Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of recombinant Ves v 2 at 2.0 Å resolution: structural analysis of an allergenic hyaluronidase from wasp venom

Lars K. Skov,[a] Ulla Seppälä,[b] Jeremy J. F. Coen,[c]‡ Neil Crickmore,[c] Te P. King,[d] Rafael Monsalve,[e] Jette S. Kastrup,[a] Michael D. Spangfort[b] and Michael Gajhede[a]*

[a]Biostructural Research, Department of Medicinal Chemistry, The Danish University of Pharmaceutical Sciences, Universitetsparken 2, DK-2100, Denmark, [b]ALK-Abelló, Bøge Allé 10-12, DK-2970 Hørsholm, Denmark, [c]School of Life Sciences, University of Sussex, Falmer, Brighton, East Sussex BN1 9QG, England, [d]Rockefeller University, 1230 York Avenue, New York, NY 10021, USA, and [e]ALK-Abelló, Calle Miguel Fleta 19, Madrid E-28037, Spain

‡ Present address: Unilever Research and Development, Port Sunlight, Quarry Road East, Bebington, Wirral CH63 3JW, England.

Correspondence e-mail: mig@dfuni.dk

Received 7 March 2006
Accepted 23 March 2006

PDB Reference: Ves v 2, 2atm, r2atmsf.

Wasp venom from *Vespula vulgaris* contains three major allergens: Ves v 1, Ves v 2 and Ves v 5. Here, the cloning, expression, biochemical characterization and crystal structure determination of the hyaluronidase Ves v 2 from family 56 of the glycoside hydrolases are reported. The allergen was expressed in *Escherichia coli* as an insoluble protein and refolded and purified to obtain full enzymatic activity. Three N-glycosylation sites at Asn79, Asn99 and Asn127 were identified in Ves v 2 from a natural source by enzymatic digestions combined with MALDI–TOF mass spectrometry. The crystal structure of recombinant Ves v 2 was determined at 2.0 Å resolution and reveals a central $(\beta/\alpha)_7$ core that is further stabilized by two disulfide bonds (Cys19–Cys308 and Cys185–Cys197). Based on sequence alignments and structural comparison with the honeybee allergen Api m 2, it is proposed that a conserved cavity near the active site is involved in binding of the substrate. Surface epitopes and putative glycosylation sites have been compared with those of two other major group 2 allergens from *Apis mellifera* (honeybee) and *Dolichovespula maculata* (white-faced hornet). The analysis suggests that the harboured allergic IgE-mediated cross-reactivity between Ves v 2 and the allergen from *D. maculata* is much higher than that between Ves v 2 and the allergen from *A. mellifera*.

1. Introduction

Human type 1 allergy is known to be triggered by allergen-mediated cross-linking of IgE antibodies attached to the FcεRI receptor on the surface of mast cells (Holowka & Baird, 1996). The part of the allergen surface that is specifically involved in protein–protein interactions with the IgE molecule is called a B-cell epitope. The epitope depends on the three-dimensional spatial arrangement of amino acids on the intact allergen. Hence, a structure determination is an essential step in the mapping of epitopes and in the characterization of the allergen.

Stings from essentially all insects of the order Hymenoptera can cause allergic reactions and are a major cause of IgE-mediated anaphylaxis. Between 3 and 12% of the population have a history of large local or systemic allergic reactions following insect stings (Golden *et al.*, 1989; Settipan & Boyd, 1970; Stuckey *et al.*, 1982). Although fatalities from insect stings are rare, the annual number of deaths arising from Hymenoptera stings is estimated to be about 40–50 in the United States (Nall, 1985). Many Hymenoptera are capable of stinging, but only species belonging to the families Apidae (bees), Formicidae (ants) or Vespidae (wasps) regularly sting humans. The Vespidae family includes yellow jackets (genus *Vespula*), hornets (genera *Vespa* and *Dolichovespula*) and

paper wasps (genus *Polistes*). Vespidae venom contains three major allergens: phospholipase A_1, hyaluronidase and antigen 5 of unknown biological function. The average vespid venom sac contains 1.7–17 µg protein (Hoffman & Jacobson, 1984) with 1.5% being hyaluronidase, 3.3% phopholipase and 8.1% antigen 5 (King *et al.*, 1983). In yellow jacket (*Vespula vulgaris*) venom these allergens are denoted Ves v 1 (phospholipase), Ves v 2 (hyaluronidase) and Ves v 5 (antigen 5). The three-dimensional structure has been determined for Ves v 5 (Henriksen *et al.*, 2001), revealing an active site with no structural resemblance to previously characterized enzymes. The structure determination of Ves v 5 also enabled analyses of conformational B-cell epitopes involved in allergic cross-reactivity.

Yellow jacket allergen Ves v 2 belongs to the hyaluronidase superfamily of extracellular proteins found in mammals, insects, leeches and bacteria (Kreil, 1995). Hyaluronidases degrade hyaluronic acid (HA), a large glycosaminoglycan that contains repeating D-glucoronic *N*-acetyl glucosamine units. HA is found in the extracellular matrix of almost all tissue as a component of the substance that connects protein filaments, collagen fibres and the connective tissue cells. The hyaluronidase in the venom is thus responsible for the breakdown of the HA after the insect sting. The metabolic pathways for HA are still not fully understood (Stern, 2003), but it is known that hyaluronidase-2 degrades the ~107 kDa polymer into 20 kDa fragments and that hyaluronidase-1 (for example, Ves v 2) degrades it further until ultimately tetrasaccharides are obtained. Based on amino-acid sequence alignments Ves v 2 has been assigned to family 56 of the glycoside hydrolases (Henrissat, 1991).

The structure of the honeybee venom hyaluronidase, Api m 2, has been determined previously (Markovic Housley *et al.*, 2000). It represented the first structure determination of a glycoside hydrolase family 56 member and was found to contain a $(\beta/\alpha)_7$ barrel. Based on the structure of Api m 2, Glu113 (Glu109 in Ves v 2) was identified as the proton donor, a finding in agreement with mutagenesis studies on the human PH-20 hyaluronidase (Arming *et al.*, 1997). The hydrolases of family 56 function by a retaining mechanism (Davies & Henrissat, 1995) and have a glutamic acid as a proton donor. It is remarkable that the catalytic nucleophile is not an amino acid from the enzyme but instead the carbonyl O atom of the C-2 acetamido group of the substrate. This type of substrate-assisted reaction is not unique; it has also been seen in glycoside hydrolase families 18 and 20 (Tews *et al.*, 1997). Cocrystallization of Api m 2 with a substrate analogue mapped out some substrate-binding subsites and identified the active site of the enzyme (Markovic Housley *et al.*, 2000).

Hymenoptera venom is used both for diagnosis and treatment *via* immunotherapy of IgE-mediated allergic reactions to insect stings (Brown *et al.*, 2003; Hunt *et al.*, 1978; Muller & Mosbech, 1993; Muller, 1990). Current diagnostics are based on the detection of specific venom IgE in the skin and/or blood of the allergic patient. Diagnosis can in some instances be difficult, since positive response can arise from either true sensitization towards specific insect venoms or be a consequence of cross-reactivity towards allergens from different venoms sharing similar conformational B-cell epitopes. For example, cross-reactivity between yellow jacket and honeybee venom hyaluronidases has been considered to be responsible for double-positive diagnosis. The ability to produce and characterize recombinant forms of venom allergens (Muller, 2002) opens the possibility for an improvement of diagnosis and for a detailed understanding of cross-reactivity patterns and specificity. This could lead to modified allergens with increased efficacy in immunotherapy treatment of allergic patients. Here, we report the cloning, expression, biochemical characterization and crystal structure determination of recombinant Ves v 2 (rVes v 2). In addition, the glycosylation sites of natural Ves v 2 (nVes v 2) were investigated, since cross-reactivity between venom allergens owing to carbohydrate-specific IgE has been demonstrated (Hemmer *et al.*, 2001).

2. Materials and methods

2.1. Expression and purification

The Ves v 2 cDNA (King *et al.*, 1996) was cloned into the expression vector pET3 (Novagen). Overnight cultures of BL21 pET3-Ves v 2 were grown in Luria–Bertani broth. Large-scale growth and induction was performed in 2 l flasks containing 500 ml Terrific Broth. After inoculation with 10 ml of the overnight culture, growth was continued for approximately 3 h at 310 K, followed by induction with 0.5 m*M* isopropyl β-D-thiogalactopyranoside for 3 h. The bacteria were harvested and resuspended in 30 ml buffer *A* [20 m*M* Tris–HCl pH 8.6, 1 m*M* ethylenediaminetetraacetic acid (EDTA)] and stored at 253 K. Cells were lysed by sonication (MSE ultrasonic disintegrater MK2) for 3 × 30 s. After centrifugation at 10 000*g* for 30 min, the pellet was resuspended in buffer *A* containing 2 *M* urea using a manual homogenizer. Triton X-100 was added to a final concentration of 1%(*v*/*v*) and the inclusion bodies were re-sedimented at 25 000*g* for 30 min. Following this, the pellet was washed in buffer *A* containing 2 *M* urea and 1%(*v*/*v*) Triton X-100 and then in buffer *A* containing 2 *M* urea. The protein pellet was solubilized in 30 ml of 20 m*M* Tris–HCl pH 8.6, 6 *M* guanidine hydrochloride, 2 m*M* EDTA and 300 m*M* dithiothreitol by constant stirring at room temperature for at least 2 h. The solution was centrifuged at 25 000*g* for 30 min and the supernatant retained. The protein concentration was estimated using the Bradford method (Bio-Rad) with bovine serum albumin as a standard. Solubilized protein was rapidly frozen in 1 ml aliquots and stored at 193 K.

Based on small-scale folding screens, the highest relative activity after refolding was obtained using 15 m*M* oxidized glutathione and a pH of 7. No activity was observed prior to refolding. For the large-scale refolding process, equal aliquots of solubilized Ves v 2 were added at hourly intervals for 4 h to the folding buffer (50 m*M* Tris–HCl pH 7.0, 15 m*M* oxidized glutathione and 3 m*M* EDTA). The final Ves v 2 concentration was 120 µg ml^{-1}. This mixture was left for 12 h at 277 K,

spun at 10 000*g* for 15 min to remove insoluble material and concentrated to approximately one-tenth of its initial volume in a Vivacell concentrator. This was followed by centrifugation at 25 000*g* for 30 min to again remove insoluble material. The protein was dialysed for 16 h against 20 m*M* MES buffer pH 6.5 and a final centrifugation was performed. The soluble Ves v 2 was loaded onto a 1 ml cation-exchange column (Mono S HR5/5; AKTA FPLC system; Amersham Pharmacia Ltd) that had been equilibrated previously with the MES buffer. The protein was eluted with a linear NaCl gradient in the same buffer. Finally, Ves v 2 was purified on a 24 ml gel-filtration column (Superose 6 10/300; AKTA FPLC system; Amersham Pharmacia Ltd) that had been equilibrated previously with 100 m*M* NaCl, 20 m*M* MES pH 6.5. Peak fractions were pooled, frozen and stored at 193 K.

During the purification process, hyaluronidase activity assays were carried out according to the method of Linker (1974), with the volume adjusted to a final volume of 1 ml. Hyaluronic acid from human umbilical cord was used as a substrate. Specific activity of purified recombinant Ves v 2 (rVes v 2) was determined using the method described by Richman & Baer (1980). Bee venom was used as a reference.

2.2. Characterization and mass-spectrometric analysis

Sequencing of the rVes v 2 was performed with Hewlett Packard G1000A sequencer (Palo Alto, CA, USA) equipped with a Hewlett–Packard 1090 series II liquid chromatograph and run as described by the manufacturer.

10% Bis-Tris SDS–PAGE (NuPage, Invitrogen, Paisley, Scotland) was used to analyze 2.0 µg rVes v 2, yellow jacket venom hyaluronidase (nVes v 2) and bee venom hyaluronidases (Api m 2). The electrophoresis was performed under reducing conditions according to the manufacturer's instructions and silver-stained.

Natural Ves v 2 and rVes v 2 (500 pmol) were adjusted to 1:1 ratio in 0.06% trifluoroacetic acid (TFA; Rathburn Chemicals, Peebleshire, UK) and subjected to a 2.0 × 250 mm Jupiter, C_4 reversed-phase column (Phenomenex, Torrence, CA, USA). The column was equilibrated with 0.06% TFA and the sample was eluted with 0.05% TFA, 80% acetonitrile (Riedel-de-Hein, Seelze, Germany) in a gradient of 5–80% in nine column volumes. Both nVes v 2 and rVes v 2 eluted as single peaks and were collected and dried in a Speedvac and stored at 253 K. The fractionation of the tryptic nVes v 2 peptide mixture was performed in a 2.0 × 250 mm Jupiter C_{12} (Phenomenex) column. The peptides were eluted with 0.05% TFA, 80% acetonitrile in a gradient of 5–80% in 18 column volumes. The separated peptides (1.0 pmol µl^{-1}) were directly subjected to MALDI–TOF MS analysis.

Natural Ves v 2 and rVes v 2 were redissolved in 50 m*M* Tris–HCl pH 8.5, 0.2 *M* NaCl and 1 *M* urea and incubated with 3%(*w*/*w*) endoproteinase Lys-C (Wako GmbH, Richmond, VA, USA) and Asp-N (Calbiochem, Sunnyvale, CA, USA) at 310 K for 18 h (Seppala *et al.*, 2005). Reduction and alkylation of rVes v 2 were performed as described in Seppala *et al.* (2005). Assignment of the disulfide bridges in r Ves v 2 and the glycosylation sites in nVes v 2 were performed by enzymatic digestion employing 3%(*w*/*w*) trypsin (Sequencing Grade Modified Trypsin, Promega, Madison, WI, USA) in 50 m*M* Tris–HCl pH 8.5, 0.2 *M* NaCl and 1 *M* urea. For fingerprinting analyses, rVes v 2 and nVes v 2 digests (2.0 pmol µl^{-1}) were desalted by custom-made micro columns packed with POROS R2/50 (Applied Biosystems, Foster City, CA, USA) reversed-phase resin (Gobom *et al.*, 1999). Samples were eluted directly on MALDI targets with 1.0 µl matrix solution containing 5 µg µl^{-1} α-cyano-4-hydroxycinnamic acid (Sigma, St Louis, MO, USA) in 70% acetonitrile and 0.1% TFA.

Deglycosylation assays were assessed with *N*-glycosidase (Glyko Enzymatic deglycosylation kit, ProZyme Inc, San Leandro, USA), endoglycosidase H (Calbiochem, La Jolla, CA, USA) and *N*-glycosidase A (Roche Diagnostics, Geneva, Switzerland). Deglycosylation experiments with *N*-glycosidase were performed according to the supplier's manual. Deglycosylation experiments with endoglycosidase H and *N*-glycosidase A were performed in 50 m*M* ammonium acetate pH 5.0 at 310 K for 18 h.

Molecular-weight analysis of rVes v 2 (20 pmol) was performed by nanospray in a Q-TOF1 MS (Waters Corporation, Manchester, England) instrument. Mass fingerprinting analyses and analyses of the isolated peptides were performed on a 4700 Proteomics Analyzer and Voyager-DE STR Biospectrometry (Applied Biosystems, Foster City, CA, USA) MALDI–TOF MS instrument using an acceleration voltage of 25 kV and a nitrogen laser at 337 nm. The spectra were acquired in the positive-ion mode and were calibrated externally. Sequazyme Peptide Mass Standard Kit (Applied Biosystems, Foster City, CA, USA) was used for calibration. Theoretical masses were calculated by *GPMAW* (Lighthouse Data, Odense, Denmark) and the database searches were performed using the *MASCOT* search engine. *GPMAW* was also used to predict putative glycosylation sites in Ves v 2.

2.3. Crystallization

Crystallization conditions were screened according to the sparse-matrix method (Jancarik & Kim, 1991) using commercially available buffers (Hampton Research, Laguna Hills, CA, USA) and the hanging-drop vapour-diffusion technique (McPherson, 1992). Polyethylene glycol (PEG) 4000 was found to be a good crystallization medium. Hanging drops were prepared by mixing 2.5 µl protein solution (1–2 mg ml^{-1} in 150 m*M* NaCl and 20 m*M* MES pH 6.5) with 2.5 µl reservoir solution [25%(*w*/*v*) PEG 4000, 100 m*M* sodium acetate pH 4.6, 200 m*M* ammonium sulfate] and equilibrated against 500 µl reservoir solution at 295 K. Crystals were flash-cooled in liquid nitrogen without further addition of cryoprotectant.

2.4. Data collection, structure solution and refinement

A native data set at 2.0 Å resolution was collected at 110 K on beamline ID-29 (λ = 0.979 Å), European Synchrotron Radiation Facility (ESRF), Grenoble, France. Diffraction images were collected on an ADSC Quantum 210 detector using a crystal-to-detector distance of 180 mm, oscillations of

Table 1
X-ray data statistics.

Values for the highest resolution shell are given in parentheses.

Data collection	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 42.7, b = 95.0, c = 100.3
Resolution range (Å)	24.7–2.0 (2.07–2.0)
Total No. of reflections measured	135089
No. of unique reflections	28373 (2756)
Intensity [$I/\sigma(I)$]	17.8 (3.6)
Completeness (%)	99.8 (98.9)
Multiplicity	4.8 (4.7)
$R_{\rm merge}(I)$† (%)	8.8 (47.7)
Refinement statistics	
Resolution range (Å)	24.7–2.0
No. of reflections	27345
$R_{\rm cryst}$‡/$R_{\rm free}$§ (%)	18.6/21.5
No. of protein atoms (residues 6–329)	2696
No. of solvent molecules	334
No. of heteroatoms	17
R.m.s. deviation for ideal bonds (Å)	0.008
R.m.s. deviation for ideal angles (°)	1.2
Average protein B value (Å^2)	25.9
Average solvent B value (Å^2)	36.3
Average hetero atom B value (Å^2)	49.4
Ramachandran plot	
Most favoured regions (%)	88.3
Additional allowed regions (%)	11.7
Generously allowed regions (%)	0
Disallowed regions (%)	0

† $R_{\rm merge}(I) = \sum_{hkl}[\sum_i(|I_{hkl,i} - \langle I_{hkl}\rangle|)]/\sum_{hkl,i}\langle I_{hkl}\rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices hkl and $\langle I_{hkl}\rangle$ is the mean intensity of that reflection. ‡ $R_{\rm cryst} = \sum_{hkl}(||F_{o,hkl}| - |F_{c,hkl}||/|F_{o,hkl}|)$, where $|F_{o,hkl}|$ and $|F_{c,hkl}|$ are the observed and calculated structure-factor amplitudes. § $R_{\rm free}$ is equivalent to $R_{\rm cryst}$ but calculated with reflections omitted from the refinement process (5% of reflections omitted).

1.0° and two passes with 2 s exposure per image. The crystal belongs to space group $P2_12_12_1$ with one molecule per asymmetric unit, resulting in a solvent content of 53% ($V_{\rm M}$ = 2.6 Å^3 Da^{-1}). Data processing was performed with the *HKL* suite (Otwinowski & Minor, 1997) and details concerning the data collection are given in Table 1.

Initially, the structure was solved by molecular replacement using low-resolution data collected on a Rigaku RU300 in-house source. The diffraction resolution was subsequently improved to 2.8 Å at MAX-Lab, Lund, Sweden and finally to 2.0 Å at ESRF (data reported in Table 1). For molecular replacement at low resolution, the structure of bee hyaluronidase (PDB code 1fcq; Markovic Housley *et al.*, 2000) was used as the search model in programs from the *CNS* package (Brünger *et al.*, 1998). Further refinement of this solution was performed with *CNS*.

For the high-resolution data (2.0 Å), initial phases were obtained from the refined molecular-replacement solution and the first model building was performed with *ARP*/*wARP* (Perrakis *et al.*, 1999) using the *CCP*4*i* interface (Collaborative Computational Project, Number 4, 1994). All subsequent model building was performed using the program *O* (Jones *et al.*, 1991) with *SIGMAA*-weighted (Read, 1986) $2mF_{\rm o} - DF_{\rm c}$ and $mF_{\rm o} - DF_{\rm c}$ density maps. Refinements were performed with *CNS* using the maximum-likelihood target function against data from 24.7 to 2.0 Å using a bulk-solvent model and anisotropic *B*-factor correction. During all refinement steps, 5% of the data were set side for cross-validation and refinement steps were accepted if they produced a lowering of $R_{\rm free}$. Water molecules were picked among spherical peaks of 1.2σ in the $2mF_{\rm o} - DF_{\rm c}$ maps and were analyzed for hydrogen-bonding interactions with the protein or other water molecules. The temperature factors were refined for every atom, but restrained to the temperature factors of neighbouring atoms. Information relating to the refinement process can be found in Table 1.

2.5. Preparation of figures

Figs. 4 and 6 were produced using *PyMOL* (W. L. DeLano, DeLano Scientific, San Carlos, CA, USA), while *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997) were used for Fig. 5. Assignments of secondary-structural elements used in Figs. 4 and 7 were made with the program *DSSP* (Kabsch & Sander, 1983). *WebLab ViewerLight* (Molecular Simulations Inc.) was used for preparation of Figs. 8 and 9 and the multi-sequence alignment (Fig. 7) was made with the web-based program *STRAP* (Gille *et al.*, 2003).

3. Results

3.1. Refolding and purification of rVes v 2

Expression of rVes v 2 was demonstrated in *Escherichia coli* strain BL21. It was found to be sequestered into inclusion bodies that were solubilized in 6 *M* guanidine hydrochloride or 8 *M* urea, but not 2 *M* urea. After *in vitro* refolding and purification, enzymatically active rVes v 2 was obtained. The specific activity of the purified sample (protein isolated from two different batches) was 361 000 units mg^{-1}, which was comparable to the natural yellow jacket hyaluronidase nVes v 2. Comparison of rVes v 2 with nVes v 2 and bee venom hyaluronidase was performed by SDS–PAGE analysis (Fig. 1). The SDS–PAGE analysis showed that rVes v 2 has a lower molecular weight (∼6 kDa) than nVes v 2 and natural

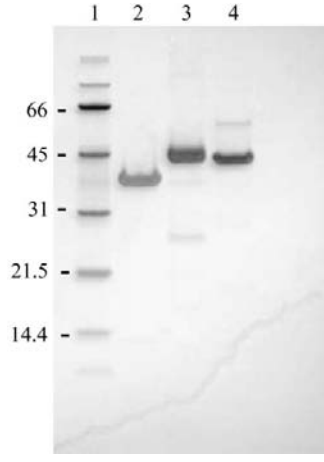

Figure 1
SDS–PAGE analysis of rVes v 2, nVes v 2 and bee hyaluronidase. Lane 1, molecular-weight markers (kDa); lane 2, 2.0 µg rVes v 2; lane 3, 2.0 µg nVes v 2; lane 4, 2.0 µg honeybee hyaluronidase.

Api m 2. The difference in molecular weight is caused by glycosylation of the nVes v 2 as described below.

3.2. Mass-spectrometric analysis of rVes v 2 and nVes v 2

The analysis of rVes v 2 showed a major peak at *m*/*z* 38 942 (data not shown). The observed value was found to be nine mass units above the theoretical value, but further sequence analysis by amino-terminal sequencing and mass fingerprinting (described below) did not reveal the nature of this difference. MALDI–TOF MS experiments on nVes v 2 revealed a broad peak in the *m*/*z* range 41 455.64–43 703.10 (data not shown) and the broad mass spectrum showed a spacing between peaks which is characteristic of glycosylation.

3.3. Proteolytic digestions and sequence analyses of rVes v 2

Reduced and alkylated rVes v 2 was first digested with endoproteinase Lys-C and endoproteinase Asp-N. The resulting peptides from the Lys-C treatment were subjected to mass fingerprinting analysis by MALDI–TOF MS and identified the protein as hyaluronidase (*V. vulgaris*), allergen Ves v 2 (Swiss-Prot code P49370) based on 64% sequence coverage. An amino-terminal fragment SERPK with *m*/*z* 616.34 was not detected from the Lys-C digest and therefore rVes v 2 was digested with endoproteinase Asp-N. Enzymatic digestion with Asp-N gave a peptide at *m*/*z* 2872.42 which was identified as the intact N-terminus (residues 1–24) of rVes v 2. This result was verified by amino-terminal sequencing, which gave a sequence of SERPKRV. Enzymatic digestion with trypsin was used to assess the positions of disulfide bridges in rVes v 2. The disulfide bridge between amino acids 185 and 197 was confirmed, but the disulfide bridge between amino acids 19 and 308 was not detected. However, both bridges were clearly seen in the crystal structure (see below). The tryptic fingerprint also verified the C-terminal sequence (*m*/*z* 2412.34) of rVes v 2 (Fig. 2*b*). The amino-acid sequence coverage of rVes v 2 obtained from the enzymatic digestions is shown in Fig. 3.

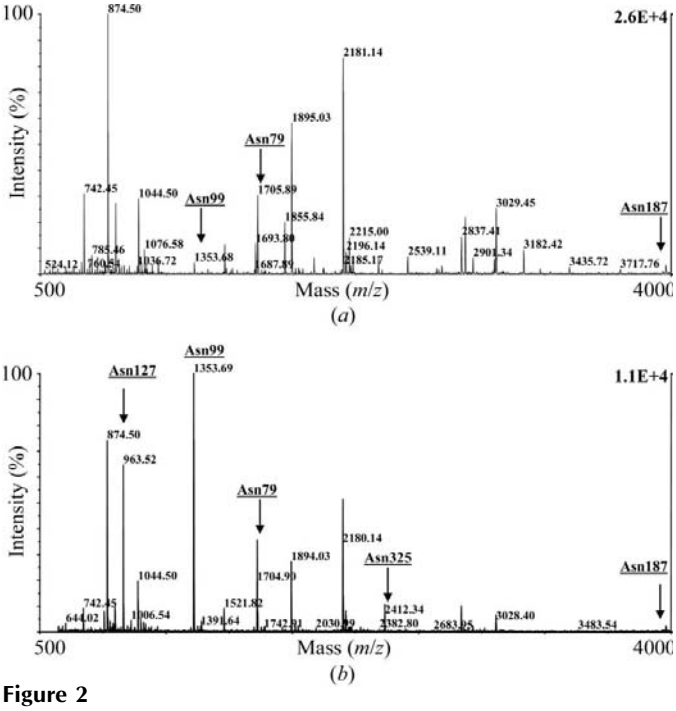

Figure 2
Mass-spectral comparison of trypsin-digested (*a*) nVes v 2 and (*b*) rVes v 2. Signals corresponding to detected non-glycosylated peptides in nVes v 2 and rVes v 2 are labelled and marked with an arrow.

3.4. Investigation of N-glycosylation sites in nVes v 2

Identification of the potential N-glycosylation sites in nVes v 2 was assessed by enzymatic digestion with trypsin. The first mass fingerprinting analysis (76% sequence coverage) again indicated hyaluronidase (*V. vulgaris*), allergen Ves v 2. In addition, *m*/*z* values corresponding to three out of five putative glycosylation sites at Asn79, Asn99 and Asn187 were observed (Fig. 2*a*). Peptides corresponding to N-glycosylation at Asn127 (*m*/*z* 963.53) and Asn325 (*m*/*z* 2256.25) were not detected by the mass fingerprinting for nVes v 2, whereas peptides covering all putative sites were found in non-glycosylated rVes v 2 (Fig. 2*b*).

To further investigate and to verify the N-glycosylation sites in nVes v 2, trypsin-digested nVes v 2 peptides were subjected to reversed-phase chromatography and separated. The intact peptides were collected in fractions and identified by MALDI–TOF and MALDI–TOF/TOF MS. Analysis of nVes v 2 peptides by MALDI–TOF MS revealed four peptides with a spectrum characteristic of glycosylation (data not shown). The glycans were funther analyzed with *N*-glycosidase and *N*-glycosidase A. It was found that the carbohydrate fragments were in the 800–1500 g mol^{-1} range. Enzymatic cleavage of the glycopeptides by *N*-glycosidase revealed peptides corresponding to glycosylation at Asn79 (*m*/*z* 1704.90) and Asn127 (*m*/*z* 1341.76 and *m*/*z* 963.53; Fig. 4). Enzymatic cleavage by *N*-glycosidase A revealed a peptide corresponding to N-glycosylation at Asn99 (*m*/*z* 1997.02). The identity of this peptide was verified by MS/MS. The fact that the glycan can be removed using *N*-glycosidase A and not by *N*-glycosidase suggests that the oligosaccharide bound to Asn99 contains α-

1	**SERPKRVFNI YWNVPTFMCH QYDLYFDEVT NFNIKRNSK**D DFQG**DKIAIF**	50
51	**YDPGEFPALL SLKDGKYKKR NGGVPQEGNI TIHLQKFIEN LDKIYPNRNF**	100
101	**SGIGVIDFER WRPIFRQNWG NMK**IHK**NFSI DLVRNEHPTW NKK**MIELEAS	150
151	KRFEK**YARFF MEETLK**LAKK TRK**QADWGYY GYPYCFNMSP NNLVPECDVT**	200
201	**AMHENDKMSW LFNNQNVLLP SVYVRQELTP DQRIGLVQGR VKEAVRISNN**	250
251	**LK**HSPK**VLSY WWYVYQDETN TFLTETDVKK TFQEIVINGG DGIIIWGSSS**	300
301	**DVNSLSKCKR LQDYLLTVLG PIAINVTEAV N**	331

Figure 3
Amino-acid sequence analysis of rVes v 2 by proteolytic digestions. The verified amino-acid sequence of rVes v 2 is marked in bold and the putative sites for N-glycosylation are underlined.

1,3-fucose linked to the core *N*-acetylglucosamine. Interestingly, MS analyses also demonstrated that nVes v 2 can occur with or without N-glycans at Asn79 and Asn99, whereas Asn127 always seems to be glycosylated. Consequently, this site was investigated in further detail. The glycosylation pattern and the effect of the removal of the glycan from a peptide including Asn127 on the spectra are shown in Figs. 4(*a*) and 4(*b*). From the spacing in Fig. 4(*a*) and the difference between *m*/*z* 1168.7 in Fig. 4(*a*) and *m*/*z* 964.5 in Fig. 4(*b*) it can be concluded that the glycan on Asn127 consists of at least three hexose rings and two *N*-acetylglucosamine moieties.

There are two additional putative glycosylation sites. However, nVes v 2 was found not to be glycosylated at Asn187 and a peptide with an *m*/*z* value corresponding to a putative glycosylation site at Asn325 was not detected.

3.5. Structure of rVes v 2

The three-dimensional structure of rVes v 2 was solved by molecular replacement using the structure of bee hyaluronidase (PDB code 1fcq) as the search model. It was subsequently refined at 2.0 Å resolution (Table 1). The structure includes 324 amino-acid residues, a sulfate ion, one MES molecule and 334 water molecules. Weak electron density was observed at the N- and C-termini. The positions of residues 1–5 and 330–331 could not be identified, and residues 6–7 and 328–329 have high *B* factors. Furthermore, some atoms in nine side chains of surface-exposed residues displayed high *B* factors ($B > 50$ Å). Met144 had additional electron density connected to S^{δ}. This was modelled as a methionine sulfoxide oxidation product and cannot explain the mass difference found by MS. The final *R* factor is 18.6%, with an $R_{\rm free}$ of 21.5%. The program *PROCHECK* (Laskowski *et al.*, 1993) was used to analyze the structure with respect to the stereochemistry of the model. In the Ramachandran plot, 88.3% of non-glycine residues were found to be in the most favoured regions, 11.7% in the additional allowed regions and none in the generously allowed or disallowed regions.

A comparison of the structure of rVes v 2 with known protein structures using the *DALI* server (Holm & Sander, 1995) showed that the bee hyaluronidase Api m 2 had the highest similarity, as expected. A total of 312 C^{α} atoms could be superimposed with an r.m.s. deviation of 1.0 Å and a *Z* score of 45.8. There is a major drop in similarity to the second protein (β-amylase; PDB code 1byb), where only 250 C^{α} atoms could be superimposed with an r.m.s deviation of 4.2 Å and a *Z* score of 12.6. The proteins further down the list (numbers 3–10) are only slightly less similar to Ves v 2 than β-amylase. All of them are enzymes with a $(\beta/\alpha)_8$-barrel structure belonging to various glycoside hydrolase families.

The single polypeptide chain (residues 6–329) is folded into a tertiary structure with a central $(\beta/\alpha)_7$ core [alternating parallel β-strands (e) and α-helices (h), see Fig. 5] related to the classical $(\beta/\alpha)_8$ TIM barrel, but with one strand–helix segment missing. Furthermore, the residues corresponding to h1 are not folded into a regular α-helix. The N-terminus seems to be flexible since the first five amino-acid residues of the sequence could not be located and residues Arg6 and Val7 have high *B* factors. The first strand (e1) of the barrel is found from Asn9 to Asn13 followed by a long loop (corresponding to the missing helix h1) to the second strand of the barrel

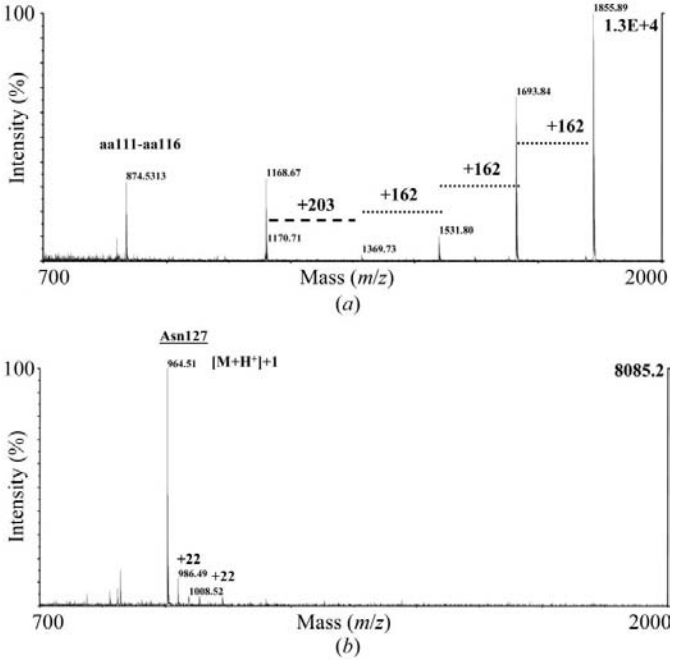

Figure 4
MALDI–TOF MS analyses of the N-glycosidic peptide from nVes v 2. Spectra of the glycopeptide, identified as amino acids 127–134 before (*a*) and after (*b*) enzymatic treatment with *N*-glycosidase. The masses referring to monosaccharides are included.

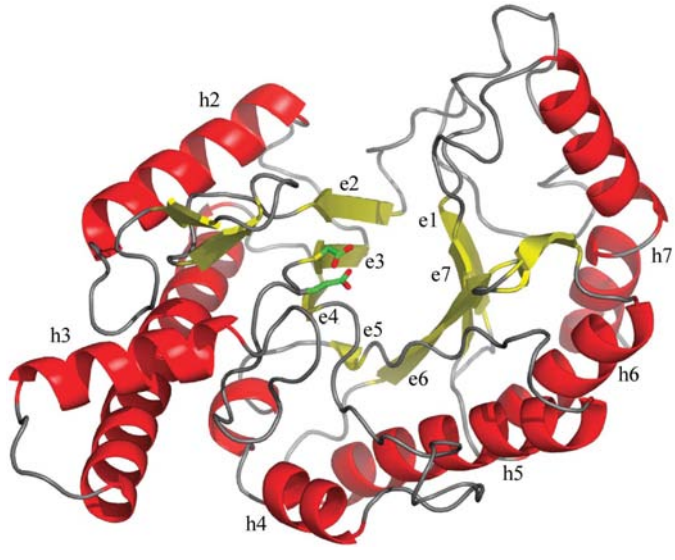

Figure 5
Overall fold of rVes v 2, including active-site residues Asp107 and Glu109. β-sheets (e1–e7) and α-helices (h2–h7) of the $(\beta/\alpha)_7$ core are labelled. The assignments of the secondary-structural elements (e for extended β-strand and h for helix) in the polypeptide chain were made with the program *DSSP* Kabsch & Sander, 1983).

(Ile47–Tyr51). Here, a long loop with a small antiparallel β-sheet (residues Ala58–Ser61 and Tyr67–Arg70) connects to the h2 helix (residues Ile80–Ile94) of the barrel. After strand e3 (residues Ile103–Asp107), a helix (residues Ile124–Glu136) is found before a sharp turn leads into the very long helix h3 (residues Lys142–Thr171). The fourth strand (e4) comprises residues Asp176–Tyr179, while the fourth helix (h4) is actually two helices (Val199–Asp206 and Ser209–212) with a small kink in the middle. Next, the very short e5 strand is found (residues Val217–Leu218) followed by helix h5 (residues Pro230–Asn250). In the loop region after strand e6 (residues Lys256–Trp262), a small β-hairpin (residues Val264–Tyr265 and Glu268–Phe272) is located. Here, only a single hydrogen bond between the two strands is found, allowing a quite twisted sheet. The sixth helix h6 (residues Glu275–Ile287) is followed by strand e7 (residues Gly292–Trp296). Lastly, helix h7 (residues Leu305–Thr317 and Leu319–Glu328) is located very close to the C-terminus. The N- and C-termini are in close proximity and like the N-terminus, the C-terminus seems to be flexible. The two last residues in the sequence (Val330 and Asn331) could not be located and Glu328 and Ala329 have high *B* factors. It should be noted that the loops connecting strands in the barrel (e1–e7) to the helices of the barrel (h1–h7) are much longer on average than those connecting helices to strands. This is a feature also found in many $(\beta/\alpha)_8$-barrel enzymes. As a consequence, all the additional structural elements found in the loops are on the same surface of the protein flanking the active-site cleft. The enzyme is further stabilized by two disulfide bonds (Cys19–Cys308 and Cys185–Cys197) that were easily identified in the electron-density map.

The active site was easily identified by comparison with the structure of the GlcA-GlcNAc-GlcA-GlcNAc complex of the bee allergen (PDB code 1fcv; Markovic Housley *et al.*, 2000). In the active site of Ves v 2, a MES molecule and a sulfate ion are bound. The MES molecule is in close contact with the catalytic residue Glu109 and is also close to the conserved Asp107 (Fig. 6), whereas the sulfate ion forms a hydrogen bond to Arg110. Other resi-

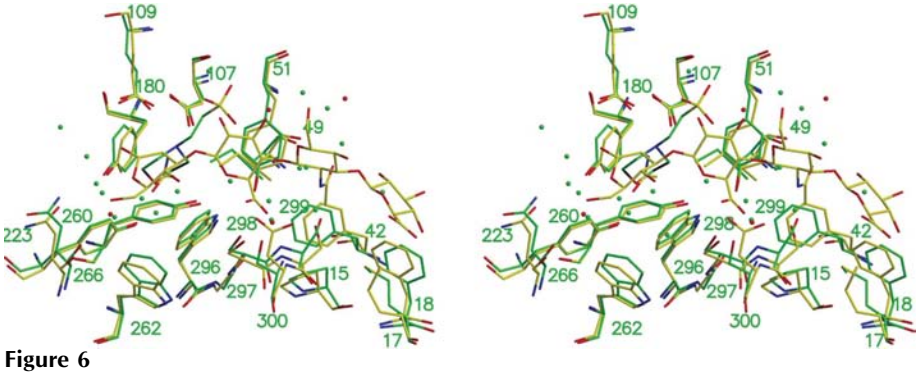

Figure 6
Stereo picture of the active-site cleft in rVes v 2 (green) and Api m 2 (coloured by atom type) with numbers indicating the Ves v 2 amino-acid sequence. The carbohydrate is from the Api m 2 complex (Markovic Housley *et al.*, 2000).

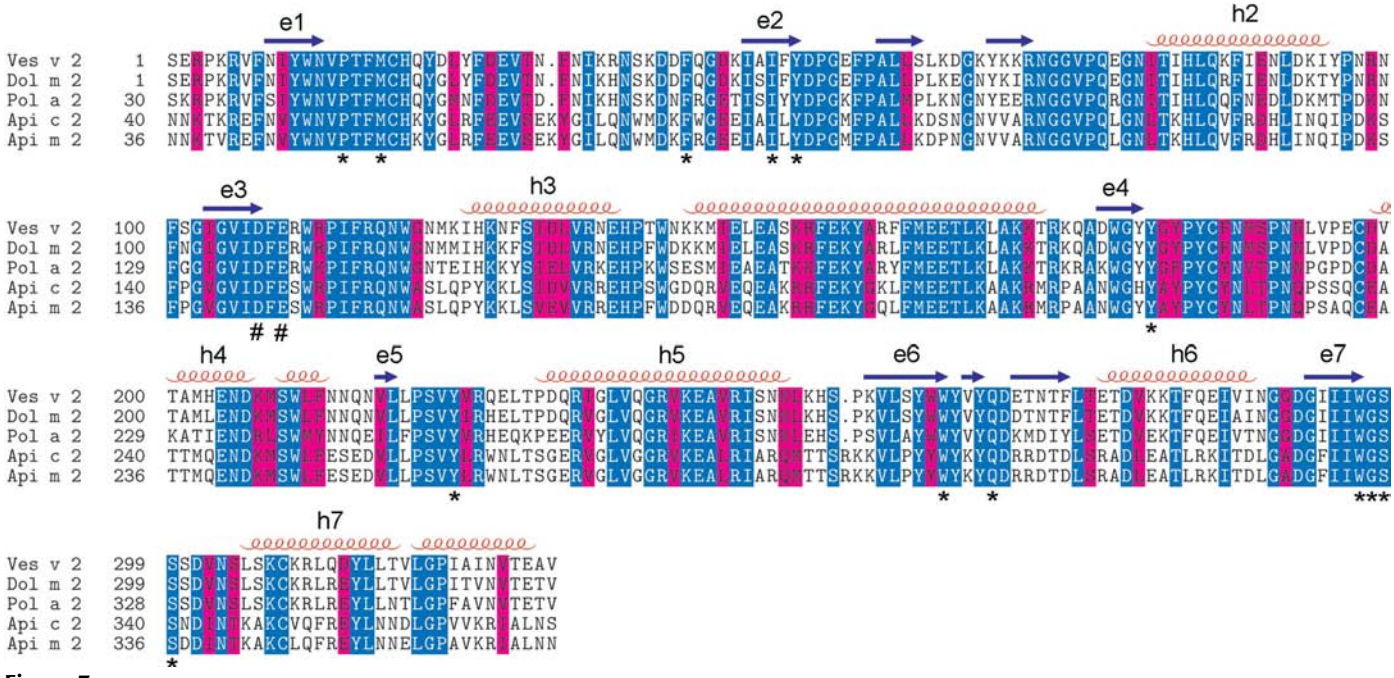

Figure 7
Amino-acid alignment of Ves v 2, Dol m 2, Pol a 2 (major group 2 allergen from *Polistes annularis*), Api c 2 (major group 2 allergen from *A. cerana*) and Api m 2. The annotation of the secondary-structural elements (e for extended β-strand and h for helix) in the $(\beta/\alpha)_7$ core (labelled e1–e7 and h2–h7) is based on the assignments made with the program *DSSP* (Kabsch & Sander, 1983). The important carboxylates Asp107 and Glu109 in the active site are marked by a # and other conserved residues in the active site are marked by *. Areas with conserved amino-acid residues are coloured blue, while similar residues are coloured magenta.

dues with contacts to the MES molecule are Tyr51, Tyr180 and Trp296.

4. Discussion

The ability to produce recombinant allergens with correct amino-acid sequence and three-dimensional structure is of major importance for their potential use as diagnostic agents and as active ingredients in vaccines. Here, we demonstrate the expression of recombinant Ves v 2 in *E. coli* that is fully active after refolding and can be crystallized. The structure determined here showed that the fold of rVes v 2 is very similar to the corresponding honeybee allergen Api m 2. For rVes v 2, the presence of both termini were confirmed by MS.

4.1. Active site of Ves v 2

The catalytic residue Glu109 is found at the same positions as Glu113 in the Api m 2 structure. Also, the type and position of the amino acids in the active-site cleft are very similar (Fig. 6). Within 4 Å of the superimposed tetrasaccharide from the Api m 2 complex, all amino acids are identical and at the same spatial positions in Ves v 2 and Api m 2 (Fig. 6). These amino acids are also conserved amongst the insect sequences included in the alignment shown in Fig. 7. The fact that the conformations of all the interacting residues are essentially identical and independent of the presence of the substrate emphasizes the importance of these four subsites as the basis for recognition. In further support of the importance of these sites, a number of water molecules from the present Ves v 2 apo structure mimic some of the tetrasaccharide hydroxyl groups. The MES molecule located in the Ves v 2 structure is also bound in this region. The N and the O atoms of the morpholine ring are found very close to the positions of the N and O atoms of the NAc unit of Glc of the tetrasaccharide. One of the O atoms of the sulfonic acid group is also seen to occupy a substrate hydroxyl position. This emphasizes the importance of ligand interactions at the catalytic nucleophile position.

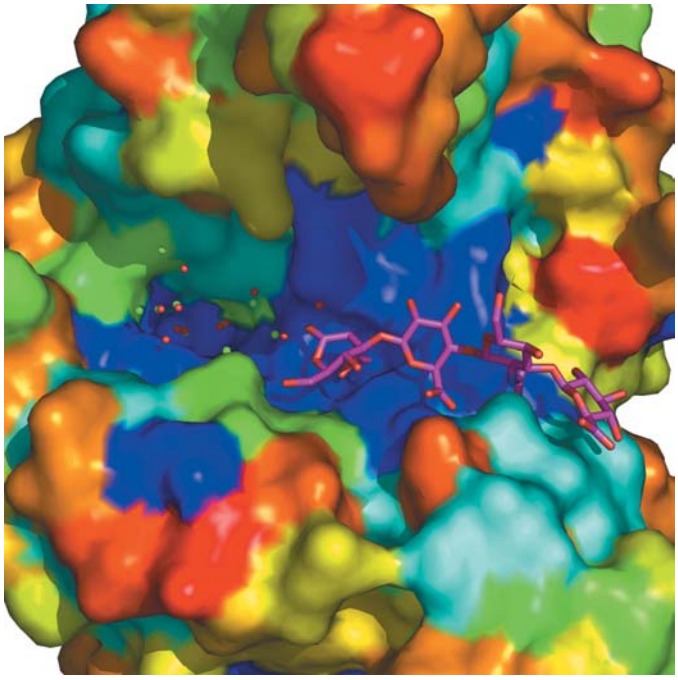

Figure 8
Solvent-accessible surface of rVes v 2. The surface is coloured by sequence entropy (dark blue means fully conserved in selected insect and mammalian enzymes, while red means not conserved). Also, conserved water molecules are shown (coloured green in Ves v 2 and red in Api m 2) as well as the tetrasaccharide (shown in magenta and red) from the Api m 2 complex (Markovic Housley *et al.*, 2000).

Colouring this region of the solvent-accessible surface of rVes v 2 according to the sequence entropy (Hamelryck & Manderick, 2003) amongst enzymes of origin spanning from insects to mammals reveals at least six fully conserved sites with potential for binding a saccharide moiety (Fig. 8). Two of these potential sites are on one side of the catalytic residue Glu109, while the other four sites are on the opposite side. The position of the four sites corresponds to the location of the tetrasaccharide identified in the Api m 2 complex structure. It seems likely that the two other conserved sites are also used in the binding of the substrate. The two sites are located in a cavity and nine water molecules are found at the same positions (within experimental error) in the Ves v 2 and the high-resolution uncomplexed Api m 2 (PDB code 1fcq) structures. The nine water molecules are in positions that could very well be mimicking the hydroxyl groups of a saccharide moiety, suggesting that the cavity is a binding pocket for two saccharide moieties. This could implicate that the enzyme acts in an exo-fashion by removing disaccharides. However, this has to be investigated further.

4.2. N-Glycosylation of natural Ves v 2

According to the Swiss-Prot database, Ves v 2 has four potential glycosylation sites (Asn79, Asn99, Asn127 and Asn325), while the program *GPMAW* predicts five (Asn79, Asn99, Asn127, Asn187 and Asn325). The primary analysis using MS fingerprinting on nVes v 2 strongly suggested glycosylation at Asn127 and Asn325. However, as described in §3.4, the N-glycosylation in nVes v 2 is more complex. Previous studies have shown that venom-allergic individuals can develop IgE antibodies against specific carbohydrate structures. The N-glycan structures containing α-1,3-linked fucose residues at the glycan core have been suggested to be involved in this type of IgE-mediated cross-reactivity and thus are considered to be a major cause of *in vitro* double positivity in honeybee and vespid-allergic individuals (Hemmer *et al.*, 2001, 2004). Interestingly, in the present study only one site at Asn99 was found to carry this type of N-glycan. Moreover, while glycosylation at Asn79 and/or Asn99 was found to be variable, Asn127, which carries a non-fucosylated N-glycan structure, always seems to be occupied. The enzymatic cleavage analyses suggest that N-glycosylation patterns within nVes v 2 are heterogeneous, containing non- and core-fucosylated glycan structures, and that more detailed analyses of the N-glycan structures are needed to explain the *in vitro* detected cross-reactivity between honeybee and vespid aller-

gens. Glycosylation and/or other post-translational modification at the C-terminal part of the molecule including the putative glycosylation site Asn325 cannot be excluded, but was not confirmed.

According to the Swiss-Prot database, Api m 2 has two putative glycosylation sites (Asn115 and Asn263). Based on the structural analysis, the only common site is Asn79 in Ves v 2, which corresponds to Asn115 in Api m 2. Recombinant Api m 2 was expressed in a baculovirus expression system and in the structure of rApi m 2 weak electron density is seen extending from Asn115 (corresponding to Asn83 in the PDB sequence; Markovic Housley *et al.*, 2000). However, as far as we know, the sugar moieties of rApi m 2 have not yet been modelled and/or characterized. In contrast, it has been shown that natural Api m 2 is glycosylated at Asn263 with a Manα6(Manα3)Manβ4GlcNAcβ4(Fucα3)GlcNAcβ-Asn type oligosaccharide containing six or seven sugar rings (Hemmer *et al.*, 2004; Kolarich & Altmann, 2000). A corresponding glycosylation site is not found in Ves v 2. Hence, a common spatial position of glycosylation site for Api m 2 (Asn115) and Ves v 2 (Asn79) seems to be present, but the locations of the other glycosylation sites within these allergens are quite different.

4.3. Implications for allergen cross-reactivity

As described above, the fold of the allergen and the arrangement of the active site is highly conserved between the rVes v 2 and Api m 2 structures. There are, however, significant structural differences between the surfaces of the proteins in charge distribution as well as in surface topology. Outside the active-site clefts, the two allergens have different electrostatic potentials (Fig. 9). In some areas, the positive-charged residues have been replaced with negative-charged residues and *vice versa*. As charges on the two counterparts in the allergen–antibody interaction are highly important, IgE cross-reactivity between Ves v 2 and Api m 2 is not likely. Also, in terms of surface topology rVes v 2 and Api m 2 (55% identity) are very different, since the sequence variation is primarily confined to the surface area distant from the active site (Fig. 10). This picture is different when comparing Ves v 2 and the major group 2 allergen from *Dolichovespula maculata* (Dol m 2; 92% identity), where large areas on the surface are also conserved (Fig. 10). Based on the structural analyses, we conclude that the IgE-mediated cross-reactivity between Ves v 2 and Dol m 2 should be much stronger than between Ves v 2 and Api m 2. Another feature that has to be considered is the glycosylation pattern of Ves v 2 and Api m 2. Here, we have shown that Ves v 2 and Api m 2 could have a common glycosylation site. This could be involved in IgE epitopes, giving rise to cross-reactivity. However, it was also shown that this site is not occupied in all nVes v 2 molecules and that in general the glycosylation pattern(s), *e.g.* no core fucosylation at Asn79, are different. In contrast, all asparagine residues assigned as putative glycosylation sites in Ves v 2 are also found in Dol m 2.

Hence, this comparison of allergen surfaces suggests that Ves v 2 does not have epitopes that could harbour IgE-mediated cross-reactivity between this allergen from wasps or bees, whereas cross-reactivity towards different wasps are

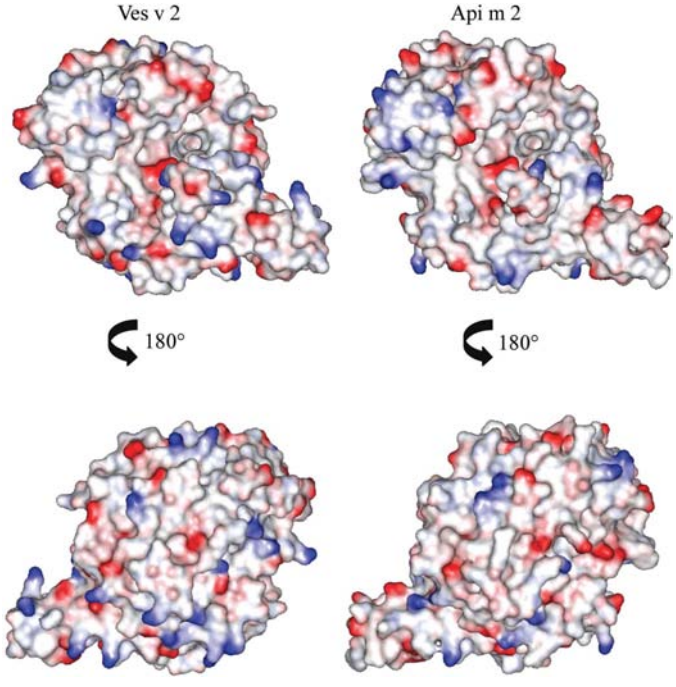

Figure 9
Surface of rVes v 2 and Api m 2 coloured by electrostatic potential (blue, positive; red, negative).

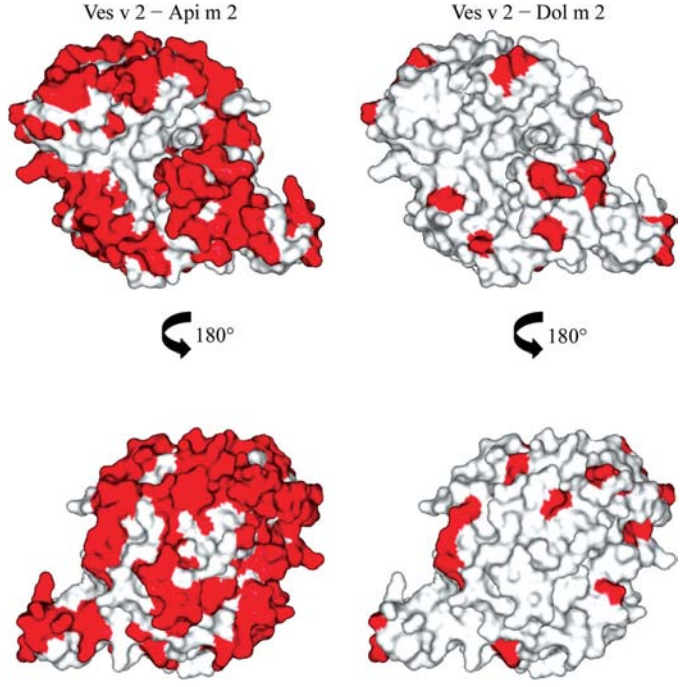

Figure 10
Solvent-accessible surface of rVes v 2 coloured by amino-acid identity to Api m 2 and Dol m 2. Identical residues are coloured white, whereas differences are shown in red.

likely to be harboured. This does not, however, address human IgE-mediated cross-reactivity between venoms from wasps and bees in general, since similar analysis of the other major allergens present in the venoms are required.

We wish to thank Dr Kudzai Mutenda from the University of Southern Denmark for her help in the MS/MS sequencing. We would also like to acknowledge ESRF and MAX-Lab for provision of synchrotron-radiation facilities. This work was supported by the EU project 'Therapeutic Recombinant Allergens from Structural Allergology' (TRAFSA, QLK-3-CT-1999-00620) and the Danish Synchrotron User Center (DANSYNC).

References

Arming, S., Strobl, B., Wechselberger, C. & Kreil, G. (1997). *Eur. J. Biochem.* **247**, 810–814.
Brown, S. G. A., Wiese, M. D., Blackman, K. E. & Heddle, R. J. (2003). *Lancet*, **361**, 1001–1006.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davies, G. & Henrissat, B. (1995). *Structure*, **3**, 853–859.
Gille, C., Lorenzen, S., Michalsky, E. & Frommel, C. (2003). *Bioinformatics*, **19**, 2489–2491.
Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R. & Roepstorff, P. (1999). *J. Mass Spectrom.* **34**, 105–116.
Golden, D. B. K., Marsh, D. G., Kageysobotka, A., Freidhoff, L., Szklo, M., Valentine, M. D. & Lichtenstein, L. M. (1989). *J. Am. Med. Assoc.* **262**, 240–244.
Hamelryck, T. & Manderick, B. (2003). *Bioinformatics*, **19**, 2308–2310.
Hemmer, W., Focke, M., Kolarich, D., Dalik, I., Gotz, M. & Jarisch, R. (2004). *Clin. Exp. Allergy*, **34**, 460–469.
Hemmer, W., Focke, M., Kolarich, D., Wilson, I. B. H., Altmann, F., Wohrl, S., Gotz, M. & Jarisch, R. (2001). *J. Allergy Clin. Immunol.* **108**, 1045–1052.
Henriksen, A., King, T. P., Mirza, O., Monsalve, R. I., Meno, K., Ipsen, H., Larsen, J. N., Gajhede, M. & Spangfort, M. D. (2001). *Proteins*, **45**, 438–448.
Henrissat, B. (1991). *Biochem. J.* **280**, 309–316.
Hoffman, D. R. & Jacobson, R. S. (1984). *Ann. Allergy*, **52**, 276–278.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Holowka, D. & Baird, B. (1996). *Ann. Rev. Biophys. Biomol. Struct.* **25**, 79–112.
Hunt, K. J., Valentine, M. D., Sobotka, A. K., Benton, A. W., Amodio, F. J. & Lichtenstein, L. M. (1978). *N. Engl. J. Med.* **299**, 157–161.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
King, T. P., Alagon, A. C., Kuan, J., Sobotka, A. K. & Lichtenstein, L. M. (1983). *Mol. Immunol.* **20**, 297–308.
King, T. P., Lu, G., Gonzalez, M., Qian, N. F. & Soldatova, L. (1996). *J. Allergy Clin. Immunol.* **98**, 588–600.
Kolarich, D. & Altmann, F. (2000). *Anal. Biochem.* **285**, 64–75.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kreil, G. (1995). *Protein Sci.* **4**, 1666–1669.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Linker, A. (1974). *Methods of Enzymatic Analysis*, edited by H. U. Bergmeyer, Vol. 2, pp. 944–948. New York: Academic Press.
McPherson, A. (1992). *J. Cryst. Growth*, **122**, 161–167.
Markovic Housley, Z., Miglierini, G., Soldatova, L., Rizkallah, P. J., Muller, U. & Schirmer, T. (2000). *Structure*, **8**, 1025–1035.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Muller, U. & Mosbech, H. (1993). *Allergy*, **48**, 37–46.
Muller, U. R. (1990). *Insect Sting Allergy: Clinical Picture, Diagnosis and Treatment*, pp. 75–91. Stuttgart: Gustav Fischer Verlag.
Muller, U. R. (2002). *Allergy*, **57**, 570–576.
Nall, T. M. (1985). *J. Allergy Clin. Immunol.* **75**, 207.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Richman, P. G. & Baer, H. (1980). *Anal. Biochem.* **109**, 376–381.
Seppala, U., Hagglund, P., Wurtzen, P. A., Ipsen, H., Thorsted, P., Lenhard, T., Roepstorff, P. & Spangfort, M. D. (2005). *J. Biol. Chem.* **280**, 3208–3216.
Settipan, G. A. & Boyd, G. K. (1970). *Acta Allergol.* **25**, 286–291.
Stern, R. (2003). *Glycobiology*, **13**, 105R–115R.
Stuckey, M., Cobain, T., Sears, M., Cheney, J. & Dawkins, R. L. (1982). *Lancet*, **2**, 41.
Tews, I., Terwisscha van Scheltinga, A. C., Perrakis, A., Wilson, K. S. & Dijkstra, B. W. (1997). *J. Am. Chem. Soc.* **119**, 7954–7959.