Mutagenic and Chemical Modification of the ABA-1 Allergen of the Nematode *Ascaris*: Consequences for Structure and Lipid Binding Properties†

Lindsay McDermott,‡,§ Joyce Moore,§ Andrew Brass,|| Nicholas C. Price,⊥ Sharon M. Kelly,⊥ Alan Cooper,‡ and Malcolm W. Kennedy*,§

Department of Chemistry and the Divisions of Infection and Immunity and of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, Scotland, and Department of Biochemistry and Molecular Biology, University of Manchester, Stopford Building, Oxford Road, Manchester M13 9PT, U.K.

Received November 22, 2000; Revised Manuscript Received February 23, 2001

ABSTRACT: The polyprotein allergens/antigens of nematodes (NPAs) are the only lipid binding proteins known to be produced as polyproteins. Cleavage of the large polyprotein precursors at regularly spaced proteinase cleavage sites produces 10 or 11 individual protein units of ∼15 kDa. The sequences of these units are highly diverse within and between species, but there are five absolutely or strongly conserved amino acid positions (Trp15, Gln20, Leu42, Cys64, and Cys120). We have tested the role of these signature amino acids by mutational or chemical alteration of the ABA-1 protein of *Ascaris*, and examined the resulting modified proteins for perturbations of their lipid binding activities and structural integrity. Substitution of Trp15 and Gln20 both affect the stability of the protein in terms of resistance to thermal or chemical denaturation, but the ligand binding function is unaffected. Mutation of Leu42, however, disrupts both the protein's structural stability and functional integrity, as does chemical disruption of the disulfide bridge formed between Cys64 and Cys120. We also find that the C-terminal, but not the N-terminal, half of the protein binds fatty acids, indicating that the binding site may be confined to this part of the protein. This also supports the idea that the NPA units are themselves derived from an ancient duplication event, and that they may comprise two functionally distinct domains.

The polyprotein allergens/antigens of nematodes (NPAs)[1] are the only lipid binding proteins known which are produced as polyprotein precursors, and they have no discernible homologues in other groups of organisms, including humans (*1*). Complete sequences from the free-living nematode *Caenorhabditis elegans* (*2*) and the highly pathogenic bovine parasite *Dictyocaulus viviparus* (*3*) show that their NPAs are produced as large precursor polypeptides which are subsequently cleaved at consensus target sites for processing proteinases into 10 or 11 functional units of ∼15 kDa (*2*). Nematode-derived or recombinant proteins representing a single unit are known to bind fatty acids and retinol at affinities similar to that expected for transporter proteins (*1, 4, 5*). No experimentally determined 3D structure yet exists for any NPA, but secondary structure predictions and circular dichroism analysis both indicate that the NPA units are dominated by helical structure, with little or no evidence of β structure (*1, 4*−*6*). They are therefore distinct from the small lipid transporter proteins of similar size and binding propensities that have been reported from vertebrates, the lipocalins and the cytosolic proteins of the FABP/P2/CRBP/CRABP family (*7*). The NPAs, therefore, not only are highly unusual in their mode of biosynthesis, but are structurally novel among small lipid binding proteins, and may be confined to nematodes. This might make them potential therapeutic targets.

The NPAs initially attracted interest because of their allergenicity (*8, 9*), and they may therefore contribute to the pathology of infections with parasitic nematodes, which could in turn be related to their structure or lipid binding activities. They are also of intrinsic interest because of the very fact that they are synthesized as polyproteins. The term polyprotein is used in the literature to denote two quite different types of entity. The first is exemplified by viral proteins and some vertebrate hormone proproteins in which a precursor protein is cleaved into polypeptides with different functions

† This work was supported by Wellcome Trust Grants to M.W.K. and A.C., and in part by a research fellowship to L.M. The CD facility is funded by the BBSRC.

* To whom correspondence should be addressed at the Division of Infection and Immunity, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, Scotland. E-mail: malcolm.kennedy@bio.gla.ac.uk.

‡ Department of Chemistry, University of Glasgow.

§ Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow.

|| Department of Biochemistry and Molecular Biology, University of Manchester.

⊥ Division of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, University of Glasgow.

[1] Abbreviations: ANS, anilinonaphthalene-8-sulfonic acid; BSA, bovine serum albumin; CD, circular dichroism; cPnA, *cis*-parinaric acid; C-ter, C-terminal half of the ABA-1A1 protein; DACA, dansyl-DL-α-aminocaprylic acid; DAUDA, 11-(((5-(dimethylamino)-1-naphthalenyl)-sulfonyl)amino)undecanoic acid; DTNB, dithiobis(nitrobenzoic acid); DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; FABP, fatty acid binding protein; Gdn-HCl, guanidine hydrochloride; GST, glutathione-*S*-transferase; IAA, iodoacetic acid; IAM, iodoacetamide; IPTG, isopropyl β-D-thiogalactopyranoside; K_{sv}, Stern−Volmer constant; NBS, *N*-bromosuccinimide; NPA, nematode polyprotein antigen/allergen; N-ter, N-terminal half of the ABA-1A1 protein; PBS, phosphate-buffered saline; rABA-1, recombinant A1 unit of the ABA-1 polyprotein; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

10.1021/bi0026876 CCC: $20.00
Published on Web 07/24/2001

(*10*−*13*). The second type is typified by certain proteins of vertebrate keratinocytes in which a large precursor containing tandemly repeated units is cleaved posttranslationally into polypeptides which are presumed (in the absence of direct testing) to have similar or identical functions (*14*−*17*). There will presumably also be cases of this type in which the different units have diverged such that they now serve different biochemical functions yet reflect their antecedents in primary, secondary, and tertiary structures. The NPAs fall into this tandemly repetitive category of polyproteins. So, what is the advantage of producing proteins as polyproteins, and why are only very few types of proteins produced in this way? A more detailed understanding of the structure of the individual units of polyproteins could be illuminating for these particular polyproteins and for polyproteins in general.

The amino acid sequences of the units of NPAs vary dramatically between species, and sometimes equally dramatically between individual units of a single polyprotein (*3*, *6*, *18*). The ligand binding properties of the different NPA units are nevertheless very similar (*6*). This sequence diversity has, however, identified a small number of absolutely conserved amino acids, which are presumably essential to the structure and function of the NPAs. Here we test the role of these NPA signature amino acids by chemically or mutationally altering them one by one, concentrating on the best understood NPA, the ABA-1 allergen of the nematode *Ascaris*. We find that the mutational alterations dramatically affect the stability of the proteins with only slight affects on their lipid binding properties, with the exception of one change, which affects both. The disruption of the single intramolecular disulfide bond is equally detrimental. We also provide evidence that the binding site is localized in the C-terminal half of an NPA unit, and our findings also support the idea that the NPA units themselves derive from the duplication of an ancient, very small, fatty acid binding protein.

EXPERIMENTAL PROCEDURES

Recombinant ABA-1A and Construction of Mutants. The recombinant proteins used were derived from the A1-type unit of the ABA-1 polyprotein of *Ascaris suum*, and produced with the terminal four arginines of the unit absent, as described previously [(*2*, *6*) and Figure 1]. This unit of the *A. suum* NPA is also designated As-NPA-A1 (*6*), but the term ABA-1A1 will be used here. *E. coli* clones expressing the recombinant polypeptide were produced as described using the pGEX expression system. In vitro mutagenesis was carried out on the above clone by half-mer PCR on clone pGEX13 containing the plasmid encoding ABA-1A1. Two mutagenesis primers, complementary to one another, were designed to cover the region to be mutated, and were used in combination to produce mutations of ABA-1A1 at W16 → R, L42 → R, and Q20 → L. Substitutions were checked by PCR and then by sequencing the insert. cDNA encoding polypeptides representing only the N- or C-terminal halves of the ABA-1A1 unit were produced by PCR from the HS10 (*19*) clone.

Recombinant Protein Expression and Purification. Bacterial clones encoding glutathione-*S*-transferase (GST) fusion proteins in pGEX vectors were purified as follows: 1 L of L-broth (or T-broth) containing ampicillin (100 μg mL^{-1}) was inoculated with 100 mL of overnight culture of *E. coli* containing the required plasmid construct, and incubated at 37 °C in an orbital shaker until the A_{550} was in the range 0.3−0.5. At this point, production of recombinant proteins was induced by the addition of isopropyl β-D-thiogalactopyranoside (IPTG) (Sigma) to a final concentration of 0.2 mM. Cultures were incubated for a further 3 h, after which the bacterial cells were harvested by centrifugation at 10000*g* for 10 min. The bacterial pellet was resuspended in 40 mL of 1 × Pharmacia PBS and sonicated on ice for 1 h using a gentle pulse. The sonicate was clarified by centrifugation (13 500 rpm for 25 min), the pelleted debris was discarded, and the supernatant was passed down a glutathione affinity column, purchased from Pharmacia and prepared following manufacturers' guidelines. Recombinant protein bound to the column was released by overnight incubation with 2.5 units of reconstituted thrombin (Sigma), yielding approximately 9 mg of total protein. Before use in analytical experiments, all recombinant proteins were passed down a Pierce ExtractGelD detergent removing column to remove any intrinsic fatty acids.

Chemical Modification of Proteins. Protein (60 μM) in 0.1 M potassium phosphate buffer, pH 7.5, 3 M Gdn-HCl, 1 mM EDTA was added to 30 μL of 3 mM DTNB. The solution was mixed thoroughly and the change in absorbance at 412 nm measured at 5 min intervals over a total of 30 min.

Lyophilized rABA-1A1 (3 mg) was dissolved in 0.1 M Tris-HCl (pH 8.4), 1 mM ethylenediaminetetraacetic acid (EDTA), 6 M guanidine hydrochloride (Gdn-HCl), and 5 mM dithiothreitol (DTT). After incubation at room temperature for 2 h, reduced rABA-1A1 was treated with either 20 mM iodoacetic acid or 20 mM iodoacetamide for 1 h at room temperature in the dark. The alkylation reactions were quenched by the addition of excess β-mercaptoethanol for 1 h.

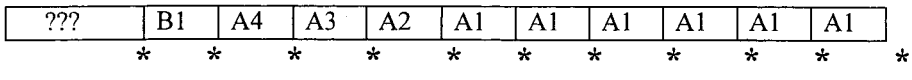
Lyophilized rABA-1A1 (2.6 mg) was resuspended in 8 M urea, pH 4.3 (pH adjusted using glacial acetic acid), and the absorbance at 280 nm measured. Additions of 5 μL of a 20 mM *N*-bromosuccinimide (NBS) (Sigma) solution were made until the A_{280} readings reached a minimum.

Protein Concentration Determination. The concentrations of protein solutions were determined spectrophotometrically using a Schimadzu UV-160A visible recording spectrophotometer, 1 cm path length quartz cuvettes, and ϵ_{280} values of 10 810 M^{-1} cm^{-1} for rABA-1A1, Q20L, and L42R; 5120 M^{-1} cm^{-1} for W15R; 10 930 M^{-1} cm^{-1} for IAA and IAM-rABA-1A1; and 6810 M^{-1} cm^{-1} for NBS-rABA-1A1. ϵ_{280} values for all proteins were calculated from the tyrosine and tryptophan content of individual proteins (*20*). For IAA- and IAM-rABA-1A1, it was assumed that all Cys residues appeared as half-cystines.

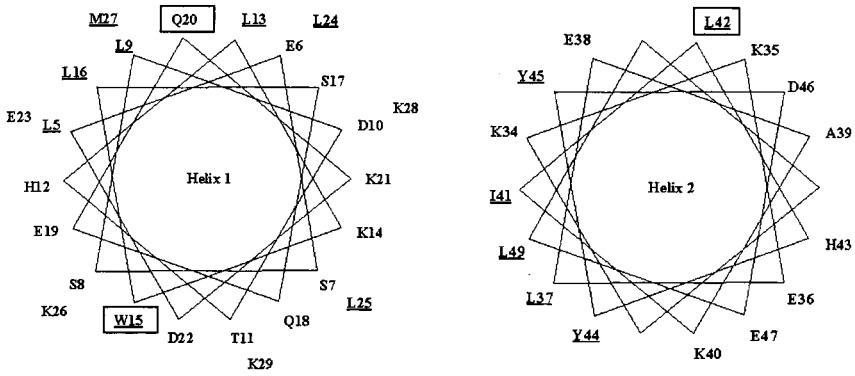
Fluorescence. All fluorescence measurements were performed at 25 °C in a Perkin-Elmer LS50 or a Spex FluoroMax fluorometer (Spex Industries, Edison, NJ) using 2 mL samples in a silica cuvette. Raman scattering by solvent was corrected where necessary using appropriate blank solutions.

The fluorescent fatty acid analogues 11-(((5-(dimethylamino)-1-naphthalenyl)sulfonyl)amino)undecanoic acid (DAUDA) and dansyl-DL-α-aminocaprylic acid (DACA)

A Ascaris suum *polyprotein (ABA-1; As-NPA)*



B



C

```
                1                                                   50
As-NPA-A1       HHFTLESSLD THLKWLSQEQ KDELLKMKKD GKAKKELEAK ILHYYDELEG
Dv-NPA-L        YEIDVDEAIS KYLTWLNEEQ KAEIKQLKEK DE.KQTIGKK IMEFFELTSG
Ce-NPA-K        SEASFQEKAA KYLDWMNEEQ LNELKRLKSE GK.KSEVMKA ILKFYDETTG
Secondary              Helix 1                          Helix 2

               51                                                  100
As-NPA-A1       DAKKEATEHL KGGCREILKH VVGEEKAAEL K....NLKDS GASKEELKAK
Dv-NPA-L        DDKEKAREQL KAACKHYVKM YVGEEKAAEL K....KLKDS GISLEEMSKK
Ce-NPA-K        DAKEKAEGKL KEACKEYSVK AFGEEKVAQF KAQYKKLKDE NAEQSEIEKL
Secondary                        Helix 3

               101                                 137
As-NPA-A1       VEEALHAVTD EEKKQYIADF GPACKKIYGV HTSRRRR
Dv-NPA-L        VTETIETIED EAVRAKARRI HSYCQRIFGI T..KARR
Ce-NPA-K        SNQYIDEIED EQKKDFAKAV VTGCKHVYAS T..RSRR
Secondary       Helix 4
```

D

```
                9                                                   51
N-terminal      LD THLKWLSQEQ KDELLKMKKD GKAKKELEAK ILHYYDELEG DAKKEATEHL KGGC
C-terminal      RE ILKHVVGEEK AAELKNLKDS GASKEELKAK VEEALHAVTD EEKKQYIADF GPAC
Consensus       .+ ...+.+++E. ..EL..+K.. G.+K.EL.AK +......+.. +.KK+..... .++C
```

FIGURE 1: (A) Organization of the polyprotein array of *Ascaris suum.* Only the central repetitive region of the polyprotein is illustrated, ignoring the short hydrophobic C-terminal extension peptide. The 5′ region of the encoding cDNA remains unresolved although complete sequences are available for two other species (*2*). The positions of the consensus $^{Lys}/_{Arg}$-Xaa-$^{Lys}/_{Arg}$-Arg cleavage sites are marked (*). Following cleavage, the proteins appear to be trimmed to remove the cleavage site amino acids; mass spectrometry on ABA-1 protein purified from the parasite yielded a value of 14 643.2 ± 1.4, which, within error, corresponds to the calculated mass (14 642.6) of the ABA-1A1 unit with the cleavage site amino acids (Arg-Arg-Arg-Arg) removed (*41*). The *Ascaris* NPA units were named according to order of discovery and similarity; the A1 units are identical in amino acid sequence, and units A2, A3, and A4 are very similar, but the B1 unit is dramatically different, albeit with similar biochemical activity (*6*). (B) Helical wheel representations of predicted helices 1 and 2 to illustrate the amphipathicity of the predicted helices. Hydrophobic amino acids are underlined, and the boxed positions are those targeted for site-directed mutagenesis. For a more complete discussion of the conserved hydrophobic positions among all NPAs then known, please see (*1*). (C) Alignment of the amino acid sequences of three NPA units to illustrate their diversity. Those selected are encoded by the most 3′ repeat unit from the NPA array of *A. suum* (As-NPA-A1, originally designated ABA-1A1), *Dictyocaulus viviparus* (Dv-NPA-L; DvA-1L), and *Caenorhabditis elegans* (Ce-NPA-K). The signature amino acids found in all NPA units known (with the single exception of the truncated unit H from *D. viviparus*) are indicated by double underlining. The terminal four amino acids which conform to the $^{Lys}/_{Arg}$-Xaa-$^{Lys}/_{Arg}$-Arg consensus cleavage site for processing enzymes, and which are probably removed after cleavage (*2, 41*), are single underlined. The regions predicted to form α helices are as indicated, and no *β*/extended structure has been predicted or detected (*1*). (D) Alignment of the N- and C-terminal halves of one unit (ABA-1A1 from *A. suum*) to illustrate the putative duplication which might have led to the modern ~15 kDa units of the NPAs; a similar effect can be observed readily for other NPA units. The consensus line shows where amino acids are identical (letters), or there are substitutions with amino acids of similar properties (+) according to the Dayhoff substitution matrix (*42*). The predicted helical sections are underlined, and the numbering corresponds to that used for the alignment in (A).

were obtained from Molecular Probes and Sigma, respectively. *all-trans*-Retinol and oleic acid were also obtained from Sigma. *cis*-Parinaric acid (cPnA) was a kind gift from Dr. Bruce Hudson of the University of Oregon, and 8-anilino-1-naphthalenesulfonic acid (ANS) was supplied by Aldrich Chemical Co. The excitation wavelengths used for DAUDA, DACA, retinol, cPnA, and ANS were 345, 345, 350, 319, and 390 nm, respectively. The fluorescent fatty acids were stored as stock solutions of approximately 1 mg mL^{-1} in ethanol in the dark at −20 °C and freshly diluted in phosphate-buffered saline (PBS: 171 mM NaCl, 3.3 mM KCl, 17 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.2) to 1 μM before use in fluorescence experiments. Free retinol is poorly soluble and unstable in aqueous solution, so it was dissolved

and diluted in ethanol immediately before use. Oleic acid was prepared as a 10 mM stock solution in ethanol and diluted in PBS or ethanol before use. A 20 mM solution of ANS was kept frozen and diluted to 198 μM in buffer before use.

With the exception of ANS, the concentrations of all stock solutions of probes used in fluorescence experiments were determined by absorbance using the following extinction coefficients: 4400, 52 480, and 7900 M^{-1} cm^{-1} at wavelengths of 335, 325, 303, and 390 nm for DAUDA (*21*), retinol, and cPnA (*22*), respectively. ANS concentration was determined by weight.

Fluorescent ligand binding titrations were performed as follows: 5 or 10 μL samples of protein at a monomer concentration of 80−100 mM were added successively to 2 mL of fluorescent fatty acid at 1 μM in PBS, and the fluorescence emission spectrum was recorded after each addition. For retinol titration experiments, 80.6 μM retinol in ethanol was added in 5 μL aliquots to 2 mL of a 3 μM solution of protein and mixed immediately. Identical blank titrations in the absence of protein were used to correct for the fluorescence of free retinol.

Fluorescence data were corrected for dilution where necessary, and fitted by standard nonlinear regression techniques (using Microcal ORIGIN software) to a single noncompetitive binding model to give estimates of the dissociation constant (K_d) and maximal fluorescence intensity ($F2$). The generalized hyperbolic fitting function (shown below) was based on the standard equilibrium binding expression for each binding site: P + L ⇔ PL with K_d = [P][L]/[PL]. $F1$ was designated the initial fluorescence intensity (without ligand), x the total ligand value, C_0 the total protein concentration, and n the number of binding sites per protein which are assumed to be independent and noninteracting. In the fitting of all fluorescence titrations, n was fixed at 1.

$$F(\text{observed}) = F1 + \frac{(F2 - F1)(x + nC_0 + K)}{(2nC_0)\left(1 - \sqrt{1 - \frac{(4nxC_0)}{(x + nC_0 + K)^2}}\right)}$$

Circular Dichroism (CD). CD spectra were recorded at 20 °C in a JASCO J-600 spectropolarimeter using quartz cells of path length 0.02 or 0.05 cm. The protein concentrations were in the range 0.07−1 mM for different experiments, as estimated by absorbance at 280 nm. Mean residue weights for each recombinant protein were calculated from their amino acid sequence, and these in turn were used to calculate molar ellipticity values. Analysis of the secondary structure of proteins was performed using the SELCON procedure (*23*) over the range 240−190 nm, unless otherwise stated. GdnHCl (ultrapure grade) was purchased from Life Technologies, Inc., Paisley, Scotland, and the concentration of solutions checked by refractive index measurements (*24*).

RESULTS AND DISCUSSION

Protein Modification Rationale. In the absence of definitive structural information, secondary structure predictions based on multiple alignments of a set of diverse NPA sequences had indicated that each NPA unit is predominantly α helix, with no evidence for β structure (*1*). This prediction was confirmed by CD on either natural or recombinant protein (*1, 6*). Four regions of amphipathic α helix were predicted, each of which exhibited strong amphipathicity (*1*). Helical wheel projections revealed three positions occupied by amino acids which were contrary in nature to the face of the particular helices they occupied. With reference to the sequence of the ABA-1A1 unit of the allergen polyprotein of *Ascaris*, these were Trp15 (a large hydrophobic side chain in a hydrophilic face of helix 1, and absolutely conserved in all NPAs), Gln20 (a polar amino acid in a strongly hydrophobic face, conserved in 94% of NPA sequences, replaced only by charged amino acids Asp or Glu), and Leu42 (a hydrophobic side chain in a strongly charged face of helix, and a conserved hydrophobic position in 69% of NPAs). Trp15 is particularly enigmatic because the expectation was that it is exposed on the outside of the protein, yet it has a fluorescence emission spectrum strongly indicative of deep burial within the protein and consequently remote from solvent water (*1*). These three positions were therefore selected for replacement by mutagenesis with amino acids of a nature akin to that of the helix face they occupied: W15→R, Q20→L, and L42→R. Trp15 was also investigated by chemical modification, which should also provide extra information on whether this position is buried within the structure of the proteins.

The remaining absolutely conserved positions are Cys64 and Cys120, the only cysteine residues in the entire sequence. Their most likely role, especially in the strongly oxidizing extracellular environment of NPAs, is tertiary structure stabilization by disulfide bridge formation. The proposal that Cys64 and Cys120 form intramolecular −S−S− links receives strong support from the fact that there are no indications of covalent oligomers resulting from intermolecular cross-linking, although there is evidence that NPAs form noncovalent dimers in solution (*1*). Free sulfhydryl groups, on the other hand, might contribute to the hydrophobicity of a binding site or the internal structure of a protein. Consequently, we examined the oxidation state and role of these Cys residues by standard reduction/alkylation methods.

Finally, sequence analysis had provided evidence of an internal duplication within NPA units (*1*), each part of which could represent the ancestral protein and be the main contributor of the binding site of the modern proteins. Very small fatty acid binding proteins have been described from other invertebrates and from plants (*25*−*30*), so it is indeed conceivable that the NPA unit is descended from a duplication of a very small fatty acid binding progenitor. Recombinant proteins representing C- and N-terminal halves of the NPA unit were therefore constructed, the site chosen for the division of the protein being defined by the limits of the presumptive duplication and coinciding with a break in the predicted helical structure in the central region of the protein (Figure 1).

The recombinant wild-type ABA-1A1, mutants thereof, and polypeptides representing N- or C-terminal halves of the wild-type protein were produced as described under Experimental Procedures and purified to yield single bands on SDS−PAGE and single symmetrical peaks on gel

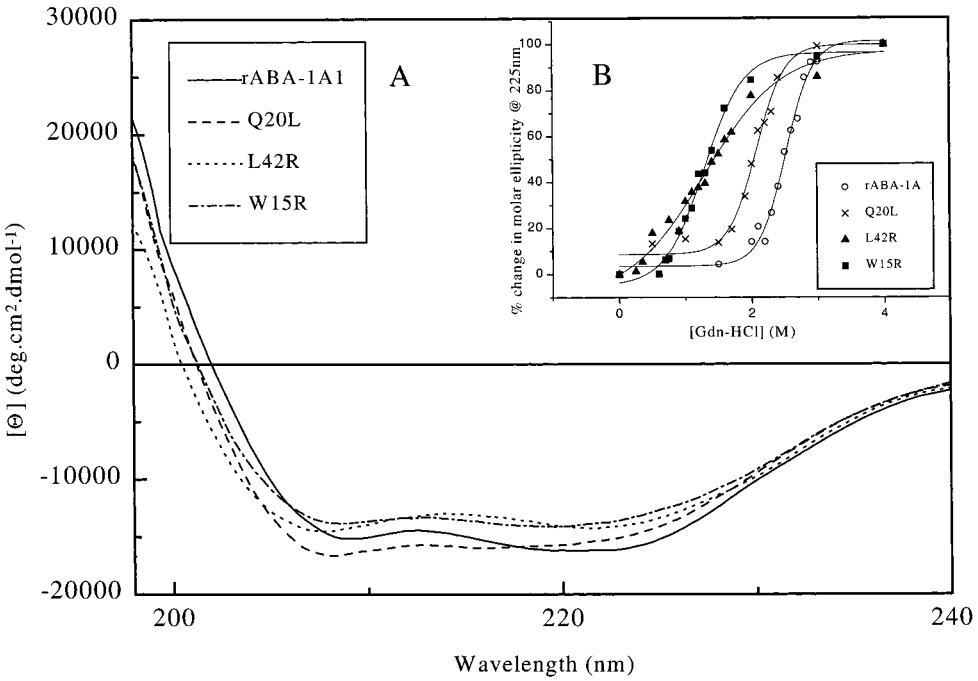


FIGURE 2: Far-UV CD analysis of wild-type and mutant rABA-1A1 proteins and sensitivity to denaturation by Gdn-HCl. (A) Far-UV CD spectra recorded at 20 °C on 70 μM protein solutions in PBS, path length 0.02 cm, providing α-helix content estimates of 50%, 43%, 47%, and 41% for rABA-1A1 (the wild-type form), W15R, Q20L, and L42R, respectively, as determined using the SELCON procedure (*23*). (B) Changes in molar ellipticity at 225 nm upon progressive addition of Gdn-HCl, path length 0.05 cm. The sigmoidal curves are drawn for guidance only.

filtration, with the exception of one which behaved anomalously under the separation conditions used (see below).

Neither Trp15 nor Gln20 Is Essential for Folding or Function. CD spectra of the wild-type protein (rABA-1A1) and the W15R and Q20L mutant proteins were essentially identical and consistent with the anticipated high content of α helix (Figure 2). Gdn-HCl denaturation experiments revealed the wild-type protein to be stable up to the addition of 2 M denaturant, whereupon unfolding proceeded steadily with increasing concentrations of Gdn-HCl, and was essentially complete by 3 M (Figure 2B). Neither W15R nor Q20L was as stable as the wild-type protein, with the former proving to be particularly sensitive to denaturant, and the latter being the most stable of the three mutant proteins.

Fatty acid binding activities were investigated using fluorescence-based assays and ligands tagged with the environment-sensitive dansyl fluorophore, as described previously (*1, 4, 5, 31*). Figure 3A illustrates the enhancement and blue-shift in the fluorescence emission of DAUDA upon interaction with the protein. In each of rABA-1A1, W15R, and Q20L, identical shifts in peak emission were obtained, moving from 542 nm (DAUDA in buffer) to 475 nm upon addition of the proteins (Figure 4). Fluorescence titrations were used to estimate the dissociation constants of a range of fluorescent fatty acids (Figure 4 and Table 1), and only minor disparities were observed in the behaviors of rABA-1A1, W15R, and Q20L. This was particularly surprising in the case of the W15R mutant given the absolute conservation of Trp in this position, and the fact that Trps are often encountered in, or close to, the binding sites of fatty acid binding proteins of diverse structures (*1, 32*−*38*).

The W15R mutant did, however, distinguish itself from the other mutants upon interaction with DACA, a fluorescent fatty acid probe in which the dansyl fluorophore is attached

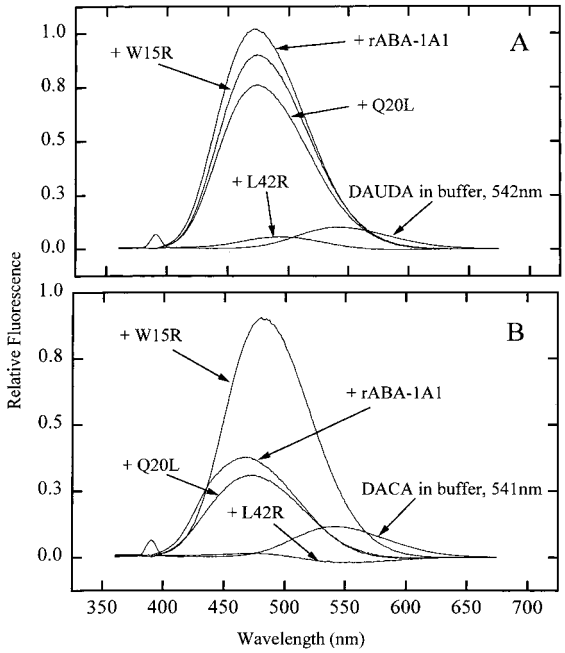


FIGURE 3: Fatty acid binding by the wild-type and mutant proteins. Binding of (A) DAUDA and (B) DACA to recombinant ABA-1A and the three site-directed mutants. Fluorescence emission spectra (λ_{Exc} = 345 nm) of approximately 1 μM solution of DAUDA or DACA alone, or with 1 μM protein. Interaction of DAUDA with each of the proteins produced shifts from 542 nm (DAUDA or DACA in buffer) to 473, 474, 475, and 493 nm upon addition of rABA-1A1, W15R, Q20L, and L42R, respectively, although L42R exhibited only minimal binding. Each of the spectra produced upon addition of the various proteins are subtracted from that for DAUDA or DACA alone to provide the emission characteristics of the two probes when in the protein binding sites.

at the α carbon, rather than to the hydrocarbon terminal as in DAUDA. While Q20L and rABA-1A1 produced similar

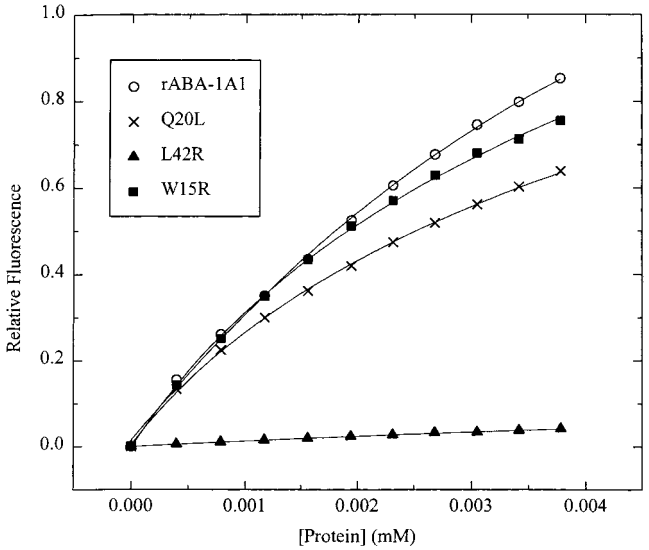


FIGURE 4: Titration curves for the binding of DAUDA to rABA-1A1 and the three site-directed mutants. Changes in relative DAUDA fluorescence intensity were recorded at the wavelength of maximum fluorescence emission produced by each protein upon incremental 10 μL additions of approximately 80 μM protein solutions to a cuvette containing 1 μM DAUDA. The solid line is the theoretical binding curve for complex formation, and is in all cases consistent with one binding site per monomer unit of protein. Dissociation constants are listed in Table 1.

Table 1: Ligand Binding Dissociation Constants and Tryptophan Accessibility of rABA-1A1 and Each of the Site-Directed Mutants[a]

		K_d (μM)			
protein	K_{SV} (M^{-1})	DAUDA	DACA	cPnA	retinol
rABA-1A1	1 ± 0.03	5 ± 1.4	8 ± 2.0	0.2 ± 1.61	0.1 ± 0.01
W15R	*NA	2 ± 0.37	2 ± 0.7	0.7 ± 3.1	0.2 ± 0.07
Q20L	0.8 ± 0.03	2 ± 1.4	17 ± 2.1	0.2 ± 0.82	0.03 ± 0.05
L42R	2 ± 0.03	*NB	*NB	*NB	*NB

[a] K_{SV}: Stern−Volmer constant for succinimide quenching of tryptophan fluorescence. For the model compound *N*-acetyltryptophanimide, $K_{SV} = 12$ M^{-1}. K_d: dissociation constant estimated by titration of different fluorescent probes with the respective proteins, estimated as described under Experimental Procedures. *NA: not applicable (no tryptophan present). *NB: no detectable binding.

blue shifts in the fluorescence emission of this probe, W15R produced an enhancement in DACA fluorescence almost double that produced by the other proteins (Figure 3B). Presumably, therefore, W15R binds DACA in an altered manner, possibly because the newly introduced Arg in some way causes an alteration in the environment of the charged headgroup of bound fatty acid, but not of the hydrocarbon tail.

Mutation L42→R Compromises both Folding and Function. The L42→R mutation proved to have substantially greater consequences for the protein than did the other mutations. First, a proportion of the protein appeared to form covalent dimers in that nonreducing SDS−PAGE gave two bands with relative mobilities commensurate with monomer and dimer, and gels run under reducing conditions gave only one band of monomer molecular size (not shown). Furthermore, two overlapping peaks were obtained on gel filtration. This indicated a less stable fold for L42R, such that untoward S−S bonding was occurring posttranslationally or following release of protein from the host bacteria.

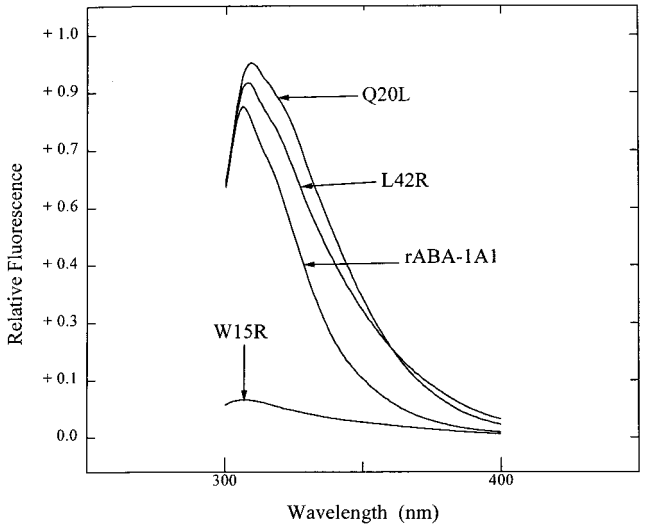


FIGURE 5: Tryptophan fluorescence emission spectra of approximately 7 μM solutions of rABA-1A1 and each of its site-directed mutants. $\lambda_{Exc} = 290$ nm.

CD analysis suggested that L42R folds to have a similar proportion of helix to the wild type and mutant proteins, and the buried environment of its Trp15 appears unchanged, as judged from the similarity in the fluorescence emission spectrum of its Trp15 (Figures 2 and 5). Stern−Volmer analysis of succinimide quenching experiments, however, revealed the protein to have a K_{SV} value double those of the wild-type protein (rABA-1A1) and the Q20L mutant (Table 1). L42R was also readily denatured by Gdn-HCl, unfolding becoming apparent at a lower concentration than for the other proteins (Figure 2B).

L42R failed to bind DAUDA (Figure 3A) or indeed any of the fluorescence fatty acid probes tested (Figure 3B and Table 1). In common with all the proteins, however, it bound ANS. This may merely be a reflection of exposure of a hydrophobic region in L42R, given that ANS is considered to be a nonspecific probe for exposed hydrophobic regions in proteins (*39*) and will not be as sensitive to alteration of a fatty acid binding site as would DAUDA, DACA, and cPnA. This interpretation is reinforced by the finding that the ANS fluorescence enhancement was reversed upon addition of free fatty acid (oleic) to a preformed complex of ANS and any of the proteins, with the single exception of L42R. The continued binding of ANS may therefore be due either to a still accessible, but corrupted, binding site and/or to exposure of new hydrophobic surfaces in the mutant protein.

Trp15 Is Deeply Buried. Helical wheel predictions suggest Trp15 should reside in a hydrophilic, water-exposed environment (Figure 1). Paradoxically, rABA-1A1's tryptophan emission spectrum is extremely blue-shifted, indicative of this single tryptophan being being deeply buried in a nonpolar environment (*40*). Figure 5 shows the intrinsic fluorescence emission spectra of rABA-1A1, Q20L, and L42R. Each spectrum has a sharp peak at approximately 307 nm, and a shoulder at 315 nm, and will be a composite of fluorescence emission by the single Trp plus the four tyrosines in each molecule. The W15R mutant, however, exhibits no shoulder at 315 nm, identifying this feature as due to Trp15 (Figure 5). Succinimide quenching confirmed the extent to which W15R's side chain is buried within each of the proteins;

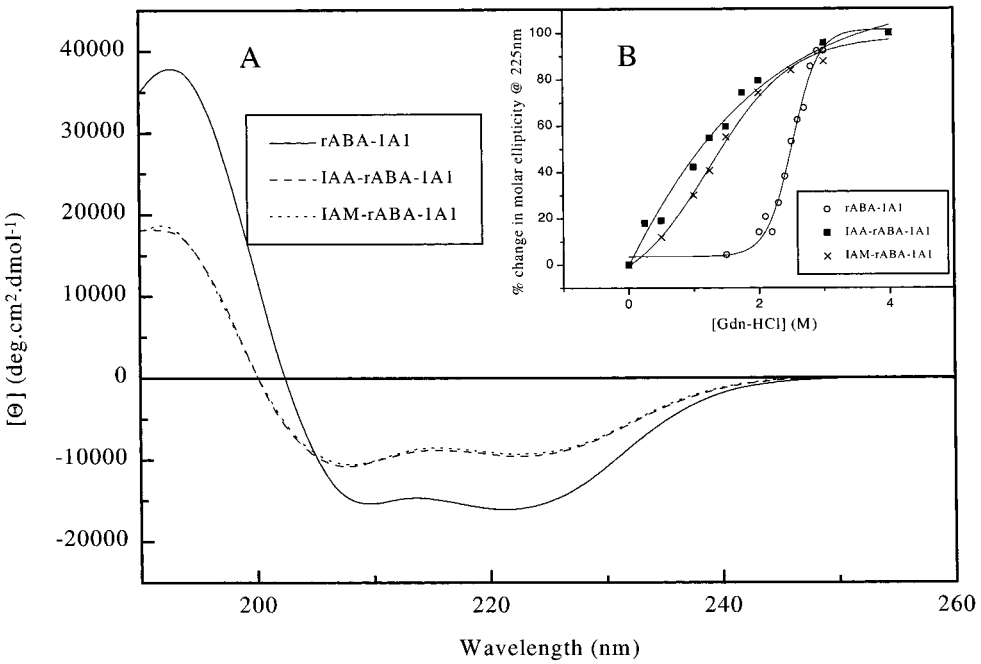


FIGURE 6: (A) Far-UV CD spectra of rABA-1A1 and its reduced and alkylated derivatives. Spectra were recorded at 20 °C using protein solutions of 66 μM, path length 0.02 cm. α helix estimates of 54%, 28%, and 28% for rABA-1A1, IAA-, and IAM-rABA-1A1, respectively, were obtained when analyzed using the SELCON procedure (*23*). (B) Change in molar ellipticity at 225 nm upon progressive additions of Gdn-HCl to an approximately 66 μM sample of each protein, path length 0.05 cm.

rABA-1A1, Q20L, and L42R (but see below) were poorly quenched in comparison to tryptophan in solution, and the K_{sv} values are appropriately low (Table 1).

It is conceivable that the unexpected emission spectrum of Trp15 is due to an unusual local environment of the side chain rather than remoteness from solvent, so the accessibility of Trp15 was further tested by chemical attack with the oxidizing agent NBS. As followed by changes in the emission spectrum and absorbance of each of the proteins, NBS was only able to attack the tryptophanyl residue in the presence of large concentrations of urea (8 M). Trp15 is therefore only chemically accessible when the protein is unfolded.

So, the paradox with Trp15 remains. It is predicted to be on the hydrophilic aspect of an amphipathic helix, and therefore exposed on the surface of the protein. All the experimental evidence, however, points to it being deeply buried within the structure of the protein. It is therefore unlikely to be in the binding site of the protein (to which water and quenching agents will get some degree of access), and its substitution does not affect the fatty binding activity of the protein. Thus, either the helix bearing Trp15 is kinked or twisted such that the side chain is buried, or the helix is involved in stabilizing inter- or intramolecular interactions such that Trp15 is isolated from solvent. The fact that the stability of the W15R mutant protein is compromised is perhaps consistent with the latter idea.

Disulfide Bridges Are Essential for Function but Not Folding. The existence of free sulfhydryls in rABA-1A1 was examined by treatment of the protein with Ellman's reagent (DTNB). The reaction produced no increase in absorbance at 412 nm, whereas identical treatment of bovine serum albumin (which possesses one free sulfhydryl) produced an appropriate increase in absorbance at 412 nm (data not shown). Moreover, SDS−PAGE analysis of reduced and nonreduced rABA-1A1 yielded single bands of M_r ~14 400 (not shown), as does the natural protein (*1*). Cys64 and Cys120 are therefore disulfide-bonded within a single polypeptide chain of rABA-1A1, and not involved in dimerization.

The role of this single disulfide bridge in the structure and biochemical activity of the protein was then examined by reduction with DTT prior to alkylation with either IAA or IAM. The modified proteins appeared as single bands on SDS−PAGE, and their molecular masses (measured using electrospray mass spectrometry) had increased by the appropriate number of mass units to indicate that the S-alkylation reactions had proceeded to completion. The Cys−Cys bond was found not to be available for reduction unless the protein was first denatured with Gdn-HCl (data not shown), presumably because the two participating Cys residues are deeply buried within the structure of the protein.

Both IAA- and IAM-rABA-1A1 retained the anomalous blue shift in their intrinsic fluorescence emission spectra (λ_{Exc} = 290 nm), so, again, the folding of the protein necessary to create the unusual environment of the Trp15 side chain appeared to proceed normally. Both derivatives were, however, markedly less stable to Gdn-HCl denaturation when compared to the wild-type protein (Figure 6B). Both began to unfold upon addition of low concentrations of denaturant, and the lack of a sharp transition in both cases indicated that unfolding was progressive rather than cooperative in both cases. Both derivatives showed little binding affinity for ligand: a minimal shift in the emission spectrum of DAUDA was obtained upon interaction of the reduced and alkylated proteins with DAUDA (from 542 to 487 nm in both cases), and no substantial increase in fluorescence emission was obtained (data not shown). The Cys64−Cys120 disulfide bond is therefore essential to both the structural integrity of the protein and its binding activity, although, curiously, the

local environment of Trp15 remains entirely unaffected by such gross alterations to the protein.

Localization of Ligand Binding Site. As yet we have no indications of where the crucial regions of the protein are for ligand binding, although it appears that the region bearing Trp15 is probably not involved. Sequence analysis has suggested that the NPA unit is itself derived from an ancient duplication event, and that the ancestral fatty acid binding protein might have been of the order of 7−8 kDa [such proteins have been described from plants and invertebrate animals (*25*−*30*)]. Subsequent evolutionary modifications of the 14 kDa protein might mean that it is impossible to identify the precise properties of the ancestor from manipulations of the modern form. Nevertheless, it may be possible to localize crucial anchor side chains and/or a region contributing to the unusually apolar binding site of ABA-1 by taking N- and C-terminal halves of the protein independently and examining their binding properties. This strategy is made all the more logical given the predicted ancestral duplication event.

cDNAs encoding the N- (N-ter) and C-terminal (C-ter) halves of ABA-1A1 were cloned and the polypeptides expressed in *E. coli* as described under Experimental Procedures. CD measurements indicated that each of the fragments folds with secondary structure estimates of 50%, 41.3%, and 33.6% α helix content for rABA-1A1, N-ter, and C-ter, respectively. rABA-1A1's conserved Trp15 residue is present within the N-terminal fragment, and the intrinsic indole fluorescence of this revealed a peak of emission at the same wavelength as the complete protein (not shown). So, once again, a dramatic alteration to the protein left the protein environment of Trp15 essentially unaffected. Denaturation experiments with Gdn-HCl showed that both fragments were considerably less stable than the parent protein, both beginning to unfold upon addition of as little as 0.5 M Gdn-HCl (not shown).

Addition of the C-terminal fragment to DAUDA solution caused an increase in the DAUDA fluorescence emission and a blue shift in the peak fluorescence to 488 nm. This shift is less than for the complete protein, but nevertheless is still substantially more than occurs with other types of fatty acid binding proteins [500 nm (*31*)]. In marked contrast, the N-terminal fragment showed no sign of interaction with DAUDA (Figure 7B).

C-ter induced a large increase in the fluorescence emission intensity and a blue shift in the wavelength of maximum emission in ANS (Figure 7A), but N-ter produced only a slight change (not shown). This suggests that only the C-terminal fragment possesses an exposed hydrophobic surface or pocket, although it remains possible that compared with the intact molecule, N-ter is altered so as to fold to obscure an otherwise exposed hydrophobic surface. Significantly, oleic acid was found partially to displace ANS from C-ter (Figure 7A).

These data indicate that the binding site for DAUDA and ANS is in C-ter and possesses the essential attributes of the ABA-1A1 binding site in terms of fatty acid binding (by displacement of fluorescence probes by oleic acid) and the degree of apolarity of the binding site (as judged by the degree of blue shift in the DAUDA fluorescence emission). It may therefore be that the C-terminal half of ABA-1A1 encompasses the ligand binding site, and that the N-terminal

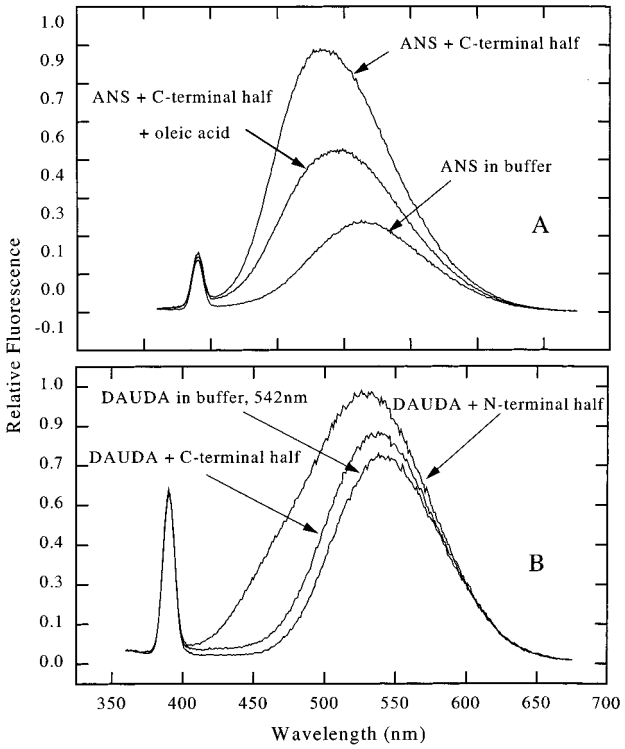


FIGURE 7: (A) Binding of ANS to the C-terminal half polypeptide (C-ter) of rABA-1A1 and its displacement by oleic acid. Fluorescence emission spectra of 2 μM ANS in PBS, and upon addition of 2 μM C-ter, and the result of subsequent addition of 1 μM oleic acid to the mixture (λ_{Exc} = 390 nm). (B) Binding of the fluorescently tagged fatty acid, DAUDA, to the N- and C-terminal halves of ABA-1A1. Shown is the emission spectra of approximately 1 μM DAUDA alone, or upon addition of approximately 2 μM protein. λ_{345} = 345 nm.

half has been modified for other purposes, although its progenitor may have, like C-ter, also been capable of binding lipids. The N-ter half of the protein, however, appears to influence the function of the C-ter given that the mutation of a single amino acid position in N-ter (L42R) abrogates lipid binding of the compete molecule. At present, it is only possible to speculate on what the functions of N-ter are, such as that it is involved in dimerization, and/or involved in interaction with other components of the worms such as cell surface receptors in the uptake or delivery of ligands. Since several species of parasitic nematodes appear to secrete these proteins, it is also conceivable that the N-terminal half has also been adapted to interact with components of the human or animals hosts. Interestingly, the DvA-1 polyprotein of the bovine lungworm *Dictyocaulus viviparus* contains an abbreviated repeat unit (unit H) which is approximately half the size of all other units (*3*), and could conceivably represent the primordial protein, although this possibility remains to be explored further.

The NPAs remain enigmatic proteins in their mode of biosynthesis, ligand binding properties, and structure. This is compounded now by our present findings in which the amino acid positions which are absolutely or highly conserved, and which appear in unexpected positions in the secondary structure, are crucial to the structure of the proteins, but not their ligand binding properties. The only cases in which we found disruption of ligand binding (in the L42R mutant or upon chemical reduction) were those in which there were also dramatic effects on structure and

stability. It is therefore likely that the changes in ligand binding may have been due to gross structural modifications rather than affecting residues that are directly involved in ligand binding. The fact that the protein is probably divisible into two functionally discrete domains is, however, important to our investigation of their function, and the genetically modified proteins now available could be most valuable in the studies of lipid binding by NPAs and the evolution of polyproteins in general.

ACKNOWLEDGMENT

We thank Dr. Andrew Pitt of the Department of Pure and Applied Chemistry at the University of Strathclyde, Glasgow, for the use of the electrospray mass spectrometry facility.

REFERENCES

1. Kennedy, M. W., Brass, A., McCruden, A. B., Price, N. C., Kelly, S. M., and Cooper, A. (1995) *Biochemistry 34*, 6700−6710.
2. Kennedy, M. W. (2000) *Biochim. Biophys. Acta 1476*, 149−164.
3. Britton, C., Moore, J., Gilleard, J. S., and Kennedy, M. W. (1995) *Mol. Biochem. Parasitol. 72*, 77−88.
4. Kennedy, M. W., Britton, C., Price, N. C., Kelly, S. M., and Cooper, A. (1995) *J. Biol. Chem. 270*, 19277−19281.
5. Kennedy, M. W., Allen, J. E., Wright, A. S., McCruden, A. B., and Cooper, A. (1995) *Mol. Biochem. Parasitol. 71*, 41−50.
6. Moore, J., McDermott, L., Price, N. C., Kelly, S. M., Cooper, A., and Kennedy, M. W. (1999) *Biochem. J. 340 (Pt. 1)*, 337−343.
7. Glatz, J. F., and Van der Vusse, G. J. (1996) *Prog. Lipid Res. 35*, 243−282.
8. Tomlinson, L. A., Christie, J. F., Fraser, E. M., McLaughlin, D., McIntosh, A. E., and Kennedy, M. W. (1989) *J. Immunol. 143*, 2349−2356.
9. Christie, J. F., Dunbar, B., Davidson, I., and Kennedy, M. W. (1990) *Immunology 69*, 596−602.
10. Yasothornsrikul, S., Toneff, T., Hwang, S. R., and Hook, V. Y. (1998) *J. Neurochem. 70*, 153−163.
11. Chou, K. C. (1996) *Anal. Biochem. 233*, 1−14.
12. Nagle. G. T., Knock, S. L., Vanheumen, W. R. A., and Kurosky, A. (1994) *Neth. J. Zool. 44*, 439−450.
13. Nedderman, P., Tomei, L., Steinkuhler, C., Gallinari, P., and Tramontano, A. D. R. (1997) *Biol. Chem. 378*, 469−476.
14. Gan, S. Q., McBride, O. W., Idler, W. W., Markova, N., and Steinert, P. M. (1991) *Biochemistry 30*, 5814.
15. Haydock, P. V., and Dale, B. A. (1990) *DNA Cell Biol. 9*, 251−261.
16. Presland, R. B., Haydock, P. V., Fleckman, P., Nirunsuksiri, W., and Dale, B. A. (1992) *J. Biol. Chem. 267*, 23772−23781.
17. Yoneda, K., Hohl, D., McBride, O. W., Wang, M., Cehrs, K. U., Idler, W. W., and Steinert, P. M. (1992) *J. Biol. Chem. 267*, 18060−18066.
18. Poole, C. B., Hornstra, L. J., Benner, J. S., Fink, J. R., and McReynolds, L. A. (1996) *Mol. Biochem. Parasitol. 82*, 51−65.
19. Spence, H. J., Moore, J., Brass, A., and Kennedy, M. W. (1993) *Mol. Biochem. Parasitol. 57*, 339−343.
20. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
21. Haughland, R. P. (1992) *Handbook of Fluorescent Probes and Research Chemicals*, 5th ed., Molecular Probes, Eugene, OR.
22. Sklar, L. A., Hudson, B. S., Petersen, M., and Diamond, J. (1977) *Biochemistry 16*, 813−819.
23. Greenfield, N. J. (1996) *Anal. Biochem. 235*, 1−10.
24. Nozaki, Y. (1972) *Methods Enzymol. 26 (Pt. C)*, 43−50.
25. Shin, D. H., Lee, J. Y., Hwang, K. Y., Kim, K. K., and Suh, S. W. (1995) *Structure 3*, 189−199.
26. Gincel, E., Simorre, J. P., Caille, A., Marion, D., Ptak, M., and Vovelle, F. (1994) *Eur. J. Biochem. 226*, 413−422.
27. Heinemann, B., Andersen, K. V., Nielsen, P. R., Bech, L. M., and Poulsen, F. M. (1996) *Protein Sci. 5*, 13−23.
28. Lee, J. Y., Min, K., Cha, H., Shin, D. H., Hwang, K. Y., and Suh, S. W. (1998) *J. Mol. Biol. 276*, 437−448.
29. Barrett, J., Saghir, N., Timanova, A., Clarke, K., and Brophy, P. M. (1997) *Eur. J. Biochem. 250*, 269−275.
30. Janssen, D., and Barrett, J. (1995) *Biochem. J. 311 (Pt. 1)*, 49−57.
31. Wilkinson, T. C., and Wilton, D. C. (1986) *Biochem. J. 238*, 419−424.
32. He, X. M., and Carter, D. C. (1992) *Nature 358*, 209−215.
33. Frapin, D., Dufour, E., and Haertle, T. (1993) *J. Protein Chem. 12*, 443−449.
34. Newcomer, M. E. (1995) *FASEB J. 9*, 229−239.
35. Banaszak, L., Winter, N., Xu, Z., Bernlohr, D. A., Cowan, S., and Jones, T. A. (1994) *Adv. Protein Chem. 45*, 89−151.
36. Sacchettini, J. C., Gordon, J. I., and Banaszak, L. J. (1989) *J. Mol. Biol. 208*, 327−339.
37. Sacchettini, J. C., and Gordon, J. I. (1993) *J. Biol. Chem. 268*, 18399−18402.
38. Curry, S., Mandelkow, H., Brick, P., and Franks, N. (1998) *Nat. Struct. Biol. 5*, 827−835.
39. Varley, P. G. (1994) *Methods Mol. Biol. 22*, 203−218.
40. Eftink, M. R., and Ghiron, C. A. (1976) *Biochemistry 15*, 672−680.
41. Christie, J. F., Dunbar, B., and Kennedy, M. W. (1993) *Clin. Exp. Immunol. 92*, 125−132.
42. Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1978) *Atlas Protein Sequence Struct. 5*, 345−362.

BI0026876