Protein Seq Data Anal (1990) 3:263–266

Protein Sequences & Data Analysis
© Springer-Verlag 1990

Structural studies of a hornet venom allergen antigen 5, *Dol m* V and its sequence similarity with other proteins*

Te Piao King, Donald Moran, De Fa Wang, Loucia Kochoumian, and Brian T. Chait

The Rockefeller University, New York, NY, 10021-6399, USA

Received November 6, 1989 / Accepted December 21, 1989

Abstract. The primary structures and the disulfide-bond pairings of two chromatographic forms A and B of a venom protein from white-faced hornet (*Dolichovespula maculata*) were determined. This protein, designated as antigen 5 (*Dol m* V), has 204–205 amino acid residues and it contains 4 disulfide bonds. This protein is an allergen in man but its biologic function in hornet is not yet known. *Dol m* V is found to have partial sequence similarity with pathogenesis-related proteins from tobacco and tomato leaves and with a sperm-coating protein from rat epididymis.

Introduction

Antigen 5 is a basic protein of about 23 kDa. It is one of the three vespid venom proteins which are known allergens in man. The vespids include hornets, yellowjackets, and wasps [1]. The other two known allergens of vespid venoms are phospholipase A_1 and hyaluronidase. The biological function of antigen 5 in vespids is not known. No enzymatic activity has been detected. Its physicochemical properties and its amino acid composition are similar to those of a presynaptic neurotoxin which has been isolated from an Asian hornet *Vespa mandarina*, and which is active at the neuromuscular junction of crustaceans [2]. The primary structures of two of the three chromatographic forms of antigen 5 from white-faced hornet (*Dolichovespula maculata*) were recently established by cDNA and protein sequencings [3]. Their sequences are shown in Fig. 1, and these two forms are designated as *Dol m* VA and VB, respectively, according to a recently proposed nomenclature for allergens [4].

We tried to investigate the biological function of *Dol m* V by searching for its sequence similarity with other proteins. *Dol m* V is found to have sequence similarity with a class of 15-kDa proteins known as pathogenesis-related proteins from tomato and tobacco leaves [5, 6] and with a 26-kDa sperm-coating glycoprotein from the rat epididymis [7]. In a 138-residue overlap of the C-terminal portion of *Dol m* V with the tobacco or tomato pathogenesis-related protein, 28–32 amino acid residues were identical (Fig. 1). Sequence alignment of *Dol m* VA and VB with rat sperm-coating protein showed 51 or 50 identical amino acid residues, respectively (Fig. 1). We also noted a limited sequence similarity of antigen 5 with scorpion neurotoxin; six to seven residues of the N-terminal region of antigen 5 were identical with a segment of a scorpion neurotoxin [1].

Pathogenesis-related proteins are produced in the leaves of many plants, such as bean, celery, cucumber, tobacco, tomato and others in response to infection or specific stresses, and they often accumulate to high levels [8]. Some of the pathogenesis-related proteins are enzymes, whereas others are of unknown biological function. The PR-1 group from tobacco and the p14 protein from tomato, which have sequence similarity with *Dol m* V, are of unknown biological function. The 26-kDa sperm-coating glycoprotein is one of several proteins which are synthesized in the rat epididymis under androgen control. The glycoproteins are believed to be important in the functional maturation of sperms, and experiments suggest the 26-kDa glycoprotein to have a role in the process of fertilization [7].

The sequence similarity of *Dol m* V with proteins from plant and mammalian sources is of interest, as it is reported that a sizable proportion of our population who have not been exposed to insect stings have low levels of venom-specific antibodies [9, 10]. This raises the possibility of antigenic cross-reactivity of these distantly related proteins.

* This work is supported in part by USPHS Grants AI-17021 and BRSG S15 CA46157. The Rockefeller University Mass Spectrometic Biotechnology Research Resource is supported by the Division of Research Resources, NIH, Grant RR-00862

Offprint requests to: T.P. King

```
Dol m VA   NNYCKIKCSR GIHTLCKFGT SMKPNCGSKI   30
Dol m VB   NNYCKIKCRK GIHTLCKFGT SMKPNCGRNV   30
Rat        DEW-----DR DLENL----S TTK-------   18

Dol m VA   VKVHGVSNDE KNEIVNRHNQ FRQKVAKGLE   60
Dol m VB   VKAYGLTNDE KNEILKRHND FRQNVAKGLE   60
Rat        -----LSVQE --EIINKHNQ LRRTVS----   37
Tobacco                               QNSQQD    6
Tomato                                QNSPQD    6

Dol m VA   TRGNPGPQPP AKNMNVLVWN DELAKIAQTW   90
Dol m VB   TRGKPGPQPP AKNMNVLVWN DELAKIAQTW   90
Rat        --------PS GSDLLRVEWD HDAYVNAQKW   59
Tobacco    YLDAHNTARA DVGVEPLTWD NGVAAYAQNY   36
Tomato     YLAVHNDARA QVGVGPMSWD ANLASRAQNY   36

Dol m VA   ANQCSFGHDQ CRN-TEKYQV GQNVAIASTT  119
Dol m VB   ANQCDFNHDD CRN-TAKYQV GQNIAISSTT  119
Rat        ANRCIYNHSP LQHRTTTLKC GENLFMANYP   89
Tobacco    VSQLAADCNL VHS-HGQY-- GENLA---QG   60
Tomato     ANSRAGDCNL IHS--GA--- GENLA---KG   58

Dol m VA   GNSYATMSKL IEMWENEVKD FNPKKGTIGD  149
Dol m VB   ATQFDRPSKL IKQWEDEVTE FNYKVGLQNS  149
Rat        AS---WSS-V IQDWYDESLD FVFGFGPKKV  115
Tobacco    SGDFMTAAKA VEMWVDEKQY YDHDSNTCAQ   90
Tomato     GGDF-TGRAA IQLWVSERPS YNYATNQCVG   87

Dol m VA   NNFSKVGHYT QMVWGKTKEI GCGSVKYIEN  179
Dol m VB   N-FRKVGHYT QMVWGKTKEI GCGSIKYIED  178
Rat        G-V-KVGHYT QVVWNSTFLV ACGVAECPDQ  143
Tobacco    GQVC--GHYT QVVWRNSVRV GCARVKCNNG  118
Tomato     GKKC--RHYT QVV-----RL GCGRARCNNG  110

Dol m VA   NWHTHYLVCN YGPAGNYMDQ PIYERK      205
Dol m VB   NWYTHYLVCN YGPGGNDFNQ PIYERK      204
Rat        P-LKYFYVCH YCPGGN-YVG RLYSPY      167
Tobacco    G-YV--VSCN YDPPGNVIGQ SPY         138
Tomato     WWF----SCN YDPVGNWIGQ RPY         129
```

Fig. 1. Amino acid sequences of antigen 5 from *Dolichovespula maculata* (*Dol m*) VA and B, rat sperm-coating glycoprotein, tobacco leaf protein PR-1b and tomato leaf protein p14. Data are taken from [3, 5, 6, 7]. *Hyphens* indicate gaps introduced for maximal alignment of the sequences. Residues which are identical for *Dol m* V and rat or leaf proteins are enclosed by *solid-lined* boxes. Complete sequences are given for *Dol m* V and leaf proteins, and only residues 5–167 are shown for the rat protein

In this report we describe our sequencing studies of *Dol m* V and the determination of its disulfide bonds. Also we have compiled, from the literature, several allergens which share sequence similarity with other known proteins.

Materials and methods

cDNA clones encoding antigen 5 of white-faced hornet (*D. maculata*) were reported previously [3]. These clones in λgt 11 vector were subcloned into Bluescript II SK− or SK+ (Stratagene, San Diego, Calif.) for generating single- or double-stranded DNA [11]. Sequence analysis of the single- or double-stranded DNA was carried out using the commercially available Sequenase kit [12].

Dol m VA, VB and VC were isolated from venom of white-faced hornet as reported previously [1]. Separation of CNBr fragments of *Dol m* V also had been reported [3]. The large CNBr fragment (about 1 mg) was digested with 25 μg of trypsin in 0.5 ml of 0.05 *M* NH_4HCO_3 at ambient temperature for 3 h and the digestion was continued for a further 3 h after adding 25 μg of chymotrypsin. The digest was lyophilized then separated by gel filtration on Sephadex G-25 followed by reverse-phase chromatography as described in the text. The purified peptides were characterized by amino acid analysis and by ^{252}Cf fission fragment ionization mass spectrometry [13].

Results

Amino acid sequences of Dol m V

We have previously isolated five clones of cDNA encoding the full length of antigen 5 from a cDNA library for white-faced hornet venom proteins in the phage vector λgt 11 [3]. Two of the clones, f5 and f10, were shown by sequencing studies to encode *Dol m* VA and VB, respectively. We have now sequenced the remaining three clones, f12, f13 and f17, in an unsuccessful attempt to establish the structure of *Dol m* VC. In the coding region f17 has an identical sequence with f5, while f12 and f13 each differs from f5 by a single base substitution. The protein encoded by f12 contains an alanine residue in place of a valine residue at position 31 of *Dol m* VA. Clone f13 encodes the same protein sequence as f5 does, as the single nucleotide change is in the third position of the codon which encodes a threonine residue at position 104 of *Dol m* VA. (These results are not shown). It is unlikely that *Dol m* VC is encoded by f12 as *Dol m* VA, VB and VC have different charge properties on ion exchange chromatography. *Dol m* Vs are probably not glycoproteins as they do not bind affinity adsorbent with concanavalin A (unpublished observation).

Disulfide bonds of Dol m V

The scheme used for determining the four disulfide bonds of *Dol m* VA is shown in Fig. 2. Following CNBr cleavage at the six methionine residues of antigen 5, two cystine-containing fragments were obtained. The small fragment was shown previously to represent residues 162–197 [3], and it contains two half-cystine residues. In the present work, mass spectral analysis of this peptide indicated its measured mass value to be 4083.0 as compared to the calculated value of 4083.4. Therefore, the results suggest that the two half-cystine residues at position 171 and 188 are linked by a disulfide bond.

The large fragment was shown to be a complex of three CNBr peptides which are linked by the disulfide

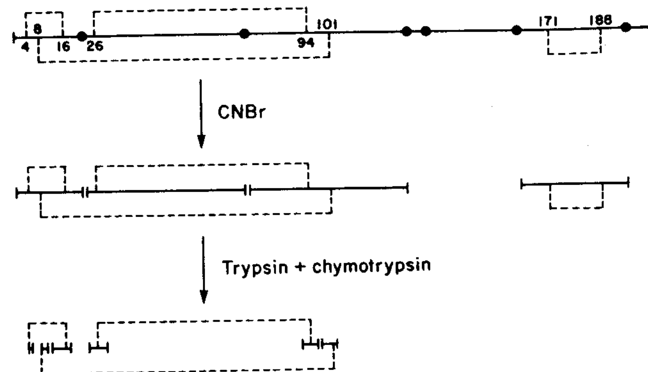

Fig. 2. Scheme for determining the disulfide bonds of *Dol m* VA. *Solid* and *dashed lines* represent peptide chain and disulfide bonds, respectively and *filled circles* represent methionine residues. The *numbers* indicate positions of half-cystine residues

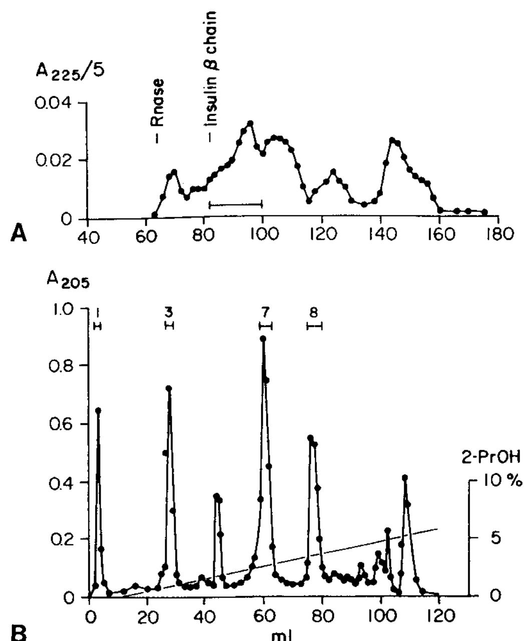

Fig. 3A, B. Isolation of cystine-containing peptides of *Dol m* VA. In **A** the combined tryptic and chymotryptic digest of the large CNBr fragment of *Dol m* VA (about 70 nmol) was separated by gel filtration on a 200 × 0.9 cm column of Sephadex G-25. The column was eluted with 1 *M* acetic acid at 12 ml/h. In **B** the indicated cut from pattern A was applied to a C18 reverse-phase column (PepRPC, HR5/5: Pharmacia). The column was eluted at 30 ml/h with a linear 2-propanol gradient of 0.08% per ml containing 0.1% trifluoroacetic acid

bonds of the remaining six half-cystine residues of antigen 5. This complex was digested with trypsin and chymotrypsin. The digest was first fractionated by gel filtration on Sephadex G-25, then the fraction containing peptides in the size range of 10–30 residues was separated by reverse-phase chromatography (Fig. 3). The data on their amino acid compositions and mass spectra (Table 1) indicate that peak 3 from Fig. 3B represents the cystine-containing peptides of residues 8–10 linked to residues 97–102, and that peak 8 represents the cystine peptide of residues 23–29 linked to residues 91–96.

Data from amino acid analysis and mass spectrometry suggest that peak 7 in Fig. 3B is a mixture containing 6% of a cystine-containing peptide of residues 4–5 joined to residues 11–17, and 94% of a decapeptide of residues 63–72 (Table 1). Amino acid analysis of peak 1 suggested it to be a mixture of a cystine-containing tetrapeptide (70%) of residues 4–5 joined to residues 16–17, and other unknown peptides (30%). Such a tetrapeptide can result from chymotryptic cleavage of the cystine-containing peptide in peak 7. The presence of tetrapeptide in peak 1 could not be confirmed by mass spectrometry; this failure may be related to the poor absorption of some small peptides to the nitrocellulose surface of sample probe which is required for mass spectrometry (B. Chait, unpublished observations). The combined yield of the cystine peptides in peaks 1 and 7 is 25%, providing support that the two half-cystines at position 4 and 16 are joined to each other.

Table 1. Analytical data of cystine-containing peptides from trypsin and chymotryps in digest of the large CNBr fragment of antigen 5

Peak[a]	Peptide	Mass values[b]		
		Calculated	Found[c]	Yield[d]
1	CK (4–5) | CK (16–17)	–	–	21%
3	CSR (8–10) | GHDQCR (97–102)	1077.2	1077.3	37%
7	GNPGPQPPAK (63–72)	962.1	962.2	58%
	CK (4–5) | GIHTLCK (11–17)	1018.3	1018.6	4%
8	KPNCGSK (23–29) | ANQCSF (91–96)	1399.4	1400.0	26%

[a] The peak numbers refer to those shown in Fig. 3B. Peak 3 and 8 showed amino acid compositions which are in good agreement (±7%) with those of the indicated peptides. The amino acid composition of peak 1 suggested it to contain about 70% of the indicated peptide and 30% of other peptides, and that of peak 7 corresponded to a mixture of the two indicated peptides

[b] The calculated and measured mass values are averages of the six most abundant isotopic components

[c] The experimental uncertainty of the mass determinations is less than 0.5 dalton

[d] Yield was determined by amino acid analysis

CNBr cleavage of *Dol m* VB followed with trypsin cleavage was performed. Two cystine-containing peptide fractions were obtained; one having the composition of residue 1–5 and 11–17 linked by a disulfide bond and the other of residue 8–9, 23–28 and 86–102 linked by two disulfide bonds. These results are not shown. The results together strongly support the scheme of disulfide bond pairings *Dol m* V given in Fig. 2.

Discussion

Dol m VA and VB have 204 and 205 amino acid residues, respectively. They differ by 23% in their sequences (Fig. 1) and they have the same disulfide bonds (Fig. 2). Their sequence differences are reflected by their antigenic activities in that mouse monoclonal antibodies differ in their binding to solid-phase *Dol m* VA and VB but they bind to *Dol m* VB and VC identically [3]. This sequence difference is also noted with sera from vespid-sensitive patients in that they had higher antibody titers for *Dol m* VB than *Dol m* VA (unpublished results). This is interesting because following a sting a person is sensitized to all three forms of *Dol m* V and the relative abundance of *Dol m* VA, VB and VC in the venom is in a ratio of 3:5:1 approximately [1].

Table 2. Some allergens and proteins of known sequence similarity and/or antigenic cross reactivity

Allergens	Common proteins	Sequence similarity/ antigenic cross reactivity[a]	References
Honey bee venom phospholipase A_2 (*Api m* I)	Bovine, porcine and human pancreatic phospholipase A_2	Yes/unknown	[14, 15]
Honey bee venom melittin (*Api m* III)	Calmodulin-binding protein	Unknown/yes	[14, 16]
	Mammalian protein with phospholipase A_2-stimulatory activity	Unknown/yes	[17]
	Human complement C9	Yes/yes	[18]
White-faced hornet venom antigen 5 form 2 (*Dol m* VB)	Tobacco and tomato leaf pathogenesis-related proteins	Yes/yes[b]	[3, 5, 6]
	Rat sperm-coating glycoprotein	Yes/unknown	[7]
Major mite allergen (*Der p* I)	Human cathepsin B and other cysteine proteases	Yes/unknown	[19, 20]
Bovine and rat serum albumins	Human serum albumin	Yes/unknown	[21, 22]

[a] Antigenic cross reactivity can be for T and B cell-specific epitopes and only examples for B cell-specific epitopes are given in this table

[b] A weak cross reactivity of hornet and tobacco proteins was detected with a tobacco protein-specific rabbit sera (unpublished results)

In Table 2 several protein allergens of known sequence are listed. Susceptible people can become sensitized after inhaling or being injected with μg amounts of these protein allergens; i.e., they produce allergen-specific IgE responses. These proteins have different sequences, properties, and functions and they do not have any obvious relationship with each other. Their one common denominator is that in each case there is sequence similarity and/or antigenic cross reactivity with another protein which is present in our body or to which we are likely to be exposed. In Table 2, 25 of the 128 residues of bee venom phospholipase A_2 are identical to those of the human enzyme, 27 of the 222 residues of *Der p* I are identical to those of human cathepsin B, and 427 of 585 residues of rat serum albumin are identical to those of human serum albumin. Bee venom melittin of 26 amino acid residues is the smallest allergen molecule known and it is reported to have sequence similarity and antigenic cross reactivity with human complement C9. Melittin is also reported to have antigenic cross reactivity with a calmodulin-binding protein and with a phospholipase-stimulatory protein.

Allergen-specific immune responses are like other immune responses in that they are subject to genetic control of the host. It is also possible that allergens have unique properties which contribute to their allergenicity in susceptible people. Continued structural studies of allergens will help to clarify this second aspect.

Acknowledgements. We are grateful to Mr. Miles Guralinick of Vespa Laboratory, Division of ALK Inc., for generous gifts of white-faced hornet venom sac extracts. We are also grateful to Dr. Winona C. Barker of Protein Identification Resource for informing us of the sequence similarities of *Dol m* V and the rat glycoprotein.

References

1. King TP, Kochoumian L, Lam T (1987) Mol Immunol 24:857–864
2. Abe T, Kawai N, Niwa A (1982) Biochemistry 21:1693–1697
3. Fang KSF, Vitale M, Fehlner P, King TP (1988) Proc Natl Acad Sci USA 85:895–899
4. Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TAE (1987) J Allergy Clin Immunol 80:639–645
5. Lucas J, Camacho Henriques A, Lottespeich F, Henschen A, Sanger HL (1985) EMBO J 4:2475–2479
6. Cornelissen BJC, Hooft van Huijsduijnen RAM, Van Loon LC, Bol JF (1986) EMBO J 5:37–40
7. Brooks DE, Means AR, Wright EJ, Singh SP, Tiver KK (1986) Eur J Biochem 161:13–18
8. Van Loon LC (1985) Plant Mol Biol 4:111–116
9. Reisman RE, Georgitis JW (1985) J Allergy Clin Immunol 76:803–807
10. Zora JA, Swanson MC, Yunginger JW (1988) J Allergy Clin Immunol 81:77–82
11. Russel M, Kidd S, Kelley MR (1985) Gene 45:333–338
12. Tabor S, Richardson CC (1987) Proc Natl Acad Sci USA 84:4767–4771
13. Chait B, Field FH (1986) Biochem Biophys Res Commun 134:420–426
14. King TP, Sobotka AK, Kochoumian L, Lichtenstein LM (1976) Arch Biochim Biophys 172:661–671
15. Verheij HM, Westermann J, Stenby B, De Haas GH (1983) Biochim Biophys Acta 747:93–99
16. Kaetzel MA, Dedman JR (1987) J Biol Chem 262:3726–3729
17. Clark MA, Conway TM, Shorr RGL, Cooke ST (1987) J Biol Chem 262:4402–4406
18. Laine RD, Esser AF (1989) J Immunol 143:553–557
19. Platts-Mills TAE, Chapman MD (1987) J Allergy Clin Immunol 80:755–775
20. Chua KY, Stewart GA, Thomas WR, Simpson RJ, Dilworth RJ, Plozza TM, Turner KJ (1988) J Exp Med 167:175–182
21. Wahn U, Peters T Jr, Siraganian RP (1980) J Immunol 125:2544–2549
22. Wahn U, Peters T Jr, Siraganian RP (1981) Mol Immunol 18:19–28