Major Venom Allergen of Yellow Jackets, Ves v 5: Structural Characterization of a Pathogenesis-Related Protein Superfamily

Anette Henriksen,[1*] Te P. King,[2] Osman Mirza,[1,3] Rafaél I. Monsalve,[2] Kåre Meno,[1,3] Henrik Ipsen,[3] Jørgen N. Larsen,[3] Michael Gajhede,[1] and Michael D. Spangfort[3]

[1]*Protein Structure Group, Department of Chemistry, University of Copenhagen, Copenhagen, Denmark*
[2]*Rockefeller University, New York, New York*
[3]*ALK-Abelló Group, Hørsholm, Denmark*

ABSTRACT **Ves v 5 is one of three major allergens found in yellow-jacket venom: phospholipase A_1 (Ves v 1), hyaluronidase (Ves v 2), and antigen 5 (Ves v 5). Ves v 5 is related by high amino acid sequence identity to pathogenesis-related proteins including proteins from mammals, reptiles, insects, fungi, and plants. The crystal structure of Ves v 5 has been solved and refined to a resolution of 1.9 Å. The majority of residues conserved between the pathogenesis-related proteins can be rationalized in terms of hydrogen bonding patterns and hydrophobic interactions defining an α-β-α sandwich core structure. A small number of consensus residues are solvent exposed (including two adjacent histidines) and located in an elongated cavity that forms a putative active site. The site has no structural resemblance to previously characterized enzymes. Homologous antigen 5's from a large number of different yellow jackets, hornets, and paper wasps are known and patients show varying extents of cross-reactivity to the related antigen 5's. The structure of Ves v 5 allows a detailed analysis of the epitopes that may participate in antigenic cross-reactivity, findings that are useful for the development of a vaccine for treatment of insect allergy. Proteins 2001;45:438–448.**

Key words: Ves v 5; antigen 5; insect allergy; X-ray crystallography; pathogenesis-related proteins

INTRODUCTION

Ves v 5[1] is the name in allergen nomenclature[1] of the yellow-jacket (*Vespula vulgaris*) venom protein antigen 5. Ves v 5 belongs to a superfamily of extracellular proteins found in mammals, reptiles, insects, fungi, and plants that share high sequence identity. The proteins are either pathogenesis related,[2] venom related,[3,4] or produced in the reproductive organs and cells.[5–7] So far the only direct evidence of the biological function of this protein superfamily has been the report of weak trypsin-inhibiting effect of a 25-kDa protein secreted by human glioblastoma cells[8] and reversible ryanodine receptor blocking by the toxin helothermine from Mexican beaded lizard venom.[4] For Ves v 5, no trypsin-inhibiting effect has been shown (T.P. King, data not shown). The ryanodine receptor, RyR, is the channel responsible for the release of Ca^{2+} from the sarcoplasmic reticulum in muscle cells, but it also plays a role in Ca^{2+} regulation in non-muscle cells.[9–12] Helothermine, however, having 22% sequence identity to Ves v 5, has a cysteine-rich C-terminal also found in the mammalian proteins but absent in the insect, plant, and fungal proteins.

Patients allergic to insect stings often show clinical symptoms of allergy against multiple insects.[13,14] This can be the result of consecutive sensitization by exposure to individual allergens from different insect venoms, or it may be caused by cross-reactivity towards one or several allergens. The fact that homologous allergens from different species can give rise to clinically observed allergic cross-reactivity suggests that clinical cross-reactivity in some cases is due to structural homology of the allergens. Structurally conserved allergen molecular surfaces have been identified among the major tree pollen allergens,[15] providing the structural basis for allergic cross-reactivity.[16] It has been suggested that conserved allergen molecular surfaces among inhalant allergens harbor major IgE-

Abbreviations: Dol m 5, *Dolichovespula maculata* venom antigen 5; FOM, figure of merit; MIRAS, Multiple isomorphous replacement with anomalous scattering; PDB, Protein Data Bank; Pol a 5, *Polistes annularis* venom antigen 5; RAST, radioallergosorbent test; r.m.s.d., root mean square deviation; RyR, ryanodine receptor; Sol i, *Solenopsis invicta* venom allergen; Ves s 5, *Vespula squamosa* venom antigen 5; Ves v 5, *Vespula vulgaris* venom antigen 5; Ves vi 5, *Vespula vidua* venom antigen 5.

The atomic coordinates reported in this paper have been deposited in the Protein Data Bank (ID code 1qnx).

Grant Sponsor: NIH; Grant number: AI-17021 Grant Sponsor: Danish Natural Science Research Council.

Anette Henriksen's present address is Carlsberg Laboratory, Department of Chemistry, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark.

Rafaél I. Monsalve's present address is Departamento de Bioquimica y Biologia Molecular, Facultad de Quimica, Universidad Complutense, Madrid 28040, Spain.

*Correspondence to: Anette Henriksen, Carlsberg Laboratory, Department of Chemistry, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark. E-mail: anette@crc.dk; or Michael D. Spangfort, ALK-Abello A/S, Boge Allé 6-8, DK-2970 Hørsholm, Denmark. E-mail: Michaels@inet.uni2.dk

Received 27 April 2001; Accepted 17 July 2001

TABLE I. Crystallographic Data and Model Refinement

Data collection and analysis	Native	$HgCl_2$	K_2PtCl_4
Resolution range, Å (outer shell)	30–1.9 (1.97–1.90)	40–2.00 (2.07–2.00)	60–2.00 (2.07–2.00)
Wavelength, Å	0.959	1.015	1.015
Completeness, % (outer shell)	92.0 (98.9)	94.6 (88.0)	99.0 (97.9)
Redundancy (outer shell)	8.1 (3.6)	9.0 (3.9)	7.2 (2.8)
R_{sym}, % (outer shell)[a]	6.7 (17.8)	7.0 (12.0)	4.8 (15.0)
$I/\sigma I$ (outer shell)	13.4 (6.9)	15.5 (11.3)	24 (6.2)
Number of sites		2	2
Phasing power acentric/centric[b]		1.4/1.0	1.7/1.6
R_{cullis} acentric/centric[c]		0.84/0.83	0.78/0.80
FOM after SHARP acentric/centric		0.41/0.31	
FOM after Solomon		0.90	

Refinement statistics No. of reflections (free)	R_{conv}[d]/R_{free}[e] (%)	No. of non-H atoms	No. of waters	R.m.s.d. dihedrals (°)	R.m.s.d. bond lengths (Å)
17,556 (845)	20.9/23.5	1,676	253	23.3	0.006

[a] $R_{sym} = \Sigma_{hkl}(\Sigma_i(|I_{hkl,i} - \langle I_{hkl}\rangle|))/\Sigma_{hkl,i}\langle I_{hkl}\rangle$, where $I_{hkl,I}$ is the intensity of an individual measurement of the reflection with Miller indicies h, k, and l, and $\langle I_{hkl}\rangle$ is the mean intensity of that reflection.
[b] $R_{cullis} = \Sigma_{hkl}\|F_{PH,hkl} \pm F_{P,hkl}| - F_{H,calc,hkl}|/\Sigma_{hkl}|F_{PH,hkl} - F_{P,hkl}|$, where F_{PH} is the structure factor of the heavy atom derivative, F_P is the structure factor of the native protein and $F_{H,calc}$ is the calculated structure factor for the heavy atom.
[c] Phasing power $= \Sigma_{hkl}F_{H,hkl}/\Sigma_{hkl}|F_{PH,obs,hkl} - F_{PH,calc,hkl}|$
[d] $R_{conv} = \Sigma_{hkl}(\|F_{obs,hkl}| - |F_{calc,hkl}\|)/\Sigma_{hkl}|F_{obs,hkl}|$, where $|F_{obs,hkl}|$ and $|F_{calc,hkl}|$ are the observed and calculated structure factor amplitudes.
[e] R_{free} is equivalent to the R_{conv}, but calculated with 5% of the reflections omitted from the refinement process.

binding epitopes,[16] but it remains to be established whether this is also the case for stinging insect allergens. A characterization of the molecular surface of these allergens will require determination of their three-dimensional structures.

Ves v 5 is one of the three major protein components of vespid venom with allergenic properties: Ves v 1 (phospholipase A_1),[17] Ves v 2 (hyaluronidase),[17] and Ves v 5.[3] Phospholipase A_1 splits off a fatty acid moiety from the 1 or 2 position of phosphatidyl compounds. Hyaluronidase hydrolyzes 1,4-linkages between N-acetyl-(-D-glucosamine and D-glucoronate residues in hyaluronate, the mucopolysaccharide constituting the bulk of animal connective tissue, but the biochemical role of Ves v 5 is unknown.

The crystal structures of bee venom hyaluronidase with 51% sequence identity to Ves v 2[18] and rat pancreatic lipase with 32% sequence identity to Ves v 1[19] have been solved. In addition, NMR structures of the pathogenesis-related protein from tomato, p14a with 27% sequence identity to Ves v 5 have been described.[20] Here we report the X-ray structure of Ves v 5 and analyze structural and possible functional relationships to the other proteins from the pathogenesis-related superfamily. Furthermore, the allergological consequences of conserved surface patches found by sequence alignment of antigen 5's from the *Vespula* species and the related proteins from *Dolichovespula, Polistes* and *Solenopsis* are discussed.

MATERIALS AND METHODS

Crystallization and Data Collection

Recombinant Ves v 5 was expressed in the yeast *Pichia pastoris* and purified as previously described.[21] The recombinant protein was provided with an N-terminal leader sequence consisting of the EAEAEF sequence. Crystallization was achieved using the sitting drop vapor diffusion method with a precipitant composed of 16–20% w/v PEG 6,000 and 100 mM citrate-buffer pH 5.0–6.0, and a drop size of 5 μL protein solution (5 mg/mL) + 5 μL precipitant. Crystals were obtained in 2 to 14 days and belong to the orthorhombic space group $P2_12_12_1$, with cell dimensions a = 46.0 Å, b = 62.3 Å, and c = 83.2 Å. Crystals were soaked for ~10 sec in a cryo-buffer containing 30% w/v PEG 6000, 100 mM citrate pH 5.6, and rapidly cooled in liquid nitrogen. Heavy atom soaks with K_2PtCl_4 and $HgCl_2$ were done in the crystallization drops by the addition of 1 μl 5 mM heavy atom salt dissolved in water. The soaking time was approximately 3 h for both salts. Data was collected at beam line 711, MAXLABII Synchrotron Facility, University of Lund, Sweden. Diffraction data were collected with a Mar345 image plate and integrated and scaled with the Denzo and Scalepack programs.[22] Data collection was optimized to get both Friedell mates for derivatives and the Friedell mates were kept separate during integration and scaling. Data collection statistics are listed in Table I.

Structure Determination and Refinement

MIRAS phases were calculated based on two heavy atom derivatives: $HgCl_2$ and K_2PtCl_4. The heavy atom positions were identified from peaks in the anomalous and isomorphous difference Patterson maps using the CCP4 program suite.[23] Heavy atom positions were refined and the phases calculated with the program Sharp[24] followed by solvent density modification with a solvent content of 43% by the program Solomon.[25] The resulting electron density map was easily traced using O.[26] The model was refined with CNS[27] with the mli target function against data from 30 to 1.9 Å using a bulk solvent model and anisotropic B-factor

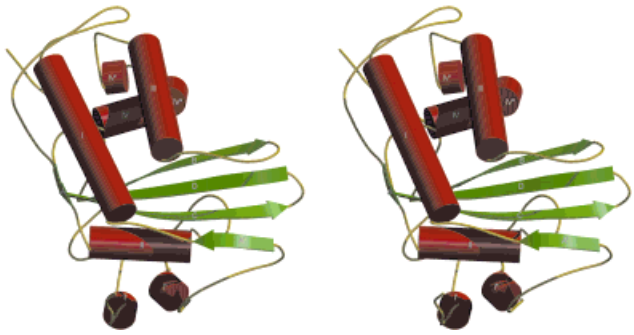

Fig. 1. Overview of the Ves v 5 structure with secondary structural elements labeled. The figure is generated using Molscript[48] and Raster3D.[49,50]

correction. Refinement steps were accepted if they produced a lowering of R_{free}. The temperature factors were refined for every atom but restrained to the temperature factors of neighboring atoms. The refinement statistics are listed in Table I. The Ramachandran plot as calculated by the program PROCHEC[28] shows no residues in disallowed or generously allowed regions and 88.5% in the most favorable regions of the phi-psi plot. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB)[29] with the accession code 1qnx.

RESULTS

Overall Structure of Ves v 5

Ves v 5 posses a unique fold among structurally characterized allergens. The structure determination reveals an α-β-α sandwich protein (Fig. 1). The upper layer consists of three α-helices, with one helix (IV, residues 155–160, omitting the leader sequence from the numbering) running parallel to the β-sheet and two helices running perpendicular to the β-sheet (I and III, residues 38–56 and residues 126–139). Two 3_{10} helices (IV' and IV") additionally contribute to the upper layer of the sandwich. The β-sheet is a four-stranded anti-parallel sheet (A': residues 30–34; B: residues 110–118; C: residues 167–177; D: residues 180–190). The lower layer consists of two α-helices (I" and II, residues 15–18 and 81–92) and the 3_{10} helix I' (residues 3–5). The secondary structure classification is based on the Kabsch and Sander algorithm.[30]

Only one protein with a similar structure is found in the PDB, which are the NMR structures[20] of a pathogenesis-related protein from tomato named p14a. Structural comparison identifies a core structure composed of three major α-helices (I–III), a minor α-helix (IV) and four β-strands (A–D). A superimposition of the four α-helices and three central β-strands (B–D) from Ves v 5 and p14a gives an r.m.s.d. of 1.5 Å for 112 superimposable C_α-atoms.

Characterization of the Ves v 5-Like Protein Superfamily

A SWISS-PROT search[31] with the Ves v 5 sequence identifies a protein superfamily in the PROSITE database,[32] which shares the [GDER]-H-[FYWH]-T-Q-[LIVM]-[LIVM]-W-x-x-[STN] and the [LIVMFYH]-[LIVMFY]-x-C-[NQRHS]-Y-x-[PARH]-x-[GL]-N- [LIVMFYWDN] consensus sequence patterns. Additional consensus sequences in the superfamily can be identified by structural comparison of p14a and Ves v 5 and alignment of primary sequences from the superfamily identified by a BLAST search.[33] A total of seven consensus sequences can be mapped in this way, where the consensus mapped out by PROSITE represents the consensus patterns number 6 and 7.

The alignment in Figure 2 includes representatives from some of the different protein families included in the Ves v 5-like protein superfamily. The original alignment, from which the consensus sequences presented here were extracted, is not shown, but it included 91 non-redundant full-length primary sequences. The term consensus residue covers a conservation of the physicochemical class of a residue. The consensus groups are defined as: alcohol, o (S,T); aliphatic, l (I, L, V); aromatic, a (F, H, W, Y); negative, − (D, E); positive, + (H, K, R); charged, c (D, E, H, K, R); tiny, u (A, G, S); small, s (A, C, D, G, N, P, S, T, V); turnlike, t (A, C, D, E, G, H, K, N, Q, R, S, T); hydrophobic, h (A, C, F, G, H, I, K, L, M, R, T, V, W, Y); polar, p (C, D, E, H, K, N, Q, R, S, T).[34,35] A consensus level of 90% is used in the discussion. We have not considered residues falling in the consensus groups p, h, t, and s to be consensus residues in a strict sense and they are not included in the detailed discussion of consensus patterns but they are mapped out in Figures 2 and 3. The structural comparison of Ves v 5 and p14a gives the structural background for the consensus sequences (using Ves v 5 numbering unless stated otherwise).

The first consensus sequence is found in the C-terminal of helix I: l_{45}-p-x-H-N-t-h-R-t-x-h_{55} (sub indices refers to Ves v 5 sequence numbers). Leu45 is stacking with Trp79 forming a conserved hydrophobic solvent exposed patch outside the putative active site cleft [Fig. 4(A)]. This patch covers 21% of the conserved solvent exposed surface of the molecule. The N_{E2} of the conserved residue His48 participates in a hydrogen bond to the O_H atom of another consensus residue, Tyr189 in strand D. The O_{D1} and N_{D2} from the likewise conserved Asn49 is engaged in hydrogen bonds with the backbone N and O of Leu77 from strand A, which is contained in the second consensus sequence: h_{74}-x-h-x-W-s-x-p-h-t-t-x-A-t-x-a-u-p_{92}. The function of these hydrogen bonds is to turn the protein chain towards helix II after the variable insert between helix I and strand A.

The length of the loop inserts between helix I and strand A vary in the superfamily, but the diverging backbones converge at Leu77 in the beginning of strand A. The side chain of the conserved Arg52 from helix I seems to be important for the formation and stabilization of the tertiary structure. It is positioned at the C-terminus of helix IV and participates in no less than five hydrogen bonds including bonds to the backbone oxygens of Lys75, Met160, Trp162, and Thr165. Met160, Trp162, and Thr165 are all included in the $[GDER]_{155}$-H-[FYWH]-T-Q-[LIVM]-[LIVM]-W-x-x-$[STN]_{165}$ consensus sequence defined in PROSITE. Lys75 is solvent exposed and the side chain is not conserved in the superfamily.

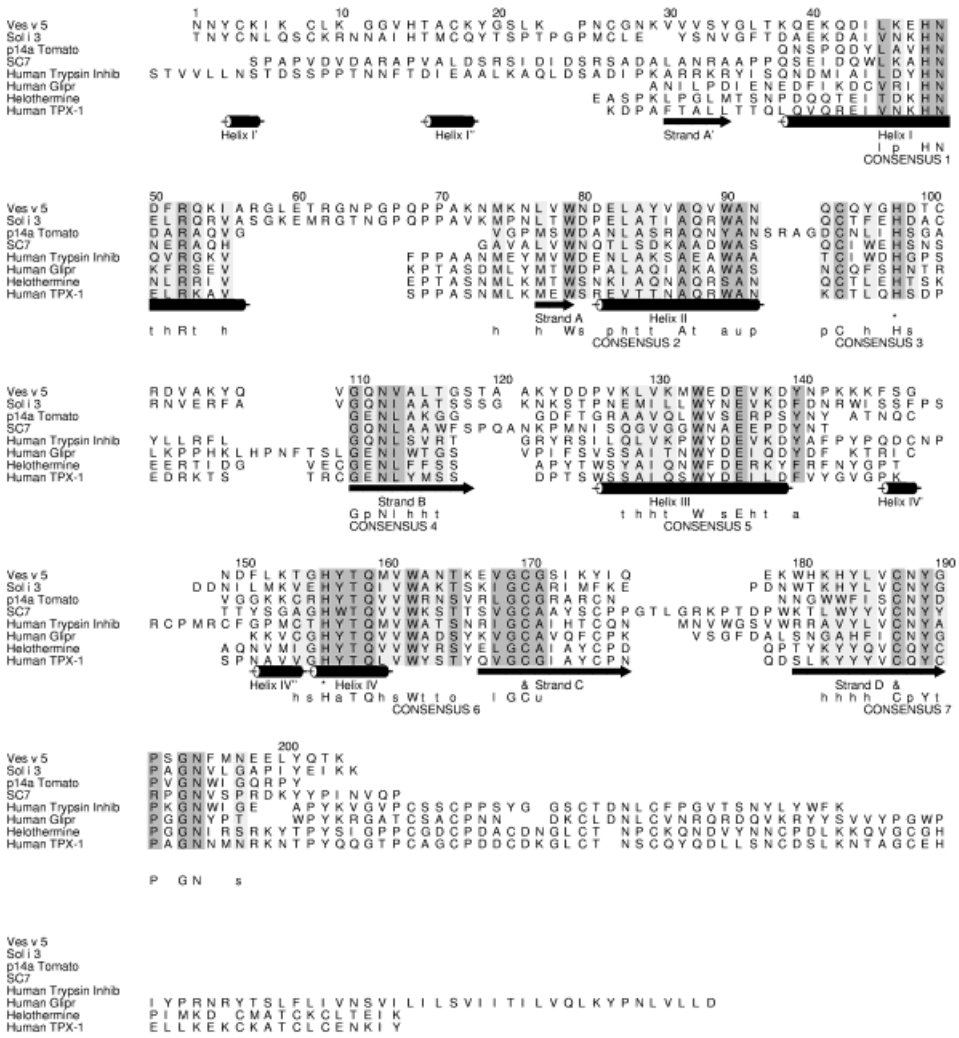

Fig. 2. Alignment of Ves v 5-like proteins. The consensus residues (having amino acid residue property conserved in > 90% of the protein superfamily) have shaded background. The consensus groups are defined in the results and discussion section. The background of residues belonging to the consensus groups o, l, a, −, +, c, and u is shaded in a darker shade of gray. The secondary structural elements of Ves v 5 are included as well as Ves v 5 sequence numbers. Sol i 3, *Solenopsis invicta* venom allergen[51]; p14a tomato, pathogenesis-related leaf protein 6 from tomato[52]; SC7, fruiting body protein SC7 from *Schizophyllum commune* (bracket fungus)[7]; human trypsin inhib., 25 kDa human trypsin inhibitor[53]; human glipr, glioma pathogenesis-related protein from human[54]; helothermine from the salivary glands of Mexican beaded lizard[4]; human TPX-1, testis-specific protein from human.[6] The alignment was made with multialign[55] and alscript.[56] The program MView was used in consensus pattern analysis.[35] Asterisks mark the putative active site residues and the & sign marks the only conserved disulfide bridge in the superfamily.

From the second consensus sequence: h_{74}-x-x-h-x-W-s-x-p-h-t-t-x-A-t-x-a-u-p_{92} [Figs. 2 and 4(B)], Trp79 has hydrophobic contacts to residues from helix I. The exact structural role of this residue is not clear as the limited variability of this solvent exposed residue suggest it to have a role extending the creation of a packing environment for helix I.

The side chain of the consensus residue Asn80 is participating in hydrogen bonds with backbone atoms of residue 82 and 83, thereby stabilizing helix II. The hydrophobic consensus residues Ala87 and Ala91 make helix II amphipatic. The residues create the packing environment between helix II and strand C and D of the β-sheet. Trp90 has a hydrogen bond to the backbone of Gly97, thereby contributing to the positioning of His98 in the central cavity.

Superposition of Ves v 5 and p14a shows that p14a has a one turn extension of helix II, and the alignment in Figure 2 suggests that helix II in the majority of family members have the same length as in Ves v 5. The alignment also shows that Cys94 and His98 following helix II are conserved in the third consensus sequence: p_{93}-C-x-h-x-H-s_{99}. In the insect proteins, Cys94 is partitioning in a disulfide bond with Cys26 found N-terminal to helix I, while a cysteine in the position corresponding to Cys94 forms a disulfide bond with a cysteine found at the C-terminal end of the C-strand in the other members of the protein superfamily. Both disulfide linkages, however, place His98

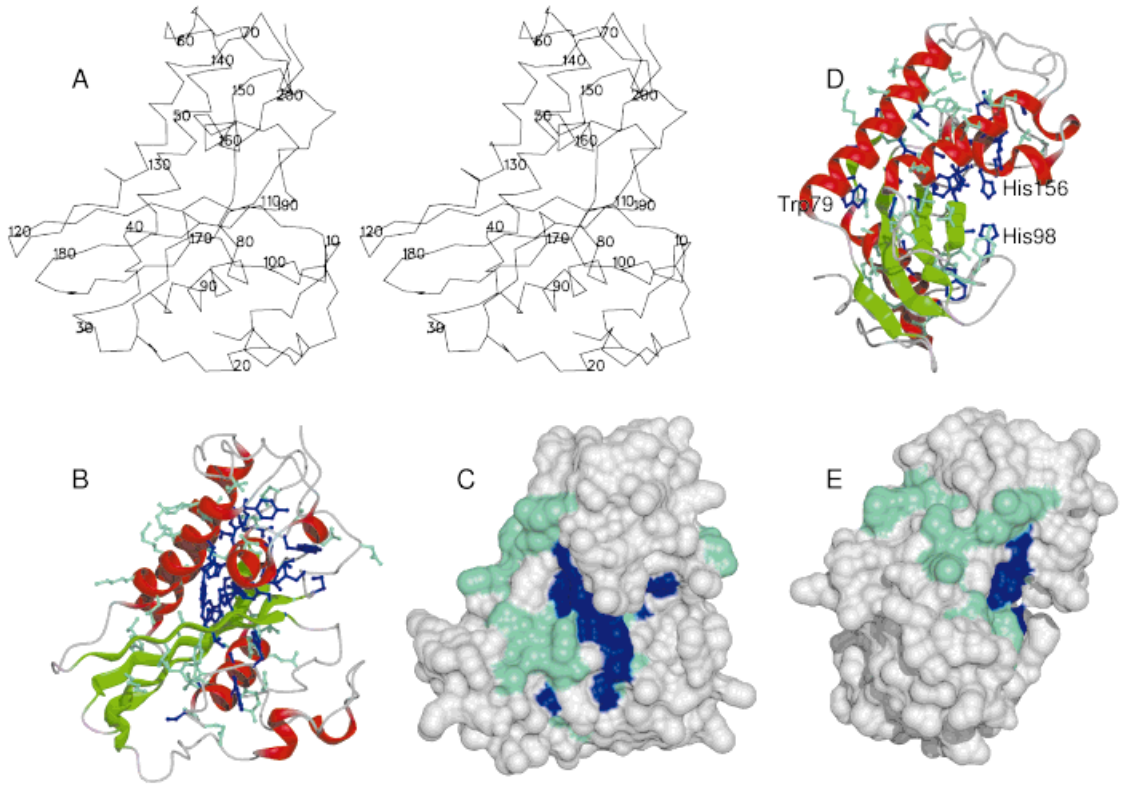

Fig. 3. **A:** Stereographic C_α trace of Ves v 5 with sequence numbers for every 10'th residue. **B,C:** Richardson representation and solvent accessible surface of Ves v 5 looking into the cleft area and (**D,E**) with the conserved cleft area on the right. The stereographic picture is in the same orientation as the representations in B and C. Dark blue residues are consensus residues belonging to the physiochemical groups o, l, a, −, +, c, or u, while light green residues belong to the groups s, t, h, or p. The stereographic picture is made with Molscript,[57] while the surface accessible pictures were made with msms[58] and dino.[59]

in the same position relative to the PROSITE consensus sequence: $[GDER]_{155}$-H-[FYWH]-T-Q-[LIVM]-[LIVM]-W-x-x-$[STN]_{165}$.

In the structure of Ves v 5, His98 is situated on one side of a cleft pointing directly towards His156 at the other side of the cleft [Fig. 4(C)]. The distance between His98 N_{E2} and His156 N_{E2} is 5.2 Å. The situation is more complicated in the NMR structures of p14a, where no sequential connectivity information is observed for the histidine residue corresponding to His98. In 11 of the 20 deposited models of p14a, this histidine residue is pointing into the cleft area, and in 9 models it is pointing away from it. The position of His98 in the Ves v 5 structure suggests that this residue is a candidate for an active site residue along with His156. His156 has already been pointed out as a possible active site residue by Fernández et al.[20]

Inserts of varying length are found in the various members of the protein superfamily between the position of His98 and the beginning of strand B. The residues forming the beginning of strand B form the fourth consensus sequence for the superfamily: G_{110}-p-N-l-h-h-t_{116}. Gly110 is required at the beginning of the β-strand due to the restricted space available at this end of strand B. The side chain of Asn112 has a hydrogen bond to the side chain of Tyr189, a conserved residue in strand D, and is solvent exposed in the central cavity with an additional hydrogen bond to the backbone of His156. In this way, the central cavity residues are linked by side chain hydrogen bonds between conserved structural elements to the interior of the molecule (Asn112-Tyr189-His48).

The backbone oxygen of Val113 is hydrogen bonded to Trp133 N_{E1} from helix III, a residue that is conserved throughout the superfamily and points into the cavity close to His156 (3.5 Å). The side chain of Val113 forms a part of the solvent exposed wall in the central cavity.

Strand B and the loop to helix III is seen to be of varying lengths within the superfamily. However, structural alignment of P14a and Ves v 5 demonstrates helix III to start at superimposable positions: Pro126 in Ves v 5.

In helix III, the residues: t_{128}-h-h-t-x-W-x-s-E-h-t-x-a_{140} form the fifth consensus sequence [Fig. 4(D)]. Trp133 is stacking with the possible active site residue His156 and has a side chain hydrogen bond to the backbone oxygen of Val113. The Glu136 side chain has a short hydrogen bond (2.5 Å) to His156 N_{D1} and hydrogen bonds with more usual lengths to the backbone nitrogen of the same residue, hereby both positioning the histidine in the central cavity and influencing its pKa. Tyr140 contributes to the hydrophobic packing environment between helix I, helix III, and helix IV.

After helix III, the alignment (Fig. 2) shows a coil region of very variable lengths. In some of the superfamily members, the region is stabilized by one (human glipr and P14a) or two (human trypsin inhibitor) disulfide bonds,

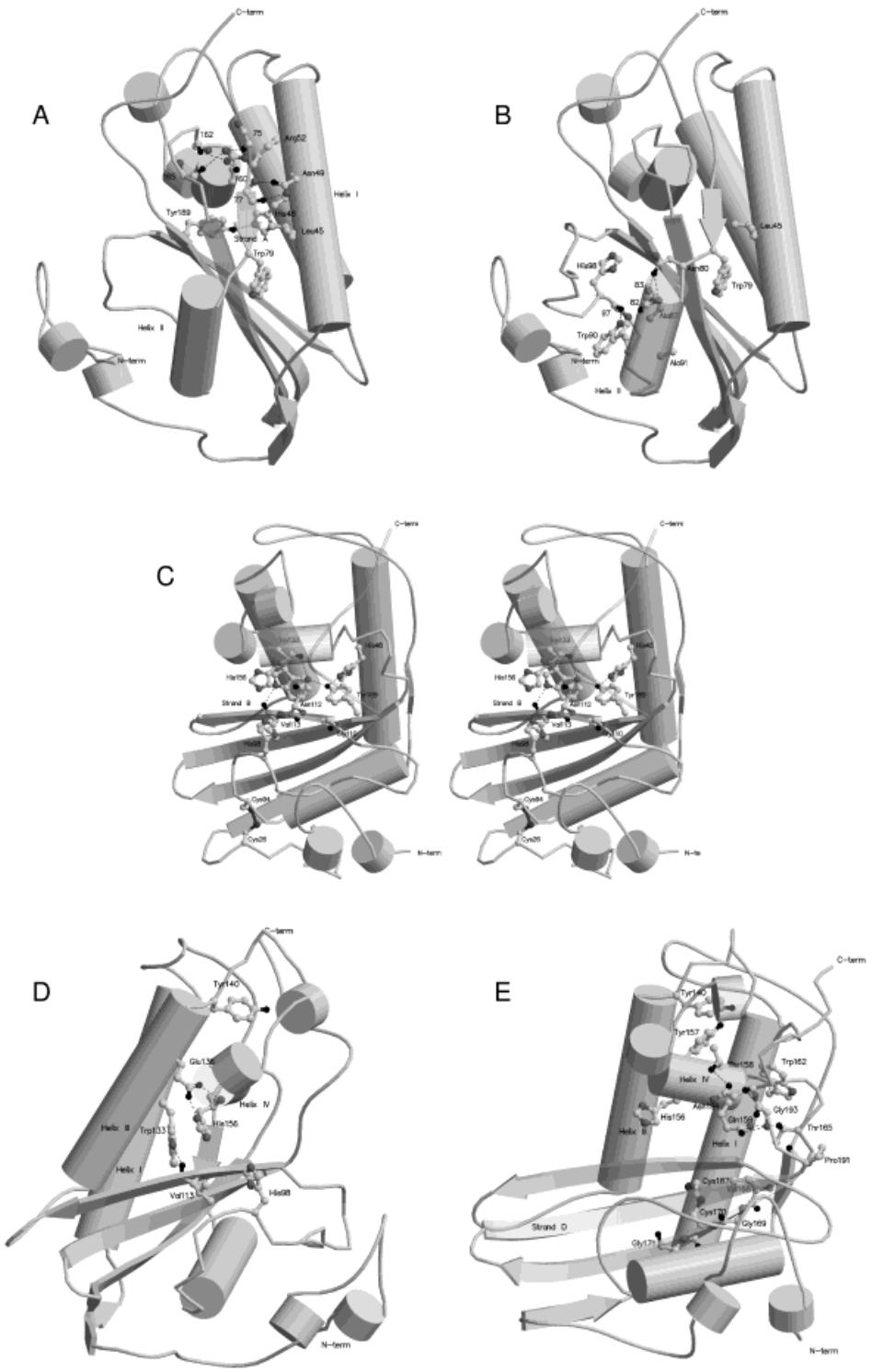

Fig. 4. Richardson representation of Ves v 5 including ball and stick models of the discussed consensus residues. Dotted lines represent hydrogen bonds. Residue number and backbone atoms are included when the interaction is to backbone atoms, while residue name and number and side chain atoms are included when the side chain is participating in the interaction. Some secondary structural elements are made transparent to give a better view of the interactions. **A:** Interactions involving residues from consensus pattern 1. **B:** Interactions involving residues from consensus pattern 2. **C:** Stereographic figure of the interactions involving residues from consensus pattern 3 and 4. **D:** Interactions involving residues from consensus pattern 5. **E:** Interactions involving residues from consensus pattern 6.

while disulfide bonds are absent in the helothermine protein, human TPX-1, Sol i, Ves v 5, and the *S. commune* protein 7. In Ves v 5, two one-turn helices precede helix IV, but none of the residues forming these helices are conserved. In contrast, the sixth consensus pattern: h_{154}-s-H-a-T-Q-h-s-W-t-t-o-x-x-l-G-C-u_{171} is highly conserved and forms a motif characterizing the protein family.

Tyr157 is part of the hydrophobic packing arrangement between helices I, III, and IV [Fig. 4(E)]. A hydrogen bond is found between Thr158 O_{G1} and Asn194O_{D1}. Asn194 found C-terminal to strand D is conserved throughout the superfamily. The side chain of Gln159 is hydrogen bonded to Thr165O_{G1} and to the backbone N of Gly193, a consensus residue following the D strand. The structural role of the conserved Trp162 is to provide a hydrophobic interface behind the conserved Gly193-Asn194 segment, which in turn determines the orientation of helix IV including His156. The consensus residues in strand C, Val168, Gly169, and Gly171, are packing towards the hydrophobic side of the amphipatic helix II.

A disulfide bond is found between Cys170 in strand C and C187 in strand D. It is the only disulfide bond that is conserved throughout the superfamily and it further stabilizes the β-sheet. Although the proteins of the superfamily generally have several disulfide bridges, the majority of the disulfide bridges must have been introduced after the first evolutionary divergence events. A high degree of divergence is also found in the region between strand C and strand D. This is where the cysteine bridging to the His98-region is located in all but the insect proteins.

The seventh consensus pattern is: h_{182}-h-h-h-x-C-p-Y-s-P-x-G-N-h-h-s_{197}. Of these residues, the roles of Cys187, Tyr189, Gly193, and Asn194 has already been mentioned. Pro191 is positioned at the tip of the turn after strand D that dictates the direction of the C-terminal of the protein.

Conserved B-cell Epitopes Among Ves v 5 Homologs

Vespids from different species have a similar protein composition in their venom including three groups of allergens, Vespula group 1, 2, and 5.[36] Skin testing of allergic patients using extracts from different vespid species show a high proportion of multiple reactivities, in agreement with RAST-based serum IgE antibody assays.[14,37] RAST inhibition studies show that multiple reactivities are caused by cross-reactivity in the majority of patients, i.e., the presence of structurally similar epitopes on homologous allergens, and sensitization to more than one species occurs only in a minority of patients.[38]

All Ves v 5 homologs from the *Vespula* genus are serologically cross-reactive, although some sera from allergic patients are less reactive towards Ves s 5.[39,40] No serological data are available for Ves vi 5. Partial cross-reactivity has been observed between *Vespula* and *Dolichovespula* antigen 5's and between *Vespula* and *Vespa* antigen 5's,[41] essentially reflecting the phylogenetic relationship and primary sequence identity between the *Vespula* and *Vespa* genera, whereas *Dolichovespula* and *Vespula* are not as closely related (see Fig. 5). In a study using mouse sera a low cross-reactivity between the allergens Ves v 5 and Pol a 5 was suggested, but the study was less conclusive about the relation between Ves v 5 and Dol m 5.[3]

An alignment of selected antigen 5 sequences from *Vespula* (yellow jackets), *Dolichovespula* (white-face hornets, yellow hornets), *Vespa* (hornets, European hornets), *Polistes* (paper wasps), and *Solenopsis* (fire ants) shows a high degree of sequence identity (Fig. 5). *Vespula* , *Dolichovespula*, *Vespa* and *Polistes* all belong to the *Vespidae* family. When considering only the *Vespula* antigen 5's, a very high degree of surface conservation is observed (Fig. 6), the conservation of residues being almost evenly distributed with only a few non-conserved residues scattered over the molecule.

In contrast, the surfaces conserved, when comparing sequences from the *Vespula* and *Polistes* genera, are restricted to 5 regions with solvent accessible areas of 392 Å^2, 585 Å^2, 589 Å^2, 673 Å^2, and 1053 Å^2, respectively (the solvent accessibility was calculated using the NACCESS program[42] with a probe radius of 1.4 Å). Similarly, five surface patches corresponding to the 5 surface patches conserved between *Vespula* and *Polistes*, are conserved between *Vespula* and *Vespa /Dolichovespula*. In the latter case, the areas are 280 Å^2, 496 Å^2, 730 Å^2, 803 Å^2, and 1,043 Å^2. In Figure 6, the conserved surface areas are highlighted in different colors.

DISCUSSION

Core Fold of the Superfamily

An overview of the secondary structural elements, with the side chains of the consensus sequences in blue, is shown in Figure 3. The figures demonstrate that the consensus residues primarily are found in the core of the protein. Overall, the majority of hydrogen bonds and hydrophobic interactions involving consensus residues are with other consensus residues. The alignment in Figure 2 shows that most of the secondary structural elements in Ves v 5 (strand A–D and helix I, II, III, and IV) must be expected to be characteristic for all the Ves v 5-like proteins, as the consensus residues are found in these elements. The interactions between the consensus residues assist the packing of secondary structural elements hereby defining the α-β-α sandwich core fold of the superfamily.

Localization of a Putative Active Site

Only a very limited number of the consensus residues contribute to the molecular surface of Ves v 5, as can be seen from the solvent accessibility surface (Fig. 3). These residues form the central cavity of the molecule, with His98 and 156 opposing each other at the sides of the cavity. One Hg^{2+} ion was bound between His98 and His156 in the Ves v 5-$HgCl_2$ derivative suggesting that the cleft in vivo could harbor an active site with a divalent cation and two histidine ligands, but attempts to soak Zn^{2+}, an ion found in many di-histidine sites, into the crystals were unsuccessful.

The largest cavity on enzyme surfaces often corresponds to catalytic centers.[43] The fact that the majority (65%) of

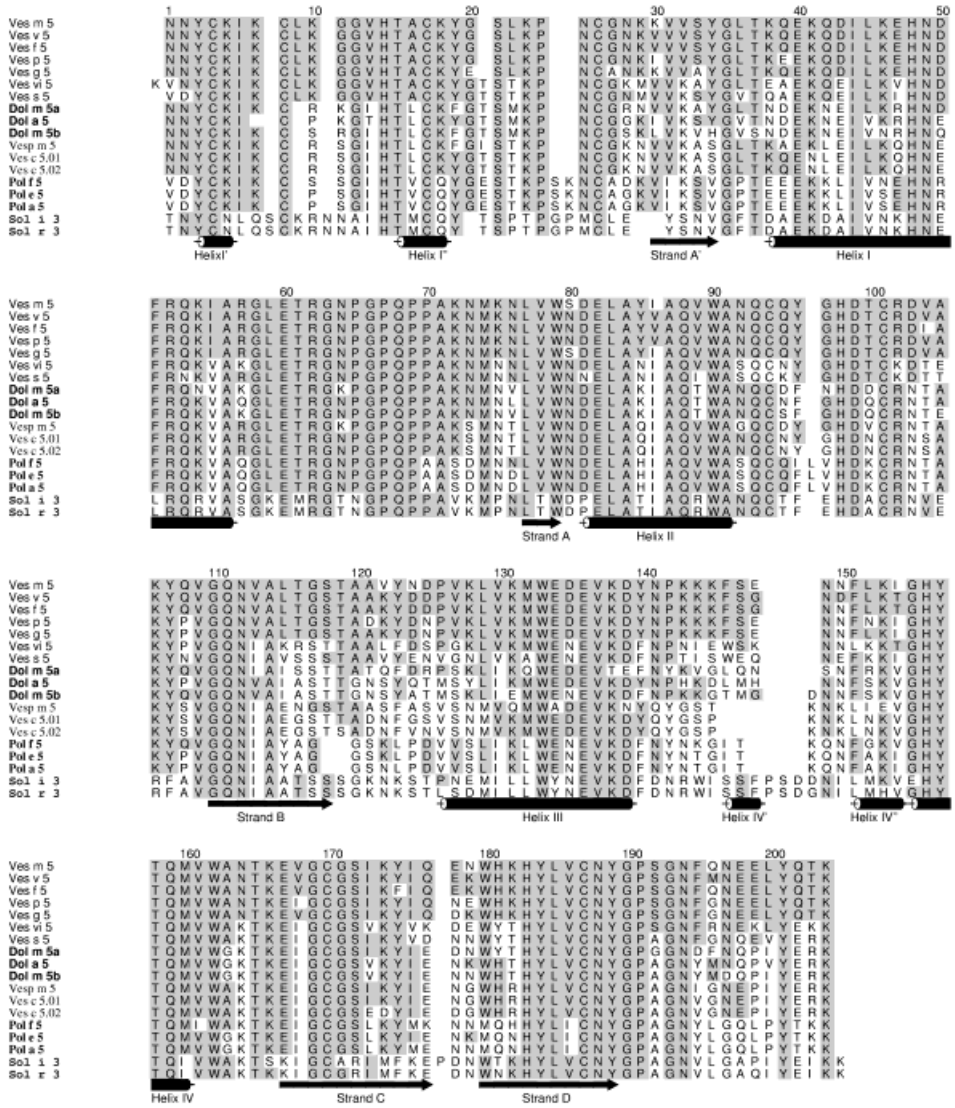

Fig. 5. Structural alignment of Ves v 5 homologous proteins from insect venoms. The secondary structural elements found in Ves v 5 are included as well as the Ves v 5 sequence numbers. Residues with gray background are identical to residues found at corresponding positions in Ves v 5. Ves m 5, *Vespula maculifrons* antigen 5[3]; Ves v 5, *V. vulgaris* antigen 5[3]; Ves f 5, *V. flavopilosa* antigen 5[60]; Ves p 5, *V. pensylvanica* antigen 5[60]; Ves g 5, *V. germanica* antigen 5[60]; Ves vi 5, *V. vidua* antigen 5[60]; Ves s 5, *V. squamosa* antigen 5[60]; Dol m 5.01, *Dolichovespula maculata* antigen 5[61]; Dol a 5, *D. arenaria* antigen 5[3]; Dol m 5.02, *D. maculata* antigen 5[61]; Vesp m 5, *Vespa mandarinia* antigen 5 (Swiss-Prot entry P81657); Vesp c 5.01, *V. crabo* antigen 5[60]; Vesp c 5.02, *V. crabo* antigen 5[60]; Pol f 5, *Polistes fuscatus* antigen 5[60]; Pol e 5, *P. exclamans* antigen 5[3]; Pol a 5, *P. annularis* antigen 5[3]; Sol i 3, *Solenopsis invicta* allergen[51]; Sol r 3, *S. richteri* allergen.[62]

the solvent exposed molecular surface of consensus residues is located in the central cavity of Ves v 5 makes us suggest that not only do the Ves v 5-like proteins form a superfamily with the same overall fold, but they are also likely to share biological function and have a similar enzymatic mechanism.

The trypsin-inhibiting effect observed for one of the proteins may be a coincidence rather than the primary function of the protein. The fact that helothermine can function as a reversible inhibitor of the ryanodine receptor indicates a role of Ves v 5-like proteins in the blocking of RyR Ca^{2+} currents, a suggestion consistent with the tissue expression pattern of the proteins. A definitive assignment of this function to the protein superfamily would require the determination of the ryanodine receptor blocking potential of more than one member of the protein family.

Conserved B-Cell Epitopes Among Ves v 5 Homologs

Antigen 5's from different species are expected to have identical protein fold due to the high degree of sequence similarity.[44] The conserved surface areas, therefore, are likely to constitute structurally similar surface contours, which are candidates for epitopes defined by cross-reactive antibodies. Studies of antigen-Fab complexes by X-ray crystallography indicate that the surface area covered by a

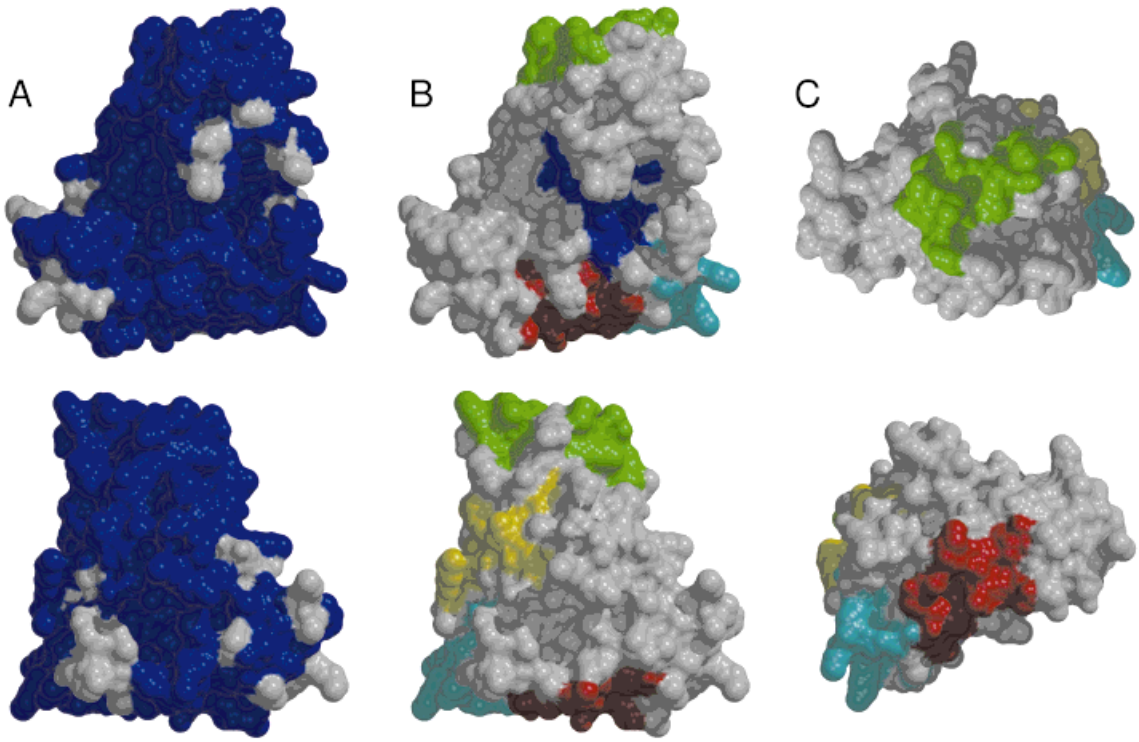

Fig. 6. Solvent accessible surface of Ves v 5. **A:** The blue patches are surface areas conserved within the *Vespula* genus but omitting Ves vi 5 and Ves s 5. **Left:** The same orientation the stereographic representation in Figure 3. **Right:** The backside of the molecule relative to the left side depiction. **B:** Green, blue, red, yellow, and cyan patches are surface regions, which are conserved between the *Vespula* and the *Vespa/Dolichovespula* genera. The orientation of the molecule is as in the upper panel. **C:** The same as in Figure 4(B), but this time seen from the top and bottom of the molecule relative to the orientations in Figure 4(B).

contact region in an antibody binding epitope is 800–900 Å^2.[45] Allergic patients cross-reactivity may be explained by the cross-linking of receptor bound IgE on the surface of mast cells and basophils, through simultaneous interaction with two conserved surface areas. Any conserved surface area of less than 800–900 Å^2 must thus be expected to cause a decrease in the binding constant for the IgE's.

This model supports the observed pattern of clinical cross-reactivity with extensive cross-reactivity within the *Vespula* genus, and variable cross-reactivity between *Vespula* and *Vespa /Dolichovespula* genera and between *Vespula* and *Polistes*. Furthermore, comparison of antigen 5 sequences from *Vespula* and homologous sequences from the *Solenopsis* genus identify conserved surface areas restricted to four small regions with solvent accessible areas of 220 Å^2, 235 Å^2, 266 Å^2 and 450Å^2, respectively. One of these areas is the conserved central cavity (Fig. 3). In agreement with the clinical observations, these conserved surface areas are too small to predict cross-reactivity of allergic patients' IgE.[46]

This structurally based model of IgE cross-reactivity could be further substantiated by experiments based on skin prick test of allergic individuals or assays based on histamine release from human basophils, experiments that are in progress. The model is similar to a model previously described for the tree pollen allergen Bet v 1.[15] Patients allergic to Bet v 1 show clinical cross-reactivity to the homologous allergens from alder and hazel, which may be explained by the occurrence of conserved surface areas representing epitopes defined by cross-reactive antibodies. It was proposed that exposure to pollens from the related trees through consecutive pollination and antibody affinity maturation would target high-affinity IgE antibodies to the conserved surface areas.[16] Thus, the conserved surface areas were proposed to constitute major IgE binding epitopes being candidates for site-directed mutagenesis aiming at the development of safer vaccines for allergen immunotherapy. Exposure to multiple species is a relatively rare event in vespid venom allergic patients and the conserved surface areas are less likely to represent dominating IgE binding epitopes. As a consequence, cross-reactivity is expected to be more variable from patient to patient, in agreement with clinical observations.[47] The concept of reducing IgE binding by substituting highly solvent exposed amino acids located centrally in conserved surface areas, however, could still be applicable for vespid venom allergens, since such artificial vaccines would be expected to reduce IgE binding irrespective of the sensitizing species, which only rarely can be determined.

ACKNOWLEDGMENTS

Part of this work has been carried out at the MAXII Synchrotron, at the Swedish National Electron Accelerator Laboratory for Nuclear Physics, and Synchrotron Radiation Research, University of Lund. T.P.K. was supported

by NIH grant AI-17021 and this research has been supported by the DANSYNC grant from The Danish Natural Science Research Council.

REFERENCES

1. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TAE, Thomas W. Allergen nomenclature. Bull WHO 1994;72:797–806.
2. Szyperski T, Fernández C, Mumenthaler C, Wüthrich K. Structure comparison of human glioma pathogenesis-related protein GliPR and the plant pathogenesis-related protein P14a indicates a functional link between the human immune system and a plant defence system. Proc Natl Acad Sci 1998;95:2262–2266.
3. Lu G, Villalba M, Coscia MR, Hoffman DR, King TP. Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets. J Immunol 1993;150:2823–2830.
4. Morrissette J, Krätzschmar J, Haendler B, el-Hayek R, Mochca-Morales J, Martin BM, Patel JR, Moss RL, Schleuning W-D, Coronado R. Primary structure and properties of Helothermine, a peptide toxin that blocks ryanodine receptors. Biophys J 95 A.D.;68:2280–2288.
5. Mizuki N, Kasahara M. Mouse submandibular glands express an androgen-regulated transcript encoding an acidic epididymal glycoprotein-like molecule. Mol Cell Endocrinol 1992;89:25–32.
6. Kasahara M, Gutknecht J, Brew K, Spurr N, Goodfellow PN. Cloning and mapping of a testis specific gene with sequence similarity to a sperm-coating glycoprotein gene. Genomics 1989;5: 527–534.
7. Schuren FH, Asgeirsdottir SA, Kothe EM, Scheer JM, Wessels JG. The Sc7/Sc14 gene family of Schizophyllum commune codes for extracellular proteins specifically expressed during fruit-body formation. J Gen Microbiol 1993;139:2083–2090.
8. Koshikawa N, Nakamura T, Tsuchiya N, Isaji M, Yasumitsu H, Umeda M, Miyazaki K. Purification and identification of a novel and four known serine proteinase inhibitors secreted by human glioblastoma cells. J Biochem (Tokyo) 1996;119:334–339.
9. Coronado R, Morrissette J, Sukhareva M, Vaughan DM. Structure and function of ryanodine receptors. Am J Physiol 1994;266:C1485–C1504.
10. Meissner G. Ryanodine receptor Ca^{2+} releasechannels and their regulation by endogenous effectors. Ann Rev Physiol 1994;56:485–508.
11. Ogawa Y. Role of ryanodine receptors. Crit Rev Biochem Mol Biol 1994;29:229–274.
12. Sorrentino V. Ryanodine receptors. Boca Raton: CRC Press; 1996.
13. Loveless MH, Fackler WR. Wasp venom allergy and immunity. 1956. Allergy Proc 1989;10:157–160.
14. Lichtenstein LM, Valentine MD, Sobotka AK. Insect allergy: the state of the art. J Allergy Clin Immunol 1979;64:5–12.
15. Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, Joost van Neerven RJ, Schou C, Løwenstein H, Spangfort MD. X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. Nat Struc Biol 1996;3:1040–1045.
16. Mirza O, Henriksen A, Ipsen H, Larsen JN, Wissenbach M, Spangfort MD, Gajhede M. Dominant epitopes and allergic cross-reactivity: Complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. J Immunol 2000;165:331–338.
17. King TP, Lu G, Gonzalez M, Qian N, Soldatova L. Yellow jacket venom allergens, hyaluronidase and phospholipase: sequence similarity and antigenic cross-reactivity with their hornet and wasp homologs and possible implications for clinical allergy. J Allergy Clin Immunol 1996;98:588–600.
18. Markovic-Housley Z, Miglierini G, Soldatova L, Rizkallah PJ, Müller U, Schirmer T. Crystal structure of hyaluronidase, a major allergen of bee venom. Structure 2000;8:1025–1035.
19. Roussel A, Yang Y, Ferrato F, Verger R, Cambillau C, Lowe M. Structure and activity of rat pancreatic lipase-related protein 2. J Biol Chem 1998;273:32121–32128.
20. Fernández C, Szyperski T, Bruyère T, Ramage P, Mösinger E, Wüthrich K. NMR solution structure of the pathogenesis-related protein P14a. J Mol Biol 1997;266:576–593.
21. Monsalve RI, Lu G, King TP. Expressions of recombinant venom allergen, antigen 5 of yellowjacket (*Vespula vulgaris*) and paper wasp (*Polistes annularis*), in bacteria or yeast. Protein Expr Purif 1999;16:410–416.
22. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Meth Enz 1997;276A:307–326.
23. Collaborative Computing Project n4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr 1994;D50:760–763.
24. de La Fortelle E, Bricogne G. Maximum-likelyhood heavy-atom parameter refinement for multiple isomouphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol 1997;276A:472–494.
25. Abrahams JP, Leslie AGW. Methods used in the structure determination of bovine mitochondrial F_1 ATPase. Acta Crystallogr 1996;D52:30–42.
26. Jones A, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these maps. Acta Crystallogr 1991;A47:110–119.
27. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. Acta Crystallogr 1998;D54:905–921.
28. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
30. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. Biopolymers 1983;22:2577–2637.
31. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 2000;28:45–48.
32. Hofmann K, Bucher P, Falquet L, Bairoch A. The PROSITE database, its status in 1999. Nucleic Acids Res 1999;27:215–219.
33. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25: 3389–3402.
34. Taylor WR. The classification of amino acid conservation. J Theor Biol 1986;119:205–218.
35. Bork P, Brown NP, Hegyi H, Schultz J. The protein phosphatase 2C (PP2C) superfamily: Detection of bacterial homologues. Protein Sci 1996;5:1412–1425.
36. King TP, Sobotka AK, Alagon A, Kochoumian L, Lichtenstein LM. Protein allergens of white-faced hornet, yellow hornet, and yellow jacket venoms. Biochemistry 1978;17:5165–5174.
37. Hoffman DR. Allergens in Hymenoptera venom XV: The immunologic basis of vespid venom cross-reactivity. J Allergy Clin Immunol 1985;75:611–613.
38. Panzani R, Blanca MSF, Juarez C. Sensitivity to European Wasps in a group of allergic patients in Marseille: preliminary results. J Investig Allergol Clin Immunol 1994;4:42–46.
39. Hoffman DR, McDonald CA. Allergens in hymenoptera venom. VII. Species specific reactivity to yellow jacket venoms. Ann Allergy 1981;47:23–27.
40. Hoffman DR, McDonald CA. Allergens in hymenoptera venom. VIII. Immunologic comparison of venoms from six species of Vespula (yellow jackets). Ann Allergy 1981;48:78–81.
41. Hoffman DR, Jacobson RS, Zerboni R. Allergens in hymenoptera venom. XIX. Allergy to Vespa crabro, the European hornet. Int Arch Allergy Appl Immunol 1987;84:25–31.
42. Hubbard SJ, Thornton JM. NACCESS. (2.1.1). 1993. Department of Biochemistry and Molecular Biology, University College London. 1992. Ref Type: Computer Program
43. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. Protein clefts in molecular recognition and function. Protein Sci 1996;5:2438–2452.
44. Chothia C, Finkelstein AV. The classification and origins of protein folding patterns. Annu Rev Biochem 1990;59:1007–1039.
45. Davies DR, Padlan EA, Sheriff S. Antibody-antigen complexes. Annu Rev Biochem 1990;59:439–473.
46. Hoffman DR, Dove DE, Moffitt JE, Stafford CT. Allergens in Hymenoptera venom. XXI. Cross-reactivity and multiple reactivity between fire ant venom and bee and wasp venoms. J Allergy Clin Immunol 1988;82:828–834.
47. Hoffman DR. Allergens in hymenoptera venom. VI. Cross reactiv-

ity of human IgE antibodies to the three vespid venoms and between vespid and paper Wasp venoms. Ann Allergy 1981;46:304–309.
48. Kraulis PJ. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. J Appl Crystallogr 1991;24:946–950.
49. Bacon DJ, Anderson WF. A fast algorithm for rendering space-filling molecule pictures. J Mol Graphics 1988;6:219–220.
50. Merritt EA, Murphy MEP. Raster3D Version 2.0: A program for photorealistic molecular graphics. Acta Crystallogr 1994;D50:869–873.
51. Hoffman DR. Allergens in Hymenoptera venom XXIV: the amino acid sequences of imported fire ant venom allergens Sol i II, Sol i III, and Sol i IV. J Allergy Clin Immunol 1993;91:71–78.
52. Lucas J, Camacho Henriquez A, Lottspeich F, Henschen A, Sanger HL. Amino acid sequence of the 'pathogenesis-related' leaf protein p14 from viroid-infected tomato reveals a new type of structurally unfamiliar proteins. EMBO J 1985;4:2745–2749.
53. Yamakawa T, Miyata S, Ogawa N, Kashikawa N, Yasumitsu H, Kanamori T, Miyazaki K. cDNA cloning of a novel trypsin inhibitor with similarity to pathogenesis-related proteins, and its frequent expression in human brain cancer cells. Biochim Biophys Acta 1998;1395:202–208.
54. Murphy EV, Zhang Y, Zhu W, Biggs J. The human glioma pathogenesis-related protein is structurally related to plant pathogenesis-related proteins and its gene is expressed specifically in brain tumors. Gene 1995;159:131–135.
55. Barton GJ, Sternberg MJ. Evaluation and improvements in the automatic alignment of protein sequences. Protein Eng 1987;1:89–94.
56. Barton GJ. ALSCRIPT. A tool to format multiple sequence alignments. Protein Engineering 1993;6:37–40.
57. Kraulis PJ. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. J Appl Crystallogr 1991;24:946–950.
58. Sanner MF, Spehner J-C, Olson AJ. Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 1996;38:305–320.
59. DINO: Visualizing structural biology. Basel: http://www.bioz.unibas.ch/~xray/dino; 2001.
60. Hoffman DR. Allergens in Hymenoptera venom. XXV: The amino acid sequences of antigen 5 molecules and the structural basis of antigenic cross-reactivity. J Allergy Clin Immunol 1993;92:707–716.
61. Fang KS, Vitale M, Fehlner P, King TP. cDNA cloning and primary structure of a white-face hornet venom allergen, antigen 5. Proc Natl Acad Sci USA 1988;85:895–899.
62. Hoffman DR. Reactions to less common species of fire ants. J Allergy Clin Immunol 1997;100:679–683.