

Crystal Structure of the Dog Lipocalin Allergen Can f 2: Implications for Cross-reactivity to the Cat Allergen Fel d 4

Chaithanya Madhurantakam^{1†}, Ola B. Nilsson^{2†}, Hannes Uchtenhagen¹, Jon Konradsen³, Tiiu Saarne², Erik Högbom⁴, Tatyana Sandalova⁵, Hans Grönlund^{6†} and Adnane Achour^{1*†}

¹Centre for Infectious Medicine, F59, Department of Medicine Huddinge, Karolinska University Hospital Huddinge, Karolinska Institutet, Stockholm, Sweden

²Department of Medicine, Clinical Immunology and Allergy Unit, Karolinska Institutet, Stockholm, Sweden

³Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden

⁴Phadia AB, Uppsala, Sweden

⁵Department of Medical Biophysics and Biochemistry, Karolinska Institutet, Stockholm, Sweden

⁶Center for Allergy Research, Department of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Received 4 January 2010;
received in revised form
17 May 2010;
accepted 18 May 2010
Available online
26 May 2010

The dog lipocalin allergen Can f 2 is an important cause of allergic sensitization in humans worldwide. Here, the first crystal structure of recombinant rCan f 2 at 1.45 Å resolution displays a classical lipocalin fold with a conserved Gly-Xaa-Trp motif, in which Trp19 stabilizes the overall topology of the monomeric rCan f 2. Phe38 and Tyr84 localized on the L1 and L5 loops, respectively, control access to the highly hydrophobic calyx. Although the rCan f 2 calyx is nearly identical with the aero-allergens MUP1, Equ c 1 and A2U from mouse, horse and rat, respectively, no IgE cross-reactivity was found using sera from five mono-sensitized subjects. However, clear IgE cross-reactivity was demonstrated between Can f 2 and the cat allergen Fel d 4, although they share less than 22% sequence identity. This suggests a role for these allergens in co-sensitization between cat- and dog-allergic patients.

© 2010 Elsevier Ltd. All rights reserved.

Edited by I. Wilson

Keywords: dog allergen; crystal structure; recombinant Can f 2; epitope; lipocalin fold

*Corresponding author E-mail address: adnane.achour@ki.se.

† C.M. and O.B.N. shared first authors. H.G. and A.A. shared last authors.

Present address: C. Madhurantakam, Department of Biochemistry, University of Zurich, Zurich, Switzerland.

Abbreviations used: A1M, α 1-microglobulin; A2U, α _{2u} globulin/rat urinary protein; ApoD, apolipoprotein D; AP, alkaline phosphatase; Bos d 2 and Bos d 5, bovine allergens 2 and 5, respectively; Can f 1, Can f 2 and Can f 3, dog allergens 1, 2 and 3, respectively; Equ c 1, horse allergen; Fel d 1 and Fel d 4, cat allergens; HC8 γ , human complement component-8 γ ; MUP1, mouse urinary protein 1; NGAL, neutrophil gelatinase-associated lipocalin; RBP, retinal-binding protein; SCR, structurally conserved region; Tlc, tear lipocalin; VEGP, von Ebners gland protein.

Introduction

Allergy to mammals is globally prominent and chronic exposure to allergens can lead to allergenic sensitization and development of hyper-reactivity and asthma.¹ Crystal structures of some mammalian aero-allergens have been determined, including cat allergen Fel d 1,^{2–6} mouse urinary protein (MUP1)⁷ and horse allergen Equ c 1.⁸ However, no 3D structure has been determined for the few dog allergens identified so far.^{9,10} The two best studied dog allergens Can f 1 (Dog allergen 1 or Dog 1) and Can f 2 (Dog allergen 2 or Dog 2) cause worldwide allergic disease. Can f 1 is present in tongue epithelial tissue and Can f 2 is produced in both tongue and parotid gland.⁹ It has been demonstrated that serum IgE from 75% of dog-allergic patients can bind to recombinant forms of Can f 1 and Can f 2. The IgE-binding feature can be used in the diagnosis of dog allergy.^{10–12} Both dog allergens were found to induce histamine release from blood basophils of allergic subjects.⁹

Most of the major mammalian respiratory allergens, e.g. from mouse, rat, cow and horse, are lipocalins.^{10,13,14} The dog allergens Can f 1 and Can f 2, which share only 24.5% sequence homology, have been predicted to belong to this protein family.⁹ Although they most often share only 18–25% sequence identity, lipocalins are defined by the striking overall similarity of their 3D fold as well as the presence of one to three structurally conserved regions (SCR) that act as a family signature. Lipocalins typically have a cup-shaped cavity within the central β -barrel formed by eight anti-parallel β -strands (β_A – β_H) with a loop scaffold that might regulate access to the ligand-binding site. Both ends of the β -barrel are topologically distinct. The top end of the β -barrel comprises the extended loop L1 along with three β -hairpin-forming loops (L3, L5 and L7) that open to the ligand-binding site, while the bottom end of the β -barrel is formed by three β -hairpin-forming loops (L2, L4 and L6) and the conserved N-terminal 3_{10} -helix that effectively closes the barrel (closed end of the molecule). Finally, the short helix H1 is localized adjacent to the core of the β -barrel.

Lipocalins can be divided into two structural sub-families named kernel lipocalins and outliers, on the basis of the presence or the absence of three structurally conserved regions (SCR1–SCR3) that are localized within the closed end of the barrel.^{15–17} While SCR1 is formed by the β_A strand and the 3_{10} -helix, SCR2 comprises sections of the β_F and β_G strands as well as the entire L6 loop that connects them. Finally, SCR3 is formed by the C-terminal end of the β_H strand and adjacent residues, conserved among most lipocalins, localized on the unique helix H1. While all three SCR are present in kernel lipocalins, outlier lipocalins have a maximum of two SCR.

Ligand specificity of lipocalins is regulated by the amino acid residues that line their binding pocket.^{15,18} A large variety of ligands, most of them

hydrophobic or amphiphilic, binds to lipocalins. In some cases, binding of the ligand is critical for the biological functions of the lipocalin.¹⁹ While some lipocalins display high specificity towards a ligand (e.g. retinal-binding protein with vitamin A),^{19,20} others are much more versatile and bind several different compounds.²¹ Importantly, a common biological function shared by most lipocalins is to transport small hydrophobic compounds such as pheromones and vitamins.¹⁴

The lipocalin family comprises at least 50 different proteins present in all branches of life (bacteria, fungi, plants and animals).²² The functions that have been ascribed to lipocalins are many and diverse. Most of the extracellular lipocalin allergens, including mouse (mouse urinary protein, MUP1),⁷ horse (Equ c 1),^{8,23} cow (Bos d 2)²⁴ and cat (Fel d 4),²⁵ retinol-binding protein (RBP),^{26,27} β -lactoglobulin^{28–30} and apolipoprotein D (ApoD)³¹ are thought to function as carriers of ligands. Other lipocalins, such as von Ebner's gland protein (VEGP),^{32,33} neutrophil gelatinase-associated lipocalin (NGAL)^{34,35} and α 1-microglobulin (A1M)³⁶ are scavengers of fatty acids, iron and heme, respectively, in extracellular fluids. The lipocalin family includes redox enzymes such as prostaglandin D synthase and plant epoxidases, as well as proteins regulating cell differentiation and growth, such as fatty acid-binding protein and glycodefin.¹⁵

In an effort to understand the structural basis underlying the allergenicity of the major dog allergen Can f 2, we have determined the crystal structure of recombinant (r)Can f 2 to 1.45 Å resolution. The 3D structure of rCan f 2 reveals a classical lipocalin fold with a highly hydrophobic calyx occupied by a glycerol moiety. The shape of the calyx and the properties of the residues lining the cavity are nearly identical with the aero-allergens MUP1, Equ c 1 and A2U (α_{2u} -globulin). However, using sera from five subjects sensitized to Can f 2, no IgE cross-reactivity could be demonstrated between rCan f 2 and these lipocalin allergens or any of the other human lipocalins tested. Conversely, clear IgE cross-reactivity was demonstrated between rCan f 2 and the cat lipocalin allergen Fel d 4, which share less than 22% identity. Structural and sequential comparison suggested a solvent-exposed region, negatively charged in both rCan f 2 and Fel d 4, and formed by three independent stretches of residues surrounding the α -helical H1 domain, as a potential common conformational epitope.

Results and Discussion

Recombinant Can f 2 displays allergenicity equivalent to that of natural Can f 2

We have demonstrated that rCan f 2 displayed similar if not identical IgE-binding properties when compared to its natural counterpart (Fig. 1a). Quantification of specific IgE reactivities towards

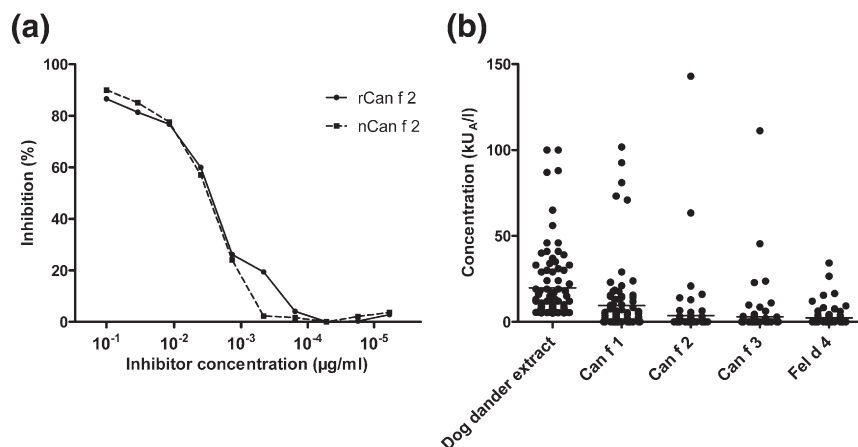


Fig. 1. The recombinant and the naturally occurring Can f 2 exhibited equivalent IgE binding properties. (a) Pooled sera from three Can f 2-sensitized individuals were pre-incubated with threefold dilutions (*x*-axis) of natural (■) and recombinant (●) Can f 2, followed by detection of IgE-responses by ELISA. Inhibition was calculated as responses (in%) after addition of the antigen (*y*-axis). (b) The sensitization profile to the three common dog allergens Can f 1, Can f 2 and Can f 3 as well as to the cat allergen Fel d 4 was identified by quantification of specific IgE reactivity. The prevalence of IgE responses to Can f 1, Can f 2, Can f 3 (albumin) and Fel d 4 in 89 dog-sensitized individuals was 67%, 40%, 25% and 54%, respectively, as measured by ELISA.

the three dog allergens Can f 1, Can f 2 and Can f 3 as well as the cat allergen Fel d 4 in 89 dog-sensitized subjects revealed responses in 67%, 40%, 25% and 54%, respectively. Furthermore, it should be noted that five Can f 2-sensitized individuals without IgE-antibodies to Can f 1 were identified (Fig. 1b; Supplementary Data Table I). This observation is crucial, as previous studies did not determine specific IgE-antibodies towards Can f 2 in the absence of Can f 1 sensitization.^{9,10} These results were confirmed by the analysis of sera from 21 asthmatic children sensitized to dog using the allergen component chip Immuno Solid-phase Allergen Chip (Phadia, Uppsala Sweden), where sera from two subjects reacted exclusively to Can f 2 and not to Can f 1 (data not shown).

Determination of the crystal structure of rCan f 2

The 3D structure of the allergen rCan f 2 was determined to a resolution of 1.45 Å (Table 1). The model displays good stereochemistry and the R_{cryst} and R_{free} values for the refined structure are 15.75% and 16.83%, respectively. The asymmetric unit includes one rCan f 2 molecule (residues 7–157), one glycerol molecule and 129 water molecules. The final electron density map is of high quality with well defined polypeptide chains. The six N-terminal residues were not defined by electron density, indicating disorder in this region. The average B -factor of the final model, 13.7 Å², is very close to the B -factor calculated from the Wilson plot (Table 1).

The dog allergen rCan f 2 belongs to the lipocalin structural family

The 3D structure of rCan f 2 displays a characteristic lipocalin fold¹⁷ with a β -barrel structure with eight central, anti-parallel β -strands (β_A – β_H) inter-

rupted by intermittent β -hairpin-forming loops (L1–L7) and with an adjacent α -helix (H1) localized at the C-terminus of the protein. The overall 3D structure, punctuated by four 3_{10} helices, is stabilized through the formation of a disulfide bridge between Cys64 and Cys157 localized on the β_D -strand and on the extended C-terminal loop, respectively (Fig. 2a and b). This is a characteristic feature of the highly conserved lipocalin fold.¹⁵ The crystal structure of rCan f 2 shows a topological polarity with two distinct regions. The first region, which we call the open end, might act as a ligand-binding site and as an access to the calyx/cavity of rCan f 2, while the second region, the closed end, might act as a binding site for surface receptors (Fig. 2a and c). It should be noted that the open end region of rCan f 2, composed of the four loops L1, L3, L5 and L7, is highly variable when compared to other lipocalins (Figs. 2a and c and 3). Conversely, the sequence of the N-terminal 3_{10} helix, which effectively closes the barrel in the closed end region, is conserved among all lipocalins.

It has been demonstrated in other lipocalin homologues, such as the horse allergen Equ c 1,²³ aphrodisin³⁷ and the boar salivary lipocalin,³⁸ that an asparagine residue (Asn50) localized on the L2 loop can be glycosylated.¹⁴ However, although Asn50 is conserved in rCan f 2, the amino acid motif required for efficient glycosylation of this position is missing. Conversely, predictions for glycosylation sites in the rCan f 2 sequence using the NetNGlyc server 1.0[‡] suggested the stretch of residues Asn27–Ser29 as a potential site that is localized on the tip of the L1 loop.

[‡] <http://www.cbs.dtu.dk/services/NetNGlyc/>

Table 1. Data collection and refinement statistics of the crystal structure of rCan f 2

rCan f 2	
A. Data collection	
Wavelength (Å)	0.93
Space group	C222 ₁
<i>a</i> (Å)	68.18
<i>b</i> (Å)	76.90
<i>c</i> (Å)	64.99
Beamline	ID 14-2/ESRF
Resolution (Å)	34.50–1.45 (1.53–1.45)
No. observed reflections	196,193 (28,265)
No. unique reflections	30,546 (4370)
Redundancy	6.4 (6.5)
Completeness (%)	99.8 (99.6)
<i>R</i> _{merge} (%)	8.4 (47.1)
<i>I</i> / σ (<i>I</i>)	13.1 (3.9)
B. Refinement statistics	
Resolution (Å)	32.50–1.45
<i>R</i> _{cryst} (%)	15.75
<i>R</i> _{free} (%)	16.83
No. protein atoms	1250
No. water molecules	129
No. glycerol molecule	1
No. chains:	1
rmsd from ideal geometry	
Bond lengths (Å)	0.010
Bond angles (deg.)	1.426
Ramachandran plot	
Preferred regions (%)	93.5
Allowed regions (%)	5.8
Outliers (%)	0.7
Wilson <i>B</i> -value (Å ²)	13.6
Average <i>B</i> -value	
Protein (Å ²)	13.7
Water (Å ²)	29.3
Glycerol (Å ²)	40.7

Values in parentheses are for the highest resolution shell.

$R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations *i* of reflection *hkl*.

$R_{\text{cryst}} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively, of a particular reflection and the summation is over 95% of the reflections in the specified resolution range. The remaining 5% of the reflections were selected randomly (test set) before the structure refinement and were not included in the structure refinement.

*R*_{free} was calculated over these reflections with the same equation used for *R*_{cryst}.

The highly conserved tryptophan residue Trp19 plays a key role in the formation and stabilization of the lipocalin fold

Specific residues that are highly conserved among lipocalins and that are localized within the core of the β -barrel are crucial for their overall stabilization.³⁹ Most of these residues are also conserved in rCan f 2. Trp19, localized on the β_A -strand, stabilizes the overall fold of rCan f 2 through the formation of a hydrogen bond and van der Waals interactions with the side chain of Arg122 localized on the β_H -strand (Fig. 4; Table 2). It has been demonstrated that substitution of the homologous tryptophan and arginine residues in the lipocalin

retinal-binding protein (RBP) led to loss in stability and lower yields of proteins.³⁹ Furthermore, many hydrogen bonds are formed between the centrally localized key tryptophan residue Trp19 and the surrounding residues Leu15, Arg18, Ile42 and His43 (Fig. 4). The overall fold of rCan f 2 is further stabilized through the formation of a net of hydrogen bond interactions between the side chain of Arg122 and the main-chain carbonyl groups of Glu14, Ser16 and Pro97 (Fig. 4).

A glycerol moiety is bound within the hydrophobic ligand-binding cavity of rCan f 2

The central elliptical cavity (or calyx) of rCan f 2, which is dominated by hydrophobic residues, suggesting that the dog allergen is able to bind and carry small hydrophobic molecules (Fig. 5), covers an area of 386.8 Å² and a volume of 454.2 Å³ (Table 2). Strong residual electron density that cannot be attributed to side chains or ordered water molecules indicates a bound ligand within the cavity of rCan f 2. Although the nature of this ligand in the recombinant protein is unknown, it has been modeled as a glycerol molecule that is present at high concentration in the crystallization conditions.⁴⁰ In addition, two well ordered water molecules were found buried within the calyx of rCan f 2 (Fig. 6a). The size and the overall hydrophobicity of the calyx of rCan f 2 are very similar to the calyces of the murine MUP1, the equine Equ c 1 and the rat A2U allergens (Fig. 5).⁴¹ Both MUP1 and A2U, with slightly more prominent and spherical hydrophobic cavities when compared to rCan f 2, act as pheromone carriers.^{7,21} The relative conservation of specific aromatic and hydrophobic residues such as Phe38, Val40, Leu69, Leu82 and Tyr84 in rCan f 2 when compared to MUP1 and A2U indicates that similar molecules may bind to all three allergens (Fig. 6b). Interestingly, it has been suggested that Tyr84 might play an important role in providing access and binding ligands in the lipocalin allergen aphrodisin.³⁷ Thus, one could speculate that, like the other aero-allergens, Can f 2 might be able to bind and carry pheromones.

Lipocalin conserved residues Phe38 and Tyr84 control access to the calyx of rCan f 2

The internal buried calyx of rCan f 2 is not exposed to the solvent directly, raising the question of how the ligand enters the core. One can speculate that rCan f 2 refolds around the ligand upon binding, or that an entry (or many entries) might allow passage of the ligand into the calyx. Our analysis of the crystal structure of rCan f 2 suggests that specific residues, conserved among most known lipocalins, might be able to regulate the access of ligands to the calyx in the fully refolded rCan f 2. In particular, loops L1 and L5 form a lid, which might take different conformations upon binding/release of the ligand by rCan f 2 (Fig. 7a). Residues positioned on both L1 and L5, such as Tyr84 and Phe38 as well as

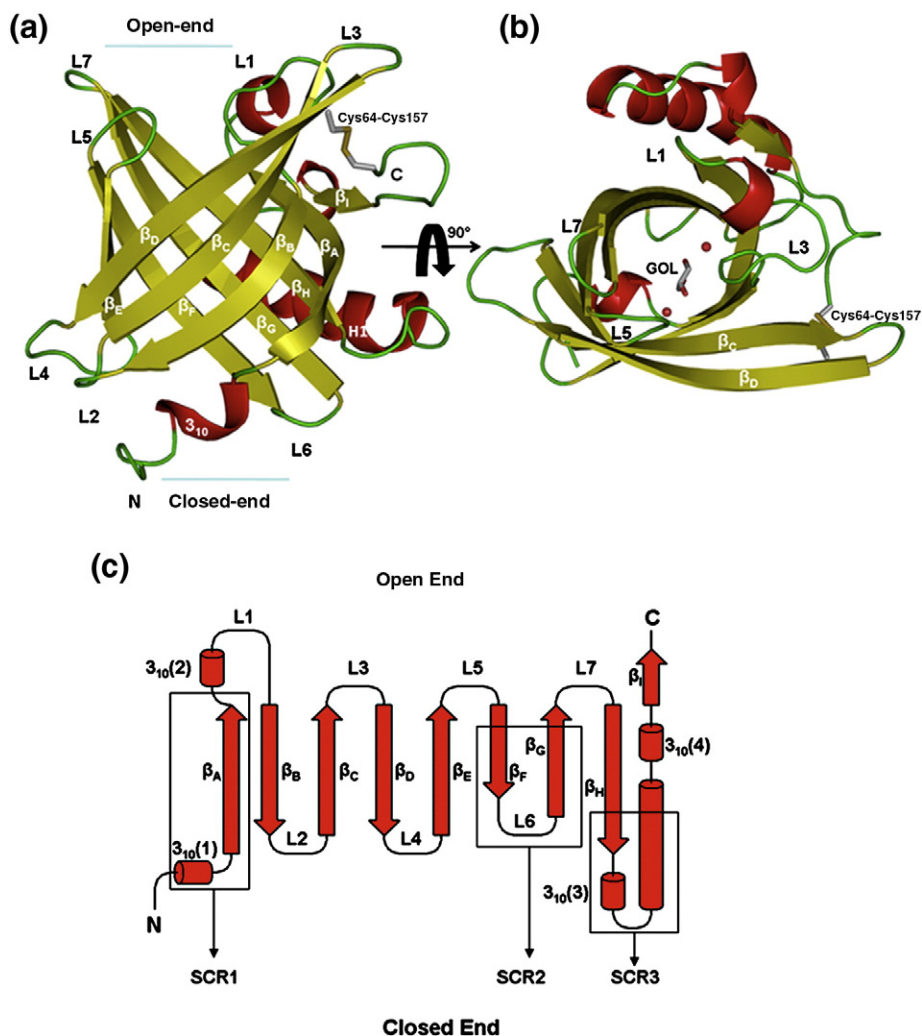


Fig. 2. The three-dimensional structure of rCan f 2 displays a characteristic lipocalin fold. (a) The eight anti-parallel β -strands that compose the lipocalin β -barrel are labeled β_A – β_H . An additional β -strand (β_I) is localized towards the C-terminal end of the molecule. Loops (L) and α -helices are shown green and red, respectively. A disulfide bridge links the C-terminal end of the molecule to the β_D -strand indicated in sticks. The open and closed ends of rCan f 2 are indicated. (b) Top view of the molecule through the open end, showing access to the calyx as well as the bound glycerol moiety (GOL). Red spheres represent the two water molecules present in the calyx. (c) Topological representation of rCan f 2 labeled with structurally conserved regions at the closed end and variable loop regions at the open end of the ligand-binding site. Structurally conserved regions characteristic of the lipocalin fold are indicated in boxes.

the adjacent residues Gly86, Asn88 and Asn107, are conserved among lipocalins. Hydrogen bonds are formed between the side chains of Asn88 and Asn107, localized both beneath the L5 loop, and the carbonyl atoms of Tyr84 and Trp85, respectively. Similarly, the L1 loop is stabilized by hydrogen bonds formed between the conserved arginine residue Arg39 and residues Ile32 and Asp155 (Fig. 7a). Finally, van der Waals interactions between the side chains of the key residues Phe38 and Tyr84 and between the relatively conserved residues His37 and Trp85 seal the access to the calyx (Fig. 7a).

It should be noted that the conformation of the L1 and L5 loops in rCan f 2 is very conserved in the crystal structures of other aero-allergens with a bound ligand (Fig. 5; Table 2), suggesting that the residues described above might also play an important role in their ligand specificity. Modeling of the

mutation of tyrosine residues at position 84 to glycine in bovine and porcine odorant-binding proteins created a wider access to the internal cavity.³⁷ Furthermore, although the electron density for Phe38 demonstrates that this side chain takes only one conformation in rCan f 2, another conformation is possible (no steric hindrance was observed) that might open access to the calyx and allow for passage of small hydrophobic compounds.

The dog allergen rCan f 2 is a monomer at neutral pH

Previous structural studies revealed that some lipocalins are found as dimers, although there is no consistent pattern in their formation. However, it should be noted that similar to rCan f 2 in this study, other lipocalins such as the bovine allergen Bos d 2,⁴²

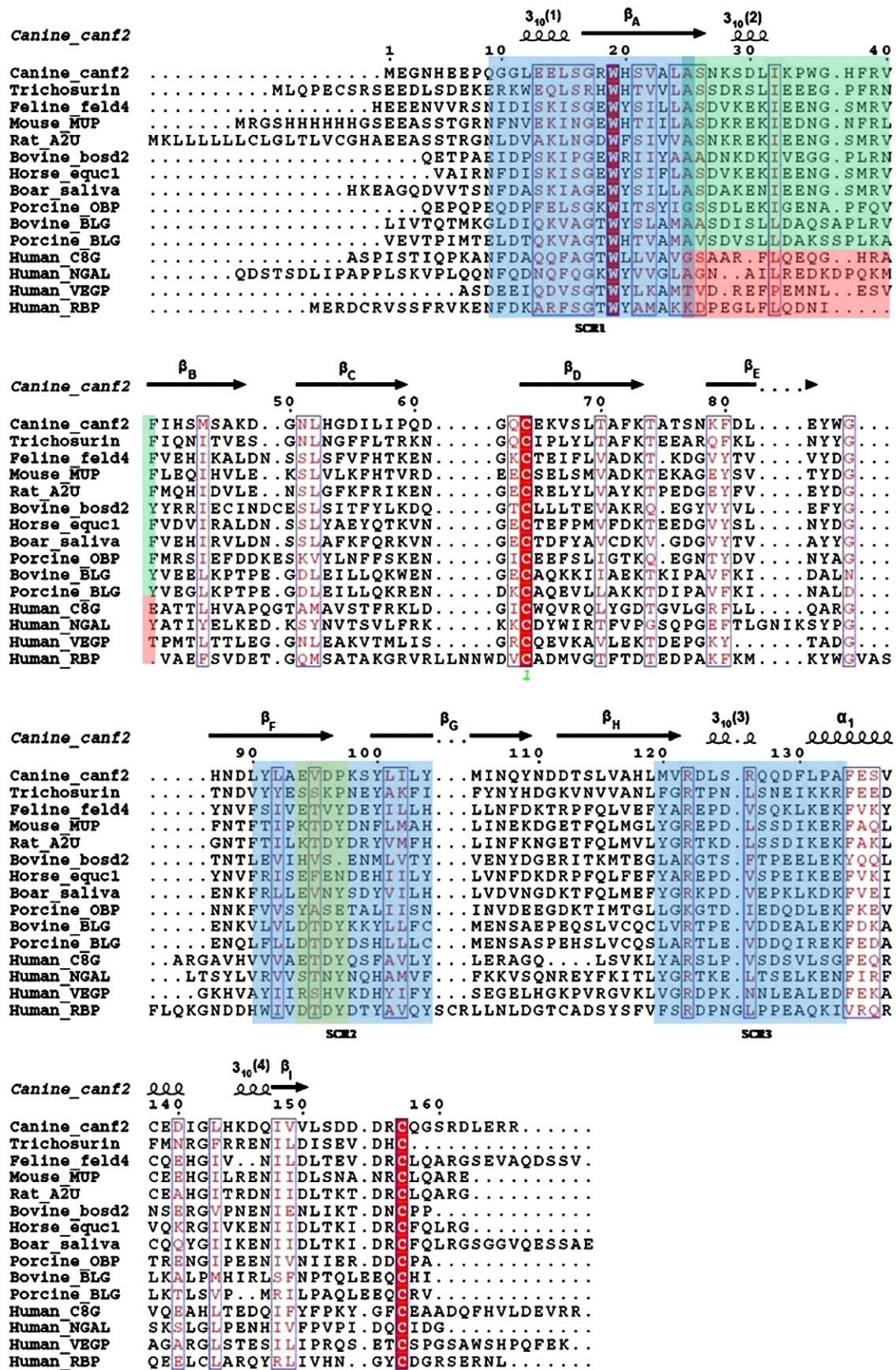


Fig. 3 (legend on next page)

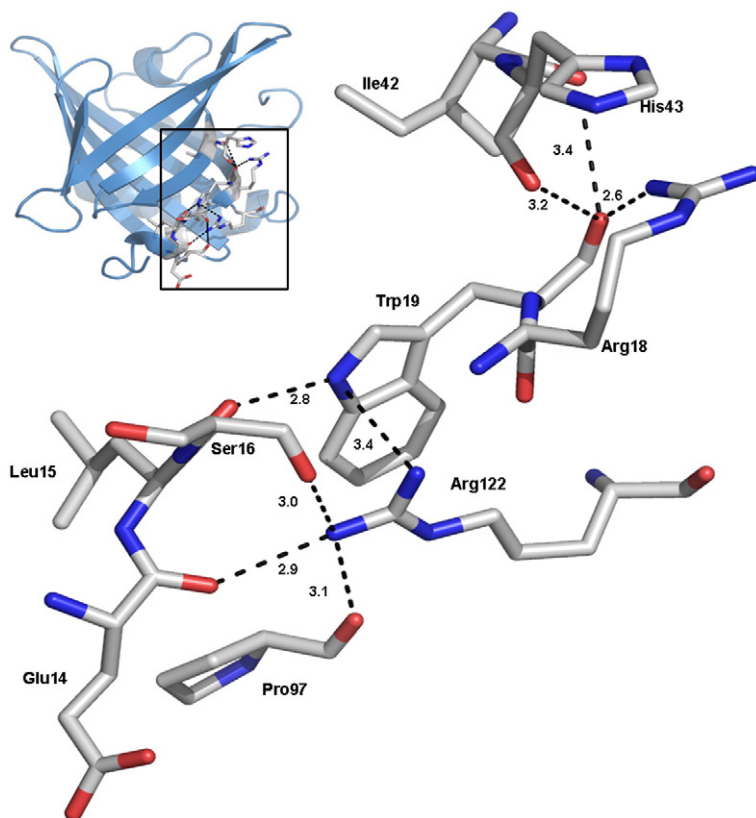


Fig. 4. Trp19, highly conserved among lipocalins, stabilizes the lipocalin fold. Residue Trp19 localized on the β_A -strand stabilized the overall fold of rCan f 2 through the formation of hydrogen bond and van der Waals interactions with the side chain of Arg122 localized on the β_H -strand. An intricate network of hydrogen bond interactions is formed between the centrally localized key residue Trp19, Arg122 and the surrounding residues Glu14, Leu15, Ser16, Arg18, Ile42, His 43 and Pro97.

the boar salivary lipocalin³⁸ and aphrodisin,³⁷ are found as monomers. Also, while some lipocalins can be found in both monomeric and dimeric forms, no clear correlation can be attributed to pH. Thus, while the bovine β -lactoglobulin forms a dimer at physiological pH and remains monomeric at acidic pH, the porcine β -lactoglobulin homologue is a dimer at acidic pH and exists as a monomer at neutral pH.^{38,43,44} Finally, in the case of odorant-binding proteins, the swapping of α -helices between monomers results in dimer formation.⁴⁵ Thus, we cannot exclude the possibility that the dog allergen exists as a dimer, potentially formed at lower pH values.

IgE from Can f 2-sensitized subjects cross-reacts with Fel d 4 from cat

A search for the highest level of structural homology between Can f 2 and other lipocalins was done with the program Dali.⁴⁶ The highest Z-scores (23.4–21.4) were provided only by other lipocalin allergens, demonstrating the structural homology of rCan f 2 to these aero-allergens (Table 2). However, the resemblance of rCan f 2 to MUP1,

Equ c 1 and A2U is striking due to the overall similarity of their calyces (Figs. 5 and 6b). In contrast, the structural similarity of rCan f 2 to human proteins was much lower, although some interesting hits were sampled for HC8 γ (human complement component-8 γ) (Z 16.4),⁴⁷ NGAL (Z 15.3),^{34,35} RBP (Z 14.1)⁴⁸ and VEGP, also known as tear lipocalin Tlc (Z 14.9).^{32,33} Comparative structural analysis reveals that although rCan f 2 and these human proteins have very similar folds, the shape of the calyces and the properties of the bound ligands are very different. Furthermore, the surface of these human proteins does not share any clear common feature with rCan f 2 (Fig. 5).

IgE cross-reactivity (using sera from five subjects with high Can f 2 reactivity) was tested towards a panel of mammalian aero-allergens including Fel d 4, Equ c 1 and Can f 1 as well as human lipocalin homologues such as RBP and NGAL. Although Can f 2 shares less than 22% sequence identity with the cat allergen Fel d 4,²⁵ our assay clearly demonstrated a patient-dependent cross-reactivity (up to 58%) between these two lipocalins (Fig. 8a). In contrast, no cross-reactivity was observed with Can f 1, Equ c 1 or any of the human homologues RBP and NGAL. Interestingly, although the aero-allergens Equ c 1

Fig. 3. Multiple sequence alignment of select members of the lipocalin family comprising trichosurin, major aero-allergens and human homologues to rCan f 2. Blue shading indicates the structurally conserved regions (SCR1–SCR3). Similarity scores are represented according to ESPript:⁷² white font with red frame, strictly identical residues (Trp19, Cys64 and Cys157); red font with blue frame, similar residues. Green shading indicates the major probable epitope region (25–41) that is conserved in all allergens. Red shading indicates the same region in human lipocalin homologues. Yellow shading indicates the second major epitope region (94–97) identified earlier in lipocalins. Secondary structure elements are indicated above the sequences.

Table 2. Comparative analysis of rCan f 2 with related lipocalins

Related lipocalin structures based on secondary structural superpositions	PDB code	Z-score	rmsd (Å)	Alignment length/no. residues	Pair-wise sequence identity (%)	Central cavity / calyx		Aromatic – amine interaction	Hydrogen bond (indole – hydrogen bond)
						Area (Å ²)	Volume (Å ³)		
Dog allergen, rCan f 2	2R73	21.5	1.70	142	34	386.8	454.2	Trp19 – Arg 122 Trp28 – Arg 131	Trp19 (NE1) – Ile 15(O) Trp28 (NE1) – Leu24 (O)
<i>A. Aero-allergens</i>									
MUP (mouse urinary protein)	1ZNL	23.4	1.24	142	27	418.6	499.6	Trp23 – Arg126	Trp23(NE1) – Ile19 (O)
α ₂ -globulin (major rat allergen)	2A2G	22.9	1.29	145	30	347.6	429.9	Trp19 – Arg122	Trp19(NE1) – Leu15 (O)
Bos d 2 (bovine lipocalin allergen)	1BI7	22.2	1.45	141	17	244.1	248.5	Trp16 – Lys119	Trp 16(NE1) – Ile12 (O)
Equ c 1 (major horse allergen)	1EW3	21.4	1.66	110	25	204.1	186.6	Trp37 – Arg141	Trp37(NE1) – Ile33 (O)
Boar salivary protein	1GM6	22.7	1.47	146	25	279.9	341.2	Trp22 – Arg125	Trp22(NE1) – Ile18 (O)
Porcine odorant-binding protein	1DZK	21.5	1.54	140	23	373.1	497.6	Trp16 – Lys120	Trp16(NE1) – Leu12 (O)
<i>B. Human structural homologues</i>									
C8γ (human complement component)	2QOS	16.4	1.73	125	23	1019.8	1888.2	Trp31 – Arg133	Trp31 (NE1) – Ph 27 (O)
NGAL (nutrophil gelatinase-associated lipocalin)	1X71	15.6	1.81	125	–	868.5	1752.4	Trp31 – Arg140	Trp31 (NE1) – Phe27 (O)
VEGP/Tlc (von Ebner's gland protein)	1XK1	14.9	1.95	115	22	758.2	1255.7	Trp17 – Arg118	Trp17 (NE1) – Val13 (O)
RBP (human retinol-binding protein)	1JYD	14.0	2.13	121	–	426.8	542.7	Trp24 – Arg139	Trp24 (NE1) – Phe20 (O)

The dimensions of the calyx are given comparing the major aero-allergens and human proteins to Can f 2. Aromatic – amine interactions are highly conserved in all the lipocalins. Residues that differ between lipocalins are in italics.

and Fel d 4 share more than 70% sequence identity, Can f 2-specific antibodies cross-reacted only with the latter.

The allergenic activity of rCan f 2 and rFel d 4 was also tested *in vitro* for the ability to induce basophil degranulation upon allergen cross-linking of FcεRI-bound allergen-specific IgE. Flow cytometric analysis of allergen-activated basophils from three asthmatic children sensitized to dog and cat revealed similar degranulation levels by both allergens in all three patients, resulting in up-regulation of CD63 with antigen concentrations as low as 10 ng/ml (Fig. 8b).

Although several crystal structures of lipocalins have been determined,^{7,8,23,42,43} the exact structural basis underlying the allergenicity of lipocalins remains unknown. Furthermore, the basis of IgE binding of many of these allergens remains obscure. As lipocalins display low levels of sequence homology to each other, the pairwise sequence identity often falls below 20%, which represents a minimal threshold level for a more reliable alignment.^{15,17} Initial multiple sequence alignment of most lipocalins used within this study indicated that the stretch of solvent-exposed residues 25–41 localized on the large loop L1, which acts as a lid to the calyx, is highly conserved in all the homologous aero-allergens (Figs. 3 and 7b). In contrast, this region varied significantly in the identified human homologues (HC8γ, NGAL,^{34,35} RBP^{20,48} and VEGP^{32,33}; Fig. 3). Indeed, two major epitope regions, corresponding to the stretches of residues 25–40 and 92–100, have been identified in the major bovine milk allergen Bos d 5,⁴⁹ which is also homologous to Can f 2. However, as mentioned above, our functional assays demonstrated that IgE from Can f 2-sensitized individuals did cross-react only with rFel d 4, impairing our initial epitope predictions. Careful further comparative structural and sequential analysis of rCan f 2 with the cross-reacting Fel d 4 and the non cross-reacting Equ c 1 combined with epitope predictions using the program Discotope (available from the Immune Epitope Database^{50,51}) suggested a conformational epitope formed by the three stretches of residues:

94–104 (EVDPKSYLILY)
122–132 (RDLSRQQDFLP)
134–141 (FESVCEDI)

as a potential site for cross-reactivity between Can f 2 and Fel d 4. This solvent-exposed region, predominantly basic in Equ c 1, is more negatively charged in both rCan f 2 and Fel d 4 (Fig. 8c). Can f 2 and Fel d 4 share a set of five solvent-exposed residues that differ with Equ c1 within this region. In contrast to Equ c 1, where the side chain of the phenylalanine residue Phe114 occupies the center of the presumptive epitope, the side chains of the glutamate residues Glu94 and Glu95 in Can f 2 and Fel d 4, respectively,

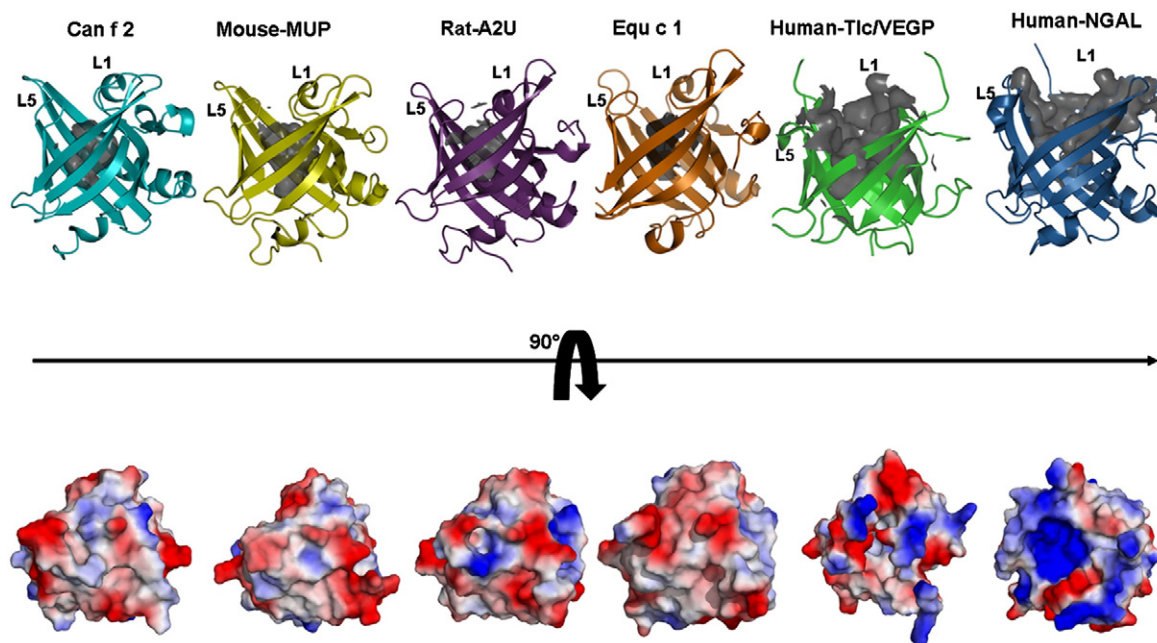


Fig. 5. The central elliptical calyx and the overall fold of rCan f 2 are very similar to other aero-allergens in shape and properties. The calyx of rCan f 2, indicated in gray, is very similar in size and shape to other aero-allergens, such as MUP1, A2U and Equ c 1. In contrast, the human homologues VEGP and NGAL display a much wider access to the central shallow calyx. rCan f 2 displays a high level of structural homology to other aero-allergens but its surface electrostatic potential varies drastically when compared to human homologues. Comparison of the surface electrostatic potential at the entrance to the calyx (top end) reveals that, in contrast to all the aero-allergens, the entry to the calyx in the human homologues is widely accessible to the solvent. Furthermore, the residues that line the calyx in the human homologues are highly charged. It should be noted that the hydrophobic internal calyx in the allergens presented is hidden by the L1 and L5 loops, which cover and close its entry. All molecules were rotated 90° in order to be viewed from the open end of the calyx.

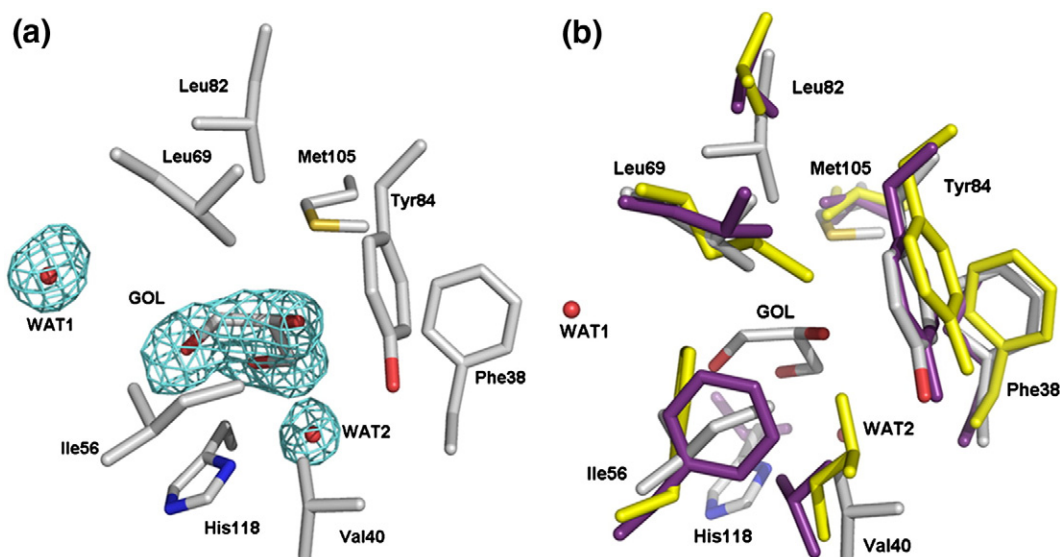


Fig. 6. A glycerol moiety is bound within the hydrophobic ligand-binding cavity of rCan f 2. (a) A glycerol moiety at the centre of the calyx is represented within the electron density, contoured at 1.2σ . The hydrophobic residues that line the calyx and surround the glycerol molecules are also represented. Two well-ordered water molecules, WAT1 and WAT2, are also present within the calyx of rCan f 2. (b) The relative conservation of hydrophobicity within the calyx of rCan f 2, MUP-1 and A2U indicates that similar molecules might bind to all three allergens. Residues lining the calyces of rCan f 2 (white atomic color mode), MUP1 (yellow) and A2U (purple) are very well conserved between these three aero-allergens, indicating that rCan f 2 might also act as a carrier of small hydrophobic molecules and pheromones. Phe38 and Tyr84, which control access to the calyces, are also highly conserved among these allergens. The glycerol moiety (GOL) and the water molecules (WAT1 and WAT2) present in the calyx of rCan f 2 are indicated.

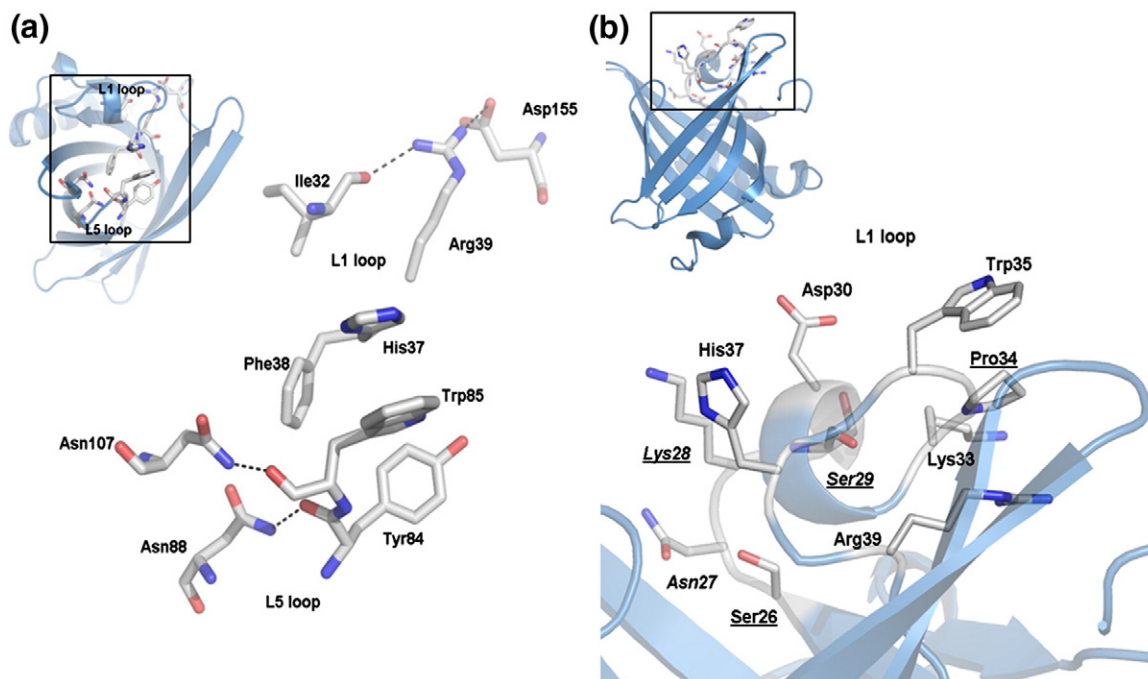


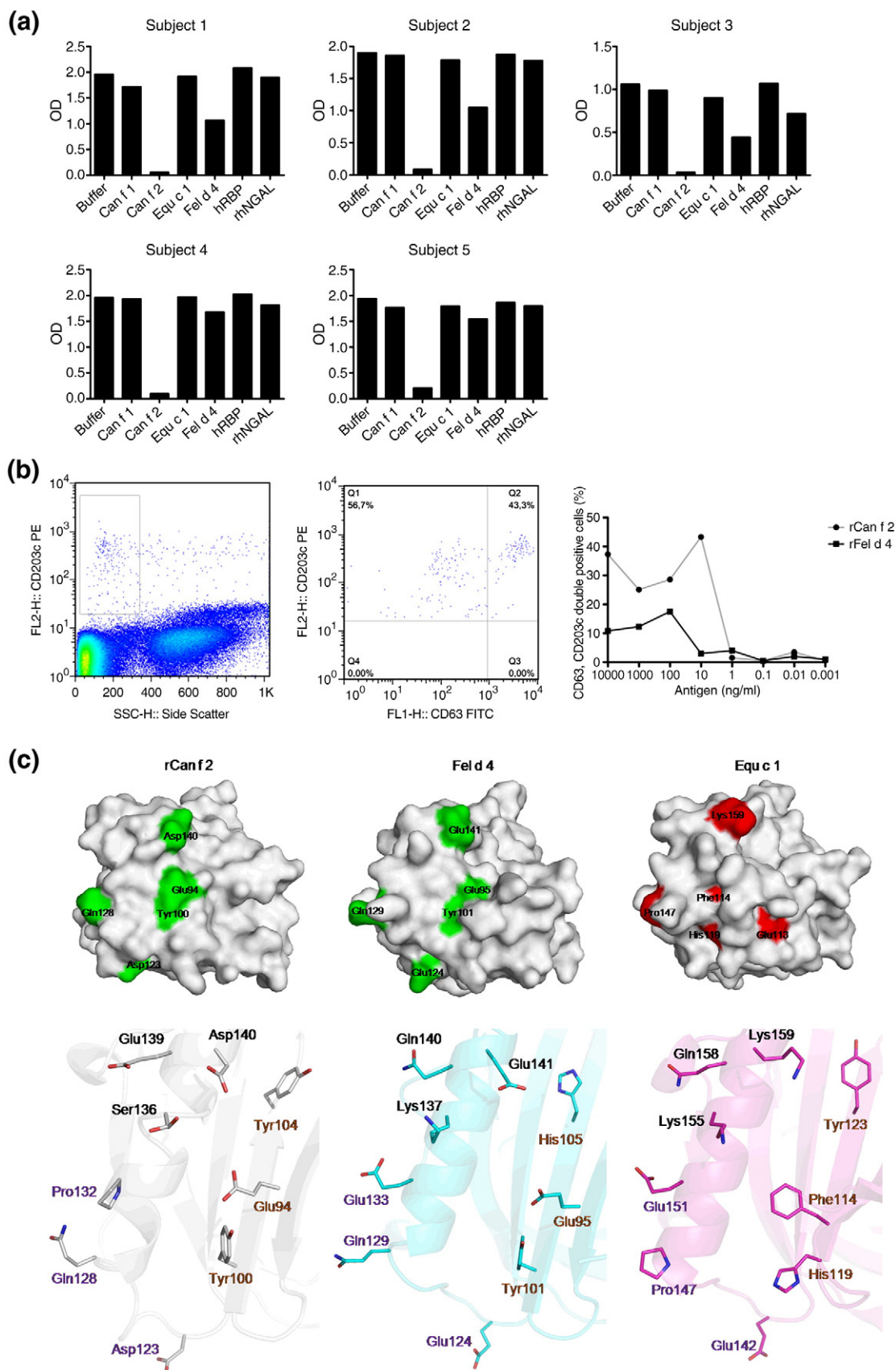
Fig. 7. (a) Loops L1 and L5 control access to the calyx of rCan f 2. Lipocalin-conserved residues Phe38 and Tyr84 control access to the calyx of rCan f 2. Aromatic–aromatic interactions are formed between the side chains of Tyr84 and Phe38 as well as between Trp85 and His37, effectively closing the passage to the calyx of rCan f 2. The conformation of the L1 and L5 loops is further stabilized by hydrogen bond interactions between Arg39, Ile32 and Asp155, as well as between the side chains of Asn88 and Asn107 and the carbonyl atoms of Tyr84 and Trp85. (b) The L1 loop acts as a lid over the open end of the calyx. The solvent-exposed side chains from the stretch of residues 25–41 (represented as sticks in white atomic color mode) that are localized on loop L1 and might represent a common epitope are represented. The schematic representation of the rest of the rCan f 2 molecule is in blue. Residues that are highly conserved among lipocalin allergens are underlined and the residues potentially involved with glycosylation are in italics.

render this region more negatively charged. The four additional sets of residues Tyr100/Tyr101, Asp123/Glu124, Gln128/Gln129 and Asp140/Glu141 shared between Can f 2 and Fel d 4, respectively, might play a key role in the formation of this potential conformational epitope. Importantly, the negatively charged side chains of Asp140 and Glu141 are replaced by the positively charged side chain of the lysine residue Lys159 in Equ c 1. Similarly, the glutamine residues at position 128 and 129 in Can f 2 and Fel d 4, respectively, are replaced by a proline residue in Equ c 1. Distances between surface-exposed residues from adjacent stretches of linear epitopes are <6 Å, which is one of the pre-requisites for a conformational epitope (Fig. 8c).

In conclusion, at least four non-exclusive hypotheses could explain the observed cross-reactivity of Can f 2-specific IgE to rFel d 4. First, as suggested

above, a solvent-exposed IgE conformational epitope common only to rCan f 2 and Fel d 4 could provide a template for IgE cross-reactivity. Second, one could speculate that the quality of the bound ligand could result in a conformational change that would create a *de novo* epitope on the surfaces of some of the lipocalins but not on others. This hypothesis is, in our opinion, viable between rCan f 2 and Fel d 4 because the residues that line the calyces are conserved between these two lipocalins. However, this does not explain why other aero-allergens, such as Equ c 1, with a very similar calyx do not cross-react. Third, the dimerization of the aeroallergen may result in the creation of a new epitope region. Interestingly, Equ c 1 forms a dimer whereas both rCan f 2 and rFel d 4 are monomers. It should be noted also that the Equ c 1 residues Lys159 and Arg160 are part of the dimerization

Fig. 8. rCan f 2 cross-reacts with the cat lipocalin allergen Fel d 4. (a) IgE antibodies to Can f 2 in sera from five sensitized subjects were incubated with lipocalin aero-allergens and human lipocalin proteins. IgE antibody cross-reactivity was observed with rFel d 4 but not with the aero-allergens rCan f 1 and Equ c 1 nor with the human lipocalins RBP and NGAL. (b) Whole blood cells from one representative asthmatic child sensitized to dog and cat were stimulated with rCan f 2 and rFel d 4. A flow cytometric analysis of gated CD203c⁺ basophils is presented (upper left panel). Activated basophils are displayed as the percentage of CD63⁺ cells among 200 gated CD203c⁺ cells (upper right panel). Induction of dose-dependent allergen activation (*y*-axis) of basophils (lower panel). (c) Upper: Molecular surface representation of the crystal structures of Can f 2 and Equ c 1 compared to a molecular model of Fel d 4. The solvent-exposed residues that are similar between Can f 2 and Fel d 4 are in green. The corresponding residues that differ in Equ c 1 are in red. Lower: Comparative structural analysis suggests the presence of a conformational epitope formed by the three different stretches of residues 94–104 (EVDPKSYLILY), 122–132 (RDLRQQDFLP) and 134–141 (FESVCEDI). The side chains of the solvent-exposed residues from each stretch of residues are displayed as sticks.



interface. Thus, the dimerization of Equ c 1 might hinder the cross-reactivity with rCan f 2. Finally, the last possibility involves the potential polyreactivity of antibody molecules.⁵² At least 20% of B-lymphocytes in peripheral blood make polyreactive antibodies, the role of which is still unclear when it comes to the initiation of autoimmunity and allergic responses. It should be noted that at least two studies have demonstrated the ability of antibodies to recognize poorly related⁵³ or totally unrelated⁵⁴ targets. Thus, one could imagine a scenario in which Can f 2-specific IgE antibodies could also bind to Fel d 4 through structural re-arrangements.⁵⁵

Concluding remarks

The present study provides a structural characterization of the major dog aeroallergen rCan f 2 and a comparative analysis of other major aero-allergens and human proteins. The crystal structure of rCan f 2 shows a classical lipocalin fold, exhibiting all the typical features ascribed to this protein family. Residue Trp19 in the highly conserved SCR1 region acts as a pillar residue, keeping the overall topology of the monomeric rCan f 2 intact. The conserved residues Phe38 and Tyr84 localized on the L1 and L5 loops, respectively, control access to the highly hydrophobic calyx. Although the shape of the rCan f 2 calyx as well as the properties of the residues lining the cavity are nearly identical with the aero-allergens MUP1, Equ c 1 and A2U, functional analysis using sera from five Can f 2-sensitized subjects demonstrates that IgE antibodies to Can f 2 cross-reacted exclusively with the cat lipocalin allergen Fel d 4, suggesting a potential role for these allergens in co-sensitization between cat and dog. Careful structural comparison of the structures of Can f 2 with Fel d 4 and Equ c 1 suggested a negatively charged conformational epitope surrounding the H1 domain as a potential site for cross-reactivity.

Materials and Methods

Sera

Sera from 89 patients with IgE responses >5 kU_A/L (ImmunoCAP™, Phadia, Uppsala, Sweden) to dog dander extract, collected from the diagnostics department at the Karolinska University Hospital, Solna, were used to analyze IgE binding to the dog allergens Can f 1, Can f 2, Can f 3 and the cat allergen Fel d 4. Sera from 21 asthmatic children were analyzed for specific IgE levels using immuno solid-phase allergen chip 103 technology (Phadia).⁵⁶ The collection of sera for this study was approved by the local ethics committee at the Karolinska University Hospital.

Production and analysis of polyclonal anti-Can f 2 antibodies

Four female BALB/c mice (Charles River Sulzfeld, Germany) were injected three times with 25 µg of Can f 2 in alum, blood was collected and allergen-specific IgG

antibodies were purified essentially as described.^{57,58} Animal experiments were approved by the Swedish local ethics committee for animal welfare. ELISA was performed as described.² Briefly, 96-well ELISA plates (Nunc, Roskilde, Denmark) were coated with 100 µl of 5 µg/ml of either Can f 1 or Can f 2/well and blocked with 1% (w/v) BSA in PBS. Purified mouse antibodies (diluted 1/1000) were added in fivefold dilutions to the wells. After 2 h, IgG was detected using AP-conjugated rabbit anti-mouse IgG (diluted 1:1000, Southern Biotech, Birmingham, AL, USA). AP tablets (Sigma-Aldrich, Steinheim, Germany) were used for color development followed by measurement of absorbance at 405 nm. No binding of Can f 2 antibodies to rCan f 1 was detected.

Allergen preparations and human lipocalins

Recombinant Can f 2 was produced and purified as described.³⁷ A synthetic gene corresponding to rCan f 1 (SwissProt accession no **O18873**) was produced by PCR using overlapping oligonucleotides and DNAWorksII, while the gene corresponding to rFel d 4 (SwissProt accession no **Q5VFH6**) was purchased from Geneart (Regensburg, Germany). The sequenced fragments of Can f 1 and Fel d 4 were subcloned into the vector pET20b (Novagen, EMD Chemicals Inc., Darmstadt, Germany) using the NdeI and XhoI restriction sites, and expressed in *Escherichia coli* BL21(DE3)pLysS (Novagen). The His₆-tagged proteins were purified as described.³⁷ Natural (n) Can f 2 was prepared by aqueous extraction from dog dander (Allergon, Ängelholm, Sweden).⁵⁹ The crude extract 50 ml was concentrated to 3 ml using a YM-3 membrane (Amicon, Inc, Beverly, MA, USA) and nCan f 2 was isolated by size-exclusion chromatography using a 16/60 Superdex 200 pg column (GE Healthcare, Uppsala, Sweden) equilibrated with PBS pH 7.4 on an ÄKTA Purifier. The fractions containing nCan f 2 were detected using the previously obtained polyclonal mouse anti-rCan f 2 IgG-antibodies. Natural (n) Equ c 1 was prepared through aqueous extraction of dander and hair from 10 horses.⁵⁹ The aqueous extract (500 ml) was concentrated to 10 ml and 2 ml was diluted 10-fold in 20 mM Tris-HCl pH 7.4 and nEqu c 1 was purified on a 10 ml Q-Sepharose HP column (GE Healthcare). Proteins were separated with a linear gradient of 15 column volumes of elution buffer (20 mM Tris-HCl, 1 M NaCl, pH 7.4). The fractions were analyzed by SDS-PAGE under reducing and non-reducing conditions, and a pure 22 kDa band corresponding to nEqu c 1,⁶⁰ was isolated and subjected to size-exclusion chromatography. Protein identity was confirmed by electrospray mass spectrometry following tryptic digestion.⁶⁰ Natural Can f 3 (albumin) was prepared from dog serum (a kind gift from Statens Veterinärmedicinska Anstalt, Uppsala, Sweden) using size-exclusion chromatography (16/60 Superdex 200 pg, GE Healthcare) equilibrated with PBS, pH 7.4. The fractions corresponding to the 66 kDa albumin were collected, pooled and concentrated. nRBP was purchased from Sigma-Aldrich, Steinheim, Germany and rNGAL was purchased from ProtEra, Firenze, Italy.

Circular dichroism (CD) spectroscopy

Protein concentrations were determined by measuring absorbance at 280 nm (NanoDrop ND-1000 spectrophotometer; SaveenWerner, Limhamn, Sweden) using calculated

extinction coefficients.⁶¹ CD experiments for all allergens were done at 20 °C in 20 mM KH₂PO₄/K₂HPO₄ (pH 7.5) using 0.12 mg/ml of purified protein. The spectra were recorded in a 1 mm pathlength cell with an JASCO spectrometer equipped with a thermoelectric temperature controller. The shapes of the CD spectra for the soluble proteins used in this study indicated that they were all well folded (Supplementary Data Figure 1).

Quantification of specific IgE antibodies towards dog and cat allergens

Quantitative ELISA measurements were done as follows. Duplicates of seven calibrators prepared from an IgE myeloma with 640 kU/l of IgE (threefold dilutions, range 0.035–25.5 kU_A/l), two control samples and patient samples were added to a 96-well plate and incubated for 2 h. The wells containing IgE-calibrators and control samples were coated with 100 µl of 2.5 µg/ml monoclonal anti-human IgE (a kind gift from Phadia, Uppsala, Sweden) and wells containing patient sera were coated with 100 µl of 5 µg/ml of the respective allergen (Can f 1, Can f 2 or Can f 3) in 0.1 M carbonate buffer, pH 9.6. Secondary antibody, rabbit anti-human IgE (MIAB, Uppsala, Sweden) and conjugates alkaline phosphatase (AP)-labeled goat anti-rabbit antibodies (Dako, Glostrup, Denmark) were incubated for 2 h and 1 h, respectively. AP tablets (Sigma-Aldrich) were used for color development followed by measurement at 405 nm.

Comparison of IgE cross-reactivity between rCan f 2 and other mammalian lipocalins by ELISA inhibition

A serum pool from three dog-sensitized patients with IgE-reactivity to rCan f 2 was used to examine the allergenic similarities between the recombinant and natural forms of Can f 2 by ELISA inhibition. The serum pool was diluted to a final concentration of 1 kU_A/l IgE to dog extract (ImmunoCAP, Phadia) and 10 threefold dilutions of the respective antigen (range 10⁻¹–10⁻⁵ µg/ml) were added 1:2 (v/v) and incubated for 2 h at room temperature. IgE binding was analyzed by ELISA using rCan f 2-coated plates. Sera from 10 sensitized individuals with a confirmed reactivity to rCan f 2 were diluted to 1 kU/l. To each serum, 110 µl of 250 µg/ml (1:2 v/v) of the respective lipocalin, (rCan f 1, rCan f 2, rFel d 4, nEqu c 1 and the human homologues RBP and NGAL) were added and IgE was measured as described above. Can f 3 was used as a control allergen and did not show any inhibition (data not shown).

Basophil activation test

Allergen-specific basophil degranulation was analyzed by monitoring the basophil activation markers CD203c and CD63 as described.⁶² Briefly, 10-fold serial dilutions of either rCan f 2 or rFel d 4 (10 µg/ml to 10⁻⁷ µg/ml), medium (negative control) and rabbit anti-human IgE (positive control) were added to venous blood samples from three asthmatic children determined to be sensitized to dog and cat. The samples were further incubated with anti-CD63 and anti-CD203c monoclonal antibodies (Immunotech, Marseille, France) and analyzed by flow cytometry using a FACS Calibur flow cytometer (BD Biosciences, San Jose, CA, USA). Basophils were identified by gating for CD203c⁺ cells and the magnitude of allergen-

activation was calculated as the percentage of CD63⁺ cells among the gated basophils.

Production and crystallization of rCan f 2

Protein production and purification were performed as described.⁴⁰ Briefly, recombinant Can f 2 (rCan f 2) was produced in *E. coli* BL21(DE3)pLysS with a His-tag localized at the C-terminus. It was isolated using a combination of immobilized metal-affinity, gel-filtration and ion-exchange chromatography. A minimum of 70 mg of highly purified monomeric rCan f 2 was obtained from 2 l of bacterial culture. Crystals of rCan f 2 (4.6 mg/ml) were obtained by the hanging-drop, vapor-diffusion method as described.⁴⁰

Data collection, structure determination and refinement

The crystals were cryoprotected in 20% (v/v) glycerol and flash-frozen in a cold nitrogen stream. The crystals belonged to space group C22₁ with unit cell dimensions *a* 68.2 Å, *b* 76.9 Å and *c* 65.0 Å. The asymmetric unit cell contained one rCan f 2 molecule with a molecular mass of 18.9 kDa, giving a solvent content of 51.6% (v/v). The crystals diffracted to 1.45 Å resolution. A total of 360 images were processed using beam line ID14-2 at the synchrotron radiation facility at ESRF (Grenoble, France) with 0.5° oscillation/image. The space group and unit cell parameters were determined using the auto-indexing option of MOSFLM.⁶³ Scaling and reduction of data was done with SCALA from the CCP4 suite of programs (Table 1).⁶⁴ The 3D structure of rCan f2 was determined by molecular replacement using the program PHASER.⁶⁵ The atomic coordinates of the metatherian *Trichosurus vulpeculus* milk whey protein (PDB ID 2R73), which shares 34% identity with rCan f2 were used as the search model. For validation, 5% of reflections were chosen at random and set aside. The solution obtained was fed into ARP-wARP⁶⁶, resulting in initial *R*_{cryst} of 18.1% and *R*_{free} of 20.5%. After each round of refinement, errors in the model were identified through careful examination of the 2*F*_o – *F*_c and *F*_o – *F*_c electron density maps with the program Coot.⁶⁷ Missing residues were added in successive cycles of manual building followed by restrained refinement cycles in REFMAC5.⁶⁸ Anisotropic refinement within the last stage of refinement resulted in a decrease of *R*_{cryst} and *R*_{free} values of almost 1%. Riding hydrogens were generated in the last round of REFMAC refinement. Water molecules were added to well defined peaks (2.0 σ and greater in *F*_o – *F*_c maps) found between 2 Å and 4 Å from O or N atoms. The final refinement parameters are given in Table 1. A total of 93.5% and 6.5% of residues lies within the preferred and allowed regions of a Ramachandran plot as defined by the program Coot. The figures were created using the programs PyMOL[†] and TopDraw.⁶⁹ Secondary structural elements were assigned using the program DSSP.⁷⁰ The program CASTp⁷¹ was used to calculate voids. Multiple sequence alignment was performed using ESPript.⁷²

Molecular modelling of Fel d 4

The molecular model of Fel d 4 was created on the basis of its sequence homology to Equ c 1 (69% sequence

[†]www.pymol.org

identity) using the SWISS-MODEL protein modelling server^{a,73}

Data Bank accession numbers

The co-ordinates of the molecular model of Fel d 4 have been registered within the PMDB protein Database Model with accession code **PM0076251**. The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ, USA with accession code 3L4R.

Acknowledgements

We gratefully acknowledge access to synchrotron beamlines 14.2 at ESRF (Grenoble, France), MX-41 at BESSY (Berlin, Germany) and I911 at MAX II (Lund, Sweden). This work was supported by grants from the Swedish Asthma and Allergy Research and Konsul Th C Bergh's Foundations, Center for Allergy Research and the Bernard Osher Initiative for Research on Severe Asthma at the Karolinska Institute. The research leading to these results has received funding from the European community's stretch framework program (FP7/2007-2013) under grant agreement no. 226716.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2010.05.043](https://doi.org/10.1016/j.jmb.2010.05.043)

References

- Munir, A. K., Einarsson, R., Schou, C. & Dreborg, S. K. (1993). Allergens in school dust. I. The amount of the major cat (Fel d 1) and dog (Can f 1) allergens in dust from Swedish schools is high enough to probably cause perennial symptoms in most children with asthma who are sensitized to cat and dog. *J. Allergy Clin. Immunol.* **91**, 1067–1074.
- Kaiser, L., Velickovic, T. C., Badia-Martinez, D., Adedoyin, J., Thunberg, S., Hallen, D. *et al.* (2007). Structural characterization of the tetrameric form of the major cat allergen Fel d 1. *J. Mol. Biol.* **370**, 714–727.
- Kaiser, L., Grönlund, H., van Hage-Hamsten, M. & Achour, A. (2005). Crystallization and preliminary crystallographic data of a Fel d 1 (1+2) construct corresponding to the major allergen from cat. *Acta Crystallogr. F*, **61**, 232–234.
- Kaiser, L., Grönlund, H., Sandalova, T., Ljunggren, H. G., Achour, A., Schneider, G. & van Hage-Hamsten, M. (2003). Three-dimensional structure of Fel d 1, the major allergen in cat. *Int. Arch. Allergy Immunol.* **132**, 25–26.
- Kaiser, L., Grönlund, H., Sandalova, T., Ljunggren, H. G., van Hage-Hamsten, M., Achour, A. & Schneider, G. (2003). The crystal structure of the major cat allergen Fel d 1, a member of the secretoglobulin family. *J. Biol. Chem.* **278**, 37730–37735.
- Kaiser, L., Grönlund, H., Sandalova, T., Ljunggren, H. G., Schneider, G., van Hage-Hamsten, M. & Achour, A. (2003). Production, crystallization and preliminary crystallographic study of the major cat allergen Fel d 1. *Acta Crystallogr. D*, **59**, 1103–1105.
- Bacchini, A., Gaetani, E. & Cavagioni, A. (1992). Pheromone binding proteins of the mouse, *Mus musculus*. *Experientia*, **48**, 419–421.
- Lascombe, M. B., Gregoire, C., Poncet, P., Tavares, G. A., Rosinski-Chupin, I. & Rabillon, J. (2000). Crystal structure of the allergen Equ c 1. A dimeric lipocalin with restricted IgE-reactive epitopes. *J. Biol. Chem.* **275**, 21572–21577.
- Konieczny, A., Morgenstern, J. P., Bizinkauskas, C. B., Lilley, C. H., Brauer, A. W., Bond, J. F. *et al.* (1997). The major dog allergens, Can f 1 and Can f 2, are salivary lipocalin proteins: cloning and immunological characterization of the recombinant forms. *Immunology*, **92**, 577–586.
- Saarelainen, S., Taivainen, A., Rytönen-Nissinen, M., Auriola, S., Immonen, A., Mantylarvi, R. *et al.* (2004). Assessment of recombinant dog allergens Can f 1 and Can f 2 for the diagnosis of dog allergy. *Clin. Exp. Allergy*, **34**, 1576–1582.
- de Groot, H., Goei, K. G. H., van Swieten, P. & Aalberse, R. C. (1991). Affinity purification of a major and a minor allergen from dog extract: serologic activity of affinity-purified Can f I and of Can f I-depleted extract. *J. Allergy Clin. Immunol.* **87**, 1056–1065.
- Schou, C., Svendsen, U. G. & Lowenstein, H. (1991). Purification and characterization of the major dog allergen, Can f I. *Clin. Exp. Allergy*, **21**, 321–328.
- Mantylarvi, R., Rautiainen, J. & Virtanen, T. (2000). Lipocalins as allergens. *Biochim. Biophys. Acta*, **1482**, 308–317.
- Virtanen, T., Zeiler, T. & Mantylarvi, R. (1999). Important animal allergens are lipocalin proteins: why are they allergenic? *Int. Arch. Allergy Immunol.* **120**, 247–258.
- Flower, D. R. (1996). The lipocalin protein family: structure and function. *Biochem. J.* **318**, 1–14.
- Flower, D. R. (2000). Experimentally determined lipocalin structures. *Biochim. Biophys. Acta*, **1482**, 46–56.
- Flower, D. R., North, A. C. & Sansom, C. E. (2000). The lipocalin protein family: structural and sequence overview. *Biochim. Biophys. Acta*, **1482**, 9–24.
- Flower, D. R. (1995). Multiple molecular recognition properties of the lipocalin protein family. *J. Mol. Recognit.* **8**, 185–195.
- Zanotti, G., Berni, R. & Monaco, H. L. (1993). Crystal structure of liganded and unliganded forms of bovine plasma retinol-binding protein. *J. Biol. Chem.* **268**, 10728–10738.
- Zanotti, G., Ottonello, S., Berni, R. & Monaco, H. L. (1993). Crystal structure of the trigonal form of human plasma retinol-binding protein at 2.5 Å resolution. *J. Mol. Biol.* **230**, 613–624.
- Bocskai, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E., Cavagioni, A. *et al.* (1992). Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature*, **360**, 186–188.

^a <http://swissmodel.expasy.org/workspace>

22. Akerstrom, B., Flower, D. R. & Salier, J. P. (2000). Lipocalins: unity in diversity. *Biochim. Biophys. Acta*, **1482**, 1–8.
23. Gregoire, C., Tavares, G. A., Lorenzo, H. K., Dandeu, J. P., David, B. & Alzari, P. M. (1999). Crystallization and preliminary crystallographic analysis of the major horse allergen Equ c 1. *Acta Crystallogr. D*, **55**, 880–882.
24. Rautiainen, J., Auriola, S., Rouvinen, J., Kauppinen, J., Zeiler, T., Novikov, D. *et al.* (1998). Molecular and crystal properties of Bos d 2, an allergenic protein of the lipocalin family. *Biochem. Biophys. Res. Commun.* **247**, 746–750.
25. Smith, W., Butler, A. J., Hazell, L. A., Chapman, M. D., Pomes, A., Nickels, D. G. & Thomas, W. R. (2004). Fel d 4, a cat lipocalin allergen. *Clin. Exp. Allergy*, **34**, 1732–1738.
26. Cowan, S. W., Newcomer, M. E. & Jones, T. A. (1990). Crystallographic refinement of human serum retinol binding protein at 2 Å resolution. *Proteins: Struct. Funct. Bioinformatics*, **8**, 44–61.
27. Newcomer, M. E., Jones, T. A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L. & Peterson, P. A. (1984). The three-dimensional structure of retinol-binding protein. *EMBO J.* **3**, 1451–1454.
28. Monaco, H. L., Zanotti, G., Spadon, P., Bolognesi, M., Sawyer, L. & Eliopoulos, E. E. (1987). Crystal structure of the trigonal form of bovine beta-lactoglobulin and of its complex with retinol at 2.5 Å resolution. *J. Mol. Biol.* **197**, 695–706.
29. Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C., Findlay, J. B., Sivaprasadarao, R. *et al.* (1986). The structure of beta-lactoglobulin and its similarity to plasma retinol-binding protein. *Nature*, **324**, 383–385.
30. Brownlow, S., Morais Cabral, J. H., Cooper, R., Flower, D. R., Yewdall, S. J., Polikarpov, I. *et al.* (1997). Bovine beta-lactoglobulin at 1.8 Å resolution—still an enigmatic lipocalin. *Structure*, **5**, 481–495.
31. Eichinger, A., Nasreen, A., Kim, H. J. & Skerra, A. (2007). Structural insight into the dual ligand specificity and mode of high density lipoprotein association of apolipoprotein D. *J. Biol. Chem.* **282**, 31068–31075.
32. Breustedt, D. A., Korndorfer, I. P., Redl, B. & Skerra, A. (2005). The 1.8 Å crystal structure of human tear lipocalin reveals an extended branched cavity with capacity for multiple ligands. *J. Biol. Chem.* **280**, 484–493.
33. Breustedt, D. A., Chatwell, L. & Skerra, A. (2009). A new crystal form of human tear lipocalin reveals high flexibility in the loop region and induced fit in the ligand cavity. *Acta Crystallogr. D*, **65**, 1118–1125.
34. Coles, M., Diercks, T., Muehlenweg, B., Bartsch, S., Zolzer, V., Tschesche, H. & Kessler, H. (1999). The solution structure and dynamics of human neutrophil gelatinase-associated lipocalin. *J. Mol. Biol.* **289**, 139–157.
35. Holmes, M. A., Paulsene, W., Jide, X., Ratledge, C. & Strong, R. K. (2005). Siderocalin (Lcn 2) also binds carboxymycobactins, potentially defending against mycobacterial infections through iron sequestration. *Structure*, **13**, 29–41.
36. Sansom, C. E., North, A. C. & Sawyer, L. (1994). Structural analysis and classification of lipocalins and related proteins using a profile-search method. *Biochim. Biophys. Acta*, **1208**, 247–255.
37. Vincent, F., Lobel, D., Brown, K., Spinelli, S., Grote, P., Breer, H. *et al.* (2001). Crystal structure of aphrodisin, a sex pheromone from female hamster. *J. Mol. Biol.* **305**, 459–469.
38. Spinelli, S., Vincent, F., Pelosi, P., Tegoni, M. & Cambillau, C. (2002). Boar salivary lipocalin. Three-dimensional X-ray structure and androsterol/androstenedione docking simulations. *Eur. J. Biochem.* **269**, 2449–2456.
39. Greene, L. H., Chrysina, E. D., Irons, L. I., Papageorgiou, A. C., Acharya, K. R. & Brew, K. (2001). Role of conserved residues in structure and stability: tryptophans of human serum retinol-binding protein, a model for the lipocalin superfamily. *Protein Sci.* **10**, 2301–2316.
40. Madhurantakam, C., Nilsson, O. B., Jonsson, K., Grönlund, H. & Achour, A. (2009). Production, crystallization and preliminary X-ray diffraction analysis of the allergen Can f 2 from *Canis familiaris*. *Acta Crystallogr. F*, **65**, 467–471.
41. Chaudhuri, B. N., Kleywegt, G. J., Bjorkman, J., Lehman-McKeeman, L. D., Oliver, J. D. & Jones, T. A. (1999). The structures of alpha 2u-globulin and its complex with a hyaline droplet inducer. *Acta Crystallogr. D*, **55**, 753–762.
42. Rouvinen, J., Rautiainen, J., Virtanen, T., Zeiler, T., Kauppinen, J., Taivainen, A. & Mantylarvi, R. (1999). Probing the molecular basis of allergy. three-dimensional structure of the bovine lipocalin allergen Bos d 2. *J. Biol. Chem.* **274**, 2337–2343.
43. Spinelli, S., Ramoni, R., Grolli, S., Bonicel, J., Cambillau, C. & Tegoni, M. (1998). The structure of the monomeric porcine odorant binding protein sheds light on the domain swapping mechanism. *Biochemistry*, **37**, 7913–7918.
44. Hoedemaeker, F. J., Visschers, R. W., Alting, A. C., de Kruif, K. G., Kuil, M. E. & Abrahams, J. P. (2002). A novel pH-dependent dimerization motif in beta-lactoglobulin from pig (*Sus scrofa*). *Acta Crystallogr. D*, **58**, 480–486.
45. Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S. & Cambillau, C. (1996). Domain swapping creates a third putative combining site in bovine odorant binding protein dimer. *Nat. Struct. Biol.* **3**, 863–867.
46. Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480.
47. Lovelace, L. L., Chiswell, B., Slade, D. J., Sodetz, J. M. & Lebiada, L. (2008). Crystal structure of complement protein C8γ in complex with a peptide containing the C8γ binding site on C8α: implications for C8γ ligand binding. *Mol. Immunol.* **45**, 750–756.
48. Naylor, H. M. & Newcomer, M. E. (1999). The structure of human retinol-binding protein (RBP) with its carrier protein transthyretin reveals an interaction with the carboxy terminus of RBP. *Biochemistry*, **38**, 2647–2653.
49. Ball, G., Shelton, M. J., Walsh, B. J., Hill, D. J., Hosking, C. S. & Howden, M. E. (1994). A major continuous allergenic epitope of bovine beta-lactoglobulin recognized by human IgE binding. *Clin. Exp. Allergy*, **24**, 758–764.
50. Haste, A. P., Nielsen, M. & Lund, O. (2006). Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.* **15**, 2558–2567.
51. Vita, R., Zarebski, L., Greenbaum, J. A., Emami, H., Hoof, I., Salimi, N. *et al.* (2010). The immune epitope database 2.0. *Nucleic Acids Res.* **38**, D854–D862.
52. Notkins, A. L. (2004). Polyreactivity of antibody molecules. *Trends Immunol.* **25**, 174–179.
53. James, L. C., Roversi, P. & Tawfik, D. S. (2003). Antibody multispecificity mediated by conformational diversity. *Science*, **299**, 1362–1367.

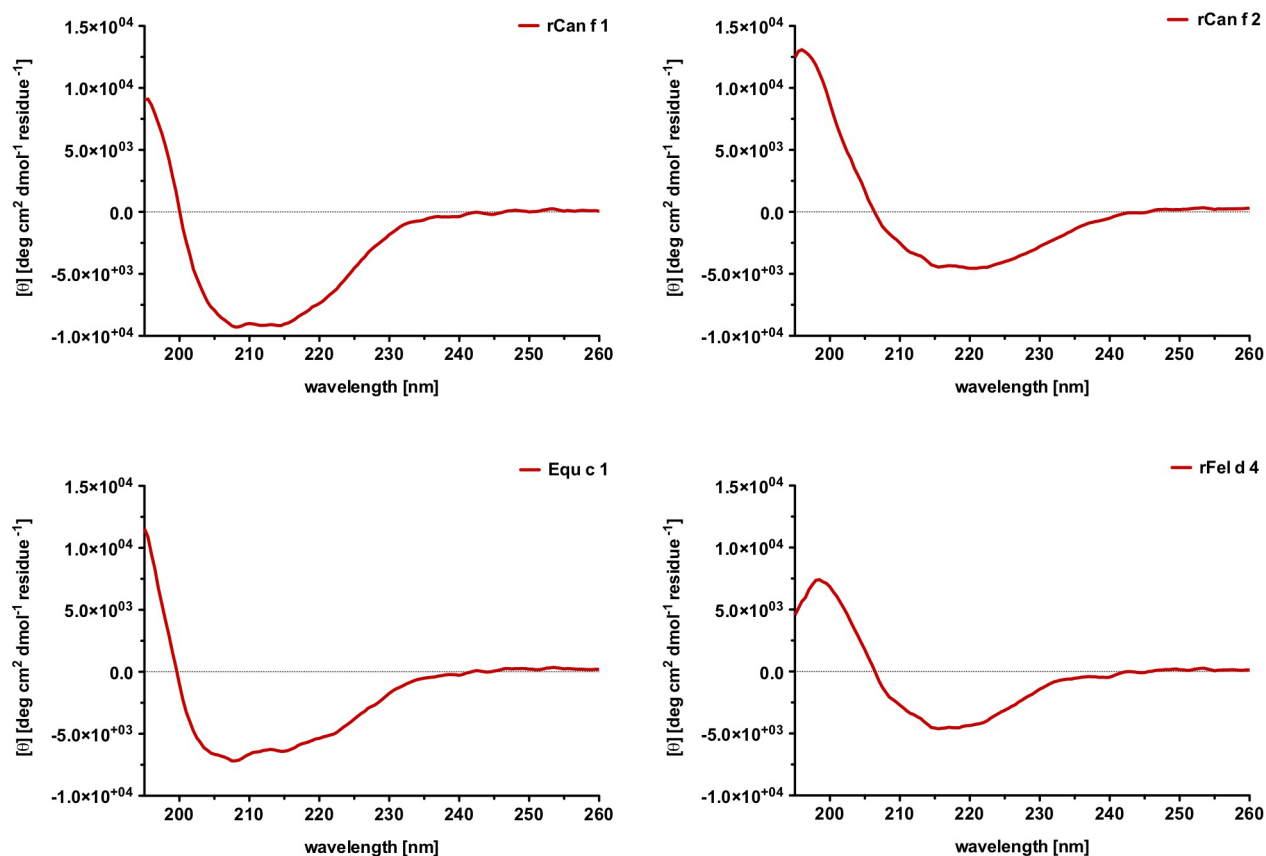
54. Bostrom, J., Yu, S. F., Kan, D., Appleton, B. A., Lee, C. V., Billeci, K. *et al.* (2009). Variants of the antibody herceptin that interact with HER2 and VEGF at the antigen binding site. *Science*, **323**, 1610–1614.
55. Parren, P. W. & Burton, D. R. (2009). Immunology. Two-in-one designer antibodies. *Science*, **323**, 1567–1568.
56. Harwanegg, C. & Hiller, R. (2005). Protein micro-arrays for the diagnosis of allergic diseases: state-of-the-art and future development. *Clin. Chem. Lab Med.* **43**, 1321–1326.
57. Grönlund, H., Vrtala, S., Wiedermann, U., Dekan, G., Kraft, D., Valenta, R. & van Hage-Hamsten, M. (2002). Carbohydrate-based particles: a new adjuvant for allergen-specific immunotherapy. *Immunology*, **107**, 523–529.
58. Grönlund, H., Bergman, T., Sandstrom, K., Alvelius, G., Reininger, R., Verdino, P. *et al.* (2003). Formation of disulfide bonds and homodimers of the major cat allergen Fel d 1 equivalent to the natural allergen by expression in *Escherichia coli*. *J. Biol. Chem.* **278**, 40144–40151.
59. Johansson, E., Johansson, S. G. & van Hage-Hamsten, M. (1994). Allergenic characterization of *Acarus siro* and *Tyrophagus putrescentiae* and their crossreactivity with *Lepidoglyphus destructor* and *Dermatophagoides pteronyssinus*. *Clin. Exp. Allergy*, **24**, 743–751.
60. Goubran, B. H., Poncet, P., Rabillon, J., Fontaine, T., Laval, J. M. & David, B. (2001). Biochemical characterization and surfactant properties of horse allergens. *Eur. J. Biochem.* **268**, 3126–3136.
61. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
62. Saarne, T., Kaiser, L., Rasool, O., Huecas, S., van Hage-Hamsten, M. & Gafvelin, G. (2003). Cloning and characterisation of two IgE-binding proteins, homologous to tropomyosin and alpha-tubulin, from the mite *Lepidoglyphus destructor*. *Int. Arch. Allergy Immunol.* **130**, 258–265.
63. Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr. D*, **62**, 48–57.
64. The CCP4 suite: programs for protein crystallography. (1994). *Acta Crystallogr. D*, **50**, 760–763.
65. 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.
66. Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallogr. D*, **49**, 129–147.
67. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, **60**, 2126–2132.
68. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321.
69. Bond, C. S. (2003). TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics*, **19**, 311–312.
70. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
71. Binkowski, T. A., Naghibzadeh, S. & Liang, J. (2003). CASTp: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **31**, 3352–3355.
72. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
73. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201.

Supplementary Table I

Specific IgE reactivities to the dog allergens Can f 1, Can f 2, Can f 3 and to the cat allergen Fel d 4 were measured among 89 dog-sensitized subjects. Serum samples from individuals determined to have an IgE reactivity above 5 kU/L to a commercially available dog extract, were collected from the diagnostics department at Karolinska University Hospital. Specific reactivities were quantified by ELISA, using a standard curve derived from an IgE myeloma.

Patient number	EE5	Can f 1	Can f 2	Can f 3	Fel d 4
1	34	35.45	0.3	0	1.2
2	8.1	0	0	0	0
3	33	6.5	0	0	0
4	88	3.5	10.4	11.4	1.8
5	5.6	0	0	0	0
6	12	7.4	0	3.1	26.5
7	39	0.9	0.5	0	0
8	19	1.3	0.3	0	3.9
9	19	2.1	0	0	0
10	29	0	0	0	0
11	10	3.2	0.1	0	0.7
12	37	7.4	0	0	8.2
13	16	0	0	0	0
14	5.2	0	0	0.6	1.5
15	5.7	0.5	0.3	0	0.9
16	18	11.5	0.2	0	3.9
17	33	9.2	0	0	0
18	9.2	6.6	1.3	1	0.4
19	35	2.7	1.4	0	7.6
20	46	7.7	6.4	1.6	0
21	41	7.2	0.4	0	2.4
22	9.9	0	0.2	0	2.8
23	5.6	1.4	0	0	1
24	6.3	0	0	0	0
25	30	5	0.2	0	1.6
26	6.5	0.8	1	0	0
27	12	0	0	0	0.4
28	12	1	0.8	0	1.5
29	12	7.1	0	0	NT
30	19	8.9	0	0	3.5
31	5.6	0.8	0	0.5	NT
32	>100	40.5	31.7	0	NT
33	15	2.8	1.84	0	0
34	5.5	0.4	0	0	0.2
35	11	0	0	0	0
36	11	2.5	0	0	4.2
37	9.3	0.4	0.3	0.4	0
38	5.1	0	0.3	0	NT
39	22	0	3.3	0	1.1
40	10	1.5	0	0	1.5
41	5.4	0	0	0	0

42	8.4	0.3	0	0	0
43	7	0	0	0	0
44	8	0	0	0	0
45	>100	46.3	71.5	55.6	16.5
46	12	2.5	0	0	1
47	13	0.3	0.2	0	0
48	8.6	0.2	0	0	0.2
49	18	4.8	0.2	0	1
50	17	0	0.7	0	9.3
51	56	36.6	0	0	0
52	9.5	0	0.2	0	0
53	5.8	0	0	0	0
54	11	7.8	0	2.2	12
55	6.5	0	0	0.4	0
56	8.1	0	0	0	6.3
57	40	11.9	7	0	NT
58	17	5	0	0	2.5
59	6.6	0	0	0	0
60	46	14.5	0.1	22.7	34.3
61	8.7	2.4	0.6	0.7	1.8
62	13	0.3	2.8	0.3	NT
63	87	50.9	8	5.5	6
64	30	5.9	3.2	0	0.1
65	6.1	0.6	0.3	0	1.1
66	15	0	0	0	0
67	6.4	0	0	0	3.5
68	9.3	1	0	4.3	1.2
69	5	2.6	0	0	0
70	9.1	3.8	0	1.7	1
71	7.1	0.1	0	0	0.1
72	12	0.5	0	0	0
73	24	8.9	0.1	0	0
74	31	0	0	0	0
75	5.3	0	0	0	0
76	29	4.2	0.5	0	2.7
77	6.8	1.8	0	0	0
78	8.6	3	0	0	0
79	5.9	0.8	0	0.3	1.6
80	10	0	0	0.4	0
81	6.3	0	0	0	0
82	7.7	0	0	1.4	1
83	5.4	3.2	0	0	0
84	24	10.5	0	1.1	15.5
85	65	2.3	1.6	11.8	0
86	41	1.4	1.5	4.9	0.1
87	6.7	0	0	0	0.2
88	8.7	0	0	0	0
89	12	0.3	0	0	0



Supplementary Figure 1. Circular dichroism analysis of the allergens used in this study. The circular dichroism for UV spectra of the allergens Can f 1, Can f 2, Fel d 4 and Equ c 1 demonstrate that they are well folded.