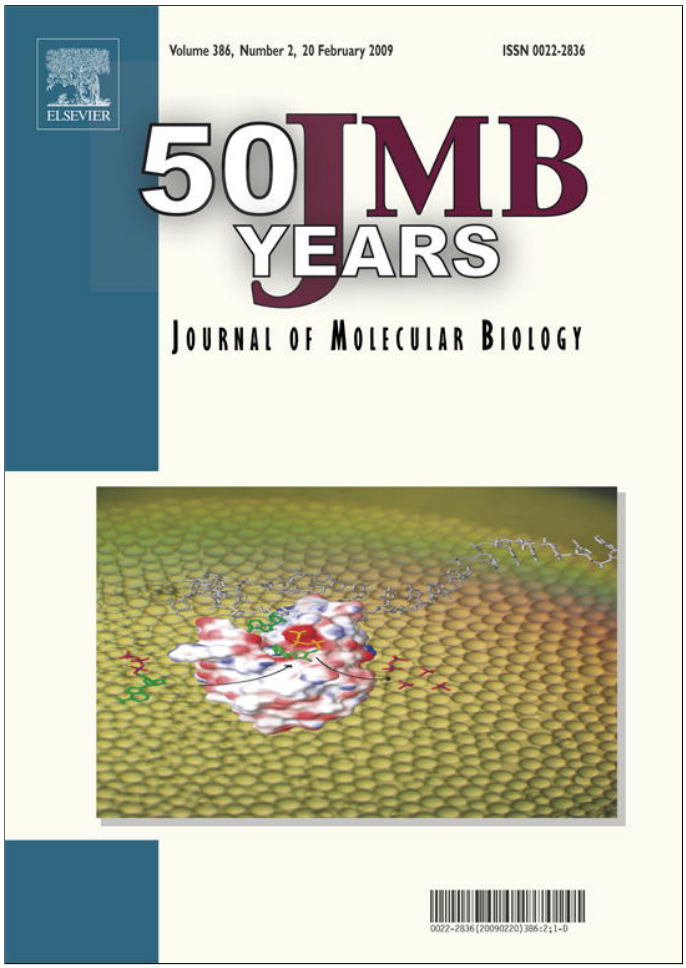

doi:10.1016/j.jmb.2008.12.049 J. Mol. Biol. (2009) 386, 520–530

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structures of Mite Allergens Der f 1 and Der p 1 Reveal Differences in Surface-Exposed Residues that May Influence Antibody Binding

Maksymilian Chruszcz[1], Martin D. Chapman[2], Lisa D. Vailes[2], Enrico A. Stura[3], Jean-Marie Saint-Remy[4], Wladek Minor[1]* and Anna Pomés[2]*

[1]*Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA*

[2]*INDOOR Biotechnologies, Inc., 1216 Harris Street, Charlottesville, VA 22903, USA*

[3]*Département d'Ingénierie et d'Etudes des Protéines, Commissariat à l'Energie Atomique Saclay, Gif-sur-Yvette, France*

[4]*Center for Molecular and Vascular Biology, University of Leuven, Leuven, Belgium*

Received 2 October 2008; received in revised form 17 December 2008; accepted 18 December 2008
Available online 30 December 2008

Edited by I. Wilson

The group 1 mite allergens Der f 1 and Der p 1 are potent allergens excreted by *Dermatophagoides farinae* and *Dermatophagoides pteronyssinus*, respectively. The human immunoglobulin E antibody responses to the group 1 allergens show more cross-reactivity than the murine immunoglobulin G antibody responses, which are largely species specific. Here, we report the crystal structure of the mature form of Der f 1, which was isolated from its natural source, and a new high-resolution structure of mature recombinant Der p 1. Unlike Der p 1, Der f 1 is monomeric both in the crystalline state and in solution. Moreover, no metal binding is observed in the structure of Der f 1 despite the fact that all amino acids involved in Ca^{2+} binding in Der p 1 are completely conserved in Der f 1. Although Der p 1 and Der f 1 share an extensive sequence identity, comparison of the crystal structures of both allergens revealed structural features that could explain the differences in murine IgG and human IgE antibody responses to these allergens. There are structural differences between Der f 1 and Der p 1 that are unevenly distributed on the allergens' surfaces. This uneven spatial arrangement of conserved *versus* altered residues could explain both the specificity and cross-reactivity of antibodies against Der f 1 and Der p 1.

Keywords: mite allergy; Der f 1; Der p 1; antibody; asthma

Introduction

Inhalation of house dust mite allergens is one of the most important risk factors associated with the development of allergic disease, including rhinitis, atopic dermatitis, and asthma. The two principal mite species *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae* belong to the *Pyroglyphidae* family and excrete allergens, including Der p 1 and Der f 1 (group 1), respectively, which are carried to the lungs in mite fecal particles. Humidity is the most important limiting factor for mite growth, and deserts or areas of high attitude have low levels of mites.[1] *D. pteronyssinus* is broadly distributed in houses in Western Europe, Japan, North America, New Zealand, and Australia. In continental regions of Europe and most of the United States, both species frequently coexist, but *D. farinae* can tolerate drier environmental conditions.[2] Most mite allergic patients (>80%) have immunoglobulin E (IgE) antibodies against group 1 mite allergens.[3,4] Der p 1 and Der f 1 are cysteine proteases of the clan CA[5] and family C1. They share both 81% sequence identity and antigenic cross-reactivity.

**Corresponding authors.* E-mail addresses: wladek@iwonka.med.virginia.edu; apomes@inbio.com.

Abbreviations used: IgE, immunoglobulin E; mAb, monoclonal antibody; CDR, complementarity-determining region; PBS, phosphate-buffered saline; IgG, immunoglobulin G.

Despite the high amino acid sequence identity between group 1 allergens, most of the monoclonal antibodies (mAbs) raised against either Der p 1 or Der f 1 are species specific (~95%).[6–8] In contrast, the degree of cross-reactivity of human IgE antibody responses to group 1 allergens, although variable, is higher (34%–90%).[8] One murine cross-reacting epitope was identified using anti-Der f 1 mAb 4C1.[7] This mAb inhibited IgE antibody binding to Der p 1 by ~40%, suggesting that the epitopes for 4C1 mAb and human IgE antibodies on Der p 1 overlap. These mAbs provide tools to study the antigenic determinants in group 1 allergens by X-ray crystallography.

The crystal structures of the proenzyme and mature forms of recombinant Der p 1 were recently determined.[9,10] Here, we report the crystal structure of mature natural Der f 1 obtained from mite culture and a new high-resolution structure of recombinant Der p 1. Both allergens are secreted with an N-terminal proregion that is autocatalytically cleaved under acidic conditions upon enzyme maturation. The proregion blocks not only the catalytic activity but also conformational IgE antibody binding epitopes.[11] Reports have indicated that proteolytic activity contributes to allergenicity, mostly in the case of Der p 1. Disruption of tight junctions in lung epithelium and cleavage of receptors (CD23, CD25) favor a Th2 response and induction of release of pro-inflammatory cytokines from bronchial epithelial cells, mast cells, and basophils.[12] These effects may promote IgE antibody synthesis and inflammation on lung epithelium, which could explain why mite allergens are strongly associated with asthma. Although a reduction of skin barrier function by proteolytic activity of Der f 1 has been reported, much less is known about its pro-inflammatory effects.[13]

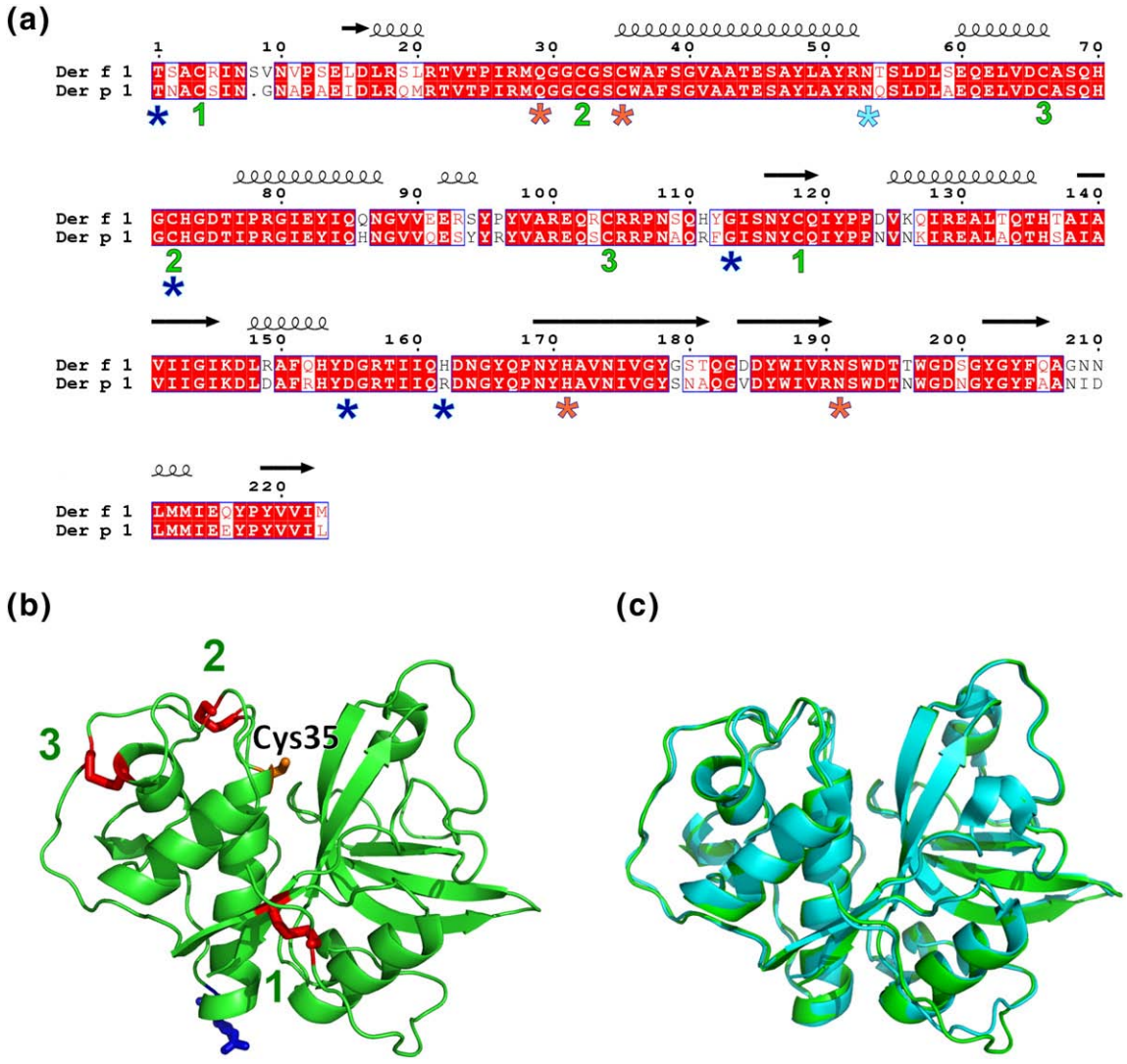

Fig. 1. (a) Sequence alignment of mature Der f 1 and Der p 1. The alignment was done with CLUSTALW,[14] and the figure was prepared using ESPRIPT.[15] Blue stars show residues mutated in different Der f 1 variants. Catalytic residues are marked with orange stars, while disulfide bond-forming cysteines are labeled using green numbers. Light blue star shows the N-glycosylation site for Der f 1. (b) Model of Der f 1 shown in ribbon representation. Cys35 (orange), Asn53 (blue), and cysteines forming disulfide bonds (red) are shown in stick representation. The disulfide bonds are labeled as in (a). (c) Superposition of the crystal structures of Der f 1 (PDB code 3D6S; green) and Der p 1 (PDB code 2AS8; cyan).

Results and Discussion

Overall structure of Der f 1

Der f 1 was crystallized in space group $P4_1$ with three protein molecules (chains A, B, and C) in the asymmetric unit. The protein is monomeric. The overall fold of Der f 1 is characteristic for papain-like cysteine proteases and similar to that observed for Der p 1, as expected from their high sequence identity (Fig. 1). The Der f 1 molecule consists of two globular domains connected by a flexible linker. Residues 1–223 could be traced in the electron density of all protein chains, with the exception of Ala3 from chain C. Superposition (using secondary-structure matching[16] as implemented in COOT[17]) of Der f 1 (chain A) on mature Der p 1 [Protein Data Bank (PDB) code 2AS8; chain A] gave a C^{α} RMSD value of 0.6 Å (over 222 residues) (Fig. 1c), while superposition of Der f 1 and proDer p 1 (PDB code 1XKG) gave a C^{α} RMSD value of 0.5 Å (over 221 residues).

The pattern of disulfide bonds observed in Der f 1 (Cys4–Cys118, Cys32–Cys72, and Cys66–Cys104; numbering of amino acids in the whole text is based on the Der f 1 sequence) is the same as that in Der p 1. Der f 1 shows only 5 polymorphisms (Fig. 1a), compared with the 23 in Der p 1.[18] Analysis of the Der f 1 structure reveals that the amino acid differences between crystallized Der f 1.0101 and other isoforms are located on the surface of the allergen. Almost all mutations, except His162Arg, modify residues identical in Der f 1 and Der p 1. However, the mutation His162Arg (variant 1.0102) increases the surface/sequence similarity between the two allergens. Analysis of the electron density shows that the form of Der f 1 crystallized is that of the most abundant variant (i.e., variant 1.0101).

Der f 1 has one N-glycosylation motif in its mature form. Electron density observed near residue Asn53 of both chain A and chain B was interpreted as an *N*-acetylglucosamine; for chain C, the density was too weak to build a model for the carbohydrate moiety. According to mass spectrometry results (data not shown), the molecular mass of the Der f 1 used for crystallization (25.462 Da) is about 270 Da greater than the calculated molecular mass of mature Der f 1, which corresponds to the approximate mass for one *N*-acetylglucosamine residue. In two of the three Der f 1 molecules in the asymmetric unit (chains A and B), the Asn53 residue is exposed to solvent such that carbohydrate groups fit without disrupting crystal packing. In chain C, the space for a carbohydrate is limited by crystal packing. Thus, spatial constraints may limit the population of molecules that form the crystal to those with shorter carbohydrates. It may also explain why the mean temperature factor of chain C is higher and its corresponding electron density is of lower quality compared with chains A and B. The results presented here, along with those previously published, suggest that different types of short carbohydrates are probably present at the N-glycosylation site in Der f 1. The N-glycosylation site in the prodomain of Der p 1 is absent in Der f 1. The role of

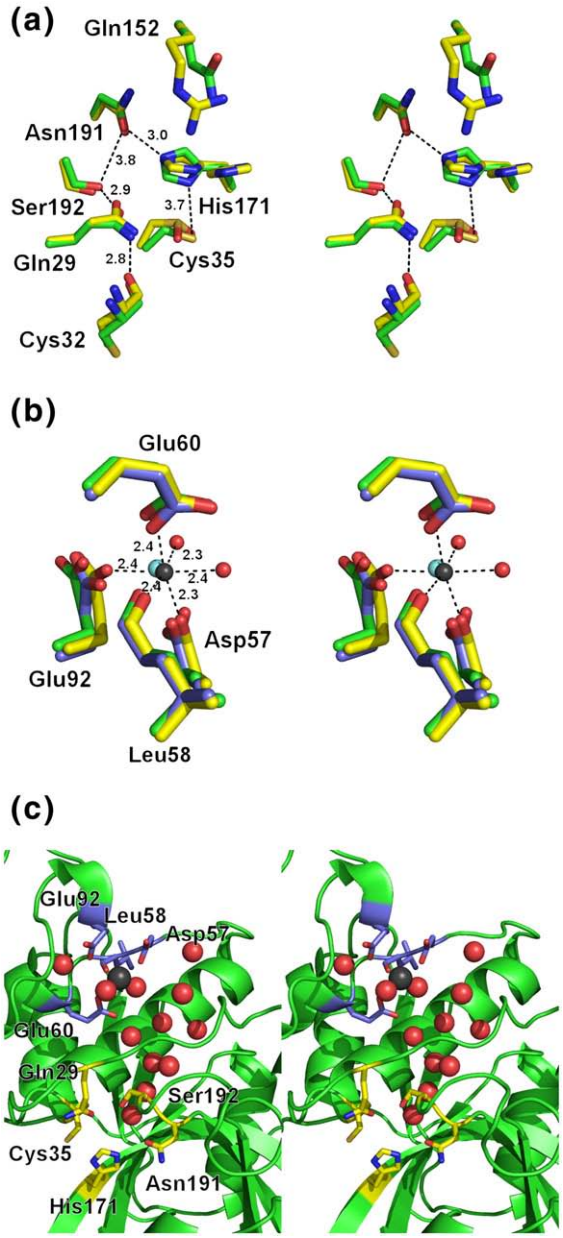

Fig. 2. (a) Superposition of the Der f 1 (green) and Der p 1 (yellow) active sites. Interactions between residues are marked with dashed lines, and interatomic distances (see Table 1) are in Å. (b) Superposition of the putative Der f 1 metal binding site (green) with the metal binding site observed in the proDer p 1 (blue; PDB code 1XKG) and mature Der p 1 (yellow) structures. Residues are numbered according to the Der f 1 sequence. The water molecules coordinating Y^{3+} (cyan sphere) were omitted for the sake of picture clarity. Coordination of Ca^{2+} (black sphere) in the Der p 1 structure (PDB code 3F5V) is shown using dashed lines. (c) Water molecules (red spheres) conserved in the Der f 1 (PDB code 3D6S), proDer p 1 (PDB code 1XKG) and mature Der p 1 (PDB code 3F5V) crystal structures. Black sphere shows the position of the metal ion in the structure of Der p 1.

carbohydrate content is not well understood for either allergen. Different effects on enzyme maturation have been reported,[19–21] but all the studies agree that nonglycosylated mutants appear to have lower solubility and that *N*-glycosylation does not influence antibody binding.[20,22,23]

New crystal form of Der p 1

In addition to the previously reported orthorhombic crystal form[10] of Der p 1, a new crystal form that also contains two molecules in the asymmetric unit (space group *C*2) was obtained. The relative arrangements of the allergen molecules in the asymmetric units of the crystal forms are similar, and the structures superpose with a C^{α} RMSD value of 0.2 Å. The overall structure is the same except for the conformation of a few surface amino acids. The higher-resolution data have allowed us to better identify the metallic divalent cation as Ca^{2+},[24] rather than Mg^{2+} as previously reported (Fig. 2).[7] Moreover, in the new crystal form, the catalytically active Cys35 residue is oxidized. This observation is consistent with the need for a reducing agent to activate the enzyme before measuring its catalytic activity. Oxygen atoms from the sulfinic group interact with $N^{\varepsilon 2}$ from the Gln28 side chain and nitrogen atom from the His170 main chain. Additionally, a significant conformational change of His171 is observed (Fig. 2a). The change of His171 conformation could contribute to the catalytic mechanism of the protease as explained below or, it could be caused by the oxidation of Cys35.

Structural comparison between Der f 1 and Der p 1

Active site

The active site of Der f 1 lies between the two globular domains, as in Der p 1, and other enzymes of clan CA[5] and family C1. The substrate binding and catalytic residues (Gln29, Cys35, His171, and Asn191) lie in a cleft formed by juxtaposition of both domains (Figs. 1b and 2a). Although group 1 mite allergens were purified and cloned over 20 years ago,[1–3] detailed mechanism of their action is still not fully understood. Residues Cys35 and His171, which form a thiolate–imidazolinium ion pair (Table 1), are probably most critical for the enzymatic activity, as this ion pair is conserved in other cysteine proteases.[25,26] The roles of the other amino acids located proximally to Cys35 and His171 have not yet been determined.

The active sites of Der f 1 and Der p 1 are very similar, as are their patterns of substrate specificity.[27,28] Both enzymes prefer to bind small aliphatic residues in position P2, charged residues in position P1, and small hydrophobic or hydrophilic residues in position P1′.[29,30] The most significant amino acid difference near the catalytic site is Gln152 in Der f 1 *versus* Arg151 in the equivalent position of Der p 1. This difference may explain why a plant cystatin, a proteinaceous cysteine protease inhibitor isolated from chestnut, inhibits Der f 1 but not Der p 1.[30] Cystatins block access to the catalytic cleft but do not bind covalently to the cysteine residue from the active center, as do small molecular inhibitors, such as E-64. The proteolytic activity of Der f 1 and Der p 1 has been reported to significantly influence immunological signaling and contribute to allergic response.[12] Their action is especially visible in the lungs,[12,28] and Der f 1 has also been reported to affect skin.[13]

Table 1. Distances between amino acids near the Der f 1 (PDB code 3D6S) and Der p 1 (PDB code 3F5V) active sites

	Der f 1			Der p 1[a]	
	A [Å]	B [Å]	C [Å]	A [Å]	B [Å]
Cys35 $S^{\gamma}\cdots N^{\delta 1}$ His171	3.7	3.6	3.4	3.0/5.6	3.5/5.4
His171 $N^{\varepsilon 2}\cdots O^{\delta 1}$ Asn191	3.0	2.7	2.7	2.8/7.2	2.9/7.1
Gln29 $N^{\varepsilon 2}\cdots$O Cys32	2.8	3.0	2.7	2.8	3.0
Gln29 $O^{\varepsilon 1}\cdots O^{\gamma}$ Ser192	2.9	2.7	2.6	2.7	2.7
Ser192 $O^{\gamma}\cdots H_2O$	2.5	2.9	2.7	2.8	2.8

Amino acids are numbered with respect to the Der f 1 sequence.
[a] Two values refer to the double conformation of His171.

The active site in the proenzyme forms of Der f 1 and Der p 1 is blocked by the allergen N-terminal prodomain. The prodomains were shown to be competitive inhibitors of the mature forms of the enzymes,[31] but in contrast to other cysteine proteases, Der p 1 was able to completely degrade its propeptide. The nonstandard features of the Der f 1 and Der p 1 prodomains (as compared with other papain-like proteins), namely, their length and lack of well-conserved ERFNIN motifs, [6,9,31] suggest that group 1 mite allergens may form a new C1 subfamily of cysteine proteases.

Metal binding site

Unlike in Der p 1 structures, no metal binding was observed in the structure of Der f 1 (Fig. 2b). If the Der f 1 initially bound a metal ion, it may have been removed from the protein during crystallization as the crystallization solution contained both L-arginine and ammonium sulfate ions at high concentrations. Moreover, during optimization of crystallization conditions, EDTA (ethylenediaminetetraacetic acid) was added to the solution from which the best crystals were obtained. The amino acids involved in metal binding in Der p 1 (Asp57, Leu58, Glu60, and Glu92; Der f 1 numbering) are all conserved in Der f 1. The differences in conformations of the metal binding residues in Der p 1 compared with the equivalent residues in Der f 1, which lacks a metal, are small (Fig. 2b). These results suggest that the presence or absence of metal ions does not influence the overall protein architecture.

The metal ion is located approximately 12 Å from Cys35 and 10 Å from Gln29, and the role of the metal ion in Der p 1 is unknown. Both Mg^{2+} and calcium ions were reported [32] to increase the catalytic activity of papain from *Carica papaya*, and the presence of the metal ions, according to CD measurements, also did

not influence the secondary structure of the enzyme. Surprisingly, the superposition of the Der f 1, proDer p 1, and mature Der p 1 models reveals a large cluster of structurally conserved water molecules (Fig. 2c) being located opposite (relative to the catalytic residues) to the substrate binding site. The water molecules fill a funnel-shaped cavity that has its broader end localized near the metal binding site. The narrow end of the funnel is located near Ser192 which forms a hydrogen bond with the hydroxyl group of the serine. Ser192 also interacts through a hydrogen bond with Gln29 (Table 1). In the case of Der f 1, the water molecule bound near Ser192 is 4.3–4.7 Å apart from Cys35 and 3.8–4.5 Å from His171. It is possible that this water molecule may be used during catalytic reactions, especially during the deacetylation step, in which the cysteine–substrate covalent intermediate is hydrolyzed.[33] Another interesting feature observed in the crystal structure of human cathepsin K is that the water molecule bound to the serine equivalent to Ser192 is positioned almost identically with the corresponding water molecule in Der f 1 and Der p 1. It suggests that the water molecule either assists the serine residue in its enzymatic activity or is important for the structure of the cysteine proteases.

Oligomeric state

We determined, by gel filtration at pH 7.5, Der f 1 to be monomeric in solution. Gel-filtration experiments[11] performed in slightly more acidic conditions (pH 6.5) also indicated that both Der f 1 and Der p 1 were monomeric in solution. In contrast, Der p 1 was reported to be dimeric in solution (pH 7.5 and pH 8.0) and in the crystal[10] (see Fig. 3); however, our analysis of the crystal structures of Der p 1 (PDB codes 2AS8 and 3F5V) showed that the decrease in the accessible surface area of Der p 1 upon dimerization (as defined by Ponstingl *et al.*[34,35]) is about 600–700 Å^2, which is slightly below the cutoff value (856 Å^2) proposed for discrimination between homodimeric and monomeric proteins.[35] The oligomeric analysis of the Der p 1 and Der f 1 structures with the PISA[36] and PITA[35] servers also consistently predicts that both proteins are monomeric.

The dimeric form previously proposed for Der p 1[10] most probably corresponds to the inactive form of the protein because the loop containing residues Tyr166 and Gln167 from one chain sterically blocks the catalytic cleft in the second protein molecule forming the putative oligomeric assembly (Fig. 3b). Analysis of the possible oligomerization interface in terms of amino acid conservation shows that almost all residues forming the interface in Der p 1 are also preserved in Der f 1. It is not clear if the catalytically inactive dimer previously proposed for mature Der p 1 is physiologically relevant or influences the stability of the protein. Oligomerization among cysteine proteases is extremely rare, and there is currently only one known cysteine protease that forms a physiological oligomer: the tetramer of cathepsin C.[37] All other enzymes from this family are monomeric.

Surface properties and antibody binding

As mentioned above, over 80% of the amino acid sequences of Der f 1 and Der p 1 are identical. Comparison of the crystal structures of the enzymes reveals that the sequence differences between Der f 1 and Der p 1 are not distributed evenly in relation to their molecular surfaces (Fig. 4). The differences in the surfaces of Der f 1 and Der p 1 are more dramatic due to lower sequence identity when only surface amino acids are taken into account. The residues that are different are mostly located in loop and α-helical regions, and the residues identical in Der f 1 and Der p 1 occupy around 70% of the molecules' surface areas. The surfaces of the Der f 1 and Der p 1 molecules near the catalytic Cys35 residue appear almost identical, whereas other surface regions of the molecules differ significantly in amino acid composition of the molecular surface. Significant structural differences are apparent as several amino acid patches distal from the catalytic site on the surfaces of both allergens. The four largest patches (Fig. 4) were composed of the following residues

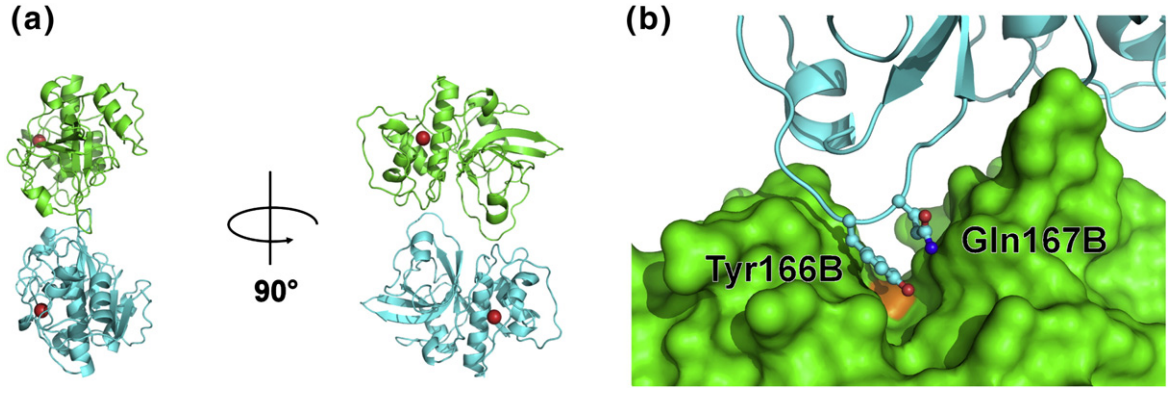

Fig. 3. (a) The putative dimer formed by mature Der p 1 (PDB codes 2AS8 and 3F5V). Der p 1 molecules are shown in ribbon representation, while calcium ions are shown as red spheres. (b) Tyr166 and Gln167 (Der f 1 numbering) are blocking the catalytic cleft in a putative dimer of Der p 1. Chain A is shown in surface representation (green); chain B is shown in blue, while catalytic Cys35A is shown in orange.

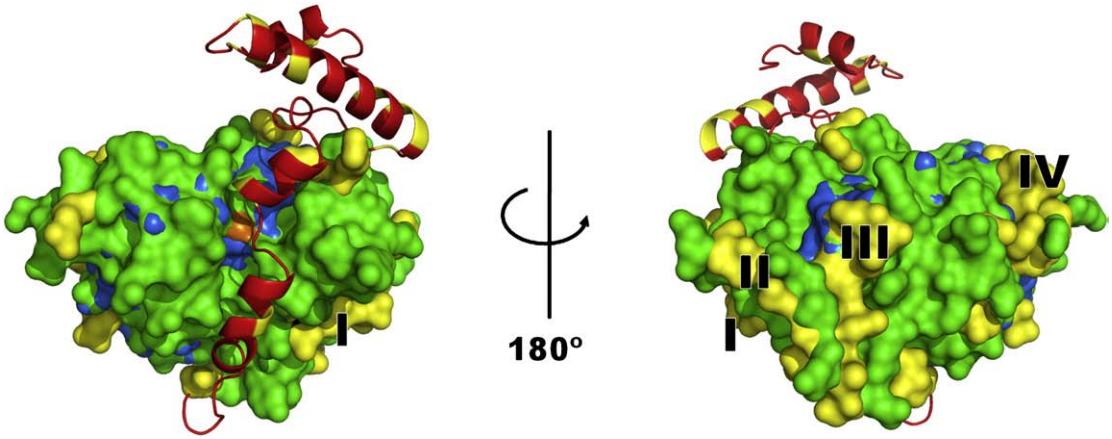

Fig. 4. The molecular surface of mature Der f 1 (PDB code 3D6S) shown in two orientations. Residues that differ in Der f 1 and Der p 1 are shown in yellow. The largest patches of different residues are labeled as I, II, III, and IV. Identical residues are marked in green, and the conserved Cys35 residue is shown in orange. Residues conserved in Der f 1, Der p 1, and human cathepsin K are shown in light blue. The N-terminal propeptide is pictured in ribbon representation (red).

(numbering according to the Der f 1 sequence): patch I: 184, 206, 208–210, and 216; patch II: 11, 13, 124, 126, 127, and 179–181; patch III: 8, 9, 15, 19, 20, and 133; and patch IV: 91, 93, 94, 96, 109, 111, and 112. Such an uneven distribution of conserved *versus* variable residues may in part explain the differences in murine and human IgE antibody responses to these allergens.

Structures of Der f 1 and Der p 1 are similar to structures of human cysteine proteases, with cathepsin K (EC 3.4.22.38) being the most similar human protein to Der f 1 in terms of overall architecture and sequence. Although the sequence identity between human cathepsin K and Der f 1 is only 36% (similarity is 50%), structures of Der f 1 and the human enzyme are similar with low RMSD values over around 200 superposed residues (e.g., 1.3 Å for PDB structure 1U9V and 1.4 Å for PDB structure 1Q6K). The degree of surface conservation between Der f 1 and Der p 1 is ~70% of identical residues, whereas for Der f 1 and cathepsin K, the surfaces are only 30% identical. Only a few residues in the surface area that are identical between Der p 1 and Der f 1 are also identical with cathepsin K (Fig. 4), occupying 10% of the Der f 1 surface. No IgE binding epitope would in general be expected in common areas between Der f 1 and cathepsin K. However, a few residues predicted to be part of the IgE binding epitopes[38,39] were also identical with cathepsin K. Epitope mapping, based on either peptide fragments or mimotopes, is not as precise as mapping based on crystal structures of allergen–antibody complexes.

Analysis of the amino acid sequences of four mAbs against group 1 mite allergens

The variable regions of the heavy and light chains of four mAbs (10B9, 5H8, 6A8, and 4C1) against group 1 mite allergens were sequenced (Fig. 5). Two mAbs recognize species-specific epitopes in Der p 1 (10B9 and 5H8), and one mAb binds to a species-specific epitope in Der f 1 (6A8).[6] The 4C1 mAb raised against Der f 1 binds to a common epitope on both allergens. Additionally, 10B9 inhibits the binding of 4C1 to Der p 1 and vice versa, indicating that the epitopes recognized by both mAbs on Der p 1 either overlap, since the binding of one mAb completely prevents the binding of the other, or lie sufficiently close to one another that binding of either mAb induces conformational changes that make the other site inaccessible.[6] Surprisingly, 10B9 is species specific and does not inhibit the binding of 4C1 to Der f 1, so structural differences between both epitopes are expected.

Amino acid sequences corresponding to the complementarity-determining regions (CDRs) from the light chain [loops L1(26–32), L2(50–52), and L3 (91–96)] and from the heavy chain [CDR1(31–35b), CDR3(50–65), and CDR3(95–102)] were compared among pairs of antibodies. The variability between the light chains of mAb pairs 4C1–6A8, 5H8–10B9, 4C1–5H8, and 4C1–10B9 is much higher (~11 times) than that between each mAb and its highest homolog in the databases, which is zero for at least one loop. The highest variability in amino acid sequence and length is in the CDR3 from the heavy chain, which is thought to mostly determine antibody diversity and contribute to antigen affinity and specificity. Curiously, the CDR2 of 4C1–10B9 has much less variability (4 of 16 residues *versus* 10–11 of 16–17) than the other three mAb pairs, which could be related to the cross-reactivity between these two mAbs (Fig. 5, underlined). The fact that antibodies 4C1 and 10B9 were raised from different mice by immunization to Der f 1 and Der p 1, respectively, makes these similarities more remarkable. Determination of the primary structure of mAb against group 1 mite allergens provides the basis for understanding antibody–allergen interaction.

Conclusions

The crystal structure of Der f 1 is the first reported structure of a natural mite allergen. Der f 1 is a

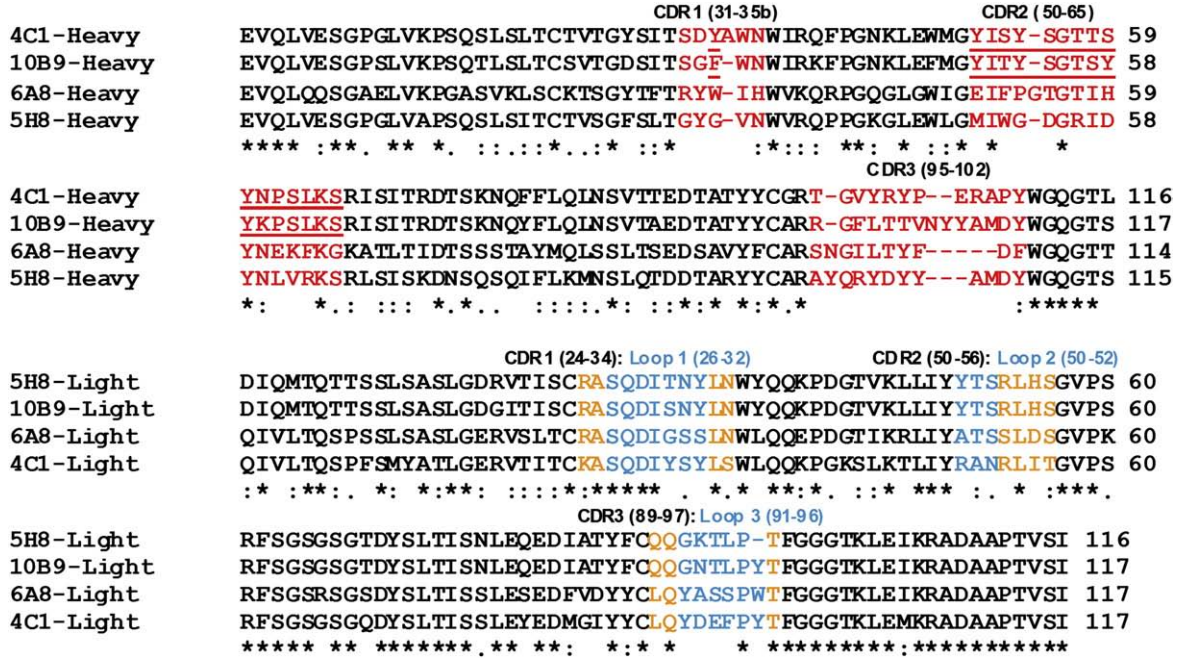

Fig. 5. Amino acid sequence alignment of the heavy and light chains of the mAbs indicating the CDR[40] (red for the heavy chains, and blue and orange for the light chains) and loops (blue for light chain).[41] The CDR2 of low variability between the cross-reacting antibodies 4C1 and 10B9 is underlined.

monomeric protein with a metal binding site that is structurally equivalent to the one described for Der p 1 but lacks a bound ion. The structure reveals four surface patches that differ between Der f 1 and Der p 1 and areas in common that are the putative basis for cross-reactivity between the two allergens. The Der f 1 and Der p 1 structures also show the presence of conserved water molecules in a funnel-shaped cavity close to the metal binding site at one end and to the catalytic site at the other. These water molecules could be biologically significant and may serve to catalyze the cysteine protease activity of the allergen.

The analysis of the variable regions of the heavy and light chains of a panel of four mAbs against Der f 1 and Der p 1 revealed parts of the variable regions that may be involved in cross-reactivity. The Der f 1 and Der p 1 structures, together with antibody binding studies, will identify the amino acids essential for IgE antibody binding and will provide a basis for further understanding of antibody–allergen interactions and, subsequently, the development of new recombinant vaccines for treatment of mite allergic disease.

Materials and Methods

Purification and crystallization

Der f 1 was purified from spent *D. farinae* mite culture extract [~100 g per 1 L of phosphate-buffered saline (PBS)] using affinity chromatography through a 4C1 mAb column. Purified Der f 1 protein (Lot 29084) in PBS was stored at −80 °C. Der f 1 (1.2 mL of 1.2 mg/mL in PBS) was incubated on ice for 1 h with E-64 (40 μL of 10 mg/mL in water; SIGMA). The solution was dialyzed overnight into 0.5 M L-arginine and 0.05 M NaCl at pH 7.5, concentrated using an Amicon Ultra concentrator (Millipore) with a 10,000-Da molecular mass cutoff, and applied on a Superdex 200 column. After gel filtration, the protein was concentrated to 11 mg/mL. The sample was passed through a 0.22 μm filter and used for crystallization. Crystallization was performed at 296 K using the hanging-drop vapor-diffusion method in NEXTAL plates. The protein solution was mixed with well solution in a 1:1 ratio. An initial hit was obtained with a Wizard I screen (Emerald Biosystems) in condition #33 containing 2 M $(NH_4)_2SO_4$, 0.2 M Li_2SO_4, and 0.1 M CAPS at pH 10.5. The conditions were subsequently optimized, and the best diffracting crystals were grown from a solution containing 2 M $(NH_4)_2SO_4$, 0.12 M Li_2SO_4, 0.004 M EDTA, and 0.1 M CAPS at pH 10.5. Crystals obtained from the initial screen and optimized crystals were grown to a similar size. Tracking and analysis of the crystallization experiments were performed with the *Xtaldb* crystallization expert system.

Der p 1 was purified and crystallized as reported previously.[7] Prior to data collection, crystals were soaked for 20 min in solution composed of 5 μL of DMSO, 1 μL of 1 M β-mercaptoethanol, 5 μL of 1 M Tris–HCl, pH 8.0, and 150 μL of 50% w/v monomethyl polyethylene glycol 2000.

Data collection, structure solution, and refinement

For data collection, crystals of Der f 1 (of approximate size 20 μm × 15 μm × 10 μm) were cooled in a cold nitrogen stream (100 K) without cryoprotectant. Data collection was performed at beamline 19-ID[42] of the Structural Biology Center at the Advanced Photon Source. Data were collected and processed with HKL-2000[43] (Table 2). Der f 1 crystallized in the $P4_1$ space group with three molecules in the asymmetric unit. Analysis of the diffraction intensities revealed that all Der f 1 crystals were partially merohedrally twinned. The Merohedral Crystal Twinning

Table 2. Data collection and refinement statistics

Protein	Der f 1	Der p 1
PDB code	3D6S	3F5V
Data collection		
Beamline	19ID (Advanced Photon Source)	ID-23-1 (ESRF)
Wavelength (Å)	0.979	1.072
Unit cell	a=91.2 Å, c=77.7 Å	a=95.2 Å, b=84.1 Å, c=75.4 Å, β=123.8°
Space group	$P4_1$	$C2$
Solvent content (%)	41	50
No. of protein chains in AU	3	2
Resolution range (Å)	50.0–2.0	50.0–1.36
Highest-resolution shell (Å)	2.02–2.00	1.38–1.36
Unique reflections	39,590 (1078)	103,499 (4795)
Redundancy	6.4 (6.6)	3.6 (2.9)
Completeness (%)	92.6 (100.0)	98.5 (91.5)
R_{merge} (%)	8.0 (51.4)	6.9 (55.1)
Average I/σ (I)	27.0 (3.3)	25.3 (2.1)
Refinement		
R (%)	21.1	15.8
R_{free} (%)	24.1	18.3
Mean B value (Å^2)	32.3	8.9
B from Wilson's plot (Å^2)	25.3	12.6
RMSD bond lengths (Å)	0.018	0.017
No. of amino acid residues	668	444
No. of water molecules	311	679
No. of metal ions	0	2
Ramachandran plot		
Most favored regions (%)	95.8	96.2
Additional allowed regions (%)	4.2	3.8

The Ramachandran plot was calculated using MOLPROBITY.

Server† was used to determine the twinning fraction, and the twinning fraction was close to 0.4 for all crystals (twinning law = 0 1 0 1 0 0 0 0 −1). Despite the small size of the crystals, they initially diffracted to around 2.5 Å, but the data resolution was quickly affected by radiation damage. To obtain a complete data set, we merged data from two crystals grown in the same drop. A partial molecular replacement solution was obtained with MOLREP[44] as incorporated into HKL-3000.[45] A search model for molecular replacement was prepared using SWISS-MODEL,[46] and the structure of Der p 1 (PDB code 2AS8) was used as a template. The partial model contained two molecules of Der f 1; after rigid-body refinement, it was discovered that additional unidentified electron density corresponded to a third Der f 1 molecule. The third molecule was localized using PHASER[47] as implemented in the CCP4 package.[48] Although E-64 was present in the crystallization conditions, no continuous density corresponding to the molecule was observed and the inhibitor could not be modeled. The initial refinement was done using CNS[49] with the protocol for twinned data. The CNS simulating annealing protocol was used to remove model bias. Further refinement was performed using SHELXL[50] and higher-resolution data. The program COOT[17] was used for manual adjustment of the model. The quality of the structure was monitored using MOLPROBITY[51] and ADIT.[52] The twin fraction was refined to 0.43. A summary of the data collection and refinement statistics for the higher-resolution data used for refinement is reported in Table 2.

Data from Der p 1 crystal were collected at ID-23-1 (ESRF, Grenoble). The structure of Der p 1 was solved using the model of the orthorhombic Der p 1 form (PDB code 2AS8) and HKL-3000 in combination with MOLREP. Refinement was done using REFMAC and COOT. In the last stages, refinement was performed using TLS groups defined with the TLMSD server.[53] Structure validation was performed using the same tools used for the structure of Der f 1. Details of data processing and refinement are summarized in Table 2.

Calculations of surface area were performed using the MSMS program at the StrucTools server‡.

Sequencing of the group 1 allergen-specific mAbs from cell lines

The cell lines producing the anti-group 1 allergen mAbs 4C1 B8 3F8, 6A8 B10 D12, 5H8 C12 D8, and 10B9 F6 A12 were grown at the Lymphocyte Culture Center (University of Virginia). The mAbs had been raised in BALB/c mice, which were reported to be the best responder strain to group 1 mite allergen.[7,54,55] Total RNA (36–100 μg/3×10^6 cells) was isolated from the cell lines using an RNeasy Mini kit (Qiagen, Valencia, CA). cDNAs encoding for the light and heavy chains of the mAbs were obtained by reverse transcription from RNA (SuperScript™ III, Invitrogen, Carlsbad, CA), and the DNA was PCR amplified using specific primers, sequenced, and analyzed.

A primer mix containing the degenerate N-terminal primers VHa and VHb [VHa: 5′-gag gtt cag ctg cag cag(ct) c-3′ and 5′-gag gtg cag ctg gtg ga(ag)tc-3′] and the C-terminal primer CH2 [CH2: 5′-tt agg agt cag agt aat ggt gag cac atc c-3′] was used to reverse transcribe the RNA and amplify the DNA encoding for the heavy chains at annealing temperatures of 40–55 °C. Middle primers were used to finalize heavy chain sequencing. To amplify the

† http://nihserver.mbi.ucla.edu/Twinning

‡ http://molbio.info.nih.gov/structbio/basic.html

light chains of the antibodies, we initially tested four pairs of primers. The N-terminal primer Vκ4 (5′-caa att gtt ctc acc cag tct cca-3′) was used for DNA amplification of the light chains of 4C1 and 6A8, and Vκ10 (5′-gat atc cag atg aca cag act aca-3′) amplified the light chains of 5H8 and 10B9, combined with the C-terminal primer (kappa constant: 5′-gat gga tac agt tgg tgc-3′). The mAbs were isotyped G1 for the heavy chain and kappa (κ) for the light chain using IsoStrip (Roche Diagnostics, Indianapolis, IN). CLUSTAL[74] and BLAST were used to align and search for homologous sequences in databases, respectively.

PDB accession codes

The coordinates and structure factors for Der f 1 were deposited in PDB with accession code 3D6S, while the coordinates and structure factors for Der p 1 were deposited with accession code 3F5V.

Acknowledgements

The work described in this article was supported by GM53163, Indoor Biotechnologies, Inc., internal grant, the University of Virginia Biomedical Engineering Industrial Internship Program, and the Virginia Biotechnology Association. The results shown in this report were derived from work performed at the Structural Biology Center of the Argonne National Laboratory Advanced Photon Source. Argonne is operated by the University of Chicago Argonne, LLC, for the U.S. Department of Energy Office of Biological and Environmental Research under contract DE-AC02-06CH11357. We thank Dr. Alex Wlodawer for valuable discussions and Kyle Williams for technical assistance.

References

1. Platts-Mills, T. A., Thomas, W. R., Aalberse, R. C., Vervloet, D. & Chapman, M. D. (1992). Dust mite allergens and asthma: report of a second international workshop. *J. Allergy Clin. Immunol.* **89**, 1046–1060.
2. Hales, B. J., Shen, H. D. & Thomas, W. R. (2000). Cross-reactivity of T-cell responses to *Dermatophagoides pteronyssinus* and *D. farinae*. Studies with group 1 and 7 allergens. *Clin. Exp. Allergy*, **30**, 927–933.
3. Chapman, M. D. & Platts-Mills, T. A. (1980). Purification and characterization of the major allergen from *Dermatophagoides pteronyssinus*—antigen P1. *J. Immunol.* **125**, 587–592.
4. Heymann, P. W., Chapman, M. D., Aalberse, R. C., Fox, J. W. & Platts-Mills, T. A. (1989). Antigenic and structural analysis of group II allergens (Der f II and Der p II) from house dust mites (*Dermatophagoides* spp.). *J. Allergy Clin. Immunol.* **83**, 1055–1067.
5. Barrett, A. J. & Rawlings, N. D. (2001). Evolutionary lines of cysteine peptidases. *Biol. Chem.* **382**, 727–733.
6. Chapman, M. D., Heymann, P. W. & Platts-Mills, T. A. (1987). Epitope mapping of two major inhalant allergens, Der p I and Der f I, from mites of the genus *Dermatophagoides*. *J. Immunol.* **139**, 1479–1484.
7. Heymann, P. W., Chapman, M. D. & Platts-Mills, T. A. (1986). Antigen Der f I from the dust mite *Dermatophagoides farinae*: structural comparison with Der p I from *Dermatophagoides pteronyssinus* and epitope specificity of murine IgG and human IgE antibodies. *J. Immunol.* **137**, 2841–2847.
8. Lind, P., Hansen, O. C. & Horn, N. (1988). The binding of mouse hybridoma and human IgE antibodies to the major fecal allergen, Der p I, of *Dermatophagoides pteronyssinus*. Relative binding site location and species specificity studied by solid-phase inhibition assays with radiolabeled antigen. *J. Immunol.* **140**, 4256–4262.
9. Meno, K., Thorsted, P. B., Ipsen, H., Kristensen, O., Larsen, J. N., Spangfort, M. D. *et al.* (2005). The crystal structure of recombinant proDer p 1, a major house dust mite proteolytic allergen. *J. Immunol.* **175**, 3835–3845.
10. de Halleux, S., Stura, E., VanderElst, L., Carlier, V., Jacquemin, M. & Saint-Remy, J. M. (2006). Three-dimensional structure and IgE-binding properties of mature fully active Der p 1, a clinically relevant major allergen. *J. Allergy Clin. Immunol.* **117**, 571–576.
11. Takai, T., Kato, T., Yasueda, H., Okumura, K. & Ogawa, H. (2005). Analysis of the structure and allergenicity of recombinant pro- and mature Der p 1 and Der f 1: major conformational IgE epitopes blocked by prodomains. *J. Allergy Clin. Immunol.* **115**, 555–563.
12. Chapman, M. D., Wunschmann, S. & Pomés, A. (2007). Proteases as Th2 adjuvants. *Curr. Allergy Asthma Rep.* **7**, 363–367.
13. Nakamura, T., Hirasawa, Y., Takai, T., Mitsuishi, K., Okuda, M., Kato, T. *et al.* (2006). Reduction of skin barrier function by proteolytic activity of a recombinant house dust mite major allergen Der f 1. *J. Invest. Dermatol.* **126**, 2719–2723.
14. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
15. Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323.
16. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2256–2268.
17. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
18. Piboonpocanun, S., Malainual, N., Jirapongsananuruk, O., Vichyanond, P. & Thomas, W. R. (2006). Genetic polymorphisms of major house dust mite allergens. *Clin. Exp. Allergy*, **36**, 510–516.
19. van Oort, E., de Heer, P. G., van Leeuwen, W. A., Derksen, N. I., Muller, M., Huveneers, S. *et al.* (2002). Maturation of *Pichia pastoris*-derived recombinant pro-Der p 1 induced by deglycosylation and by the natural cysteine protease Der p 1 from house dust mite. *Eur. J. Biochem.* **269**, 671–679.
20. Takai, T., Mineki, R., Nakazawa, T., Takaoka, M., Yasueda, H., Murayama, K. *et al.* (2002). Maturation of the activities of recombinant mite allergens Der p 1 and Der f 1, and its implication in the blockade of proteolytic activity. *FEBS Lett.* **531**, 265–272.
21. Takai, T., Mizuuchi, E., Kikuchi, Y., Nagamune, T., Okumura, K. & Ogawa, H. (2006). Glycosylation of

recombinant proforms of major house dust mite allergens Der p 1 and Der f 1 decelerates the speed of maturation. *Int. Arch. Allergy Immunol.* **139**, 181–187.
22. Best, E. A., Stedman, K. E., Bozic, C. M., Hunter, S. W., Vailes, L., Chapman, M. D. *et al.* (2000). A recombinant group 1 house dust mite allergen, rDer f 1, with biological activities similar to those of the native allergen. *Protein Expression Purif.* **20**, 462–471.
23. Yasuhara, T., Takai, T., Yuuki, T., Okudaira, H. & Okumura, Y. (2001). Biologically active recombinant forms of a major house dust mite group 1 allergen Der f 1 with full activities of both cysteine protease and IgE binding. *Clin. Exp. Allergy*, **31**, 116–124.
24. Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L. & Minor, W. (2008). Data mining of metal ion environments present in protein structures. *J. Inorg. Biochem.* **102**, 1765–1776.
25. Mellor, G. W., Thomas, E. W., Topham, C. M. & Brocklehurst, K. (1993). Ionization characteristics of the Cys-25/His-159 interactive system and of the modulatory group of papain: resolution of ambiguity by electronic perturbation of the quasi-2-mercaptopyridine leaving group in a new pyrimidyl disulphide reactivity probe. *Biochem. J.* **290**, 289–296.
26. Dardenne, L. E., Werneck, A. S., de Oliveira Neto, M. & Bisch, P. M. (2003). Electrostatic properties in the catalytic site of papain: a possible regulatory mechanism for the reactivity of the ion pair. *Proteins*, **52**, 236–253.
27. Takai, T., Kato, T., Sakata, Y., Yasueda, H., Izuhara, K., Okumura, K. & Ogawa, H. (2005). Recombinant Der p 1 and Der f 1 exhibit cysteine protease activity but no serine protease activity. *Biochem. Biophys. Res. Commun.* **328**, 944–952.
28. Deb, R., Shakib, F., Reid, K. & Clark, H. (2007). Major house dust mite allergens *Dermatophagoides pteronyssinus* 1 and *Dermatophagoides farinae* 1 degrade and inactivate lung surfactant proteins A and D. *J. Biol. Chem.* **282**, 36808–36819.
29. Schulz, O., Sewell, H. F. & Shakib, F. (1998). A sensitive fluorescent assay for measuring the cysteine protease activity of Der p 1, a major allergen from the dust mite *Dermatophagoides pteronyssinus*. *Mol. Pathol.* **51**, 222–224.
30. Pernas, M., Sanchez-Ramos, I., Sanchez-Monge, R., Lombardero, M., Arteaga, C., Castanera, P. & Salcedo, G. (2000). Der p 1 and Der f 1, the highly related and major allergens from house dust mites, are differentially affected by a plant cystatin. *Clin. Exp. Allergy*, **30**, 972–978.
31. Zhang, J., Hamilton, J. M., Garrod, D. R. & Robinson, C. (2007). Interactions between mature Der p 1 and its free prodomain indicate membership of a new family of C1 peptidases. *Allergy*, **62**, 1302–1309.
32. Kaul, P., Sathish, H. A. & Prakash, V. (2002). Effect of metal ions on structure and activity of papain from *Carica papaya*. *Nahrung*, **46**, 2–6.
33. Ma, S., Devi-Kesavan, L. S. & Gao, J. (2007). Molecular dynamics simulations of the catalytic pathway of a cysteine protease: a combined QM/MM study of human cathepsin K. *J. Am. Chem. Soc.* **129**, 13633–13645.
34. Ponstingl, H., Kabir, T. & Thornton, J. M. (2003). Automatic inference of protein quaternary structure from crystals. *J. Appl. Crystallogr.* **36**, 1116–1122.
35. Ponstingl, H., Henrick, K. & Thornton, J. M. (2000). Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins*, **41**, 47–57.
36. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
37. Dolenc, I., Turk, B., Pungercic, G., Ritonja, A. & Turk, V. (1995). Oligomeric structure and substrate induced inhibition of human cathepsin C. *J. Biol. Chem.* **270**, 21626–21631.
38. Greene, W. K. & Thomas, W. R. (1992). IgE binding structures of the major house dust mite allergen Der p I. *Mol. Immunol.* **29**, 257–262.
39. Szalai, K., Fuhrmann, J., Pavkov, T., Scheidl, M., Wallmann, J., Bramswig, K. H. *et al.* (2008). Mimotopes identify conformational B-cell epitopes on the two major house dust mite allergens Der p 1 and Der p 2. *Mol. Immunol.* **45**, 1308–1317.
40. Kabat, E. A. & Wu, T. T. (1971). Attempts to locate complementarity-determining residues in the variable positions of light and heavy chains. *Ann. N. Y. Acad. Sci.* **190**, 382–393.
41. Tomlinson, I. M., Cox, J. P., Gherardi, E., Lesk, A. M. & Chothia, C. (1995). The structural repertoire of the human V kappa domain. *EMBO J.* **14**, 4628–4638.
42. Rosenbaum, G., Alkire, R. W., Evans, G., Rotella, F. J., Lazarski, K., Zhang, R. G. *et al.* (2006). The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. *J. Synchrotron Radiat.* **13**, 30–45.
43. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
44. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
45. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 859–866.
46. Kopp, J. & Schwede, T. (2004). The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res.* **32**, D230–D234.
47. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
48. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
49. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
50. Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallogr., Sect. A*, **64**, 112–122.
51. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G. *et al.* (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, **50**, 437–450.
52. Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). Automated and accurate deposition of structures solved by X-ray

diffraction to the Protein Data Bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 1833–1839.
53. Painter, J. & Merritt, E. A. (2006). TLSMD Web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111.
54. Chapman, M. D., Sutherland, W. M. & Platts-Mills, T. A. (1984). Recognition of two *Dermatophagoides pteronyssinus*-specific epitopes on antigen P1 by using monoclonal antibodies: binding to each epitope can be inhibited by serum from dust mite-allergic patients. *J. Immunol.* **133**, 2488–2495.
55. Ovsyannikova, I. G., Vailes, L. D., Li, Y., Heymann, P. W. & Chapman, M. D. (1994). Monoclonal antibodies to group II *Dermatophagoides* spp. allergens: murine immune response, epitope analysis, and development of a two-site ELISA. *J. Allergy Clin. Immunol.* **94**, 537–546.