

Minireview

The polyprotein allergens of nematodes (NPAs) – Structure at last, but still mysterious

Malcolm W. Kennedy

Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, Graham Kerr Building, University of Glasgow, Glasgow G12 8QQ, Scotland, UK

ARTICLE INFO

Article history:

Received 9 March 2011

Received in revised form 14 June 2011

Accepted 15 June 2011

Available online 29 June 2011

Keywords:

Nematodes

Nematode polyprotein allergens (NPAs)

Allergens

Lipid-binding

Structure

ABSTRACT

We are engaged in structural and functional studies of several types of lipid binding protein that are only found in nematodes. Amongst these are the nematode polyprotein allergens (NPAs) and we now report the solution structure of ABA-1A (As-NPA-A1), the most repeated unit within the NPA array of *Ascaris suum*, which is almost identical in amino acid sequence to that of *Ascaris lumbricoides*. The protein forms a slightly flattened, compact, globular fold consisting of a long central helix that participates in two flanking helical bundles. Two pockets lined with apolar amino acid sidechains are apparent, one in the carboxy-terminal region of the protein, and another smaller one in the amino-terminal region. The former appears to be the main site of fatty acid binding, and the latter may have different, though possibly overlapping, ligand binding propensities. The structure of the binding sites indicates that lipid ligands are anchored within them with their hydrophobic tails oriented towards the core of the protein and their polar headgroups bound to charged sidechains at the mouth of the pockets. The three-dimensional architectures of the amino- and carboxy-terminal halves of ABA-1A are closely similar, thereby strengthening the long-suspected idea that the repeated units of NPAs themselves originate from an ancient duplication event.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Nematodes, both parasitic and free-living, produce several types of small lipid binding protein that are either unique to the Nematoda, or are widespread in metazoans but structurally modified only in nematodes. Examples of the former include the polyprotein allergens (NPAs) that are the subject of this article, and an unrelated group of helix-rich proteins with similar binding properties to the NPAs, the fatty acid and retinol-binding (FAR) proteins (Basavaraju et al., 2003; Garofalo et al., 2003; Kennedy, 2000; Kennedy et al., 1997). The nematode FABPs (nemFABPs) are clearly members of the cytoplasmic fatty acid and retinoid-binding proteins (FABPs) found across the animal phyla, including the vertebrates, arachnids, crustaceans and insects, but in nematodes have structural features so far found in no other taxon (Mei et al., 1997). The nemFABPs are also unusual in that they are secreted from the synthesizing cell, and are abundant in the perivitelline fluid of nematode eggs, so may be involved in some special function in the maintenance of the shell's lipid layers (Mei et al., 1997; Michalski et al., 2002). This is turn could be implications for the development of drugs against those nematodiasis

in which larval stages are the causative agent of disease, such as in onchocerciasis.

We have chosen to concentrate specifically on these three unusual types of nematode proteins for the very reason that they represent nematode-specific adaptations, and we are currently close to solving the structure of one each of them. One member of the highly divergent FAR family has been solved by another group (Jordanova et al., 2009) and we are pursuing those members of the family that tend to be secreted by parasitic nematodes or are represented prominently in transcriptome surveys. The NPAs are of pertinence not only to parasitology, but also to the fundamental reasons why some proteins are produced in such an unusual way (see below), or why it is that this method of protein synthesis is so rare.

The NPAs originally received a great deal of attention in the 1990s because cDNAs encoding them were commonly retrieved from cDNA expression libraries screened with serum antibody from humans or animals infected with nematodes (Kennedy, 2000). They have consequently been described as being immunodominant, a description that is enhanced in light of the fact that they do not feature notably in transcriptome surveys. Added to that was the frequent finding that the NPAs are also the targets of IgE antibody responses, exemplified by the ABA-1 allergen of *Ascaris* spp. and the Gp15/400 of *Brugia* spp. (Paxton et al., 1993; Tomlinson et al., 1989). This had dual implications in that,

E-mail addresses: malcolm.kennedy@glasgow.ac.uk, malken@chem.gla.ac.uk

although nematode infections do not commonly elicit allergic or anaphylactic responses, some conditions such as Loeffler's Syndrome in ascariasis (a pulmonary eosinophilia accompanied by high IgE levels; (Takenaka et al., 1975)), can be lethal, and epidemiological studies have indicated that IgE antibody responses to nematodes can be associated with the development of natural resistance to the infections (for *Ascaris lumbricoides*, see references (McSharry et al., 1999; Turner et al., 2005)). Interest in the NPAs grew with the finding that the immune repertoire to them (whether an infected host responded to them or not) is under strong genetic control by the major histocompatibility complex, and then that they comprised a rare type of protein in nature (Kennedy, 2000).

Polypeptides exist in nature in two distinct forms. First, precursor polypeptides that are cleaved posttranslationally into functional proteins or peptides of different sizes and functions. Second, large polypeptides that are posttranslationally cleaved into multiple copies of proteins of similar size and assumed functions. It is to this latter class that NPAs belong, a notable other example of this rare form of protein synthesis and processing to produce multiple copies of very similar proteins is the flaggrins of mammalian skin (Sandilands et al., 2009). NPAs have been described in free-living nematodes, and in both animal and plant-parasitic (e.g. *Globo-dera pallida*) species, including the agents of filariases and ascariasis in humans and animals, and potentially lethal nematodiasoses of domestic cattle and sheep (e.g. *Ostertagia ostertagi*, *Haemonchus contortus*, *Dictyocaulus viviparus*) (Jones et al., 2009; Kennedy, 2000; Kuang et al., 2009). NPAs appear to be found in the greatest concentrations in the body fluid and secretions of large ascaridids such as *A. lumbricoides*, *Ascaris suum* and *Toxocara canis* (Kennedy et al., 1989). The biochemical activity of the NPA of *A. suum* was first described as a binding protein for small lipids such as fatty acids and retinoids (Kennedy et al., 1995), and the structure of a single unit of the polyprotein array has now been solved (Meenan et al., 2011).

2. Structure

In *Ascaris spp.* and *Brugia spp.*, most of the tandemly repeated units within the NPA array tend to be similar, and sometimes identical, in amino acid sequence (this seems to be typical of the Ascaridida and Spirurida), whereas in other groups (e.g. Strongylida, Rhabditida) they can be highly diverse (Kennedy, 2000). cDNA encoding a unit that is repeated identically several times in the *A. suum* array was used to produce the protein in bacteria. This is a type A unit of the ABA-1 allergen polyprotein array of *Ascaris* (ABA-1A; systematic name, As-NPA-1A; reference (Moore et al., 1999)), biosynthetically labeled with ^{15}N and ^{13}C so that its structure could be solved by protein nuclear magnetic resonance (NMR) (Meenan et al., 2011). This revealed a helix-rich protein that exhibits a novel three-dimensional architecture (Fig. 1). The protein forms a compact, globular fold consisting of a long central helix that participates in two flanking helical bundles. Two pockets lined with apolar sidechains are apparent, as revealed by probing the molecule *in silico* with virtual spheres the size of a water molecule or a methyl group (Fig. 1). But, when chemical shifts in NMR were compared between samples of the bacterial recombinant protein before and after an excess of fatty acid (oleic) was added, only the side chains of amino acids in the carboxy (C)-terminal region surrounding that pocket were significantly affected (Meenan et al., 2011). This could be due either to direct interaction between those side chains and the ligand, or a change in their position or environment within the structure. So, a binding site for fatty acids at least was localized to this binding pocket, but what of the other cavity, in the amino (N)-terminal region? When the protein was

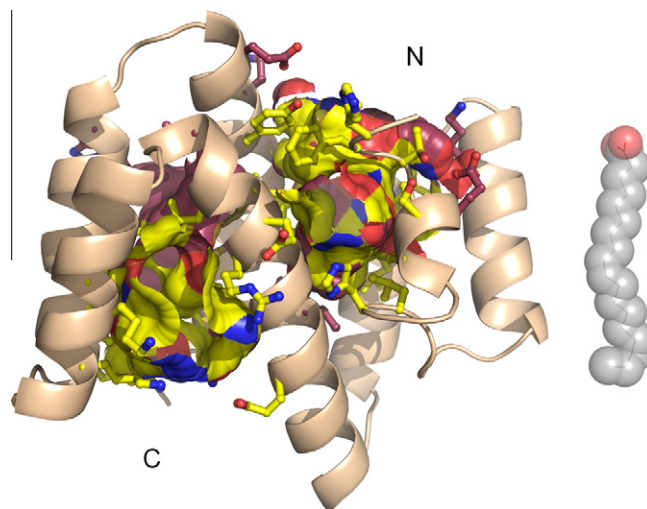


Fig. 1. Structure of ABA-1A and the position and environment of its binding cavities. ABA-1A and an oleic acid molecule to approximately the same scale. The carboxy-terminal cavity (C) and N-terminal (N) cavities were explored with a 1.4 Å ($\approx$ water) or a 1.9 Å ($\approx$ methyl) sphere, the wall boundaries of which are indicated in magenta and yellow, respectively. Amino acid sidechains surrounding the cavities are shown, all of which are apolar with the exception of positive polar sidechains at the mouth of the larger carboxy-terminal cavity (C). The bound ligand therefore probably orients with its hydrophobic region directed internally to the protein, anchored by its charged head group at the mouth of the cavity. Such an orientation is also consistent with our spectroscopic observations using fluorophore-tagged fatty acids (Kennedy et al., 1995), and is similar to that in human serum albumin (Curry et al., 1998), but opposite to that in our own cellular fatty acid binding proteins (Hellberg et al., 2010). For more detail on the whole structure and the cavities, including a restriction in one of them, see reference (Meenan et al., 2011). (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

completely stripped of resident ligand(s) (by unfolding and reverse phase chromatography) that may have been loaded into the protein in the synthesizing bacteria, then saturated with oleic acid, NMR chemical shifts were observed in both the N- and the C-terminal regions (Meenan et al., 2011). It therefore seems that this NPA indeed has two binding pockets, a C-terminal site being readily accessible to fatty acids (and some fluorescent analogues thereof), and a N-terminal site that appears to be a high affinity binding site for hydrophobic ligand(s), of as yet unknown nature, produced by bacteria. So, the NPAs may be general carriers or scavengers of small hydrophobic compounds, and their two binding pockets have different, though perhaps overlapping, binding specificities.

3. Origin

The tandem repetitive nature of the NPA array indicates that it arose from a series of duplication events amounting to arrays of about of 10 or 11 units in all. But, it was noticed at an early stage from amino acid sequence analysis that each of the NPA units themselves may have arisen from a duplication. Further sequence comparisons of the units in the NPA array of *Dirofilaria immitis* strengthened this idea (Poole et al., 1996). The new structure of a single ABA-1A NPA unit shows that the three-dimensional structures of the N- and carboxy-terminal subdomains are virtually superimposable, further increasing the likelihood that a duplication occurred to create the basic unit of an NPA. A similar process of duplication and divergence in function of progenitor binding sites is thought to have occurred in the evolution of the serum albumin of vertebrates (Curry et al., 1998), which is also a transporter of hydrophobic ligands. There remains much more to learn about the nature of the NPA binding sites, but it is possible that a fuller understanding could lead to improved design of anthelmintic

drugs, either to interfere with the function of NPA, either inside or outside the parasites, or by using them as efficient carriers of a drug to its target within the worm.

4. Why polyproteins?

One of the reasons that the structure of an NPA unit was pursued was that it might provide a clue as to the advantage of producing any protein as a tandemly repetitive polyprotein. For polyproteins in general it could be argued that it is a highly efficient means by which to produce functionally similar or dissimilar proteins quickly and in quantity. The counter question to that is why then are more proteins not made this way, particularly those that require to be made in abundance or quickly? The disadvantages could include that the genes encoding repetitive polyproteins are susceptible to disruption by homologous recombination, or that adjacent units may interact and fold inappropriately as they emerge from the ribosome during synthesis. At the moment, the structure of ABA-1A does not reveal any clues to answer these questions, and there is no indication of any surface features to indicate that there is any direct interaction, or indeed repulsion, between two adjacent repeat units before posttranslational cleavage separates them.

5. How do NPAs tend to be allergens?

The structure also reveals no indication as to how it is that ABA-1 and other NPAs tend to be preferential targets of IgE responses (or generally immunodominant). Over the past decade or so there has been considerable effort expended on attempting to predict the structural features that may be useful in assigning a probability of allergenicity to individual proteins or protein families. So far, this has met with limited success, although there is a tendency for allergens (and sometimes IgE epitopes) to exhibit high proportions of charged amino acids (Razzera et al., 2010; Thomas et al., 2005a). This applies to ABA-1A, which is enriched in lysines and glutamic acids (Fig. 2), and this is also a tendency exhibited by other NPA

units, such as those of *Brugia spp.*, *Onchocerca spp.* and *D. viviparus* (not shown), though it remains to be seen whether these biases have any relevance at all to allergenicity. Curiously, lipid binding is a relatively common feature of protein allergens (Thomas et al., 2005b), and it has been further speculated that hydrophobic features on proteins (such as in lipid binding pockets in the native fold or as exposed upon unfolding) could act to trigger innate immune responses via Toll-like receptors (TLRs) (Seong and Matzinger, 2004; Shakib et al., 2008). With nematode lipid binding proteins such as NPAs, however, it is at this stage important to consider instead that the pathogen-associated molecular pattern (PAMP) involved here could be a hydrophobic molecule of a type specific to nematodes that is delivered in the binding pockets of the proteins and it is this ligand(s), or the protein–ligand complex, that acts to trigger a relevant TLR.

6. Prospective

Knowing the structure of ABA-1A will improve our understanding of the NPAs in general, as will the new appreciation that it possesses two discrete binding sites for hydrophobic ligands. In those parasitic species that have arrays that have units with highly divergent amino acid sequences, there remains the question of whether the different units have similar structures (which seems likely), but, more interestingly, whether they have different ligand binding specificities. For the moment, the carboxy-terminal region binding site is understood the best, though still incompletely, but there is much yet to learn about the other binding cavity, which is smaller, is perhaps more specialized, and may prove interesting with respect to drug development and, more interestingly, involvement in triggering innate immune responses to parasitic nematodes.

Acknowledgments

This work is the result of collaborations that are too many to list here, but I thank all those who have been involved, especially Nicola Meenan and Brian Smith who were the principal players in solving ABA-1's structure while working closely with colleagues then at Edinburgh University, and myself and Alan Cooper working on the biophysics of the protein. Brian Smith also kindly provided Fig. 1, of which there is a simpler version in reference (Meenan et al., 2011).

References

- Basavaraju, S., Zhan, B., Kennedy, M.W., Liu, Y.Y., Hawdon, J., Hotez, P.J., 2003. AC-FAR-1, a 20 kDa fatty acid- and retinol-binding protein secreted by adult *Ancylostoma caninum* hookworms: gene transcription pattern, ligand binding properties and structural characterisation. *Molecular and Biochemical Parasitology* 126, 63–71.
- Curry, S., Mandelkow, H., Brick, P., Franks, N., 1998. Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nature Structural & Molecular Biology* 5, 827–835.
- Garofalo, A., Rowlinson, M.C., Amambua, N.A., Hughes, J.M., Kelly, S.M., Price, N.C., Cooper, A., Watson, D.G., Kennedy, M.W., Bradley, J.E., 2003. The FAR protein family of the nematode *Caenorhabditis elegans* – Differential lipid binding properties, structural characteristics, and developmental regulation. *Journal of Biological Chemistry* 278, 8065–8074.
- Hellberg, K., Grimsrud, P.A., Kruse, A.C., Banaszak, L.J., Ohlendorf, D.H., Bernlohr, D.A., 2010. X-ray crystallographic analysis of adipocyte fatty acid binding protein (aP2) modified with 4-hydroxy-2-nonenal. *Protein Science* 19, 1480–1489.
- Jones, J.T., Kumar, A., Pylypenko, L.A., Thirugnanasambandam, A., Castelli, L., Chapman, S., Cock, P.J.A., Grenier, E., Lilley, C.J., Phillips, M.S., Blok, V.C., 2009. Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology* 10, 815–828.
- Jordanova, R., Groves, M.R., Kostova, E., Woltersdorf, C., Liebau, E., Tucker, P.A., 2009. Fatty acid- and retinoid-binding proteins have distinct binding pockets for the two types of cargo. *Journal of Biological Chemistry* 284, 35818–35826.
- Kennedy, M.W., 2000. The nematode polyprotein allergens/antigens. *Parasitology Today* 16, 373–380.

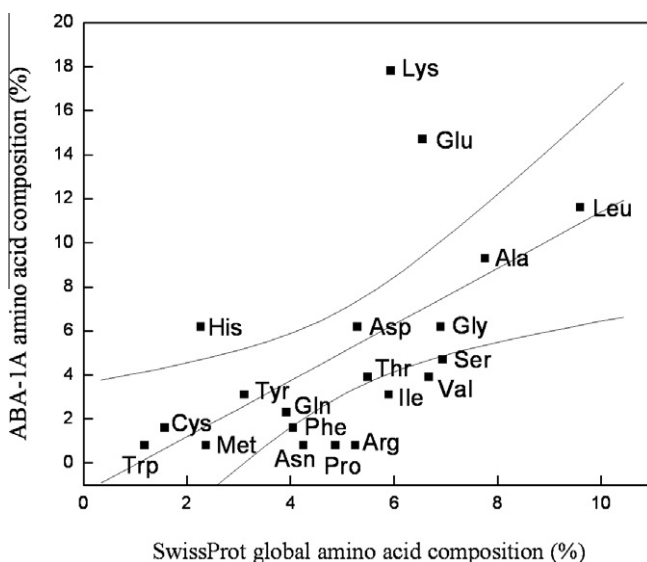


Fig. 2. Amino acid bias in ABA-1A. The amino acid composition of the ABA-1A (As-NPA-1A) unit of the NPA array of *A. suum* is plotted against the average amino acid composition of all proteins in the UniProtKB database (<http://www.expasy.ch/sprot/relnotes/relstat.html>) illustrating its relative enrichment in lysines and glutamic acids. A similar bias exists in the NPAs of *Brugia spp.*, *Onchocerca spp.* and *Dictyocaulus viviparus* (not shown). The fitted line is a simple linear fit, and the curves represent the 95% confidence limits.

- Kennedy, M.W., Brass, A., McCruden, A.B., Price, N.C., Kelly, S.M., Cooper, A., 1995. The ABA-1 allergen of the parasitic nematode *Ascaris suum* – Fatty-acid and retinoid-binding function and structural characterization. *Biochemistry* 34, 6700–6710.
- Kennedy, M.W., Garside, L.H., Goodrick, L.E., McDermott, L., Brass, A., Price, N.C., Kelly, S.M., Cooper, A., Bradley, J.E., 1997. The Ov20 protein of the parasitic nematode *Onchocerca volvulus* – A structurally novel class of small helix-rich retinol-binding proteins. *Journal of Biological Chemistry* 272, 29442–29448.
- Kennedy, M.W., Qureshi, F., Fraser, E.M., Haswell-Elkins, M.R., Elkins, D.B., Smith, H.V., 1989. Antigenic relationships between the surface-exposed, secreted and somatic materials of the nematode parasites *Ascaris-Lumbricoides*, *Ascaris-suum*, and *Toxocara-Canis*. *Clinical and Experimental Immunology* 75, 493–500.
- Kuang, L., Colgrave, M.L., Bagnall, N.H., Knox, M.R., Qian, M., Wijffels, G., 2009. The complexity of the secreted NPA and FAR lipid-binding protein families of *Haemonchus contortus* revealed by an iterative proteomics-bioinformatics approach. *Molecular and Biochemical Parasitology* 168, 84–94.
- McSharry, C., Xia, Y., Holland, C.V., Kennedy, M.W., 1999. Natural immunity to *Ascaris lumbricoides* associated with immunoglobulin E antibody to ABA-1 allergen and inflammation indicators in children. *Infection and Immunity* 67, 484–489.
- Meenan, N.A.G., Bal, G., Bromek, K., Uhrin, D., Cooper, A., Kennedy, M.W., Smith, B.O., 2011. Solution structure of a repeated unit of the ABA-1 nematode polypeptide allergen of *Ascaris* reveals a novel fold and two discrete lipid-binding sites. *Plos Neglected Tropical Diseases* 5, e1040.
- Mei, B.S., Kennedy, M.W., Beauchamp, J., Komuniecki, P.R., Komuniecki, R., 1997. Secretion of a novel, developmentally regulated fatty acid-binding protein into the perivitelline fluid of the parasitic nematode, *Ascaris suum*. *Journal of Biological Chemistry* 272, 9933–9941.
- Michalski, M.L., Monsey, J.D., Cistola, D.P., Weil, G.J., 2002. An embryo-associated fatty acid-binding protein in the filarial nematode *Brugia malayi*. *Molecular and Biochemical Parasitology* 124, 1–10.
- Moore, J., McDermott, L., Price, N.C., Kelly, S.M., Cooper, A., Kennedy, M.W., 1999. Sequence-divergent units of the ABA-1 polypeptide array of the nematode *Ascaris suum* have similar fatty-acid- and retinol- binding properties but different binding-site environments. *Biochemical Journal* 340, 337–343.
- Paxton, W.A., Yazdanbakhsh, M., Kurniawan, A., Partono, F., Maizels, R.M., Selkirk, M.E., 1993. Primary structure of and immunoglobulin-E response to the repeat subunit of Gp15/400 from human lymphatic filarial parasites. *Infection and Immunity* 61, 2827–2833.
- Poole, C.B., Hornstra, L.J., Benner, J.S., Fink, J.R., McReynolds, L.A., 1996. Carboxy-terminal sequence divergence and processing of the polypeptide antigen from *Dirofilaria immitis*. *Molecular and Biochemical Parasitology* 82, 51–65.
- Razzera, G., Gadermaier, G., de Paula, V., Almeida, M.S., Egger, M., Jahn-Schmid, B., Almeida, F.C.L., Ferreira, F., Valente, A.P., 2010. Mapping the interactions between a major pollen allergen and human IgE antibodies. *Structure* 18, 1011–1021.
- Sandilands, A., Sutherland, C., Irvine, A.D., McLean, W.H.I., 2009. Filaggrin in the frontline: role in skin barrier function and disease. *Journal of Cell Science* 122, 1285–1294.
- Seong, S.Y., Matzinger, P., 2004. Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses. *Nature Reviews Immunology* 4, 469–478.
- Shakib, F., Ghaemmaghami, A.M., Sewell, H.F., 2008. The molecular basis of allergenicity. *Trends in Immunology* 29, 633–642.
- Takenaka, T., Okuda, M., Kubo, K., Uda, H., 1975. Studies on interrelations between Eosinophilia serum immuno globulin E and tissue mast cells. *Clinical Allergy* 5, 175–180.
- Thomas, K., Bannon, G., Hefle, S., Herouet, C., Holsapple, M., Ladics, G., MacIntosh, S., Privalle, L., 2005a. In silico methods for evaluating human allergenicity to novel proteins: International Bioinformatics Workshop Meeting Report, 23–24 February 2005. *Toxicological Sciences* 88, 307–310.
- Thomas, W.R., Hales, B.J., Smith, W.A., 2005b. Structural biology of allergens. *Current Allergy and Asthma Reports* 5, 388–393.
- Tomlinson, L.A., Christie, J.F., Fraser, E.M., McLaughlin, D., McIntosh, A.E., Kennedy, M.W., 1989. MHC restriction of the antibody repertoire to secretory antigens, and a major allergen, of the nematode parasite *Ascaris*. *Journal of Immunology* 143, 2349–2356.
- Turner, J.D., Faulkner, H., Kamgno, J., Kennedy, M.W., Behnke, J., Boussinesq, M., Bradley, J.E., 2005. Allergen-specific IgE and IgG4 are markers of resistance and susceptibility in a human intestinal nematode infection. *Microbes and Infection* 7, 990–996.