Sequence-divergent units of the ABA-1 polyprotein array of the nematode *Ascaris suum* have similar fatty-acid- and retinol-binding properties but different binding-site environments

Joyce MOORE^{*1}, Lindsay McDERMOTT^{*†}, Nicholas C. PRICE[‡], Sharon M. KELLY[‡], Alan COOPER[†], and Malcolm W. KENNEDY^{*2}

^{*}Division of Infection and Immunity, Institute of Biomedical and Life Sciences, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K.,

[†]Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K., and [‡]Department of Biological Sciences, University of Stirling, Stirling FK9 4LA, Scotland, U.K.

Polyproteins comprise long polypeptides that are post-translationally cleaved into proteins of different function, or tandemly repetitive polypeptides which are processed into multiple versions of proteins which are presumed to have the same function. In the latter case the individual units of the polyprotein can differ substantially in sequence. Identity of function between the different units therefore cannot be assumed. Here we have examined the ABA-1 polyprotein allergen of the parasitic nematode *Ascaris suum* and found it to contain units which show a 50% difference in amino acid sequence. The parasite therefore produces at least two radically different forms of the allergen encoded within the polyprotein array. In fluorescence-based ligand-binding assays, recombinant polypeptides representing the two forms (designated ABA-1A1 and ABA-1B1) showed similar binding affinities for a range of fluorescent active-site probes [retinol, dansylundecanoic acid, dansyl-DL- α -amino-

octanoic acid, *cis*-parinaric acid (cPnA)] and for the non-specific hydrophobic surface probe 8-anilinoanthracene-1-sulphonic acid. However, the molecular environments in the active sites are markedly different, as indicated by disparate fluorescence emission peaks and intensities of bound probes. CD showed that the proteins have similar secondary structures but differ in susceptibility to chemical denaturation/unfolding by guanidinium chloride. Both retain a single conserved tryptophan residue in a characteristic non-polar environment, as revealed by extreme fluorescence blue shift. Thus the gross differences in sequence of the two proteins are not reflected in their ligand-binding specificities but in their binding-site environments.

Key words: ABA-1 allergen, fatty-acid-binding protein, polyproteins, retinol-binding protein.

INTRODUCTION

The term polyprotein is used to denote two different types of entity. The first is exemplified by viral proteins and certain vertebrate hormones in which a single nascent polypeptide is cleaved into proteins with completely different functions [1–5]. The second is typified by certain proteins of vertebrate keratinocytes in which a large polypeptide containing tandemly repeated units is cleaved post-translationally into polypeptides which are presumed, in the absence of direct testing, to have similar or identical functions [6–10]. 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 advantage in producing proteins in repetitive polyprotein arrays is presumably that it represents a highly efficient means by which to produce large quantities of a functional product [8]. Proproteins of terminally differentiated mammalian keratinocytes such as filaggrin, involucrin and loricin comprise peptide or polypeptide repeats that display sequence variations yet retain distinctive structural motifs, and are encoded by genes lacking introns in the coding region [6–10]. The postulated advantage of

this is that a maturing epidermal granular cell needs to mobilize rapidly, and precisely regulate, a large pool of cell type-specific proteins in the face of nuclear function which is compromised by approaching senescence [8].

However, examples of repetitive polyproteins with intronless coding regions have recently arisen which break the rule that such proteins are products of a senescing cell. These are the polyprotein allergen proteins of nematodes (NPAs) which bind fatty acids and retinol [11–18]. Some NPAs exhibit dramatic differences in the sequences of the repeated units, although each possesses conserved signature amino acid positions [15,16]. To date, the ligand-binding characteristics of these proteins have been demonstrated only for single units of the polyprotein from a variety of nematode species, but not for two different units from the same polyprotein array. It therefore remains a possibility that the various members of the same array are functionally different *in vivo*, which will be important, given that NPA proteins are secreted into the tissues inhabited by parasitic nematodes in humans and other hosts [11–13,15].

We have here selected a pair of sequence-divergent units from the same NPA array, and, using recombinant polypeptides, compared their structure, stability and ligand-binding properties. The latter were examined using environment-sensitive fluorescent

Abbreviations used: A1, recombinant protein encoded in the A1 unit of the ABA-1 protein of *Ascaris suum*; ANS, 8-anilinoanthracene-1-sulphonic acid; B1, recombinant protein encoded in the B1 unit of the ABA-1 protein of *Ascaris suum*; cPnA, *cis*-parinaric acid; DACA, dansyl-DL- α -amino-octanoic acid; DAUDA, 11-(dansylamino)undecanoic acid; FABP, fatty-acid-binding protein; GdmCl, guanidinium chloride; GST, glutathione S-transferase; retinol, all-*trans*-retinol; NPA, nematode polyprotein allergen/antigen; FABP, fatty-acid-binding protein.

¹ Present address: Department of Veterinary Parasitology, University of Glasgow, Glasgow G61 1QH, Scotland, U.K.

² To whom correspondence should be sent (e-mail malcolm.kennedy@bio.gla.ac.uk).

The sequences of the new cDNAs mentioned, but not shown, in this paper have been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number AF051702.

probes which can explore subtle differences in the nature of a binding site. The two proteins were found to have similar ligand-binding activities, but showed differences in the environments of their binding sites, a finding that may be of biological significance. The findings prompt the question of why other proteins which are required in abundance (or at short notice) are not also produced in this manner and what the selective advantage is in having sequence diversity in this kind of polyprotein array.

MATERIALS AND METHODS

cDNAs encoding ABA-1

A cDNA expression library was constructed from mRNA from the infective stage of *Ascaris suum* as previously described [14]. The library was screened with rabbit antibody against excretory/secretory products of the infective larvae of the parasite [19], and two positive clones (pJMA1 and pJMA2) were isolated, the plasmids excised, and the sequences of the entire inserts on both strands obtained using standard methods [20].

Expression and purification of recombinant proteins

Primers A-Nter (ggaattcCATCATTTACCCCTTG; lower-case letters indicate the appended *EcoRI* restriction site plus extra nucleotides to enable cutting) and A-Cter (ggaattcCCTCCTTCGTCGCGAAG) were used to amplify the 3'-most repeat unit known to be encoded within the *aba-1* array [14,21]. Primer A-Cter Δ 4R (gggaattcTCACGAAGTATGTACACCATAG) was used in place of A-Cter to produce a PCR product which did not encode the four terminal arginine residues which are probably absent in the mature protein [14,22–25]. The above combination of primers were used in PCR reactions containing 1 ng of pHS10 DNA [14] with the following conditions: 94 °C for 1 min/55 °C for 1 min/72 °C for 1 min for 30 cycles, followed by 72 °C for 5 min. The above refers to production of DNA encoding the type A (ABA-1A) unit in the NPA array. Primers B-Nter (gggaattcCATACAATGGAACACTATC) and B-Cter (gggaattcCCTCCTTCGTCGATGATG) were used similarly in PCR reactions to produce a single repeat unit corresponding to the type B (ABA-1B) unit, except that the PCR annealing temperature was reduced to 50 °C to allow for the characteristics of the primers used, and the pJMA1 plasmid was used as a template. The PCR reaction products were cut with *EcoRI* and purified on agarose gels. The purified fragments were ligated to *EcoRI*-digested and alkaline phosphatase-treated pGEX-1 λ T (Pharmacia, Milton Keynes, U.K.) expression vector. The plasmid clones thus obtained were checked by sequencing to validate the fidelity of the PCR amplification.

Escherichia coli XL1-Blue containing pGEX-1 λ T derivatives containing inserts encoding ABA-1A1 or ABA-1B1 were induced to produce the recombinant proteins with isopropyl thio- β -D-galactoside (0.2 mM) and cells were harvested by centrifugation at 15000 *g* for 10 min at 4 °C. The pellet was resuspended in PBS (171 mM NaCl/3.35 mM KCl/10 mM Na₂HPO₄/1.8 mM KH₂PO₄, pH 7.2), sonicated on ice and centrifuged at 22000 *g* for 30 min at 4 °C. The glutathione S-transferase (GST) fusion protein was purified from the supernatant by binding to GSH-agarose beads (Sigma). In order to minimize contamination with GST, the fusion protein partner, recombinant ABA-1A or ABA-1B fusion proteins with GST were cut using thrombin while still bound to the GSH-agarose beads. Approx. 5 NIH (National Institutes of Health) units of thrombin (Sigma)/mg of recombinant protein were added to the column in a volume equal to the

bed volume in buffer containing 2.5 mM CaCl₂. The column was incubated at room temperature for 24 h and the recombinant ABA-1A or ABA-1B proteins collected by addition of buffer. Residual detergent was removed from solutions of protein by passage through an Extracti-Gel D column (Pierce, Chester, U.K.).

Spectrofluorimetry and fluorescence-based ligand-binding assays

Fluorescence emission spectra were recorded at 20 °C with a SPEX FluorMax spectrofluorimeter (Spex Industries, Edison, NJ, U.S.A.), using 2 ml samples in a silica cuvette. All spectra are uncorrected for photomultiplier response. Raman scattering by solvent water was subtracted where necessary using appropriate blank solutions. The fluorescent fatty acid analogues 11-(dansylamino)undecanoic acid (DAUDA) and dansyl-DL- α -amino-octanoic (= caprylic) acid (DACA) were obtained from Molecular Probes (Eugene, OR, U.S.A.) and Sigma respectively. All-*trans*-retinol, oleic acid, 8-anilinoanthralene-1-sulphonic acid (ANS) and dansylamide were also obtained from Sigma, and cPnA was kindly given by Dr. Bruce Hudson of the University of Oregon or obtained from Molecular Probes. The excitation wavelengths used for DAUDA, DACA, retinol and cPnA were 345 nm, 345 nm, 350 nm and 319 nm respectively. The dansylated fatty acids were stored as stock solutions of approx. 1 mg/ml in ethanol, in the dark at –20 °C, and freshly diluted in PBS to 1 μ M before use in the fluorescence experiments. Competitors of fluorescent-fatty-acid binding were prepared as stock solutions in ethanol at approx. 10 mM and diluted in PBS or ethanol for use. Free retinol is poorly soluble and unstable in aqueous solution, so it was dissolved and diluted in ethanol immediately before use, and binding to proteins was tested by addition of typically 5 μ l of this directly to a cuvette containing protein in PBS.

To estimate the dissociation constant of ABA-1-fatty acid binding, successive 5 μ l or 10 μ l portions of protein at a monomer concentration of 60 μ M were added to 2 ml of DAUDA at 1 μ M and the fluorescence intensity measured at 483 nm ($\lambda_{\text{emission}}$), with $\lambda_{\text{excitation}}$ at 345 nm. The concentration of the ethanol stock solution of DAUDA was checked by measuring the absorbance of a 1:10 dilution in methanol at 335 nm, using a molar absorption coefficient, ϵ_{335} , of 4400 M⁻¹·cm⁻¹ [26]. The concentration of retinol was estimated from the absorbance of a solution of retinol in ethanol at 325 nm, using an ϵ_{325} of 52480 M⁻¹·cm⁻¹. For the retinol titration experiments, 72 μ M retinol in ethanol was added in 5 μ l aliquots to 2 ml of a 3.27 or 3.86 μ M solution of ABA-1 and mixed immediately. Correction for the fluorescence of free retinol involved addition to a cuvette containing only PBS, as previously described [27]. The concentration of the recombinant proteins was estimated from the absorbance at 280 nm, using a molar absorption coefficient, ϵ_{280} , of 10810 M⁻¹·cm⁻¹ and 16939 M⁻¹·cm⁻¹ for ABA-1A and ABA-1B respectively, based on the amino acid composition of the recombinant protein [28]. Fluorescence data were corrected for dilution where necessary and were fitted by standard non-linear-regression techniques (using Microcal ORIGIN software) to a single non-competitive binding model to give estimates of the dissociation constant (K_d) and maximal fluorescence intensity (F_{max}). Similar non-linear-regression methods were used to analyse results of competition experiments in which oleic acid was progressively added to DAUDA/ABA-1 mixtures. Typically, a stock solution of oleic acid in ethanol was freshly diluted to 1 mM in PBS and increasing concentrations of oleic acid were added to a mixture containing 1.7 μ M DAUDA and 1.94 mM ABA-1.

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 estimated from the absorbance at 280 nm and were 0.5 mg/ml with a 0.02-cm-path-length cell for the observations on native protein, and 0.2 mg/ml for the denaturation experiments using a 0.05-cm-path-length cell. Molar-ellipticity values were calculated using values of 113 Da and 117 Da for the mean residue weight calculated from the amino acid sequence of recombinant ABA-1A1 and ABA-1B1 protein respectively. Analysis of the secondary structure of the protein was performed using the CONTIN procedure [29] over the range 240–190 nm. Guanidinium chloride (GdmCl) (Ultrapure grade) was purchased from Gibco-BRL, Paisley, Refreewshire, Scotland, U.K., and the concentrations of solutions checked by refractive-index measurements [30].

RESULTS AND DISCUSSION

cDNA cloning and sequence analysis

The cDNA expression library was screened as described in the Materials and methods section, and DNA sequencing of positive clones revealed two (pJMAS1, 1.8 kb and pJMAS2, 2.1 kb) with distinct overlapping sequences. Neither of these are full-length, both having arisen by internal priming of the mRNA in first-strand cDNA synthesis, presumably due to the high frequency of adenine residues within the coding region. This was, however, fortuitous given that the full-length mRNA encoding the ABA-1 polyprotein, and that from other species, is variously estimated or known to contain between 10 and 20 units [11–13,15], which would amount to a transcript of such a size that complete copies are unlikely to be represented in the cDNA library; similar difficulties have prevented the complete sequencing of any filaggrin cDNA to date [6–9]. Translation of the open reading

Figure 1 Diversity within the polypeptide units encoded in the polyprotein array of the nematode *Ascaris suum*

(A) Structures of the three overlapping cDNAs encoding the C-terminal end of the ABA-1 polyprotein array from *Ascaris suum*. The assembled sequence encodes six polypeptide units of identical amino acid sequence designated type A1 (with 'silent' substitutions in the DNA allowing positioning of overlapping regions), three units designated A2, A3 and A4, which are unique but very similar to A1, and the divergent B1 repeat unit. The actual assembled nucleotide sequence has been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number AF051702. (B) Alignment of all the repeat unit types illustrating the abrupt difference between the A and the B-type units. Trp¹⁵, Gln²⁰, Cys⁶⁴ and Cys¹²⁰ are conserved throughout the NPA proteins [15,16]. The consensus proteinase cleavage sites are underscored and have the ^{Lys}/Arg-Xaa-^{Lys}/Arg motif identified for processing endoproteases; the arginine tetrad is removed in parasite-processed protein [22–25], and the N-terminal amino acid sequence obtained from fully processed parasite-derived ABA-1 allergen protein begins immediately after the cleavage sequence and is in **bold** [21]. Gaps (—) indicate amino acid positions which are identical with those of the A1 sequence. (C) Evidence that both type A1 and B1 units may have arisen from duplication of a single ancestral peptide of approx. 60 amino acids in length. The N- and C-terminal halves of each unit have been aligned using the MultAlin program set for the Dayhoff matrix [42]. Residues which are identical across the alignment are in UPPER CASE, those which appear three out of four times are in lower case. ! indicates where aligned residues are either I or V, \$ if either L or M, and # if any of N, D, Q or E (single-letter amino acid code).

Figure 2 Fluorescence spectral analysis of ligand binding and intrinsic tryptophan fluorescence of ABA-1 type A1 or B1 protein

(A, B) Emission spectra of DAUDA and DACA and changes upon binding to ABA-1 type A1 or B1. A 10 μ l portion of a 60.6 μ M solution of A1 or B1 protein was added to 2 ml of 1 μ M DAUDA (A) or DACA (B) in a fluorescence cuvette and the emission spectrum recorded. The wavelengths of maximum fluorescence emission were 542 nm for DAUDA in buffer, 480 nm for DAUDA + ABA-1B1, and 473 nm for DAUDA + ABA-1A1. The wavelengths of maximum fluorescence emission were 542 nm for DACA in buffer, 491 nm for DACA + ABA-1B1, and 471 nm for DACA + ABA-1A1. The spectra for DAUDA or DACA in buffer were subtracted from the appropriate spectra obtained upon addition of protein. $\lambda_{\text{excitation}}$ was 345 nm. (C) Binding of ANS to ABA-1 type A1 or B1 protein. Fluorescence emission spectrum of 20 μ M ANS in buffer alone or upon addition of 0.3 μ M A1 or B1 protein, showing enhanced fluorescence upon interaction with B1 over that produced with protein A1. In both cases, the enhanced emission of ANS was reversed by addition of oleic acid (result not shown). The wavelengths of maximum fluorescence emission were 517 nm for ANS in buffer, 477 nm for ANS + ABA-1B1, and 475 nm for ANS + ABA-1A1. (D) Fluorescence emission spectra of Trp¹⁵ in types A1 and B1 protein. Fluorescence spectra were recorded with approx. 8 μ M of proteins A1 and B1. The wavelength at the peak of emission and at the shoulder were 306 nm and 314 nm respectively in both cases. A1 has one tryptophan and four tyrosine residues, and B1 has one tryptophan and five tyrosine residues. $\lambda_{\text{excitation}}$ was 290 nm.

Table 1 Binding of lipids to ABA-1A1 and ABA-1B1 recombinant proteins

Ligand	ABA-1A1		ABA-1B1		
	λ_{max}^* (nm)	K_D (μ M) [†]	λ_{max} (nm)	K_D (μ M)	$F_A/F_B^\ddagger$
DAUDA	473	1.8	480	4.0	0.60
DACA	471	8.0	491	9.0	0.63
ANS	475	—	477	—	—
cPnA	411	1.9	411	1.3	0.62
Oleic acid	—	0.6	—	0.3	—
Retinol	465	0.1	465	0.3	0.28

* λ_{max} is the wavelength of peak fluorescence emission of ligand when bound to protein.
† All dissociation-constant (K_D) estimations were obtained by titration with the fluorescent ligands listed, or, in the case of oleic acid, the value is for apparent K_D obtained in competition experiments with DAUDA.
‡ F_A/F_B is the ratio between the maximum fluorescence intensity (F_{max}) at the wavelength of peak emission for ABA-1A1 and ABA-1B1; F_{max} was estimated by extrapolation to infinity of the titration data used for calculation of K_D . Errors in experimental determination of K_D are typically of the order of $\pm 10\%$ [16].

frame of the inserts showed that they comprise multiple head-to-tail units encoding polypeptides similar or identical with the amino acid sequence of the ABA-1 allergen of *Ascaris* which has been obtained from direct N-terminal sequencing of parasite-derived protein (see Figure 1) [21]. Figure 1(A) shows the organization of the cDNAs and how they align with a previously described cDNA (pHS10; [14]) to give a contiguous sequence containing ten NPA units, six of which encode identical polypeptide units (type A1), three more with only minor differences from these (A2, A3 and A4) and one which is substantially different from the others (B1). The N-terminal amino acid sequence of the parasite-derived ABA-1 allergen matches exactly that of the A1 units, presumably because these represent the most abundant form [21,22]. The amino acid sequence of B1 is only 49% identical with that of A1, representing an abrupt change in the pattern of repeat units. Southern-blotting analysis with genomic DNA showed that types A and B are in the same single copy gene (results not shown). The parasite therefore produces multiple forms of the allergen, which could in turn have consequences for the hypersensitivity responses engendered by the infection, particularly in light of the genetic control of allergic reactions to ABA-1 [21,31].

The differences between type A and B units are scattered throughout the sequences (Figure 1B), although Trp¹⁵, Gln²⁰, Cys⁶⁴ and Cys¹²⁰ are conserved across these and all other NPA units so far described [15,16]. Genetic or chemical modification of any of these residues variously affects the protein's structure or ligand binding (L. McDermott, J. Moore, N. C. Price, S. M. Kelly, A. Cooper and M. W. Kennedy, unpublished work). It has previously been noted that the type A unit shows evidence of having itself arisen from a duplication [14], and the same appears also to be true of unit B1, in which there are similarities between the N- and C-terminal halves of B1 and A1 (Figure 1C).

Recombinant proteins representing type A1 and B1 units were produced as described in the Materials and methods section. We have already tested A1 protein in a biological assay for allergenic activity and found that the recombinant protein is equivalent to parasite-derived material in allergenic activity *in vivo*, and that A1 and B1 are antigenically distinct. What follows is a biochemical and structural comparison of the two.

Ligand-binding functions

Both A1 and B1 bound to DAUDA, producing a dramatic shift in the wavelength of maximum emission of the dansyl fluorescent reporter group, from 542 nm for DAUDA in buffer to 473 nm and 480 nm for DAUDA interacting with A1 and B1 respectively (Figure 2A). This blue shift was similar to that observed for the parasite-derived protein (475 nm; [16]), and is extreme compared with that observed with vertebrate fatty-acid-binding proteins (FABPs) such as the β -rich FABPs and serum albumin (≈ 500 nm; [16,32,33]). However, B1 produced a substantially greater enhancement of DAUDA fluorescence on a molar basis than did its counterpart.

The disparity in the two proteins in terms of degree of blue shift and fluorescence enhancement they produce was more marked upon binding of a dansylated fatty acid in which the fluorophore is attached at the α -carbon atom (DACA), rather than at the methyl terminus (as in DAUDA). With DACA, the proteins were clearly distinguishable in that the blue shift was substantially greater for A1 (to 471 nm) than for B1 (491 nm), whereas the reverse was true for the enhancement of intensity (Figure 2B). The degree of blue shift in the emission of the dansyl fluorophore is considered to be a measure of the polarity of a protein-binding site [34], so A1 and B1 presumably differ in

Figure 3 Titration curve for the binding of retinol to ABA-1 types A1 and B1 protein

The Figure shows the change in relative fluorescence intensity at 470 nm of 4 μ M A1 or B1 monomer on addition of increasing concentrations of retinol. $\lambda_{\text{excitation}}$ was 350 nm. The data were corrected for retinol fluorescence when added to buffer alone, as described previously [27]. The curves drawn for the corrected data are the theoretical binding curves for complex formation with dissociation constants (K_D) 8.3×10^{-8} M for A1 and 3.1×10^{-7} M for B1.

terms of amino acid side chains close to the ligand or accessibility to solvent water within the binding site. Titration experiments showed that the two proteins have similar dissociation constants and also emphasized the greater fluorescence emission produced in DAUDA on binding to B1. Similar experiments were carried out with a naturally fluorescence fatty acid (cPnA) whose emission was also greater when bound to B1 than when bound to A1, but the dissociation constants were similar (Table 1).

Binding of retinol was also examined by fluorescence, showing that B1 produced a substantially greater fluorescence enhancement of retinol than did A1, although the latter proved to have the greater affinity for retinol in titration experiments (Figure 3).

The final fluorescent probe used in the comparison was ANS, which is considered to increase its fluorescence upon interaction with exposed hydrophobic regions on proteins [35], although we also find it to bind to proteins with hydrophobic lipid-binding pockets such as serum albumin and β -lactoglobulin (results not shown). The fluorescence emission of this probe increased dramatically upon addition of either A1 or B1, but the change in fluorescence intensity with B1 was approximately double that with A1 (Figure 2C). In both cases, however, the fluorescence increase was reversed upon addition of oleic acid to the mixture, indicating that ANS binds in the ligand-binding site. The wavelength of maximum fluorescence emission of ANS in the two proteins was similar (477 nm and 475 nm) and not significantly different from that when bound to BSA (473 nm) or β -lactoglobulin (476 nm) (results not shown).

The different ligand binding propensities of the two proteins are summarized in Table 1.

Structural analysis

An unusual characteristic of ABA-1 and its homologues from other species is the dramatic blue shift in the fluorescence emission of the single, conserved tryptophan (Trp¹⁵), the extent of which is matched by few proteins [36–38]. When the emission spectrum of Trp¹⁵ in the two recombinant proteins was analysed,

Figure 4 CD analysis of ABA-1 type A1 and B1 proteins

The Figure shows the CD spectrum recorded at 20 °C on a 0.5 mg/ml (approx. 30 μ M) solution of A1 or B1 in PBS; the pathlength was 0.02 cm. Using the CONTIN procedure [29], the predicted estimates of α -helix are $47 \pm 2.2\%$ and $48 \pm 1.8\%$ for ABA-1A1 and ABA-1B1 respectively. The predicted contents of β -structure are $34 \pm 3.8\%$ and $40 \pm 3.2\%$ for ABA-1A1 and ABA-1B1 respectively, although it should be noted that estimates for β -structure can be unreliable for α -helix-rich proteins. The inset shows the change in ellipticity at 225 nm of a 0.2 mg/ml (approx. 12 μ M) sample of A1 or B1 with increasing concentration of GdmCl; the pathlength was 0.05 cm. The curves drawn are merely for guidance.

the blue shifts were found to be identical (Figure 2D), indicating that the tryptophan residues of the two proteins are in similarly buried and water-inaccessible sites within the structure of the proteins, despite the disparity in their primary structures. The spectra for the two proteins have the same emission peak at 306 nm and a shoulder at 314 nm; the shoulder is absent from the emission spectrum of a deletion mutant in which Trp¹⁵ is substituted (results not shown), so the 306 nm peak is probably due to tyrosine residues, of which there are four in A1 and five in B1.

CD analysis gave very similar spectra for A1 and B1, with a similarly high predicted α -helix content (Figure 4). Neither protein showed any significant change in CD spectrum over a 50-fold concentration range (75–1.5 μ M; results not shown). When the stabilities of the folded proteins were compared in GdmCl titration experiments, the major part of the unfolding of the two proteins occurred between 1 and 3 M GdmCl (as in parasite-derived ABA-1 [16]). There was, however, a reproducible difference in the shape of the denaturation curves which was apparent in three separate experiments using different batches of the proteins, although the mid-points for the two proteins occurred at approximately the same concentration of denaturant (≈ 2.5 M; Figure 4, inset). This suggests that, although the stabilities of the folded proteins are similar, the co-operativity or mechanisms of unfolding in the presence of GdmCl may differ between the two forms. Preliminary analysis assuming a two-state model [39] of the unfolding curves gives Gibbs-free-energy-change (ΔG^0) values for the unfolding of the native proteins of 27 kJ/mol [m (slope of the plot of free energy against

GdmCl concentration) = $-11 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{M}^{-1}$] for ABA-1A1 and $16 \text{ kJ} \cdot \text{mol}^{-1}$ ($m = -6.5 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{M}^{-1}$) for ABA-1B1. In the presence of oleic acid there was essentially no change in the shape of the denaturation curve for either protein (results not shown).

These studies indicate that sequence-divergent units from the same nematode polypeptide array are functional lipid-binding proteins but possess binding sites with distinctive biophysical characteristics. Two possibly related questions are prompted, namely why are NPAs produced as polypeptides and why has the dramatic sequence diversity evolved when the end products have similar functional properties, at least as regards ligand binding? Perhaps the answer to both questions arises from a requirement for large amounts of functionally active protein to be produced rapidly and efficiently at high local concentrations at the site of synthesis or action. The polypeptide mechanism would achieve this, but the limited solubility of a structurally homogeneous protein [40] would impose limits on the concentrations safely achievable (ABA-1 concentration in *Ascaris* body fluid is approx. 21.3 mg/ml (M. W. Kennedy, unpublished work), and there are fewer species of protein present than in, say, human plasma [21]). The combination of polypeptide and diversity might overcome this problem by raising the threshold concentration of protein which could be produced before significant aggregation occurs. In relation to the solubility argument, it may be relevant to note that different components of the *Ascaris* and other nematode polypeptide arrays show significant differences in predicted pI (ranging from 5.0 to 6.4), but will all be negatively charged at neutral pH (A. Cooper and M. W. Kennedy, unpublished work).

Alternatively, the explanation might be genetic. Following the amplification of DNA encoding the ancestral NPA unit, an array comprising tandemly repeated sequences might be at risk of disruption through homologous recombination during meiosis. There would therefore be a selective advantage to diversifying the DNA, within functional constraints imposed on the encoded proteins. Instability of such a gene might, for instance, be the reason for the apparent difference between individual humans in the sequences and number of repeat units in the flaggrins, and apparent size heterogeneity of the products [9,41]. On the other hand, the amplification event(s) might merely be sufficiently ancient that the diversified NPA arrays we observe today result from the gradual divergence in the sequences of the different units in an array, but with conservation of function.

The different responses shown by the various fluorescent probes when bound to ABA-1A1 and ABA-1B1 unexpectedly showed that the sequence diversity between these two influences functionally important parts of the folded molecule, producing changes in the molecular environment of the active site. It is possible that this active-site variability merely reflects the multiplicity of ways in which these relatively non-specific fatty-acid-/retinoid-binding sites may be constructed, with some mutational latitude in the absence of other evolutionary pressures. Alternatively (but not incompatible with this), it might reflect some other additional function(s) that we have not yet identified. For example, whereas these molecules might be relatively indiscriminate in what they bind, the sites of delivery (what they deliver and where) may be more specific. The diversity in the active site and other features of the proteins might therefore reflect different mechanisms for transfer or release of their payload at different receptors or membrane locations within nematodes.

We are indebted to Fiona McMonagle for her expert help in the preparation of the recombinant proteins. This work was supported by grants from the Wellcome Trust (grant number 031559 to M.W.K. and grant number 044156 to M.W.K. and A.C.). The CD facility at the University of Stirling is supported by the Biotechnology and Biological Sciences Research Council (U.K.).

REFERENCES

- Fisher, J. M. and Scheller, R. H. (1988) *J. Biol. Chem.* **263**, 16515–16518
- Neddermann, P., Tomei, L., Steinkuhler, C., Gallinari, P., Tramontano, A. and DeFrancesco, R. (1997) *J. Biol. Chem.* **378**, 469–476
- Vogt, V. M. (1996) *Curr. Top. Microbiol. Immunol.* **214**, 95–131
- Maia, I. G., Seron, K., Haenni, A. L. and Bernardi, F. (1996) *Plant Mol. Biol.* **32**, 367–391
- Wallace, C. J. A. (1993) *Protein Sci.* **2**, 697–705
- Presland, R. B., Haydock, P. V., Fleckman, P., Nirunskisiri, W. and Dale, B. A. (1992) *J. Biol. Chem.* **267**, 23772–23781
- Haydock, P. V. and Dale, B. A. (1990) *DNA Cell Biol.* **9**, 251–261
- Rothnagel, J. A. and Steinert, P. M. (1990) *J. Biol. Chem.* **265**, 1862–1865
- Gan, S.-Q., McBride, O. W., Idler, W. W., Markova, N. and Steinert, P. M. (1990) *Biochemistry* **29**, 9432–9440
- 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
- Culpepper, J., Grieve, R. B., Friedman, L., Mika-Grieve, M., Frank, G. R. and Dale, B. (1992) *Mol. Biochem. Parasitol.* **54**, 51–62
- Tweedie, S., Paxton, W. A., Ingram, L., Maizels, R. M., McReynolds, L. A. and Selkirk, M. E. (1993) *Exp. Parasitol.* **76**, 156–164
- Poole, C. B., Granda, A. G., Maina, C. V., Jenkins, R. E., Selkirk, M. E. and McReynolds, L. A. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5986–5990
- Spence, H. J., Moore, J., Brass, A. and Kennedy, M. W. (1993) *Mol. Biochem. Parasitol.* **7**, 339–344
- Britton, C., Moore, C., Gilleard, J. S. and Kennedy, M. W. (1995) *Mol. Biochem. Parasitol.* **72**, 77–88
- Kennedy, M. W., Brass, A., McCruden, A. B., Price, N. C., Kelly, S. M. and Cooper, A. (1995) *Biochemistry* **34**, 6700–6710
- Kennedy, M. W., Allen, J. E., Wright, A. S., McCruden, A. B. and Cooper, A. (1995) *Mol. Biochem. Parasitol.* **71**, 41–50
- Kennedy, M. W., Britton, C., Price, N. C., Kelly, S. M. and Cooper, A. (1995) *J. Biol. Chem.* **270**, 19277–19281
- Christie, J. F., Fraser, E. M. and Kennedy, M. W. (1992) *Parasite Immunol.* **14**, 59–73
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Christie, J. F., Dunbar, B., Davidson, I. and Kennedy, M. W. (1990) *Immunology* **69**, 596–602
- Christie, J. F., Dunbar, B. and Kennedy, M. W. (1993) *Clin. Exp. Immunol.* **92**, 125–132
- Barr, P. J. (1991) *Cell* **66**, 1–3
- Steiner, D. F. (1998) *Curr. Opin. Chem. Biol.* **2**, 31–39
- Yasothornsrikul, S., Toneff, T., Hwang, S. R. and Hook, V. Y. H. (1998) *J. Neurochem.* **70**, 153–163
- Haughland, R. P. (1992) *Handbook of Fluorescent Probes and Research Chemicals*, 5th edn., Molecular Probes Inc., Eugene, OR
- Cogan, U., Kopelman, M., Mokady, S. and Shinitzky, M. (1976) *Eur. J. Biochem.* **65**, 71–78
- Gill, S. C. and von Hippel, P. H. (1989) *Anal. Biochem.* **182**, 319–326
- Provencher, S. W. and Glöckner, J. (1981) *Biochemistry* **20**, 33–37
- Nozaki, Y. (1972) *Methods Enzymol.* **26**, 43–50
- Tomlinson, L. A., Christie, J. F., Fraser, E. M., McLaughlin, D., McIntosh, A. E. and Kennedy, M. W. (1989) *J. Immunol.* **143**, 2349–2356
- Wilkinson, T. C. I. and Wilton, D. C. (1986) *Biochem. J.* **238**, 419–424
- Wilton, D. C. (1990) *Biochem. J.* **270**, 163–166
- Macgregor, R. B. and Weber, G. (1986) *Nature (London)* **319**, 70–73
- Stryer, L. (1965) *J. Mol. Biol.* **13**, 482–495
- Eftink, M. R. and Ghiron, C. A. (1976) *Biochemistry* **15**, 672–680
- Eftink, M. R. and Ghiron, C. A. (1984) *Biochemistry* **23**, 3891–3899
- Johnson, C. M. and Price, N. C. (1987) *Biochem. J.* **245**, 525–530
- Pace, C. N. (1986) *Methods Enzymol.* **131**, 266–280
- Arakawa, T. and Timasheff, S. N. (1985) *Methods Enzymol.* **114**, 49–77
- Thulin, C. D., Taylor, J. A. and Walsh, K. A. (1996) *Protein Sci.* **5**, 1157–1164
- Corpet, F. (1988) *Nucleic Acids Res.* **16**, 10881–10890

Received 15 October 1998/16 February 1999; accepted 8 March 1999