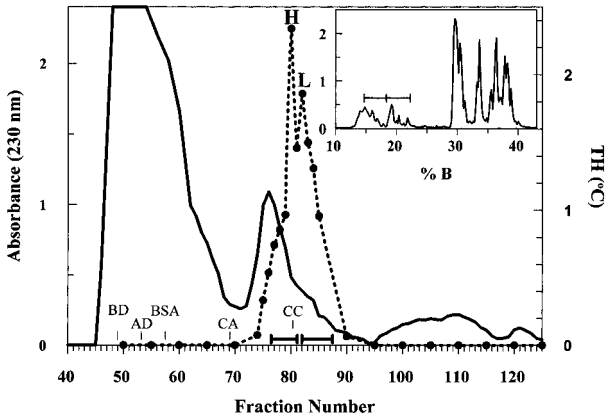

FIGURE 1: Size-exclusion chromatography and semipreparative HPLC of *T. molitor* THPs. *T. molitor* hemolymph was fractionated on Sephacryl S100 as described in Experimental Procedures. Absorbance at 230 nm is indicated by a solid line, and TH values of selected fractions (average of three readings) are indicated by solid circles (●) joined by a dotted line. Fractions in which molecular mass markers (Sigma) eluted are marked by acronyms: BD = blue dextran (2000 kDa), AD = alcohol dehydrogenase (150 kDa), BSA = bovine serum albumin (66 kDa), CA = carbonic anhydrase (29 kDa), and CC = cytochrome *c* (12 kDa). Fractions were pooled from each of the H and L peaks (indicated by letters and bars) and applied to a semipreparative C18 HPLC column. The inset shows fractionation of the H peak only, with the bar indicating TH-active fractions.

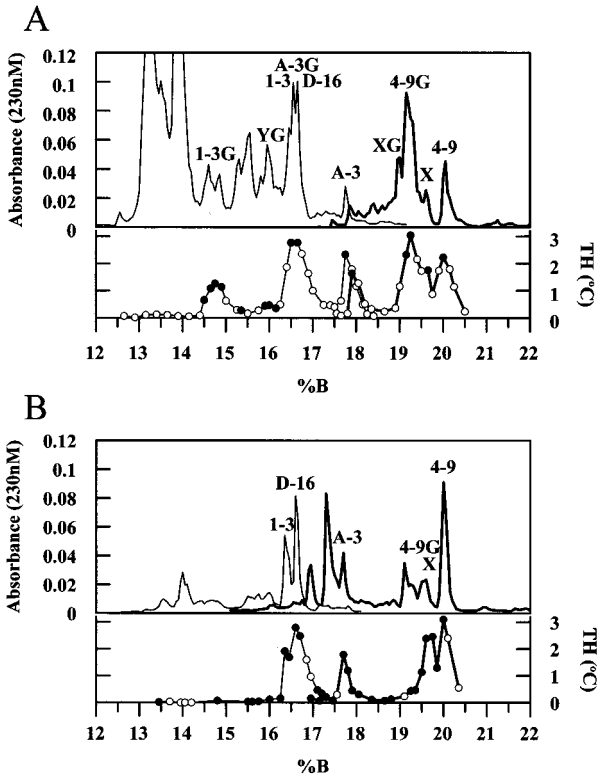

FIGURE 2: Reversed-phase HPLC of *T. molitor* THPs. (a) TH-active fractions from semipreparative C-18 reversed-phase HPLC of the H peak (Figure 1 inset) were subdivided into two groups and individually rechromatographed on an analytical C-18 reversed-phase column (thin and thick lines). Absorbance and TH activity are shown in the top and bottom panels, respectively. TH activity was measured twice in fractions marked with solid black dots (●) and once in fractions marked with open dots (○) as described in Experimental Procedures. TH activity does not increase linearly with THP concentration, especially at higher TH values, where large increases in concentration are needed to double the activity (*7*). Therefore, the disproportionate activity shown by smaller peaks does not necessarily indicate higher specific activity. These peaks are denoted in Table 1 by the prefix H and the % B at which they eluted. They are labeled here with the names of cDNA clones that were deduced to encode these proteins. Peaks without matches are labeled X or Y. G indicates that the isoform is glycosylated. (B). Conditions were as in panel A except that the L peak was rechromatographed and the resulting peaks are denoted by the prefix L.

essentially as described by Graham et al. (*22*), but the final wash was done with 0.2% SSPE and 0.1% SDS for 15 min at 60 °C. The portion of the cDNA YL-1 (*7*) encoding the mature protein sequence was used as the probe and was ^{32}P-labeled by random priming as above.

Effect of Additives on TH Activity. Hemolymph was diluted in buffer, 0.1 M NH_4HCO_3 (pH 8.0) containing 1 mM phenylthiocarbamide and 0.1 mM phenylmethane sulfonyl fluoride, to obtain a starting TH activity of approximately 1 °C. For the chelation study, the sample was dialyzed extensively against the same buffer with and without 10 mM EDTA (or 10 mM *o*-phenanthroline). To study the effects of metal ions, EDTA-free samples were incubated at 22 °C for 1 h with $ZnCl_2$ or $CaCl_2$ to a final concentration of 1 mM. To test the effects of reduction, dithiothreitol (10 mM) or water was also added to EDTA-free samples, which were then incubated for various periods of time at 37 °C. To modify any free sulfhydryl groups, 10 mM *N*-ethylmaleimide was added to THP samples, which were then incubated at 25 °C for 1 h.

RESULTS

Purification of T. molitor THP. Fractionation of 1.2 mL of diluted (1:1) hemolymph by size-exclusion chromatography resulted in two overlapping peaks of TH (Figure 1). This division into two peaks of high (H) and low (L) apparent molecular mass was reproducible and was not due to one or two aberrant measurements. The TH activity was well separated from the bulk of the hemolymph proteins, which eluted in the void volume peak, and was partially separated from a smaller peak centered at fraction 75.

Analysis of peak H by semipreparative HPLC showed that it contained two sets of proteins (Figure 1, inset). The leading set eluted before 22% B and contained virtually all the TH-active material, while the trailing set, containing the bulk of the protein, eluted between 29% and 40% B and consisted of numerous 12−13 kDa proteins. TH-active fractions (shown by the bars) were further resolved on an analytical HPLC column (Figure 2A). Numerous peaks displaying TH activity eluted between 14.5% and 22.5% B. The masses of the principal components within active peaks were determined by ESMS and ranged from 8.4 to 10.8 kDa (Table 1). Fractionation of peak L, from size-exclusion chromatography (Figure 1), by semipreparative HPLC produced a similar profile to that obtained with peak H, except that less trailing material was present (data not shown). Following analytical HPLC, a group of THPs eluting at 16−22.5% B were observed (Figure 2B). Because of overlap between the H and L peaks obtained from gel-exclusion chromatography (Figure 1), some isoforms were present in both profiles as judged by their identical mass and elution. In general, this second group (L peak) contained primarily 8.4 kDa THPs, while the H peak was enriched in larger isoforms (Table 1). All active fractions produced the same ice crystal morphol-

Table 1: Observed and Deduced Masses of Proteins in Various Fractions Following Analytical HPLC of the High (H) and Low (L) Molecular Mass-Size Exclusion Chromatography Peaks[a]

fraction	observed mass (Da ± 1 Da) and abundance[b]	observed mass −sugars (Da)	calculated mass −sugars (Da)	cDNA and *glycan*[c] assignment
L16.35	8477.5	n/c[d]	8477.2	1−3
L16.70	8399.5, 8477.5 (0.2), 8417.4 (0.2)	n/c, n/c, n/c	8399.0, 8477.2, ?	D-16, 1−3, ?
L17.30	3584.4[e]			
L17.70	3582.7[e], 6087.7[f], 9769.6	n/c	9769.6	A-3
L19.25	6073.5[f]			
L19.50	6072.5[f], 9211.6	8351.1	8351.0	4-9 + $carb_2$
L19.60	8369.5	n/c	?	X
L20.00	8351.5	n/c	8351.0	4-9
H14.50	9499.8	8477.2	8477.2	1-3 + $carb_1$
H14.75	9499.8, 9337.2 (0.9)	8477.2, 8476.7	8477.2	1-3 + $carb_{1,2}$
H14.90	9337.4	8476.9	8477.2	1-3 + $carb_2$
H16.00	10645.3	9784.8	?	Y + $carb_2$?
H16.40	8478.6, 10794.1 (0.6)	n/c, 9771.5	8477.2, 9769.6	1-3, A-3 + $carb_1$
H16.50	8478.6, 10631.2 (0.9), 10793.5 (0.4)	n/c, 9770.7, 9770.9	8477.2, 9769.6	1-3, A-3 + $carb_{2,1}$
H16.65	8399.2, 10631.0 (0.7)	n/c, 9770.5	8399.0, 9769.5	D-16, A-3 + $carb_2$
H17.65	[8398.0], [8351.9]	n/c, n/c	8399.0, 8350.9	D-16, 2−14
H17.75	9770.5	n/c	9769.5	A-3
H18.15	3582.7[e], 9768.1, 6087.7[f]	n/c	9769.5	A-3
H18.90	9389.8, 9227.4 (0.4)	8367.2, 8366.9	(8366.5)	X + $carb_{1,2}$
H19.00	9227.4, 9389.8 (0.7)	8366.9, 8367.2	(8366.5)	X + $carb_{2,1}$
H19.15	9374.1, 9212.5 (0.4), [10642.7], [10804.5]	8351.5, 8352.0, 9782.2, 9781.9	8351.0, 9781.6	4-9 + $carb_{1,2}$, 5-15 + $carb_{2,1}$
H19.25	9212.5, 9374.1 (0.85), [10642.8]	8352.0, 8351.5, 9782.3	8352.0, 8351.5, 9781.6	4-9 + $carb_{2,1}$, 5-15 + $carb_2$
H19.65	8366.5	n/c		X
H20.00	8351.4	n/c	8351.0	4-9

[a] Fractions are named for the percent B at which they eluted. In the case of glycoproteins, the deduced masses of matching cDNAs were calculated with and without (−sugars) the attached glycan. THPs without matching cDNAs are labeled X or Y. [b] Relative abundance of THPs was estimated from the height of the peaks obtained during mass spectrometry and is expressed as a fraction (in parentheses) of the peak height of the most abundant THP. Otherwise, they are listed in order of decreasing peak height with minor components enclosed in square brackets. [c] $carb_1 = GlcNAc_2 Man_2 Fuc_2$ (1022.6 Da), $carb_2 = GlcNAc_2 Man_1 Fuc_2$ (860.5 Da). [d] No change in mass as sugars were not present. [e] Not associated with TH activity. [f] Associated with weak TH activity at best.

Table 2: Amino Acid Compositions of Purified THPs and Deduced Compositions of cDNA Clones from Various Insect THPs[a]

	purified proteins		cDNAs			
amino acid	H19.15	L20.00	*T. molitor* 4-9	*T. molitor* 2-14	*D. canadensis* 1A	*C. fumiferana*
Asx	11	11	11	13	12	8
Glx	5	5	4	4	4	3
Ser	6	5	6	7	9	15
Gly	10	7	7	8	4	10
His	1	2	2	2	1	0
Arg	0	0	0	0	4	2
Thr	14	18	19	17	16	15
Ala	9	9	9	8	7	5
Pro	2	2	2	2	2	3
Tyr	1	1	1	1	3	2
Val	3	3	3	2	1	4
Met	2	1	0	0	0	0
Ile	0	0	0	0	1	6
Leu	0	0	0	0	1	2
Phe	1	1	1	1	0	2
Lys	3	3	3	3	2	5
subtotal	68[b]	68[b]	68	68	67	82
Cys	nd	nd	16	16	16	8
Trp	nd	nd	0	0	0	0
total	nd	nd	84	84	83	90
mol mass	8351 + sugars	8351	8351	8351	8678	9062

[a] Molecular masses (daltons) of proteins encoded by cDNAs were deduced, while those of purified protein were determined by ESMS (Experimental Procedures). Measured compositions were rounded to the nearest integer. [b] Number of residues estimated from molecular weight by assuming identity to deduced sequence of 4-9.

ogy as seen with hemolymph at comparable TH activity (*7*).

The amino acid contents of pooled fractions from the L peak eluting between 19.90% and 20.15% B (L20.0) and one fraction from the H peak, corresponding to the peak eluting at 19.15% B (H19.15) (Figure 2), were determined (Table 2). These proteins were rich in Thr and other small amino acids. Their compositions, in conjunction with molecular masses determined by electrospray mass spectrometry (Table 1), enabled their specific activities to be determined. Fraction L20 had a TH activity of 2.1 °C at a concentration

Table 3: TH Activity of *T. molitor* Hemolymph and Fractions in the Presence and Absence of Various Compounds[a]

sample	TH (°C) ± SD	*n*
hemolymph (1:1) prior to chromatography	3.8 ± 0.3	3
H19.15[a] (55 μg/mL)	1.6 ± 0.1	3
H20.00[a] (21 μg/mL)	1.09 ± 0.07	3
L20.0 (114 μg/mL)	2.1 ± 0.3	3
hemolymph 1 (1:14)	0.92 ± 0.06	4
hemolymph 1 (1:14) + 10 mM EDTA	0.89 ± 0.06	4
hemolymph 1 (1:14) + 2 mM $CaCl_2$	0.80 ± 0.09	5
hemolymph 1 (1:14) + 2 mM $ZnCl_2$	0.91 ± 0.02	4
hemolymph 1 (1:14) + 10 mM DTT (2 min)	0.121 ± 0.016	3
hemolymph 1 (1:14) + 10 mM DTT (5 min)	0.068 ± 0.014	3
hemolymph 1 (1:14) + 10 mM DTT (10 min)	0.019 ± 0.009	3
hemolymph 1 (1:14) + 10 mM DTT (20 min)	0 ± 0	3
hemolymph 1 (1:14) + ddH_2O (20 min)	0.92 ± 0.07	3
hemolymph 2 (1:9)	1.06 ± 0.12	4
hemolymph 2 (1:9) + 10 mM *o*-phenanthroline	1.05 ± 0.18	4
hemolymph 3 (1:9)	1.00 ± 0.02	3
hemolymph 3 (1:9) + 10 mM *N*-ethylmalemide	1.05 ± 0.10	3

[a] Previously determined (*7*).

of 114 μg/mL, consistent with our previously reported activities (Table 3).

Cloning of THP cDNAs. The amino acid sequence (ATT?TGSTD?NTAVT?TNSK) of an internal endo Lys-C peptide from an N-terminally blocked protein in a fraction corresponding to H19.15 was previously determined (*7*). Two degenerate oligonucleotide primers were designed based on this sequence (Figure 3), which, when used in combination with the vector primer (T7) in a nested PCR reaction, amplified an ~330 bp fragment from the *T. molitor* larval fat body cDNA library (Figure 3). The sequence of this PCR fragment downstream of the nested 5′ primer encoded the five amino acids at the C-terminal end of the sequenced peptide (CTNSK) and an additional 29 amino acids. Both the sequenced peptide and deduced sequence were Thr-rich, consistent with the composition of fraction H19.15. The 5′ portion (226 base pairs) of the PCR fragment was reamplified with the nested degenerate primer and a nondegenerate primer complementary to a portion of the 3′ untranslated region (Figure 3). This fragment was then labeled and hybridized to approximately 2×10^5 cDNA clones (95% recombinant), resulting in 500−600 positive signals, suggesting that THP mRNA is moderately abundant in larval fat body.

Characteristics of the THP cDNA Sequences. A total of 17 nonidentical cDNA clones encoding nine distinct THP sequences at the protein level (eight distinct mature proteins) were examined. These 17 cDNAs were derived from different mRNA transcripts as they all showed polymorphism with respect to base sequence, and/or the site of polyadenylation, and/or the length of the 5′ untranslated region. The sequence identity between the encoded mature proteins averaged 86%. One of the newly obtained clones, 4-9 (Figure 3), encoded the previously obtained peptide sequence and matched the PCR sequence derived from the nested primer sets (Figure 3). In the conceptual translation of the cDNAs (Figure 4), the first 28 amino acids encode a secretory signal peptide with a core of nine well-conserved hydrophobic residues preceded by two or three basic residues. The N-terminal amino acid of six of the eight variants is predicted to be Gln +1 with Cys at the −1 position and Cys or Tyr at the −3 position (*23*). N-Terminal blockage was previously ascribed

```
                                        M  G  F  K  -25
 5′-CAACAATATTACAAAAAACTATGGGATTCAAA  32
  T  C  G  F  S  K  K  W  L  V  T  A  -13
ACGTGTGGTTTTTCAAAAAAATGGTTAGTAACAGCA  68
  V  I  V  M  C  L  C  T  E  C  Y  C   -1
GTTATAGTTATGTGTTTGTGTACCGAGTGTTATTGC 104
  Q  C  T  G  G  A  D  C  T  S  C  T   12
CAATGCACTGGAGGTGCTGATTGTACTAGTTGTACA 140
  A  A  C  T  G  C  G  N  C  P  N  A   24
GCAGCATGCACTGGTTGTGGAAACTGTCCAAATGCA 176
  V  T  C  T  N  S  Q  H  C  V  K [A   36
GTAACGTGTACCAATTCTCAACATTGTGTCAAGGCA 212
  T  T  C  T  G  S  T  D  C  N  T  A   48
ACAACATGTACTGGGTCTACAGATTGTAATACAGCC 248
  V  T  C  T  N  S  K] D  C  F  E  A   60
GTGACGTGTACAAACTCAAAAGACTGTTTCGAAGCC 284
  Q  T  C  T  D  S  T  N  C  Y  K  A   72
CAAACATGTACTGACTCAACCAACTGTTACAAAGCT 320
  T  A  C  T  N  S  T  G  C  P  G  H   84
ACAGCCTGTACCAATTCAACAGGATGTCCCGGACAT 356
  ★
TAAGTTTTTCTATTGTCAACAATCATAAAACACAAT 392
TATTGTTATCTAAGTTAAAACATAATTGTAAGTTTA 428
CTCTGTATTGTATCCGATCTGTCTCTTTGCCTCCAA 464
GGATGATAATCTTGTACTGGGAGCGAAAGGGTTATC 500
GGACAATAATAAACTAAAATAATTGATAAAAAAAAA 536
AAAAA-3′                              542
```

FIGURE 3: Sequence of cDNA clone 4-9. The deduced amino acid sequence encoded by clone 4-9 is shown above its DNA sequence. The amino acid sequence is numbered from the start of the mature protein sequence and is organized to emphasize the repeating nature of the mature protein sequence (note the alignment of Cys in the last six 12-amino acid blocks). The secretory signal peptide is italicized, and the termination codon is indicated by a star. The internal endoprotease Lys-C peptide that was sequenced is bracketed, and the positions of the fully degenerate primers designed from this sequence are denoted by double underlining of the corresponding cDNA sequence. The underlined DNA sequence exactly matched the portion determined by nested PCR. The PCR product used for library screening extends from bp 234 to 459 inclusive, and the sequence complementary to the 3′ primer is double underlined. The polyadenylation signal is italicized.

to conversion of this Gln residue to pyroglutamate (*7*) by spontaneous cyclization or by the enzyme glutaminyl cyclase (*24*). The N-terminal residues of the other two variants are predicted to be His and Arg. The amino acid compositions of the purified fractions (H19.15 and H20.00) and the deduced mature sequences, exemplified by 2-14 and 4-9, are very similar (Table 2). As previously noted for the first four isoforms sequenced (*7*), the most abundant amino acids are Thr (21−26%) and Cys (18−19%). The overall content of hydrophilic amino acids (D, E, H, K, N, Q, R, S, and T) is greater than 50%, which is appreciably higher than that seen in fish AFPs (*25*), and there are relatively few aromatic and long-chain aliphatic amino acids that might constitute a hydrophobic core. The isoelectric points of these isoforms are predicted to vary from 4.7 to 10.2 (*26*).

The dominant structural element of all isoforms is a tandemly linked, 12-amino acid repeat, with the consensus sequence CTxSxxCxxAxT (Figure 4). Each repeat contains two Cys spaced six residues apart. Thus, Cys is repeated at six-residue intervals throughout most of the protein, and almost every second Cys is flanked by two Thr. However,

```
                                                                                R1
4-9              MGFKTCGFSKKWLVTAVIVMCLCTECYC QCTGGADCTSCTAA CTGCGNCPNAVT  26
1-3, 1-1, C-5    .A...........I.............. H..........D. ..........H.  26
3-4, E-12        .A..A........I.............. H..........D. ..........H.  26
2-14, 2-7, A-1   .A...........I.............. ...........G. ............  26
C-9, 5-1, E-2    .A...........I.............. R..........Q. ..S.R.....K.  26
D-16             .A...........I.............. .....S........ ..........H.  26
A-3, 3-8         .A...........II............. ............... .....S....H.  26
5-15             .A...........II............. ............... .....S....H.  26
2-20             .S..IST.T.I..II........N.YN. ....A......... ..........I.  26

            R2            R2'           R3           R4            R3'
4-9     CTNSQHCVKATT ------------ CTGSTDCNTAVT CTNSKDCFEAQT ------------ 62
1-3     ..D.KN....A. ------------ .....K....R. ..........K. ------------ 62
3-4     ..D.KN....A. ------------ .....K....R. ..........K. ------------ 62
2-14    ..........N. ------------ ..........Q. ..........N. ------------ 62
C-9     ........R.R. ------------ ........R.M. ..........K. ------------ 62
D-16    ..D.......A. ------------ ..........R. ..........A. ------------ 62
A-3     ..D.KN..R.E. CTDSENCVKAHT ....RN....M. ..........K. ------------ 74
5-15    .ID.KN..R.E. CTDSENCVKAHT ....RN....M. ..........K. ------------ 74
2-20    ..G.KN..R... ------------ .....N..R.T. .....G.L..T. CTGSTHCHRATT 74

            R4'           R5'           R5           R6
4-9     ------------ ------------ CTDSTNCYKATA CT▮STGCPGH  84  8351.0
1-3     ------------ ------------ ............ ..........  84  8477.2
3-4     ------------ ------------ ............ ..........  84  8477.2
2-14    ------------ ------------ ............ ....S.....  84  8350.9
C-9     ------------ ------------ ...........T ..........  84  8811.6
D-16    ------------ ------------ ............ ..H.......  84  8399.0
A-3     ------------ ------------ ............ ..........  96  9769.6
5-15    ------------ ------------ ............ ..........  96  9781.6
2-20    CTNSKDCFEATT CTGSSNCYTATT ..N......... ..........  120 12005.0
```

FIGURE 4: Sequence of *T. molitor* THP isoforms: Conceptual amino acid sequences of nine THP isoforms derived from 17 fat body cDNA clones. The names of all cDNA clones encoding each isoform are shown only at the beginning of the sequence. The numbers at the end of the sequences are the predicted masses (daltons) of the mature proteins. The first 28 amino acids (−28 to −1, italicized) represent the signal sequence. The 12-amino acid repeats of the mature protein sequence are marked by gaps and are numbered consecutively (R + number), where a prime designation indicates similarity to one of the other repeats. Conserved residues are indicated by dots and gaps by dashes, with the residues conserved between repeats shown in boldface type. The clones 2-14, 1-3, 3-8, and 2-20 were previously designated YL1−YL4 respectively (*7*). The five clones 1-3, 1-1, C-5, E-12, and 3-4 encode the same mature protein, while 2-14 and 4-9, although having very similar masses, encode different mature proteins. The Asn residue that is 8 residues from the C-terminus and is within a consensus glycosylation signal (Asn-X-Ser/Thr) is stipled. Our inability to identify the 12 kDa isoform in the chromatographic profile suggests that it is a rare isoform, and two isoforms were not encoded by the cDNAs cloned.

the first 14 amino acids do not conform to this pattern and the first repeat (residues 15−26) contains a Cys instead of the consensus Ser. Therefore, the following Cys spacing is present in the first 21 amino acids (C-X_5-C-X_2-C-X_3-C-X_2-C-X_2-C). All isoforms contain an even number of Cys, and their spacing is perfectly conserved throughout the isoform set. Three other deviations from the repeat consensus are observed. Clone 5-15 had a substitution of Thr-28 by Ile in the second repeat; all but clone C-9 have the final Thr replaced by Ala in the fifth repeat; and the final repeat of all isoforms is truncated with the sequence Pro-Gly-His following the last Cys. Although other positions within the repeat are more variable, some constraints on the amino acid present are apparent. For example, the third residue within each repeat is small and hydrophilic. Positions 5, 6, and 9 are also populated exclusively by hydrophilic residues, but the length of their side chains is more variable. Position 8 is hydrophobic in all but repeat 3; while no obvious bias is apparent in position 11. Indeed, position 11 shows the highest degree of substitution among all the positions.

Differences in length between the variants represent multiples of the 12-amino acid repeat. The most abundant THP sequences (13 cDNAs encoding five mature isoforms) contain five 12-amino acid repeats in their 84 residues. There are two examples encoded by three cDNAs (3-8, A-3, and 5-15) where there is one extra repeat (96 residues). The longest THP (2-20) has three additional repeats for a total of 120 residues. The extra repeats present in the longer variants resemble, but are not exact copies of, any of the other repeats. All isoforms except D-16 contain a consensus N-glycosylation site near the C-terminus.

Mass Analysis. Masses for most of the 84-residue cDNA-encoded isoforms range from 8350 to 8477 Da and were calculated on the basis of the predicted signal sequence cleavage and the assumptions that N-terminal Gln was cyclized and that all Cys were disulfide-bonded (Figure 4). Their low residue mass average reflects the abundance of short-chain amino acids (Cys, Thr, Ser, Ala, and Gly) with the exception of C-9 (8812 Da), which has substitutions of larger amino acids. Matches were found between some of the predicted masses and those determined by ESMS (Table 1). Clones from the 96-residue subset (A-3 and 5-15) encode proteins with predicted masses of 9770 and 9782 Da, and matching masses were observed. The predicted mass of the

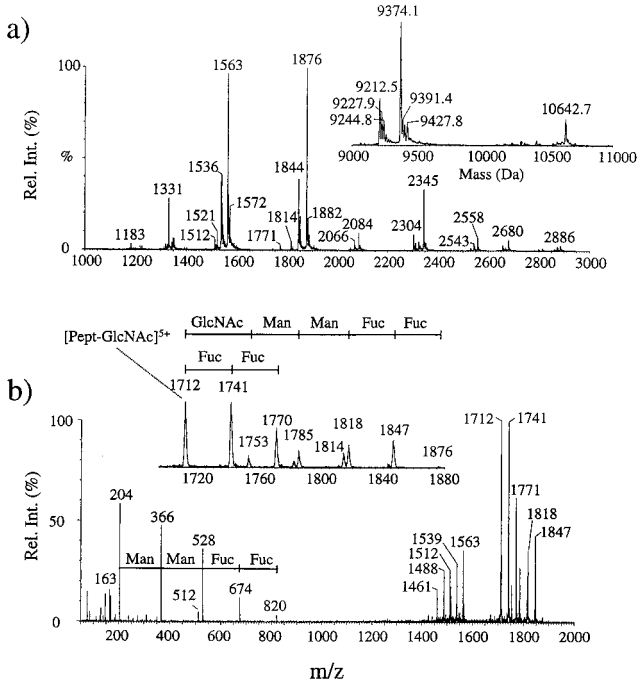

FIGURE 5: Mass spectral analysis of glycosylated THP from fraction H19.15. (a) Conventional ESMS with inset showing molecular mass profile of the principal components. (b) Product ion spectrum following tandem mass spectroscopy of *m*/*z* 1563 fragment from above. Fragment ion masses were rounded up for clarity. The inset shows an expanded view of the pentuply charged ion fragments. Bars above the peaks are used to indicate which sugars the *m*/*z* differences are attributed to, and these patterns of residue loss were used to deduce the glycan structure.

120-residue isoform (2-20) is 12 005, which is considerably larger than any observed isoform mass (Table 1).

There were several active isoforms with masses in the range of 9.2−9.5 kDa that did not match any deduced masses. For example, ESMS of fraction H19.15 (Figure 5a) revealed a number of multiply protonated ions from which masses could be deduced (Figure 5a, inset). The mass difference between the two major components of 9212.5 and 9374.1 Da was consistent with the mass of a single hexose residue, although amino acid substitution could not be ruled out at this stage. Therefore, additional experiments with tandem mass spectrometry were performed. The product ion spectrum of the $[M + 6H]^{6+}$ ion from the 9374.1 Da isoform (Figure 5b) revealed a number of oxonium fragment ions at *m*/*z* 147, 163, and 204, consistent with Fuc, Man, and GlcNAc residues, respectively. This, combined with oligosaccharide ions at *m*/*z* 366, 528, 674, and 820, provides further evidence of glycosylation. The expanded region of the pentuply charged fragment (Figure 5b, inset) showed consecutive residue losses typical of N-linked glycans (*27*), leaving a final product of the expected mass consisting of the protein containing a single GlcNAc at *m*/*z* 1712. The structure of the glycan was deduced from the order in which residues were lost from the multiply charged fragment ions. For example, fragment ions at *m*/*z* 1741 and 1771 were consistent with a difucosylated GlcNAc, and linkage to a second GlcNAc with two attached Man residues was indicated by fragment ions at *m*/*z* of 1753, 1785, and 1818. Similar analysis of the 9212.5 Da isoform (data not shown) indicated a similar glycan lacking one terminal Man. The minor peaks (Figure 5a, inset) are consistent with isoforms containing slightly different glycans and/or amino acid substitutions and were not further characterized.

The location and structure of the attached glycan and the identity of the major protein component was confirmed by separate experiments in which fraction H19.15 was reduced, alkylated, and digested with trypsin, and the resulting fragments were analyzed by liquid chromatography−ESMS (data not shown). Although 2-14 and 4-9 have the same deduced mass, five amino acid differences result in different masses for all predicted tryptic fragments. The masses of the fragments were consistent with those predicted for isoform 4-9 with the glycan attached to the C-terminal fragment (13 amino acids), which includes the consensus N-linked glycosylation signal. In addition, the mass of the N-terminal fragment was consistent with modification of the N-terminal Gln to pyroglutamate.

Subtraction of the presumed glycan mass from the observed masses of other isoforms resulted in close matches to some deduced protein masses (Table 1). Thus, in addition to glycosylated 4-9, fraction H19.15 also contained trace amounts of material consistent with both glycoforms of 5-15, to which a nonglycosylated match was not found. Masses consistent with both the glycosylated and nonglycosylated forms of A-3, 1-3, and 4-9 were also observed, and the glycoforms eluted from the column earlier than the nonglycosylated forms, as expected. A glycosylated D-16 was not observed, which is consistent with the lack of the N-glycosylation signal. We estimate that a little less than half of the THP was glycosylated.

TH Activity. The effects of various compounds on the TH activity of hemolymph was examined (Table 3). All activity was eliminated following incubation with 10 mM dithiothreitol at 37 °C for 20 min, whereas no activity was lost under the same conditions in the absence of reducing agent. No activity was lost when hemolymph was incubated with the sulfhydryl reagent *N*-ethylmaleimide (Table 3). Although the observed masses (Table 1) are consistent with full oxidation of Cys residues, the Cys spacing in the N-terminal region includes C-X_5-C and C-X_2-C, both of which appear in some zinc-binding motifs (*28*). There was, however, no loss of activity when hemolymph was exhaustively dialyzed against 10 mM EDTA or *o*-phenanthroline, and therefore no evidence of metal liganding. In addition, incubation of dialyzed, EDTA-free hemolymph with 2 mM $ZnCl_2$ or 2 mM $CaCl_2$ (Table 3) did not result in increased TH activity.

Genomic Southern Blot. Southern blot analysis of DNA from individual insects digested with *Hin*dIII and probed with one of the THP coding regions at moderate stringency indicated the presence of a large gene family (Figure 6). The bands are unlikely to contain partial gene sequences as *Hin*dIII does not cut within the coding region of any known isoform nor within PCR-amplified genomic DNA encompassing the coding region (data not shown). The number of bands ranged from 31 in insects 1 and 4 to 40 in insect 5 but is probably greater due to band overlap. These bands do not represent partial digestion products as only one or two bands/lane were obtained when the same blot was hybridized to two unrelated cDNA sequences (ref *22* and unpublished data). Although the banding patterns obtained with some insects were identical, restriction fragment length polymorphisms were seen with at least one of the two other cDNA probes. Thus, members of this THP gene family can be

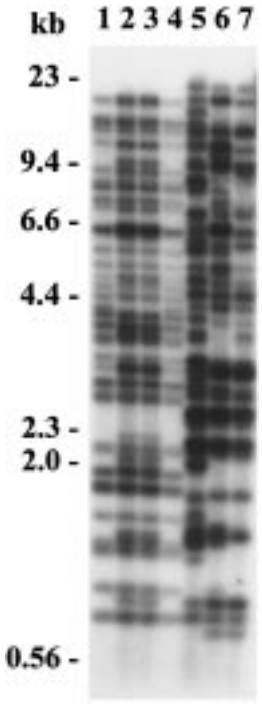

FIGURE 6: Genomic Southern blot. Genomic DNA from individual insects (1−7) was digested with *Hin*dIII, Southern blotted, and probed with the coding region encoding the mature 2-14 isoform. The positions of DNA markers derived from a *Hin*dIII digestion of λ DNA are indicated on the left.

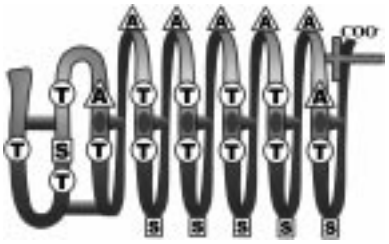

FIGURE 7: Model structure for *T. molitor* THP. Hypothetical model for the three-dimensional structure of *T. molitor* THP using the disulfide linkage pattern determined for *D. canadensis* THP (*29*). The conserved Ala, Ser, and Thr of the 12-amino acid repeats are indicated by one-letter code. Disulfide bridges are indicated by the thick bars between the ribbon representing the polypeptide backbone. The N-terminus is to the left and the glycan near the C-terminus is represented by the checkered cross.

inherited as a single unit and must therefore be closely linked (linkage disequilibrium).

T. molitor THP Model. The fact that some *T. molitor* THP isoforms differed in length by the insertion of 12-amino acid repeats at different points along the molecule suggested that each repeat is a structural unit (*7*). Moreover, the formation of a disulfide bond within each repeat was predicted as each repeat contains two Cys, six residues apart, and THP was inactivated by reduction but not by reagents that react with free sulfhydryls. Subsequently, the disulfide bond arrangement of the *D. canadensis* THP, which has Cys in analogous positions throughout the protein, was reported (*29*), along with isoform length variation due to different repeat numbers (*16*). This allowed assignment of disulfide linkages for the irregularly spaced Cys near the N-terminus and facilitated the development of a model for *T. molitor* THP (Figure 7). Each internally disulfide-bonded 12-amino acid repeat is stacked side-by-side to form a β-helix, which has arbitrarily been coiled in a right-handed manner. The essential feature of the model is that it aligns most of the Thr residues on one side of the molecule in a regular array that could potentially interact with the crystalline ice lattice.

DISCUSSION

We have previously reported the isolation of two hyperactive THP fractions from *T. molitor* hemolymph that had 10−100 times the specific activity of AFPs from fish (*7*). Here, a detailed analysis of hemolymph has shown that its 5.5 °C of TH activity is derived from multiple isoforms, which all belong to one protein type. Thus, it seems increasingly likely that earlier accounts of the purification of *T. molitor* THPs with varied properties and amino acid compositions (*4*, *9*−*14*) can be attributed to the contamination of other proteins by small quantities of these hyperactive THPs.

In retrospect, several properties of *T. molitor* THP have contributed to this ambiguity and have made the protein difficult to isolate and characterize. The high level of TH activity is associated with low hemolymph THP concentrations (∼1−2 mg/mL). Consequently, the THP is not readily identified with a protein peak during size-exclusion chromatography (Figure 1) and its presence is largely obscured by an abundant family of 12−13 kDa proteins. In contrast, *D. canadensis* THPs attain a concentration, in hemolymph, of 22 mg/mL (*16*). This is consistent with our observation that the hemolymph of overwintering *D. canadensis* collected in Kingston, Ontario, diluted 1:1, has greater TH activity than the maximum we can measure (8 °C) with our apparatus. Given that *T. molitor* is a domesticated pest of human grain stores and is unlikely to be exposed to the same temperature extremes as *D. canadensis* in the wild, it may not require high levels of THP.

A further complication to THP purification is the distribution of TH activity among a large family of isoforms that differ in mass, charge, and hydrophilicity due to amino acid substitutions, incomplete glycosylation, and a variable number of the tandem 12-amino acid repeats that make up the bulk of each THP. For all these reasons, it has been difficult to obtain more than microgram quantities of individual isoforms for further study. Moreover, these THPs tend to migrate atypically and unreliably on SDS−PAGE and do not react well with most protein stains. In addition, they are extensively disulfide-bonded and most of the isoforms are N-terminally blocked.

Glycosylation increased observed hemolymph isoform diversity approximately 3-fold as most protein isoforms showed two major glycoforms and a nonglycosylated form. The composition and sequence of the glycans on *T. molitor* THP are the same as some of the smaller glycans observed on the membrane glycoproteins of three moth cell lines (*30*). There is no obvious role for glycosylation in ice binding because the modified and unmodified forms both display activity, but it is possible that the glycans have to be small to prevent steric interference during binding to ice. Interestingly, the consensus glycosylation site is not found on the homologous *D. canadensis* THPs (*15*, *16*). Carrot AFP (*31*), which is glycosylated at one or a small number of residues, is also active with and without the attached glycan.

Genomic Southern blotting is consistent with the observed isoform diversity and suggests there could be 30+ linked THP genes. Some of these may well be duplicate copies, but the presence of multiple bands and multiple cDNAs indicates that the members of this gene family are quite polymorphic. In many respects, this gene family and its organization bears a striking resemblance to fish AFP gene families. The four well studied AFP types found in fishes are typically made up of multiple, independently active AFP isoforms encoded by large gene families (*32*, *33*), and in the case of types I and III AFPs, it has been shown that these genes are clustered (*34*, *35*). This arrangement provides opportunities for, and there is evidence of (*36*), expansion/contraction of gene copy number, presumably through unequal recombination. It has been suggested that AFP gene copy number is responsive to selective pressure, possibly associated with climate change (*37*, *38*).

The presence of repetitive units within *T. molitor* THP genes is the likely source of some of this diversity. Replicational slippage or unequal recombination following misalignment of the 36 bp repeats encoding the 12 amino acid consensus sequences has probably produced the genes containing different numbers (six, seven, or nine) of the repeats observed. This is reminiscent of the variation seen in the number of 11-amino acid repeats in the α-helical AFPs from flounders (*39*, *40*) and in the tripeptide repeats in the antifreeze glycopeptides (*41*).

The primary sequence of the THP isoforms was not significantly similar to any other sequence besides the *D. canadensis* THPs in the Genbank nonredundant databases (January 1999) as determined by blastn, blastp, psi-blast, and tblastn searches (*42*). However, some of the low-scoring matches also contain Cys-rich repeats. One of the sequences, an integral membrane protein isolated from the protozoan *Trypanosoma brucei*, contains numerous 12-amino acid repeats with Cys spaced at six-residue intervals (*43*). The function of this domain is unknown, and aside from the Cys spacing, there was not enough sequence similarity to suggest homology.

The disulfide bonding pattern and variability in repeat number introduces constraints to limit the number of three-dimensional structures possible for *T. molitor* THP. The model we have proposed (Figure 7) introduces a regularity to the surface of the protein that reflects the repeating structure. In this β-helix, every unit would have the same orientation, so additional repeats could be accommodated by extending the length of the molecule. This is similar to the way additional 11-amino acid repeats extend the length of type I AFP by three turns of the α-helix (*40*). This arrangement would result in regularly spaced Thr being clustered on one side of the protein. Thr is an attractive candidate for an ice-binding residue since it is abundant in both beetle and spruce budworm THPs (Table 2), and the putative ice-binding Thr are highly conserved between all isoforms and between *T. molitor* and *D. canadensis* THPs. In addition, Thr has been implicated in this role in a number of AFPs, most notably the α-helical type I AFP (*44*) and type III AFP (*45*). In type I AFP, up to five ice-binding Thr protrude from one side of the helix and are separated from each other by three turns of the α-helix. These residues form a match with the ice lattice along a plane of ice that is not extensively expressed in the absence of the AFP (*46*). The further study of these beetle THPs is required to determine the accuracy of the proposed model, the ice-binding plane, and the basis for such high TH activity.

ACKNOWLEDGMENT

We thank the Core Facility for Protein/DNA Chemistry at Queen's University, Kingston, Ontario, for DNA sequencing. We are grateful to the Harvard Microchemistry Facility in Cambridge, MA, for performing amino acid analysis and protein sequencing. As well, we thank T. Scott Reid for his assistance in characterizing hemolymph THPs.

REFERENCES

1. Cotton, R. T., and St. George, R. A. (1929) *Tech. Bull. U.S. Dep. Agric. 95*, 1−37.
2. Johnston, S. L., and Lee, R. E. (1990) *Cryobiology 27*, 562−568.
3. Ramsey, J. A. (1964) *Philos. Trans. R. Soc. London B 248*, 279−314.
4. Horwath, K. L., Easton, C. M., Poggioli, G. J., Myers, K., and Schnorr, I. L. (1996) *Eur. J. Entomol. 93*, 419−433.
5. Patterson, J. L., and Duman, J. G. (1978) *J. Exp. Biol. 74*, 37−45.
6. Hansen, T. N., and Baust, J. G. (1988) *Biochim. Biophys. Acta 957*, 217−221.
7. Graham, L. A., Liou, Y.-C., Walker, V. K., and Davies, P. L. (1997) *Nature 388*, 727−728.
8. Grimstone, A. V., Mullinger, A. M., and Ramsay, J. A. (1968) *Philos. Trans. R. Soc. London B 253*, 343−382.
9. Patterson, J. L., and Duman, J. G. (1979) *J. Exp. Zool. 210*, 361−367.
10. Schneppenheim, R., and Theede, H. (1980) *Comp. Biochem. Phys. B 67*, 561−568.
11. Patterson, J. L., and Duman, J. G. (1982) *J. Exp. Zool. 219*, 381−384.
12. Tomchaney, A. P., Morris J. P., Kang, S. H., and Duman, J. G. (1982) *Biochemistry 21*, 716−721.
13. Kroeker, E. M. (1989) Dsp28: a desiccation stress protein in *Tenebrio molitor* hemolymph, Ph.D. Thesis, Queen's University, Kingston, Ontario,Canada.
14. Tang, W. (1993) Biochemical and molecular biological studies of antifreeze proteins from the insect *Tenebrio molitor*. Ph.D. Thesis, State University of New York, Binghamton, NY.
15. Duman, J. G., Li, N., Verleye, D., Goetz, F. W., Wu, D. W., Andorfer, C. A., Benjamin, T., and Parmelee, D. C. (1998) *J. Comp. Physiol. B 168*, 225−232.
16. Li, N., Andorfer, C. A., and Duman, J. G. (1998) *J. Exp. Biol. 201*, 2243−2251.
17. Tyshenko, M. G., Doucet, D., Davies, P. L., and Walker, V. K. (1997) *Nat. Biotechnol. 15*, 887−890.
18. Chakrabartty, A., and Hew, C. L. (1991) *Eur. J. Biochem. 202*, 1057−1063.
19. Morris, H. R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R. S., Hoyes, J., and Bateman, R. H. (1996) *Rapid Commun. Mass Spectrom. 10*, 889−896.
20. Graham, L. A., Bendena, W. G., and Walker, V. K. (1996) *Insect Biochem. Mol. 26*, 127−133.
21. Ashburner, M. (1989) *Drosophila*: *A Laboratory Manual*. Cold Spring Harbor Press, Cold Spring Harbor, NY.
22. Graham, L. A., Bendena, W. G., and Walker, V. K. (1996) *Dev. Genet. 18*, 296−305.
23. von Heijne, G. (1986) *Nucleic Acids Res. 14*, 4683−4690.
24. Pohl, T., Zimmer, M., Mugele, K., and Spiess, J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10059−10063.
25. Sönnichsen, F. D., Sykes, B. D. and Davies, P. L. (1995) *Protein Sci. 4*, 460−471.
26. Bjellqvist, B., Basse, B., Olsen, E., and Celis, J. E. (1994) *Electrophoresis 15*, 529−539.

27. Bateman, K. P., White, R. L., Yaguchi, M., and Thibault, P. (1998) *J. Chromatogr. A 794*, 327−344.
28. Klug, A., and Schwabe, J. W. R. (1995) *FASEB J. 9*, 597−604.
29. Li, N., Chibber, B. A. K., Castellino, F. J., and Duman, J. G. (1998) *Biochemistry 37*, 6343−6348.
30. Kubelka, V., Altmann, F., Kornfeld, G., and März, L. (1994) *Arch. Biochem. Biophys. 308*, 148−157.
31. Worrall, D., Elias, L., Ashford, D., Smallwood, M., Sidebottom, C., Lillford, P., Telford, J., Holt, C., and Bowles, D. (1998) *Science 282*, 115−117.
32. Chen, L., DeVries, A. L., and Cheng, C.-H. C. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 3811−3816.
33. Hayes, P. H., Scott, G. K., Ng, N. F. L., Hew, C. L., and Davies, P. L. (1989) *J. Biol. Chem. 264*, 18761−18767.
34. Scott, G. K., Hew, C. L., and Davies, P. L. (1985) *Proc. Natl. Acad. Sci. 82*, 2613−2617.
35. Scott, G. K., Hayes, P. H., Fletcher, G. L., and Davies, P. L. (1988) *Mol. Cell. Biol. 8*, 3670−3675.
36. Scott, G. K., Davies, P. L., Kao, M. H., and Fletcher, G. L. (1988) *J. Mol. Evol. 27*, 29−35.
37. Scott, G. K., Fletcher, G. L., and Davies, P. L. (1986) *Can. J. Fish. Aquat. Sci. 43*, 1028−1034.
38. Hew, C. L., Wang, N.-C., Joshi, S., Fletcher, G. L., Scott, G. K., Hayes, P. H., Buettner, B., and Davies, P. L. (1988) *J. Biol. Chem. 263*, 12049−12055.
39. Gourlie, B., Lin, Y., Price, J., DeVries, A. L., Powers, D., and Huang, R. C. C. (1984) *J. Biol. Chem. 259*, 14960−14965.
40. Chao, H., Hodges, R. S., Kay, C. M., Gauthier, S. Y., and Davies, P. L. (1996) *Protein Sci. 5*, 1150−1156.
41. Burcham, T. S., Osuga, D. T., Rao, B. N., Bush, C. A., and Feeney, R. E. (1986) *J. Biol. Chem. 261*, 6384−6389.
42. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
43. Lee, M. G., Bihain, B. E., Russell, D. G., Deckelbaum, R. J., and Van Der Ploeg, L. H. (1990) *Mol. Cell. Biol. 10*, 4506−4517.
44. Knight, C. A., Cheng, C. C., and DeVries, A. L. (1991) *Biophys. J. 59*, 409−418.
45. Chao, H., Sönnichsen, F. D., DeLuca, C. I., Sykes, B. D., and Davies, P. L. (1994) *Protein Sci. 3*, 1760−1769.
46. Davies, P. L., and Sykes, B. D. (1997) *Curr. Opin. Struct. Biol. 7*, 828−834.

BI990613S