Structure-based prediction of the IgE epitopes of the major dog allergen Can f 1

Structure-based prediction of the IgE epitopes of the major dog allergen Can f 1

Masatoshi Nakatsuji[1], Keisuke Sugiura[1], Keisuke Suda[1], Michiko Sakurai[1], Miki Ubatani[1], Haruka Muroya[1], Rina Okubo[1], Ryo Noguchi[1], Yoichi Kamata[2], Yuma Fukutomi[3], Osamu Ishibashi[1], Shigenori Nishimura[1], Takashi Inui[1, *]

[1] Department of Applied Life Sciences, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan

[2] Department of Food and Nutrition, Senri Kinran University, Suita, Osaka, 565-0873, Japan

[3] Clinical Research Center for Allergy and Rheumatology, Sagamihara National Hospital, Sagamihara, Kanagawa 252-0392, Japan

[*]Correspondence author:

Takashi Inui, Laboratory of Biological Macromolecules, Department of Applied Life Sciences, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan. Phone number: +81-72-254-9473, E-mail address: inuit@bioinfo.osakafu-u.ac.jp

Masatoshi Nakatsuji and Keisuke Sugiura equally contributed to this work.

Running title: Structure-based prediction of IgE epitopes of Can f 1

Abbreviations

Can f 1, *Canis familiaris* allergen 1; TL, tear lipocalin

Keywords

Can f 1; crystal structure; allergen; IgE epitope; conformational epitope

Conflicts of interest

The authors declare no conflict of interest.

Data Accessibility

The coordinates and structure factors for the crystal structure of Can f 1 were deposited to PDB (PDB ID: 7DRU).

Abstract

Allergy to dogs has become increasingly prominent worldwide. Seven dog allergens have been identified, including *Canis familiaris allergen* 1 to 7 (Can f 1–7). Although Can f 1 is a major dog allergen sensitized to 50–75% of dog-allergic subjects, its IgE epitopes have not been identified. The structural analysis of an allergen is important to identify conformational epitopes. In this study, we generated a recombinant Can f 1 protein and determined its crystal structure using X-ray crystallography. Can f 1 had a typical lipocalin fold, which is composed of an eight-stranded β-barrel and α-helix, and has high similarity to Can f 2, Can f 4, and Can f 6 in overall structure. However, the localizations of surface charges on these proteins were quite different. Based on sequence alignment and tertiary structure, we predicted five critical residues (His86, Glu98, Arg111, Glu138, and Arg152) for the IgE epitopes. The relevance of these residues to IgE reactivity was assessed by generating Can f 1 mutants with these residues substituted for alanine. Although the effects of the mutation on IgE binding depended on the sera of dog-allergic patients, H86A and R152A mutants showed reduced IgE reactivity compared to wild-type Can f 1. These results suggest that Can f 1 residues His86 and Arg152 are candidates for the IgE conformational epitope.

Introduction

Dogs (*Canis familiaris*) are one of the most popular pets in the world, and support humans as guide and service dogs. However, they are a major source of human respiratory allergens, which are found in their dander, urine, and saliva. These allergens can cause severe allergic reactions, such as asthma, allergic rhinitis, and allergic conjunctivitis in 5–10% of the population [1]. As contact between dogs and humans becomes more frequent and intimate, dog allergies have become increasingly prominent worldwide, particularly in advanced nations [2-4].

Seven respiratory allergens have been identified thus far for the domestic dog, specifically Can f 1 to Can f 7. Four of these (Can f 1, Can f 2, Can f 4, and Can f 6) belong to the lipocalin protein family [5-7], whereas Can f 3, Can f 5, and Can f 7 are classified as a serum albumin, a prostatic kallikrein, and an epididymal secretory protein E1, respectively [8-10]. Many mammalian allergens, including those of dog, cat, horse, and mouse, belong to the lipocalin protein family [11, 12]. Although lipocalin proteins have low sequence identity between family members, they are characterized by a common tertiary structure consisting of an eight-stranded, antiparallel β-barrel [13]. The structural similarity and polysensitization of lipocalin proteins can explain their IgE cross-reactivity reported thus far [13]. Can f 1 is a lipocalin allergen in dog that is expressed in tongue epithelial tissue, parotid, mandibular glands, and skin. It is a major dog allergen, with 50–75% of dog-allergic subjects sensitised to it [5, 13]. These reports encourage the development of a novel hypoallergenic vaccine against Can f

1, which is a highly desired treatment for dog allergies.

Allergic reactions, including anaphylaxis, are triggered by the interactions between an allergen and specific IgE on the surface of effector cells, and the allergen epitope must be determined to develop a hypoallergenic vaccine [14, 15]. IgE epitopes are classified into two types: linear and conformational epitopes [16, 17]. Linear epitopes are short with continuous amino acid sequences, while conformational epitopes are composed of amino acids that are adjacently located in a folded protein but distantly separated in the primary sequence. Although both types of epitopes have been identified using various approaches, such as peptide microarray, *in silico* prediction based on sequence similarity, and known cross-reactivity with other allergens, only limited information on conformational epitopes is currently available [16].

The structural analysis of an allergen can help identify conformational epitopes, because tertiary structures provide information on surface-exposed residues, thus facilitating the rational design of hypoallergenic vaccines. The IgE epitopes of *Betula verrucosa* allergen 1 and *Dermatophagoides farinae* allergen 13 have been identified based on primary and tertiary structures, and potential hypoallergenic vaccines have been generated for allergen-specific immunotherapy [18, 19]. Therefore, to identify putative IgE epitopes on allergens, it is beneficial to determine a tertiary structure. Although the tertiary structures of the lipocalin dog allergens Can f 2, Can f 4, and Can f 6 have previously been determined [20-22], those of Can f 1 have not.

In this study, we determined the crystal structure of Can f 1 at 2.5 Å resolution and identified possible IgE epitopes based on sequence alignment and tertiary structure information, followed by subsequent experimental validation.

Results

Purification and structural analysis of Can f 1

Wild-type (WT) Can f 1 was expressed as a glutathione *S*-transferase (GST)-fusion protein and purified via affinity and gel filtration chromatography. The purified WT Can f 1 appeared as a single band of an expected size in SDS-PAGE (Fig. 1A). IgE reactivity of the purified WT Can f 1 was evaluated via ELISA using sera from 17 dog-allergic patients (Fig. 1B). Sera from 9 (53%) of these patients reacted with WT Can f 1, consistent with previous reports [5, 13].

The tertiary structure of the C118A mutant was analyzed instead of WT Can f 1 as the crystallization of WT protein may be inhibited by the formation of an intermolecular disulfide bond (Fig. 1A). Purified C118A mutant was as reactive to IgE as WT Can f 1 (Fig. 2A). Like WT Can f 1, the far-UV circular dichroism (CD) spectra of the C118A mutant showed a negative maximum at 209 nm (Fig. 2B). The near-UV CD spectra of the C118A mutant exhibited a negative Cotton effect at 280 nm due to tryptophan, which is similar to WT Can f 1 (Fig. 2B). Thus, we confirmed that substitution of the Cys118 residue with alanine did not affect IgE reactivity nor the secondary or tertiary structures of Can f 1.

The C118A mutant crystal diffracted to 2.5 Å resolution and contained seven molecules per asymmetric unit. The refinement statistics are summarized in Table 1. Each molecule showed high similarity with a root mean square deviation (RMSD) of 0.39 Å for Cα atoms. The structure of chain C in the C118A mutant is shown in Figure 2C and is representative

of the seven molecules. The overall structure of the C118A mutant included a classic lipocalin fold composed of an eight-stranded β-barrel (A, Lys34-Ala42; B, Asp51-Ala60; C, Leu66-Thr74; D, Gln77-Lys87; E, Lys93-Ala96; F, Gln100-Pro107; G, His113-Glu119; H, Arg128-Gly134). Cys78 and Cys169 residues formed an intramolecular disulfide bond, which is also highly conserved among lipocalins (Fig. 2C). Gln19-Val29, Lys62-Gly64, and Glu98-Gly99 of chain C, which form a loop in most lipocalin proteins, were not observed due to unclear electron density, suggesting that these regions are highly flexible.

Structural comparison between Can f 1 and other dog lipocalin allergens

Next, the tertiary structure of Can f 1 was compared to other dog lipocalin allergens. The primary sequence identity of Can f 1 was as low as 25% compared to Can f 2, Can f 4, and Can f 6. However, the overall structure of Can f 1 was similar to Can f 2, Can f 4, and Can f 6 (Fig. 3A). Structural differences were observed in the flexible loop regions (AB-, CD-, and GH-loops). The AB-loop of Can f 2, Can f 4, and Can f 6 formed a short helix, whereas Can f 1 did not form a helix due to the helix-breaking property of Pro47 (Fig. 3B).

The interaction between tryptophane and arginine residues at the bottom region of the barrel significantly contributes to folding and structural stabilization of lipocalin proteins [23]. In Can f 1, Trp35 is in the bottom region of the barrel and forms a network with Val31, Ser32, Lys34, Met56, and Arg135 through hydrogen bonds and van der Waals' interactions (Fig. 3A

and 3C). A similar network surrounding a tryptophane residue in the bottom region was also formed in Can f 2, Can f 4, and Can f 6 (Fig. 3A and 3D). These results suggest that the tertiary structure of dog lipocalin allergens is likely stabilized by a network of hydrogen bonds and van der Waals' interactions surrounding a tryptophane residue.

We also calculated the surface electrostatic charge among these proteins (Fig. 3E). The ratio of positive to negative surface charge was similar for all proteins except Can f 2, which has a strong negative surface charge. However, the localization of surface charge on these proteins varied significantly. These results may explain the poor cross-reactivity between Can f 1 and other dog lipocalin allergens [11, 24].

Predicting IgE epitopes of Can f 1

The IgE epitopes of Can f 1 were predicted based on sequence alignment and the tertiary structures of Can f 1 and human tear lipocalin (TL), a major protein in human tear fluid. TL, which shares 61% identity with Can f 1 in primary sequences (Fig. 4A), was used as a non-allergic reference because the WT Can f 1-reactive sera from three dog-allergic patients exhibited 15% or less IgE reactivities to TL compared to WT Can f 1 (Fig. 4B). Therefore, we selected amino acid residues that are not conserved between Can f 1 and TL. These residues were further screened to select charged residues located in the surface-facing region of Can f 1, because they were known to play an important role in IgE binding [19, 25]. As a result, we

predicted 15 amino acid residues that could play a key role in the IgE epitopes of Can f 1. Notably, although the overall structure of Can f 1 was highly similar to that of TL as supported by the RMSD of 0.91 Å for Cα atoms (Fig. 4C), difference was observed in the localization of surface charge between these proteins (Fig. 4D). The difference in surface charge is possibly caused by the different composition of amino acid residues that are spatially adjacent in the folded proteins; therefore, regions exhibiting the different surface charge in Can f 1 possibly form conformational epitopes. Because 5 out of the 15 amino acid residues (His86, Glu98, Arg111, Glu138, and Arg152) were located in the regions, they were finally selected as amino acid residues that possibly play a key role in the IgE epitopes of Can f 1. Residues His86 and Arg152 are located in the β_D-strand and the α-helix, respectively. Glu98, Arg111, and Glu138 are in the EF- and FG-loop region, and the loop region between β_H-strand and the α-helix. These residues are mapped on the tertiary structure of Can f 1 in Figure 4E.

IgE reactivity of Can f 1 mutants

We evaluated the validity of the predicted IgE epitopes by substituting the residues for alanine. The purity of the Can f 1 mutants was verified by the presence of a single band in SDS-PAGE analysis (Fig. 5A). ELISA was used to evaluate the IgE reactivity of the Can f 1 mutants using WT Can f 1-reactive sera from six dog-allergic patients (Fig. 5B and Table 2). The IgE reactivities of the E98A, R111A, and E138A mutants were similar to that of WT Can f 1,

whereas those of H86A and R152A were 9% and 11% lower and differed depending on the sera used. For example, that of serum from patient no. 1 to both mutants was 29% lower, while serum from patient no. 4 had similar reactivity to H86A and R152A mutants as WT Can f 1. This variation suggests that Can f 1 is a polyvalent allergen with multiple epitopes, and the degree of IgE recognition of individual epitopes is different in each Can f 1-reactive serum. Furthermore, these results indicate that His86 and Arg152 are involved in IgE binding. To ensure that the reduction of IgE reactivity was caused by the loss of IgE epitopes rather than the conformational change of the protein, we measured the CD spectra of Can f 1 mutants (Fig. 5C). The far-UV CD spectra of Can f 1 mutants, which showed a negative maximum at 209 nm, and the near-UV CD spectra of Can f 1 mutants were similar to those of WT Can f 1. These results suggest that the substitution of the selected residues to alanine did not affect the secondary and tertiary structures of Can f 1.

Discussion

The tertiary structure of Can f 1, as determined by X-ray crystallography, has a typical lipocalin fold composed of an eight-stranded β-barrel. In addition, we identified possible IgE epitopes of Can f 1. This is the first report to clarify the crystal structure of the major dog lipocalin allergen Can f 1.

The structure of Can f 1 is similar to that of other dog lipocalin allergens; however, the distribution of surface charge on these proteins varies greatly. Differences in surface charge are also observed between Can f 1 and lipocalin allergens from other animals, such as horses (Equ c 1), mice (Mus m 1), and bovines (Bos d 2) (Fig. 6), and these allergens are not cross-reactive with Can f 1 [11]. Thus, the distribution of surface charge is critical for allergen recognition by IgE, which implies that structural information is essential for identifying IgE epitopes.

We predicted that five amino acid residues are important for IgE reactivity toward Can f 1 and experimentally verified that His86 and Arg152 are involved in IgE binding. A previous study that analyzed the crystal structure of an IgE Fab fragment in complex with β-lactoglobulin, a major lipocalin allergen of bovine milk, found that the epitope-binding region of the IgE antibody was planar and consisted mainly of β-sheet structures rather than loop regions [26]. His86 and Arg152 residues of Can f 1 are contained in the β_D-strand and α-helix, respectively, of the planar structure (Fig. 7). Glu98, Arg111, and Glu138 mutations did not affect the IgE reactivity of Can f 1 and are contained in loop regions with a convex structure (Fig. 7). These

results indicate that in addition to amino acid sequences, structural elements and shapes are also important for allergen recognition by IgE. Thus, structural analysis is essential for accurate evaluation and identification of epitopes.

Curin *et al*. [27], using a synthesized Can f 1 overlapping peptide, recently demonstrated that Can f 1 likely contains conformational epitopes and not linear epitopes. They found that the N- and C-terminal regions of Can f 1 (~30 amino acid residues) are major conformational epitope-containing regions [27]. The C-terminal region, particularly the area around Arg152, is likely to contain the IgE epitopes of Can f 1. We also identified a region near His86 as a novel IgE epitope of Can f 1. Notably, the Can f 1 Arg152 residue is conserved in Fel d 7, a cat (*Felis domesticus*) lipocalin allergen (Fig. 8), which has IgE cross-reactivity with Can f 1 [28]. Therefore, this Arg residue is possibly involved in the cross-reactivity between Can f 1 and Fel d 7. TL was used for IgE epitope prediction in this study and is a non-allergenic protein. Substituting the Can f 1 sequence that composes the vicinity of Arg152 for the corresponding sequence of TL may further reduce IgE reactivity.

In conclusion, we demonstrated that His86 and Arg152 are important to IgE reactivity toward Can f 1. However, the importance of these residues in allergenic activity is still poorly understood and further studies (i.e., *in vivo* studies) are needed.

Materials and Methods

Specimen collection and ethical statement

The sera samples used in this study were collected from 17 consecutive outpatients with dog allergies who visited Sagamihara National Hospital, Japan (Table 3). Sera from three allergic donors without dog allergies were also collected as controls (Table 3). Written informed consent was obtained from all patients. This study was approved by the Ethics Committee of Sagamihara National Hospital.

Expression and purification of recombinant Can f 1

The nucleotide and amino acid sequences of Can f 1 were obtained from GenBank (accession numbers AF027177 and AAC48794, respectively). cDNA of Can f 1 was kindly provided by Prof. Yoichi Kamata (Department of Food and Nutrition, Senri Kinran University). Using the SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP), a 522-nucleotide open reading frame was found in the cDNA, and the first 54 nt sequence of the ORF was predicted to encode an 18-amino acid signal peptide. The region encoding a putative and mature Can f 1 protein was amplified by PCR using the primer sets listed in Table S1. The PCR products were digested with *Sma*I and *Xho*I restriction enzymes and inserted into the expression plasmid pGEX4T-2 to construct pGEX4T-2-WT Can f 1. After sequence validation, *Escherichia coli* BL21 (DE3) cells were transformed with pGEX4T-2-WT Can f 1. The WT Can f 1 was expressed as a GST-fused protein. The fusion protein was bound to glutathione Sepharose 4B column (Cytiva) and

238 incubated with thrombin (Sigma-Aldrich) to cleave the WT Can f 1 from GST. The recombinant
239 protein was further purified by gel filtration chromatography using HiLoad 16/600 Superdex
240 75 pg (Cytiva) in PBS.
241
242 *Site-directed mutagenesis of Can f 1*
243 Mutations were introduced at selected residues of Can f 1 to generate six single-mutant proteins
244 (H86A, E98A, R111A, C118A, E138A and R152A). The primers used for site-directed
245 mutagenesis by megaprimer PCR [29] are listed in Table S1. The PCR products were digested
246 with *Sma*I and *Xho*I restriction enzymes, and subcloned into pGEX 4T-2 vector to generate the
247 Can f 1 mutants. The expression and purification of these Can f 1 mutants was performed as
248 described earlier.
249
250 *Crystallization and X-ray structure determination*
251 Crystallization experiments were performed at 20°C using the sitting-drop vapor-diffusion
252 method. Purified C118A mutant (20 mg/mL) was prepared in 20 mM Tris-HCl (pH 7.4) and
253 mixed with an equal volume of reservoir solution containing 0.1 M imidazole (pH 6.5), 30%
254 polyethylene glycol 3350, and 6% isopropanol in 0.1 M $CaCl_2$ on crystallization plates. Crystals
255 suitable for X-ray diffraction appeared within 1 week in crystallization drops. For data
256 collection, crystals were transferred directly from their mother liquor to a nitrogen cold stream

at -173°C. Diffraction data were collected from the beam-line BL26B1 (λ = 1.0000 Å) at SPring-8 (Hyogo, Japan) using an EIGER-4M detector (Rayonix LLC) at a cryogenic temperature (-173°C). Data were integrated and scaled using XDS software [30]. The structure was determined by molecular replacement using MOLREP [31] in the CCP4i package [32] with the structure of TL (Protein Data Bank ID: 3EYC) as the search model. This template shared 55% sequence identity with Can f 1. Iterative rounds of model building and refinement were carried out using COOT [33], Phenix Refine [34], and Refmac5 [35] in the CCP4i package, respectively. Graphical representations were prepared using PyMOL software (http://www.pymol.org). Electrostatic calculation was done in PyMOL software using the Adaptive Poisson-Boltzmann Solver plug-in [36]. The statistics associated with collection, processing, and refinement are summarized in Table 1. The atomic coordinates and structure factors of Can f 1 have been deposited in the Protein Data Bank under ID code 7DRU.

Circular dichroism measurements

CD measurements were performed with a J-820 spectropolarimeter (Jasco). The temperature of the sample solution in the cuvette was controlled at 37°C by a Peltier PTC-423 L thermo-unit (Jasco). The path length of the optical quartz cuvette was 1.0 mm for far-UV range, and 10 mm for near-UV range CD measurements. The sample concentration for the far-UV and near-UV range was 5 and 50 μM in PBS, respectively. The data are expressed as molar residue ellipticity

(θ).

Evaluation of IgE reactivity by ELISA

WT Can f 1 and mutants (5 μg) were immobilized on ELISA plates (AGC TECHNO GLASS CO., LTD) at 4°C overnight. The wells were washed with 0.05% Tween-20 in PBS. After blocking with 3% skim milk in PBS for 1 h at room temperature, sera from dog-allergic patients diluted 1:1000 with Can Get Signal® Solution 1 (TOYOBO) was added to the wells, and incubated for 1 h at 37°C. After washing with 0.05% Tween-20 in PBS, the biotin-labeled goat anti-human IgE antibody (Seracare Life Sciences, Inc) diluted 1:5000 with Can Get Signal® Solution 2 (TOYOBO) was added to the wells, and incubated for 1 h at 37°C. The bound biotin labeled antibody was detected with Pierce® High Sensitivity Streptavidin-HRP (Thermo Fischer Scientific) diluted 1:10,000 with Can Get Signal® Solution 2 (TOYOBO), and TMB Substrate Reagent (BD Biosciences). Absorbance at 450 nm was measured using microplate reader Model 680 (Bio-Rad Laboratories).

Statistical analysis

The cut-off value of IgE reactivity shown in Figure 1 was calculated as the mean of non-dog-allergic donors + 3× standard deviation (SD), as described previously [37]. The statistical significance of reduction in IgE reactivity was evaluated using one-sample one-tailed *t*-test, and

differences with p values less than 0.05 were considered statistically significant.

Author contributions

MN, KSug and TI planned and designed the experiments. MN, KSug, KS, MS, MU, HM, RO, and RN performed the experiments. MN, KSug, KS and SN analyzed the data. YK, YF and OI helped supervise the project. MN, OI and TI wrote the manuscript. TI designed and directed the project.

Acknowledgments

We are thankful to Mrs. Mizuki Tabata, Shun Nakayama, Kenji Yamamoto, and Masaya Sakaguchi for their technical assistance in some experiments. This work was supported in part by Japan Society for the Promotion of Science [grant numbers 17K19329 and 25242046 (to TI)].

Reference

1. Polovic, N., Waden, K., Binnmyr, J., Hamsten, C., Gronneberg, R., Palmberg, C., Milcic-Matic, N., Bergman, T., Gronlund, H. & van Hage, M. (2013) Dog saliva - an important source of dog allergens, *Allergy.* **68**, 585-92.

2. Gergen, P. J., Arbes, S. J., Jr., Calatroni, A., Mitchell, H. E. & Zeldin, D. C. (2009) Total IgE levels and asthma prevalence in the US population: results from the National Health and Nutrition Examination Survey 2005-2006, *J Allergy Clin Immunol.* **124**, 447-53.

3. Heinzerling, L., Frew, A. J., Bindslev-Jensen, C., Bonini, S., Bousquet, J., Bresciani, M., Carlsen, K. H., van Cauwenberge, P., Darsow, U., Fokkens, W. J., Haahtela, T., van Hoecke, H., Jessberger, B., Kowalski, M. L., Kopp, T., Lahoz, C. N., Lodrup Carlsen, K. C., Papadopoulos, N. G., Ring, J., Schmid-Grendelmeier, P., Vignola, A. M., Wohrl, S. & Zuberbier, T. (2005) Standard skin prick testing and sensitization to inhalant allergens across Europe--a survey from the GALEN network, *Allergy.* **60**, 1287-300.

4. Park, H. J., Kim, E. J., Yoon, D., Lee, J. K., Chang, W. S., Lim, Y. M., Park, J. W. & Lee, J. S. (2017) Prevalence of Self-reported Allergic Diseases and IgE Levels: A 2010 KNHANES Analysis, *Allergy Asthma Immunol Res.* **9**, 329-339.

5. Konieczny, A., Morgenstern, J. P., Bizinkauskas, C. B., Lilley, C. H., Brauer, A. W., Bond, J. F., Aalberse, R. C., Wallner, B. P. & Kasaian, M. T. (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**, 577-86.

6. Mattsson, L., Lundgren, T., Olsson, P., Sundberg, M. & Lidholm, J. (2010) Molecular and immunological characterization of Can f 4: a dog dander allergen cross-reactive with a 23 kDa odorant-binding protein in cow dander, *Clin Exp Allergy.* **40**, 1276-87.

7. Hilger, C., Swiontek, K., Arumugam, K., Lehners, C. & Hentges, F. (2012) Identification of a new major dog allergen highly cross-reactive with Fel d 4 in a population of cat- and dog-sensitized patients, *J Allergy Clin Immunol.* **129**, 1149-51.

8. Spitzauer, S., Schweiger, C., Sperr, W. R., Pandjaitan, B., Valent, P., Muhl, S., Ebner, C., Scheiner, O., Kraft, D., Rumpold, H. & et al. (1994) Molecular characterization of dog albumin as a cross-reactive allergen, *J Allergy Clin Immunol.* **93**, 614-27.

9. Mattsson, L., Lundgren, T., Everberg, H., Larsson, H. & Lidholm, J. (2009) Prostatic kallikrein: a new major dog allergen, *J Allergy Clin Immunol.* **123**, 362-8.

10. Khurana, T., Newman-Lindsay, S., Young, P. R. & Slater, J. E. (2016) The NPC2 protein: A novel dog allergen, *Ann Allergy Asthma Immunol.* **116**, 440-446 e2.

11. Saarelainen, S., Rytkonen-Nissinen, M., Rouvinen, J., Taivainen, A., Auriola, S., Kauppinen, A., Kinnunen, T. & Virtanen, T. (2008) Animal-derived lipocalin allergens exhibit immunoglobulin E cross-reactivity, *Clin Exp Allergy.* **38**, 374-81.

12. Virtanen, T., Kinnunen, T. & Rytkonen-Nissinen, M. (2012) Mammalian lipocalin allergens--insights into their enigmatic allergenicity, *Clin Exp Allergy.* **42**, 494-504.

13. Hilger, C., Kuehn, A. & Hentges, F. (2012) Animal lipocalin allergens, *Curr Allergy Asthma Rep.* **12**, 438-47.

14. Bush, R. K. (2016) Advances in allergen immunotherapy in 2015, *J Allergy Clin Immunol.* **138**, 1284-1291.

15. Pfaar, O., Lou, H., Zhang, Y., Klimek, L. & Zhang, L. (2018) Recent developments and highlights in allergen immunotherapy, *Allergy.* **73**, 2274-2289.

16. Breiteneder, H. (2018) Mapping of conformational IgE epitopes of food allergens, *Allergy.* **73**, 2107-2109.

17. Dall'antonia, F., Pavkov-Keller, T., Zangger, K. & Keller, W. (2014) Structure of allergens and structure based epitope predictions, *Methods.* **66**, 3-21.

18. Spangfort, M. D., Mirza, O., Ipsen, H., Van Neerven, R. J., Gajhede, M. & Larsen, J. N. (2003) Dominating IgE-binding epitope of Bet v 1, the major allergen of birch pollen, characterized by X-ray crystallography and site-directed mutagenesis, *J Immunol.* **171**, 3084-90.

19. Chan, S. L., Ong, S. T., Ong, S. Y., Chew, F. T. & Mok, Y. K. (2006) Nuclear magnetic resonance structure-based epitope mapping and modulation of dust mite group 13 allergen as a hypoallergen, *J Immunol.* **176**, 4852-60.

20. Madhurantakam, C., Nilsson, O. B., Uchtenhagen, H., Konradsen, J., Saarne, T., Hogbom, E., Sandalova, T., Gronlund, H. & Achour, A. (2010) Crystal structure of the dog lipocalin

allergen Can f 2: implications for cross-reactivity to the cat allergen Fel d 4, *J Mol Biol.* **401**, 68-83.

21. Niemi, M. H., Rytkonen-Nissinen, M., Janis, J., Virtanen, T. & Rouvinen, J. (2014) Structural aspects of dog allergies: the crystal structure of a dog dander allergen Can f 4, *Mol Immunol.* **61**, 7-15.

22. Yamamoto, K., Ishibashi, O., Sugiura, K., Ubatani, M., Sakaguchi, M., Nakatsuji, M., Shimamoto, S., Noda, M., Uchiyama, S., Fukutomi, Y., Nishimura, S. & Inui, T. (2019) Crystal structure of the dog allergen Can f 6 and structure-based implications of its cross-reactivity with the cat allergen Fel d 4, *Sci Rep.* **9**, 1503.

23. 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**, 2301-16.

24. Nilsson, O. B., Neimert-Andersson, T., Bronge, M., Grundstrom, J., Sarma, R., Uchtenhagen, H., Kikhney, A., Sandalova, T., Holmgren, E., Svergun, D., Achour, A., van Hage, M. & Gronlund, H. (2014) Designing a multimer allergen for diagnosis and immunotherapy of dog allergic patients, *PLoS One.* **9**, e111041.

25. Sinha, N., Mohan, S., Lipschultz, C. A. & Smith-Gill, S. J. (2002) Differences in electrostatic properties at antibody-antigen binding sites: implications for specificity and cross-reactivity, *Biophys J.* **83**, 2946-68.

383 26. Niemi, M., Jylha, S., Laukkanen, M. L., Soderlund, H., Makinen-Kiljunen, S., Kallio, J.
384 M., Hakulinen, N., Haahtela, T., Takkinen, K. & Rouvinen, J. (2007) Molecular interactions
385 between a recombinant IgE antibody and the beta-lactoglobulin allergen, *Structure*. **15**, 1413-
386 21.
387 27. Curin, M., Weber, M., Hofer, G., Apostolovic, D., Keller, W., Reininger, R., Swoboda,
388 I., Spitzauer, S., Focke-Tejkl, M., van Hage, M. & Valenta, R. (2017) Clustering of
389 conformational IgE epitopes on the major dog allergen Can f 1, *Sci Rep*. **7**, 12135.
390 28. Apostolovic, D., Sanchez-Vidaurre, S., Waden, K., Curin, M., Grundstrom, J., Gafvelin,
391 G., Cirkovic Velickovic, T., Gronlund, H., Thomas, W. R., Valenta, R., Hamsten, C. & van
392 Hage, M. (2016) The cat lipocalin Fel d 7 and its cross-reactivity with the dog lipocalin Can f
393 1, *Allergy*. **71**, 1490-5.
394 29. Sarkar, G. & Sommer, S. S. (1990) The "megaprimer" method of site-directed
395 mutagenesis, *Biotechniques*. **8**, 404-7.
396 30. Kabsch, W. (2010) Xds, *Acta Crystallogr D Biol Crystallogr*. **66**, 125-32.
397 31. Vagin, A. & Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular
398 Replacement, *Journal of Applied Crystallography*. **30**, 1022-1025.
399 32. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R.,
400 Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G.
401 N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011)

Overview of the CCP4 suite and current developments, *Acta Crystallogr D Biol Crystallogr.* **67**, 235-42.

33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr.* **66**, 486-501.

34. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr.* **66**, 213-21.

35. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr D Biol Crystallogr.* **67**, 355-67.

36. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc Natl Acad Sci U S A.* **98**, 10037-41.

37. Kamata, Y., Miyanomae, A., Nakayama, E., Miyanomae, T., Tajima, T., Nishimura, K., Tada, T. & Hoshi, H. (2007) Characterization of dog allergens Can f 1 and Can f 2. 2. A comparison of Can f 1 with Can f 2 regarding their biochemical and immunological properties, *Int Arch Allergy Immunol.* **142**, 301-8.

421 ## Supporting Information

422 Table S1. The sequences of primers used in this study.

423

424 ## Tables

425 **Table 1 Data collection and refinement statistics for Can f 1 structure determination.**

Data collection	
Sample	Can f 1
Resolution range (Å)	50.00-2.50 (2.65-2.50)
Space group	P1211
Cell dimensions	
a, b, c (Å)	68.0, 128.7, 68.6
α , β, γ (deg)	90, 97.7, 90
Completeness (%) (overall/most outer shell)	99.2/99.0
Rmerge (%) (overall/most outer shell)	0.03/0.55
Number of molecules per asymmetric unit	7
Number of observed reflections	139731
Number of unique reflections	40304
Multiplicity (overall/outermost shell)	3.5/3.6
I/σ (I) (overall/outermost shell)	16.0/2.4

Refinement	
Number of protein atoms	6290
Number of amino acid residues	895
Number of water molecules	20
Rfactor (%)	24.3
Rfree (%)	28.6
Ramachandran plot (%)	
Preferred regions	96.51
Allowed regions	3.49
Outliners	0
PDB ID	7DRU

426

427 **Table 2 Serum IgE reactivity of Can f 1 mutants (%).**

Patient	H86A	E98A	R111A	E138A	R152A
No.1	71	100	102	104	71
No.4	101	101	99	87	98
No.9	95	104	100	93	92
No.14	93	97	91	91	92
No.16	93	97	85	101	86

No.17	90	96	99	104	91
Mean	91	99	96	97	89
P value*	0.037	0.286	0.104	0.148	0.015

Values represent %activity compared to WT Can f 1.

*One-sample one-tailed *t*-test vs WT Can f 1.

Table 3 Serological data of the sera used in this study.

Patient No.	Age	Gender	sIgE (kUA/ml) dog dander
1	63	F	70.1
2	53	F	>100
3	38	M	23
4	18	F	77.3
5	34	F	40.7
6	32	F	24.4
7	48	M	10.3
8	40	F	15.2
9	41	M	12.8
10	26	F	14.9

11	39	F	14.2
12	56	F	70.6
13	23	F	>100
14	22	F	>100
15	40	F	>100
16	26	F	>100
17	30	F	>100
N1	37	F	<0.35
N2	27	M	<0.35
N3	51	F	<0.35

Figure legends

Fig. 1. Purification of WT Can f 1. (A) SDS-PAGE analyses of WT Can f 1 and C118A mutant. Purified recombinant proteins of WT Can f 1 and C118A mutant were electrophoretically separated under reducing or non-reducing conditions and stained with Coomassie Brilliant Blue. (B) IgE reactivity of dog-allergic patients to recombinant Can f 1 measured using ELISA. Sera from 9 out of 17 (53%) dog-allergic patients show Can f 1-reactive IgE levels above the cut-off value (the mean of non-dog-allergic subjects + 3×SD) indicated by the dashed line. Sera from three non-allergic donors are indicated by N1 to N3 (light blue bars). The data are expressed as

441 the mean ± SD (n = 3).

442

443 Fig. 2. Structural analysis of Can f 1. (A) Relative IgE reactivity to C118A mutant compared to
444 WT Can f 1 was evaluated by ELISA. Can f 1-reactive sera from three patients were subjected
445 to this assay and the data are expressed as the mean ± SD (n = 3). (B) Far-UV (left panel) and
446 near-UV (right panel) CD spectra of WT Can f 1 (black) and C118A mutant (blue) in PBS (pH
447 7.4). (C) X-ray crystal structure of the C118A mutant and positions of typical secondary
448 structures presented along with the amino acid sequence. Cartoon representation of a typical
449 lipocalin fold with the intramolecular disulfide bond in yellow. The structure was represented
450 using PyMOL software. The secondary structure elements are colored red and blue for α-helix
451 and β-strands, respectively. The dashed lines represent the disordered regions resulting from
452 unclear electron density.

453

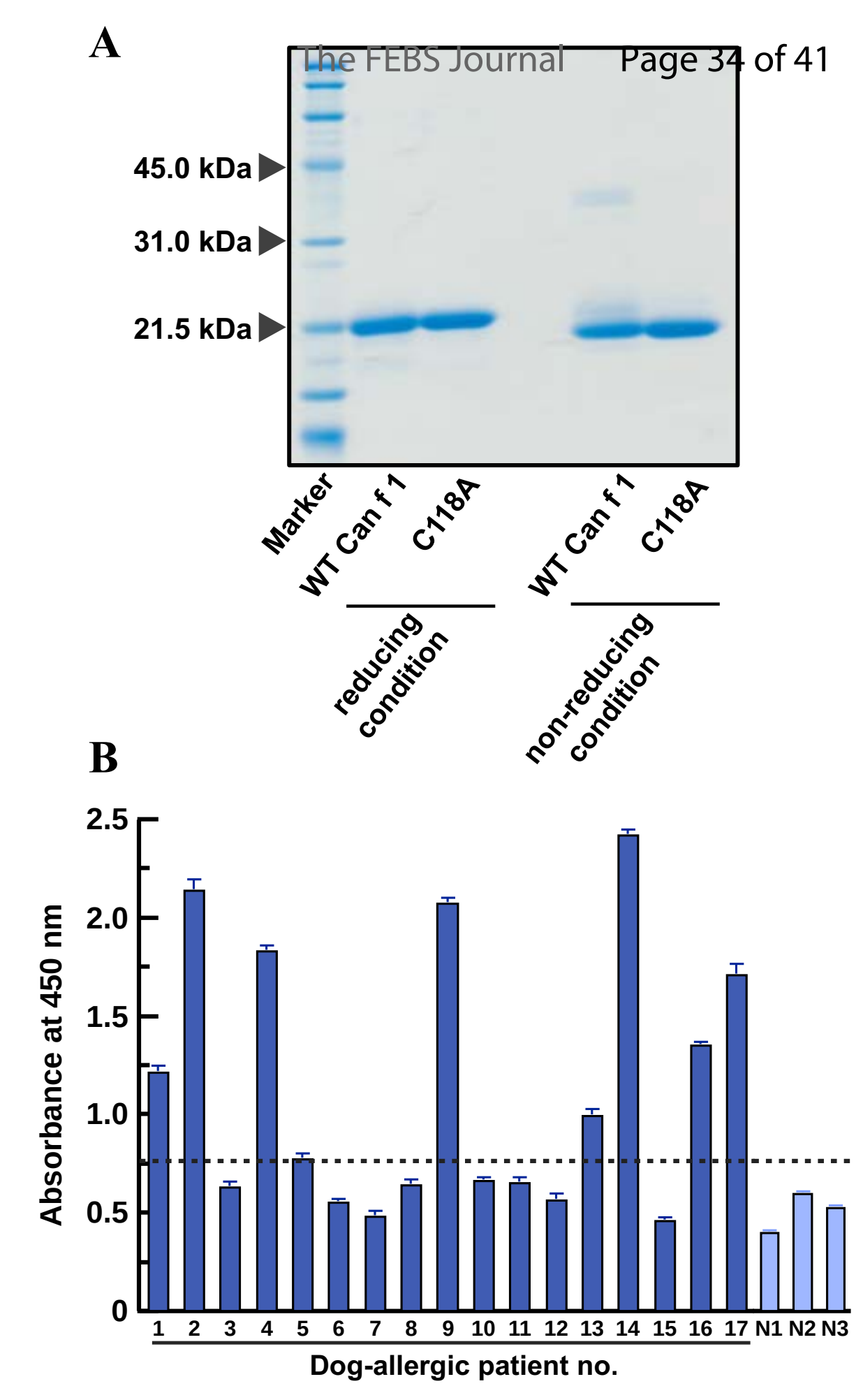
454 Fig. 3. Comparison of the Can f 1 structure with other dog lipocalin allergens. (A) Primary
455 structure alignment of dog lipocalin allergens (left). Red shading indicates the short helix on
456 the AB-loop of Can f 2, Can f 4, and Can f 6. Orange shading indicates the conserved
457 tryptophane and arginine residues at the bottom region of the barrel in lipocalin proteins.
458 Aligned residues are indicated with an asterisk (*****) for identical amino acids, colon (:) for similar
459 amino acids, and a period (.) for slightly similar amino acids. Superimposed X-ray structures

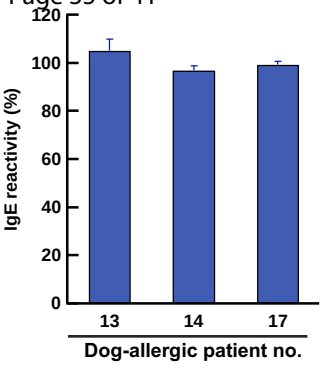
460 of dog lipocalin allergens, C118A mutant (yellow), Can f 2 (green; PDB ID: 3L4R), Can f 4
461 (purple; PDB ID: 4ODD), and Can f 6 (gray; PDB ID: 5X7Y). (B) Comparison of the AB-loop
462 among dog lipocalin allergens. Can f 2, Can f 4, and Can f 6 form a short helix in the AB-loop
463 (red), whereas Can f 1 did not form a short helix due to Pro47 (yellow). (C) The network of
464 hydrogen bonds and van der Waals' interactions surrounding the tryptophane residue at the
465 bottom region of the barrel in Can f 1. The hydrogen bonds are depicted as dashed lines. (D)
466 The network of hydrogen bonds and van der Waals' interactions surrounding the tryptophane
467 residue at the bottom region of the barrel of Can f 2, Can f 4, and Can f 6. The hydrogen bonds
468 are depicted as dashed lines. (E) Comparison of the surface electrostatic charge among dog
469 lipocalin allergens. Electrostatic calculation was done in PyMOL software using the Adaptive
470 Poisson-Boltzmann Solver plug-in. The regions of positive and negative electrostatic charge
471 are colored in blue and red, respectively. The structures in panels A-E were represented using
472 PyMOL software.
473
474 Fig. 4. Can f 1 IgE epitope prediction. (A) Primary sequence alignment of Can f 1 and TL. Blue
475 shading indicates the conserved residues between Can f 1 and TL. Red shading indicates the
476 five residues predicted to be IgE epitopes. (B) Relative IgE reactivity to TL compared to WT
477 Can f 1 was evaluated by ELISA. Can f 1-reactive sera from three patients were subjected to
478 this assay and the data are expressed as the mean ± SD (n = 3). (C) Superimposed X-ray

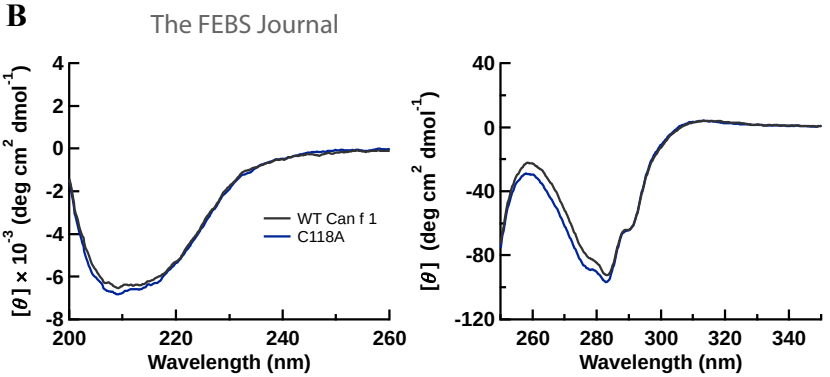
479 structure of Can f 1 (green) and TL (gray; PDB ID: 3EYC). (D) Comparison of the surface
480 electrostatic charge between Can f 1 and TL. The regions of positive and negative electrostatic
481 charge are colored in blue and red, respectively. (E) Mapping conserved regions between Can
482 f 1 and TL (shown in blue). The residues selected for mutation are shown as stick
483 representations. Glu98 was not observed due to unclear electron density. The structures in
484 panels C-E were represented using PyMOL software.
485
486 Fig. 5. IgE reactivity of Can f 1 mutants. (A) SDS-PAGE analyses of the Can f 1 mutants.
487 Purified recombinant proteins of Can f 1 mutant were electrophoretically separated and stained
488 with Coomassie Brilliant Blue. (B) Relative IgE reactivity toward Can f 1 mutants and WT was
489 evaluated by ELISA. Can f 1-reactive sera from six dog-allergic patients was used in this assay.
490 Lines in individual columns denote the mean value. (C) Far-UV (left panel) and near-UV (right
491 panel) CD spectra of WT and Can f 1 mutants in PBS (pH 7.4).
492
493 Fig. 6. Comparison of the surface electrostatic charge among Can f 1 and Mus m 1 (PDB ID:
494 1QY0), Equ c 1 (PDB ID: 1EW3), and Bos d 2 (PDB ID: 4WFU). Electrostatic charges were
495 calculated using PyMOL software with the Adaptive Poisson-Boltzmann Solver plug-in.
496 Positive and negative charges are blue and red, respectively.
497

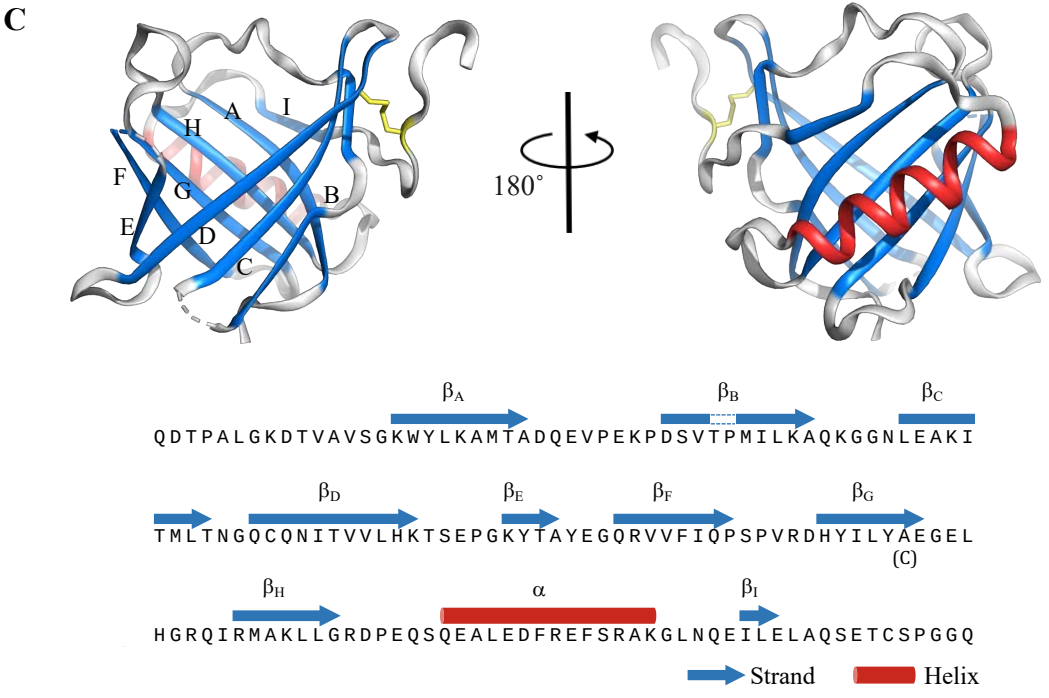
Fig. 7. Structural shapes of the predicted IgE epitope residues for Can f 1. Red circles represent the location of each residue. The structure was represented using PyMOL software.

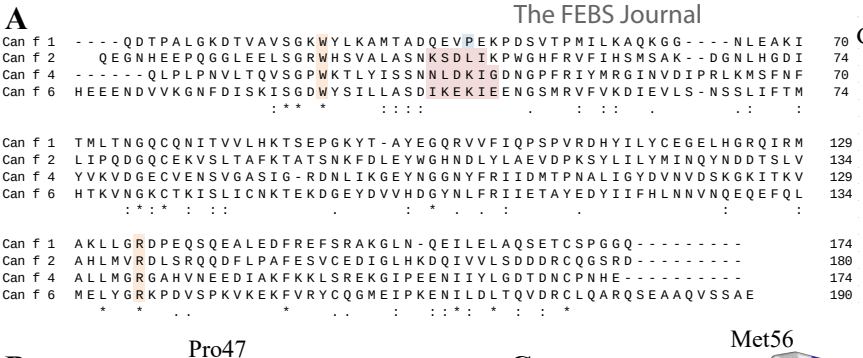
Fig. 8. Primary sequence alignment of Can f 1, Fel d 7, and TL. Blue shading indicates the conserved region between Can f 1 and the other proteins. Residue Arg152 is shaded red.

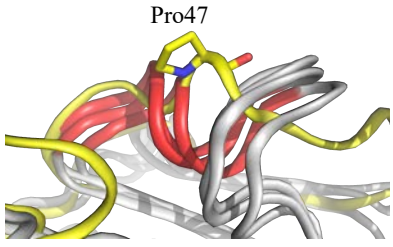

A
45.0 kDa
31.0 kDa
21.5 kDa
Marker
WT Can f 1
C118A
WT Can f 1
C118A
reducing condition
non-reducing condition
B
Absorbance at 450 nm
2.5
2.0
1.5
1.0
0.5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 N1 N2 N3
Dog-allergic patient no.

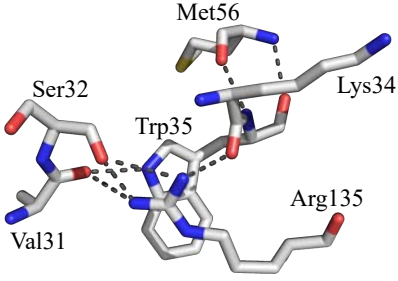

A
IgE reactivity (%)
120
100
80
60
40
20
0
13
14
17
Dog-allergic patient no.

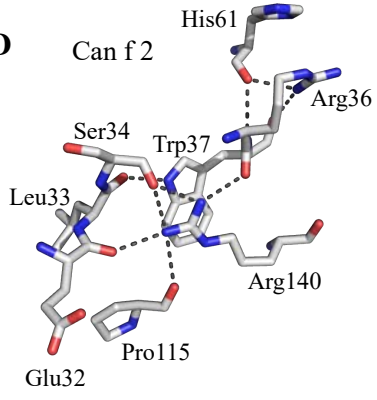

B
[θ] × 10^-3 (deg cm^2 dmol^-1)
WT Can f 1
C118A
Wavelength (nm)
[θ] (deg cm^2 dmol^-1)
Wavelength (nm)

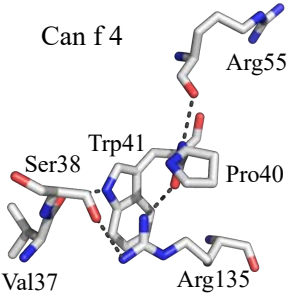

C
A
I
H
F
G
B
E
D
C
180°
βA
βB
βC
QDTPALGKDTVAVSGKWYLKAMTADQEVPEKPDSVTPMILKAQKGGNLEAKI
βD
βE
βF
βG
TMLTNGQCQNITVVLHKTSEPGKYTAYEGQRVVFIQPSPVRDHYILYAEGEL
(C)
βH
α
βI
HGRQIRMAKLLGRDPEQSQEALEDFREFSRAKGLNQEILELAQSETCSPGGQ
Strand
Helix

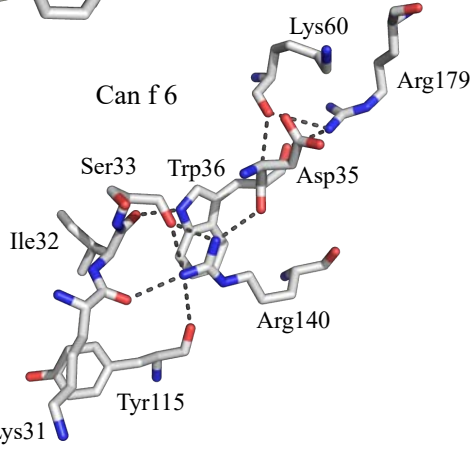

A
Can f 1 ----QDTPALGKDTVAVSGKWYLKAMTADQEVPEKPDSVTPMILKAQKGG----NLEAKI 70
Can f 2 QEGNHEEPQGGLEELSGRWHSVALASNKSDLIKPWGHFRVFIHSMSAK--DGNLHGDI 74
Can f 4 ------QLPLPNVLTQVSGPWKTLYISSNNLDKIGDNGPFRIYMRGINVDIPRLKMSFNF 70
Can f 6 HEEENDVVKGNFDISKISGDWYSILLASDIKEKIEENGSMRVFVKDIEVLS-NSSLIFTM 74
Can f 1 TMLTNGQCQNITVVLHKTSEPGKYT-AYEGQRVVFIQPSPVRDHYILYCEGELHGRQIRM 129
Can f 2 LIPQDGQCEKVSLTAFKTATSNKFDLEYWGHNDLYLAEVDPKSYLILYMINQYNDDTSLV 134
Can f 4 YVKVDGECVENSVGASIG-RDNLIKGEYNGGNYFRIIDMTPNALIGYDVNVDSKGKITKV 129
Can f 6 HTKVNGKCTKISLICNKTEKDGEYDVVHDGYNLFRIIETAYEDYIIFHLNNVNQEQEFQL 134
Can f 1 AKLLGRDPEQSQEALEDFREFSRAKGLN-QEILELAQSETCSPGGQ---------- 174
Can f 2 AHLMVRDLSRQQDFLPAFESVCEDIGLHKDQIVVLSDDDRCQGSRD---------- 180
Can f 4 ALLMGRGAHVNEEDIAKFKKLSREKGIPEENIIYLGDTDNCPNHE---------- 174
Can f 6 MELYGRKPDVSPKVKEKFVRYCQGMEIPKENILDLTQVDRCLQARQSEAAQVSSAE 190
GH loop
AB loop
(short helix)
CD loop
Conserved
Trp residue

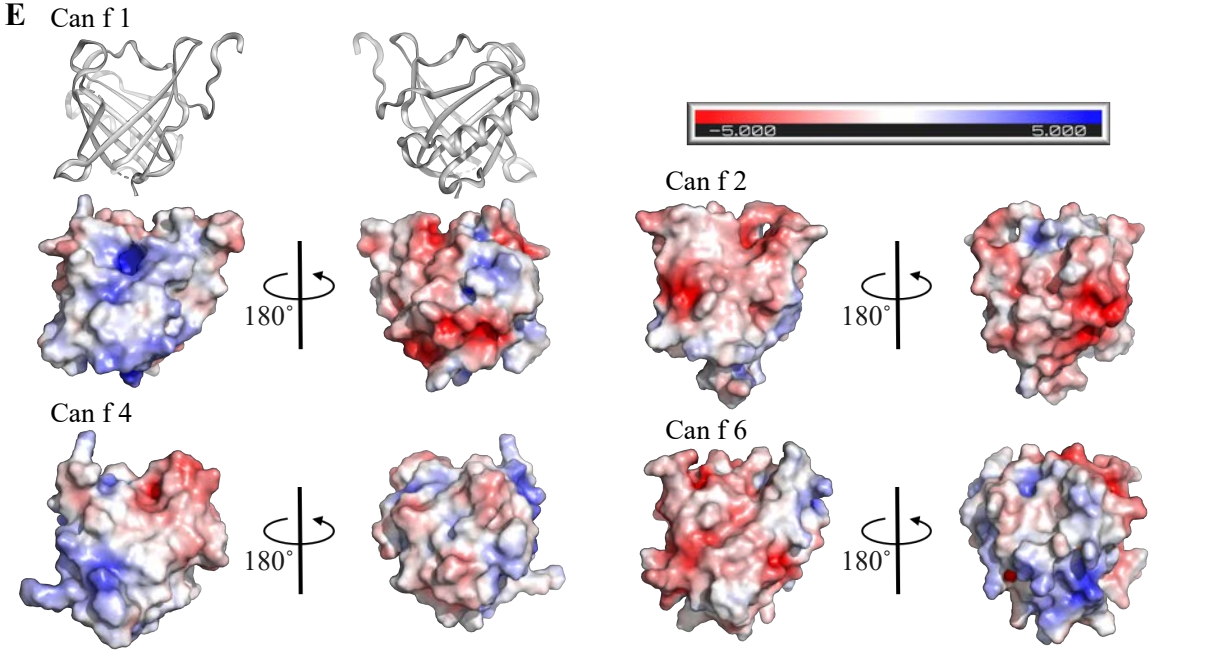

B
Pro47

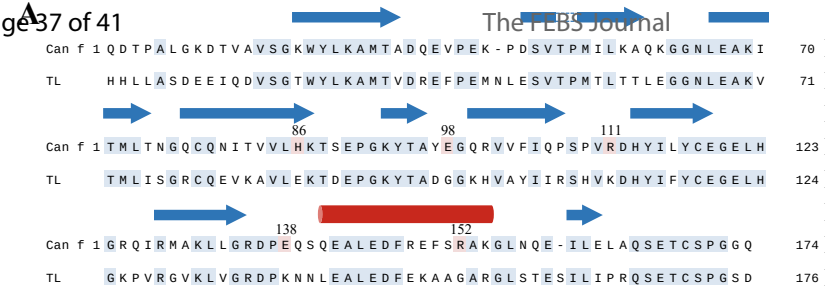

C
Met56
Ser32
Lys34
Trp35
Arg135
Val31

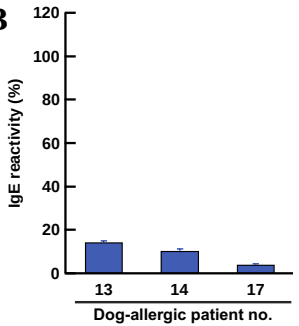

D
Can f 2
His61
Arg36
Ser34
Trp37
Leu33
Arg140
Pro115
Glu32

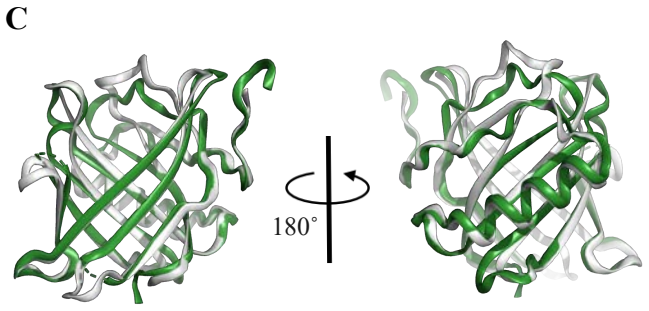

Can f 4
Arg55
Trp41
Ser38
Pro40
Val37
Arg135

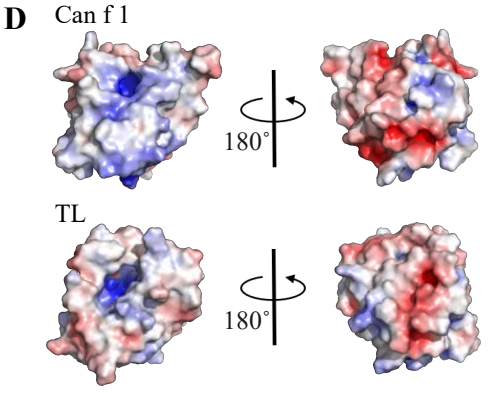

Can f 6
Lys60
Arg179
Ser33
Trp36
Asp35
Ile32
Arg140
Tyr115
Lys31

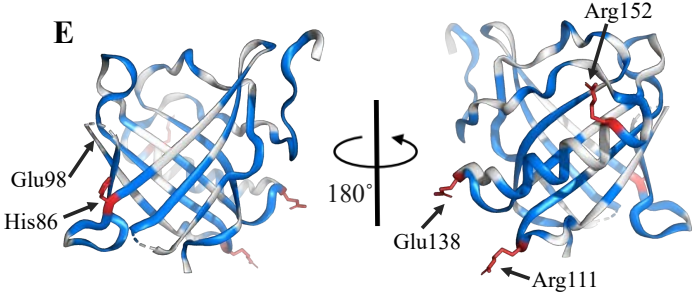

E
Can f 1
-5.000
5.000
180°
Can f 2
180°
Can f 4
180°
Can f 6
180°

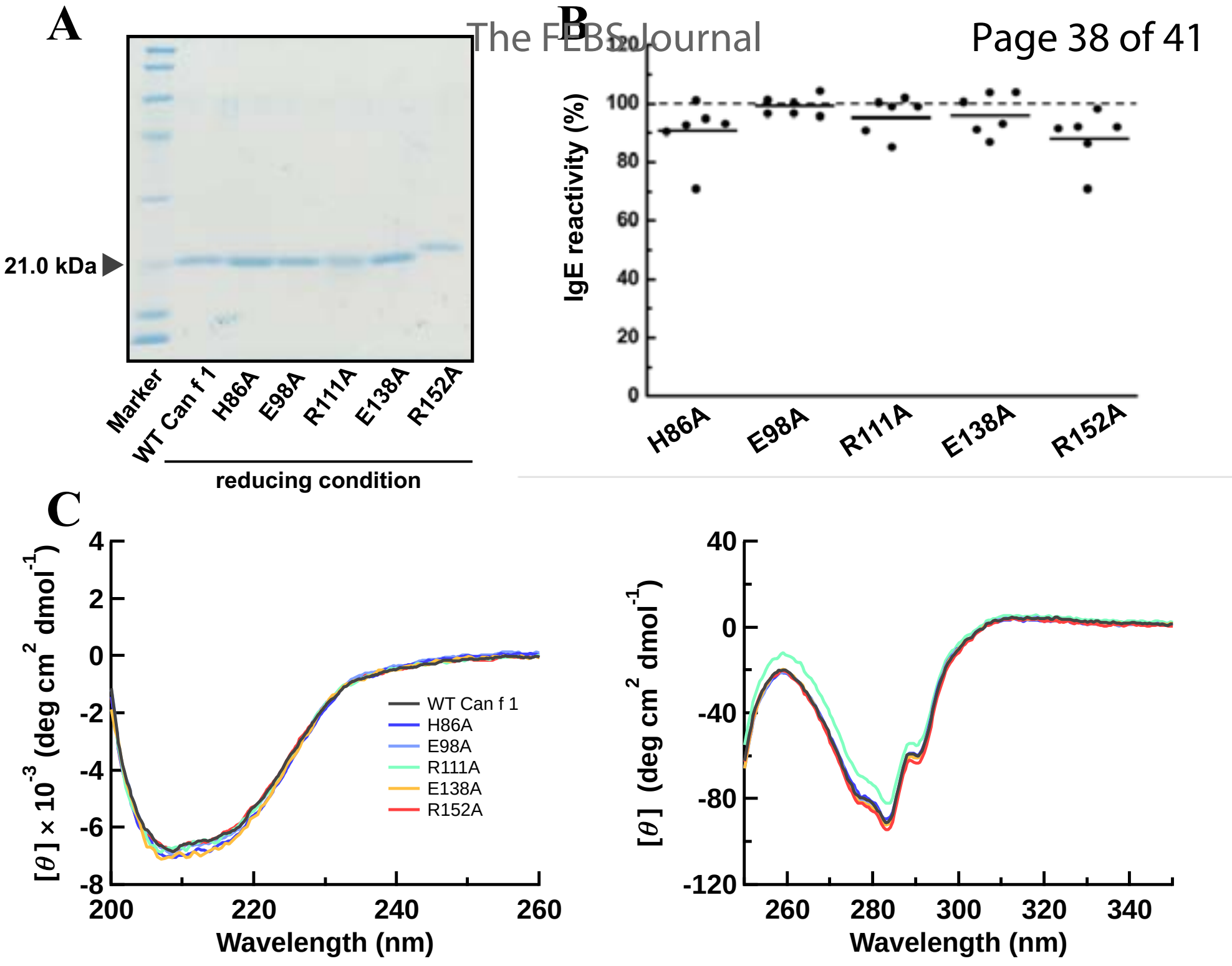

A
Can f 1 QDTPALGKDTVAVSGKWYLKAMTADQEVPEK-PDSVTPMILKAQKGGNLEAKI 70
TL HHLLASDEEIQDVSGTWYLKAMTVDREFPEMNLESVTPMTLTTLEGGNLEAKV 71
86 98 111
Can f 1 TMLTNGQCQNITVVLHKTSEPGKYTAYEGQRVVFIQPSPVRDHYILYCEGELH 123
TL TMLISGRCQEVKAVLEKTDEPGKYTADGGKHVAYIIRSHVKDHYIFYCEGELH 124
138 152
Can f 1 GRQIRMAKLLGRDPEQSQEALEDFREFSRAKGLNQE-ILELAQSETCSPGGQ 174
TL GKPVRGVKLVGRDPKNNLEALEDFEKAAGARGLSTESILIPRQSETCSPGSD 176

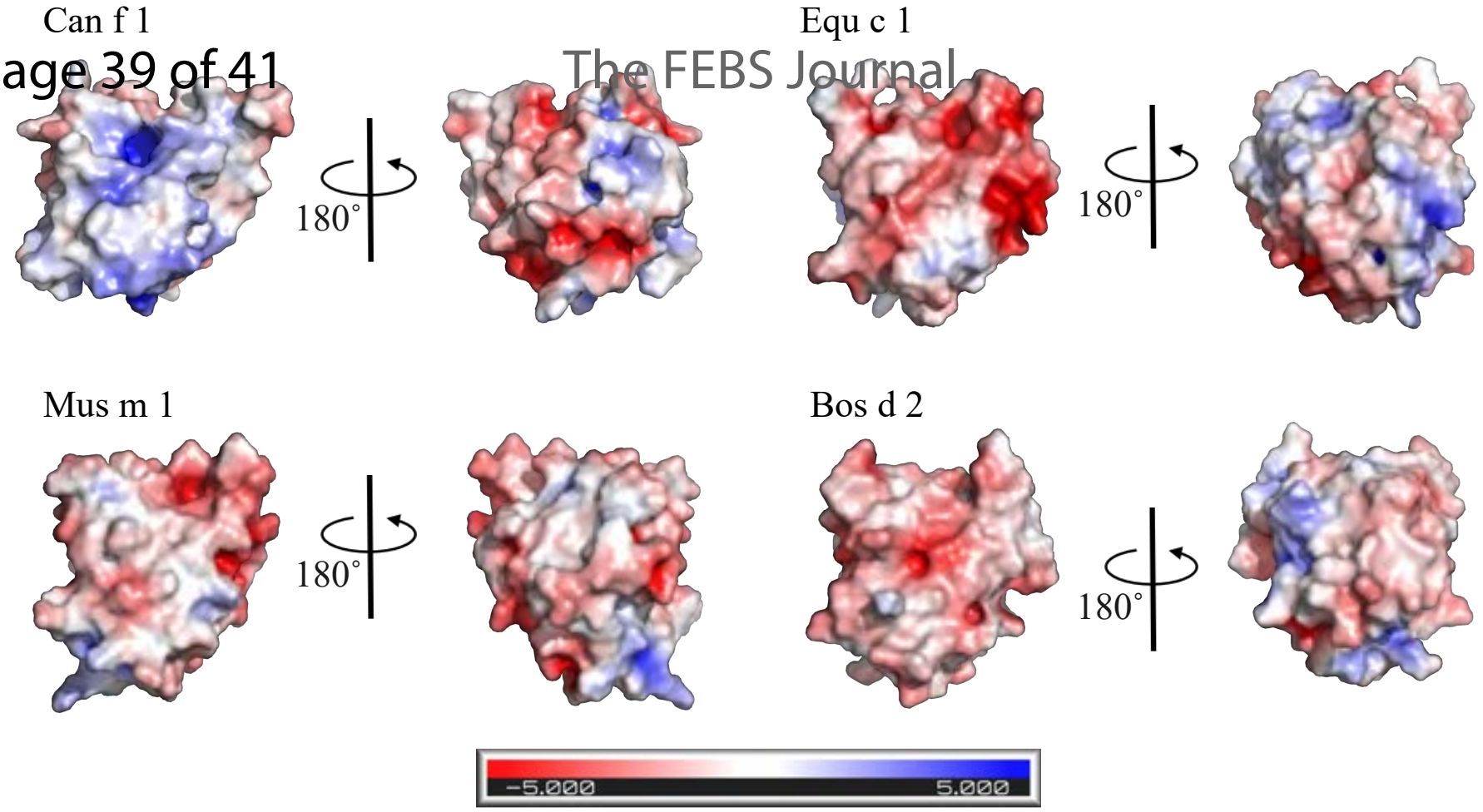

B
IgE reactivity (%)
120
100
80
60
40
20
0
13
14
17
Dog-allergic patient no.

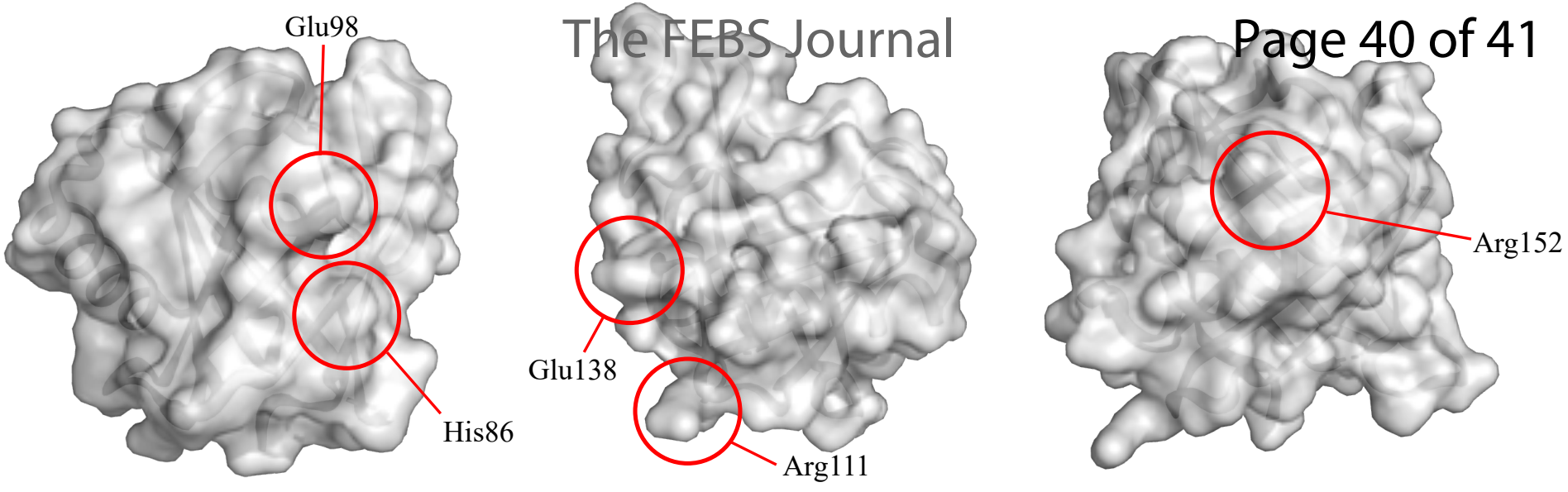

C
180°

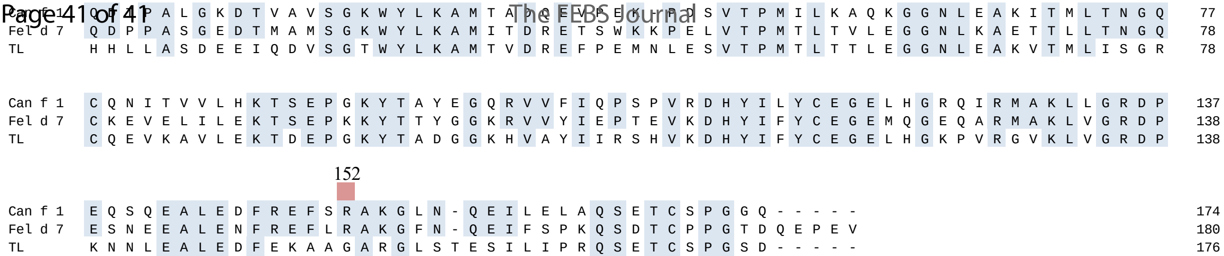

D
Can f 1
180°
TL
180°


E
Glu98
His86
180°
Arg152
Glu138
Arg111


A
21.0 kDa
Marker
WT Can f 1
H86A
E98A
R111A
E138A
R152A
reducing condition
B
IgE reactivity (%)
120
100
80
60
40
20
0
H86A
E98A
R111A
E138A
R152A
C
[θ] × 10^-3 (deg cm^2 dmol^-1)
4
2
0
-2
-4
-6
-8
200
220
240
260
Wavelength (nm)
WT Can f 1
H86A
E98A
R111A
E138A
R152A
[θ] (deg cm^2 dmol^-1)
40
0
-40
-80
-120
260
280
300
320
340
Wavelength (nm)


Can f 1
180°
Equ c 1
180°
Mus m 1
180°
Bos d 2
180°
−5.000
5.000


Glu98
His86
Glu138
Arg111
Arg152


Can f 1 Q E I P A L G K D T V A V S G K W Y L K A M T A D Q E V P E K - P D S V T P M I L K A Q K G G N L E A K I T M L T N G Q 77
Fel d 7 Q D P P A S G E D T M A M S G K W Y L K A M I T D R E T S W K K P E L V T P M T L T V L E G G N L K A E T T L L T N G Q 78
TL H H L L A S D E E I Q D V S G T W Y L K A M T V D R E F P E M N L E S V T P M T L T T L E G G N L E A K V T M L I S G R 78
Can f 1 C Q N I T V V L H K T S E P G K Y T A Y E G Q R V V F I Q P S P V R D H Y I L Y C E G E L H G R Q I R M A K L L G R D P 137
Fel d 7 C K E V E L I L E K T S E P K K Y T T Y G G K R V V Y I E P T E V K D H Y I F Y C E G E M Q G E Q A R M A K L V G R D P 138
TL C Q E V K A V L E K T D E P G K Y T A D G G K H V A Y I I R S H V K D H Y I F Y C E G E L H G K P V R G V K L V G R D P 138
152
Can f 1 E Q S Q E A L E D F R E F S R A K G L N - Q E I L E L A Q S E T C S P G G Q - - - - - 174
Fel d 7 E S N E E A L E N F R E F L R A K G F N - Q E I F S P K Q S D T C P P G T D Q E P E V 180
TL K N N L E A L E D F E K A A G A R G L S T E S I L I P R Q S E T C S P G S D - - - - - 176