Adam Mickiewicz University

Faculty of Biology

Institute of Molecular Biology and Biotechnology

Tomasz Osiński

Determination of the spatial structure and comparative analysis of selected inhaled allergens and their complexes with antibodies

Supervisor:
Professor Janusz M. Bujnicki
The International Institute of Molecular
and Cell Biology, Warsaw
and
Institute of Molecular Biology
and Biotechnology
Faculty of Biology
Adam Mickiewicz University

Co-supervisor:
Professor Władek Minor
Department of Molecular Physiology
and Biological Physics,
University of Virginia, USA

Poznań, 2015

Acknowledgements

Prof. Władek Minor and Prof. Janusz M. Bujnicki – guidance, patience

Prof. Maksymilian Chruszcz – discussion of ideas, showing how things should be done properly, patience

Dr. David Cooper – proofreading

Dr. Matthew Zimmerman – proofreading

Dr. Karolina Majorek – help with crystal growth, wet lab support

Dr. Przemyslaw Porebski – help with programming

Mgr Szymon Osinski – help with structure refinement

…and special thanks to Karolina Majorek for all of her support and encouragement.

Grant support

Der p 1

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01AI077653 (to Maksymilian Chruszcz and initially to Władek Minor), and in part by NIH grant GM53163 (to Władek Minor). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Bla g 4

The work described in this paper was supported partially by GM053163 (to Władek Minor) grant and funds from University of South Carolina (to Maksymilian Chruszcz).

Alt a 1

Supported by National Institutes of Health grant GM53163 (to Władek Minor). The structural results shown in this report are derived from work performed at Argonne National Laboratory at the Structural Biology Center of the Advanced Photon Source. Argonne is operated by University of Chicago Argonne, LLC, for the US Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (grant 085P1000817).

Results presented in this dissertation are included in:

Research papers:

Structural analysis of Der p 1-antibody complexes and comparison with complexes of proteins or peptides with monoclonal antibodies

Tomasz Osiński, Anna Pomés, Karolina A. Majorek, Jill Glesner, Lesa R. Offermann, Lisa D. Vailes, Martin D. Chapman, Władek Minor, Maksymilian Chruszcz

J Immunol. 2015 Jul 1;195(1):307-16. doi: 10.4049/jimmunol.1402199. Epub 2015 May 29. PubMed PMID: 26026055

***Alternaria alternata* allergen Alt a 1: a unique β-barrel protein dimer found exclusively in fungi.**

Maksymilian Chruszcz, Martin D. Chapman, **Tomasz Osiński**, Robert Solberg, Matthew Demas, Przemysław J. Porebski, Karolina A. Majorek, Anna Pomés, Władek Minor

J Allergy Clin Immunol. 2012 Jul;130(1):241-7.e9. doi: 10.1016/j.jaci.2012.03.047. Epub 2012 Jun 2. PubMed PMID: 22664167; PubMed Central PMCID: PMC3391610.

Molecular determinants for antibody binding on group 1 house dust mite allergens.

Maksymilian Chruszcz, Anna Pomés, Jill Glesner, Lisa D. Vailes, **Tomasz Osiński**, Przemysław J. Porebski, Karolina A. Majorek, Peter W. Heymann, Thomas A. E. Platts-Mills, Władek Minor, Martin D. Chapman

J Biol Chem. 2012 Mar 2;287(10):7388-98. doi: 10.1074/jbc.M111.311159. Epub 2011 Dec 30. PubMed PMID: 22210776; PubMed Central PMCID: PMC3293536.

The major cockroach allergen Bla g 4 binds tyramine and octopamine.

Lesa R. Offermann, Siew Leong Chan, **Tomasz Osiński**, Yih Wan Tan, Fook Tim Chew, Jayaraman Sivaraman, Yu-Keung Mok, Władek Minor, Maksymilian Chruszcz

Mol Immunol. 2014 Jul;60(1):86-94. doi: 10.1016/j.molimm.2014.03.016. Epub 2014 Apr 24.

PDB deposits

4PP1

The crystal structure of Der p 1 allergen complexed with Fab fragment of mAb 5H8

Tomasz Osiński, Karolina A. Majorek, Anna Pomes, Lesa R. Offermann, Szymon Osiński, Jill Glesner, Lisa D. Vailes, Martin D. Chapman, Władek Minor, Maksymilian Chruszcz

4PP2

The crystal structure of Der p 1 allergen complexed with Fab fragment of mAb 10B9

Tomasz Osiński, Karolina A. Majorek, Anna Pomes, Lesa R. Offermann, Szymon Osiński, Jill Glesner, Lisa D. Vailes, Martin D. Chapman, Władek Minor, Maksymilian Chruszcz

4POZ

Fab fragment of Der p 1 specific antibody 10B9

Tomasz Osiński, Karolina A. Majorek, Anna Pomes, Lesa R. Offermann, Szymon Osiński, Jill Glesner, Lisa D. Vailes, Martin D. Chapman, Władek Minor, Maksymilian Chruszcz

4N7C

Structural re-examination of native Bla g 4

Lesa R. Offermann, Siew Leong Chan, **Tomasz Osiński**, Yih Wan Tan, Fook Tim Chew, Jayaraman Sivaraman, Yu-Keung Mok, Władek Minor, Maksymilian Chruszcz

4N7D

Selenomethionine incorporated Bla g 4

Lesa R. Offermann, Siew Leong Chan, **Tomasz Osiński**, Yih Wan Tan, Fook Tim Chew, Jayaraman Sivaraman, Yu-Keung Mok, Władek Minor, Maksymilian Chruszcz

3V0R

Crystal structure of Alternaria alternata allergen Alt a 1

Maksymilian Chruszcz, Robert Solberg, **Tomasz Osiński**, Martin D. Chapman, Władek Minor

Poster presentations:

Bioinformatic analysis of selected allergen families

Tomasz Osiński, Karolina A. Majorek, Maksymilian Chruszcz, Janusz M. Bujnicki and Wladek Minor

Presented at:

- *Protein & RNA Structure Prediction Conference on Protein Structure Prediction & Refinement, Mexico, 3 – 7 December 2011*

- *14th International Conference on the Crystallization of Biological Macromolecules, Huntsville, Alabama, USA, 23 – 28 September 2012*

- *42nd Mid-Atlantic Protein Crystallography Workshop, Charlottesville, VA, USA, 31 May – 2 June 2012*

Structural Analysis Reveals Molecular Basis for Interactions of Group 1 Allergens with Species Specific and Cross-Reactive Antibodies

Maksymilian Chruszcz, Anna Pomés, **Tomasz Osiński**, Karolina A. Majorek, Jill Glesner, Władek Minor, Lisa D. Vailes, Martin D. Chapman

Presented at:

American Academy of Allergy, Asthma and Immunology, Annual Meeting 2013; San Antonio, Texas; 22-26 February 2013

Structural analysis of complexes formed by the major house dust mite allergen Der p 1 with 5H8 and 10B9 antibodies.

Tomasz **Osiński**, Anna Pomés., Karolina A. Majorek, Jill Glesner, Lisa D. Vailes, Martin D. Chapman, Władek Minor, Maksymilian Chruszcz

Presented at:

- *43rd Mid-Atlantic Macromolecular Crystallography Meeting; Durham, North Carolina; 30 May – 1 June 2013*

Analysis of complexes formed by the major house dust mite allergen Der p 1 with 4C1, 5H8 and 10B9 antibodies

Tomasz Osiński, Anna Pomés, Karolina A. Majorek, Przemyslaw J. Porebski, Lisa D. Vailes, Martin D. Chapman, Wladek Minor, Maksymilian Chruszcz

Presented at:

American Crystallographic Association Meeting, Honolulu, Hawaii, USA, 20 - 24 July 2013

Major allergen from house dust mite Der p 1 complexed with Fab fragments of monoclonal antibodies 5H8 and 10B9

Tomasz Osiński, Anna Pomés, Karolina A. Majorek, Jill Glesner, Lesa Offermann, Lisa D. Vailes, Martin D. Chapman, Wladek Minor, Maksymilian Chruszcz

Presented at:

Multi-Pole Approach to Structural Science, Warsaw, Poland, 10 - 13 May 2015

Table of Contents

1. Streszczenie

Alergie towarzyszą człowiekowi niemal od zarania dziejów. Już w starożytnym Egipcie, Mezopotamii i Grecji znano reakcje alergiczne na pewne substancje. Alergią nazywamy stan, w którym organizm reaguje na substancję niegroźną w sposób gwałtowny i nieadekwatny do poziomu rzeczywistego zagrożenia. Reakcja alergiczna może występować pod wpływem różnych czynników i pod różnymi postaciami. Astma, wysięk z nosa, wysypka, problemy pokarmowe to tylko główne schorzenia związane z alergią. Astma jest jedną z najczęstszych (Masoli et al., 2004) i jedną z najpoważniejszych chorób dróg oddechowych, która może być powodowana przez alergeny wziewne (Busse and Lemanske, 2001). Przebieg astmy może być ostry lub chroniczny i ekspozycja na pyłek, ślinę i naskórek zwierzęcy, odchody roztoczy kurzu domowego, różne substancje karalusze, a także substancje niebędące alergenami mogą spowodować atak astmy. Poznanie struktury pneumoalergenów oraz próba zrozumienia molekularnych podstaw oddziaływań pomiędzy badanymi alergenami i przeciwciałami może przyczynić się do opracowania w przyszłości odpowiedniej terapii immunologicznej a tym samym zmniejszenia objawów astmy u osób nią dotkniętych.

Głównym elementem tezy doktorskiej było zbadanie molekularnych podstaw oddziaływania głównego alergenu Grupy 1 z roztocza kurzu domowego pochodzącymi z Europy Der p 1 (skórożarłoczek skryty - *Dermatophagoides pteronyssinus*) oraz roztoczy kurzu domowego pochodzącymi z Ameryki (*Dermatophagoides farinae*) Der f 1 z przeciwciałami monoklonalnymi 4C1, 5H8 oraz 10B9. Została przeprowadzona szczegółowa analiza powierzchni oddziaływania epitopu oraz paratopu w otrzymanych kompleksach alergenu z przeciwciałami.

Zbadanie alergenu Bla g 4 pochodzącego z karaczana prusaka (*Blatella germanica*) było kolejną częścią projektu poznania pneumoalergenów. Alergen ten należy do lipokalin – rodziny

białek wiążących małe cząsteczki. Lipokaliny charakteryzują się stosunkowo niewielkim zachowaniem ewolucyjnym na poziomie sekwencji, ale silnym na poziomie struktury. Funkcja Bla g 4 oraz ligand wiązany przez to białko są nieznane. Rozwiązanie struktury Bla g 4 umożliwiło identyfikację ligandu, a przeprowadzenie analiz strukturalnych, sekwencyjnych oraz filogenetycznych najbliższego homologa – Per a 4 oraz innych spokrewnionych alergenów umożliwi w nieodległej przyszłości poznanie ich wzajemnych relacji.

Poznanie struktury oraz zbadanie homologów alergenu Alt a 1 było ostatnią częścią projektu. Alergen Alt a 1 pochodzi z pleśni *Alternaria alternata* występującej powszechnie w klimacie umiarkowanym. Alt a 1 jest alergenem o nieznanej strukturze i funkcji. Rozwiązanie struktury, oraz analizy sekwencyjne i filogenetyczne będą pierwszym krokiem do przyszłych badań nad tym alergenem.

Kompleksy alergenu Der p 1 z przeciwciałami 10B9 oraz 5H8, a także przeciwciała 10B9 w niezwiązanej formie dostarczyły unikatowej okazji do przeanalizowania sposobu wiązania regionów determinujących komplementarność przeciwciał do epitopów. Bardzo bliskie pokrewieństwo Der p 1 oraz Der f 1 oraz różnorodność otrzymanych struktur umożliwiła porównanie zmian zachodzących podczas wiązania przeciwciał, a także rodzajów uwarunkowań molekularnych do ich wystąpienia. Umożliwiło to porównanie zmian zachodzących podczas wiązania przeciwciała 10B9 do Der p 1 w kontekście otrzymanej uprzednio struktury alergenów Der p 1 i Der f 1 z podwójnie swoistym przeciwciałem 4C1. Otrzymane wyniki prowadzą do wniosku, że nawet takie same lub prawie identyczne epitopy mogą zachowywać się zgodnie zarówno z modelem „klucza i zamka” jak i modelem indukowanego dopasowania Identyfikacja reszt aminokwasowych odgrywających znaczącą rolę w oddziaływaniach alergenu z przeciwciałami oraz zrozumienie strukturalnych podstaw komplementarności między nimi może zostać wykorzystane w projektowaniu alergenów o

epitopach charakteryzujących się obniżoną siłą wiązania przeciwciał do celów immunoterapii alergenowej.

Dzięki krystalografii rentgenowskiej możliwe było poznanie szczegółów oddziaływania alergenu Bla g 4 z tyraminą, a dzięki analizie struktur oraz sekwencji białek homologicznych, będących także alergenami, poznanie zachowanego ewolucyjnie miejsca i sposobu wiązania tego ligandu wśród pokrewnych alergenów. Okazuje się, że nawet najbliższy homolog Bla g 4 - Per a 4 pochodzący z karalucha amerykańskiego (*Periplaneta americana*) nie ma zachowanych kluczowych aminokwasów odpowiedzialnych za wiązanie tyraminy i oktopaminy, więc najprawdopodobniej wiąże inne ligandy oraz pełni inną funkcję.

Poznanie struktury alergenu Alt a 1, jako unikalnej dimerycznej β-baryłki, a także jako pierwszej z całej rodziny białek z grzybów o nieznanej funkcji jest pierwszym krokiem w celu dalszych badań nad funkcją oraz powiazaniem struktury z funkcja, co może doprowadzić do opracowania nowych form immunoterapii dla osób uczulonych na ten alergen.

Uzyskanie struktur krystalicznych za pomocą rentgenografii krystalograficznej oraz analizy molekularnych podstaw oddziaływania alergenów z przeciwciałami; analizy strukturalnej wraz z sekwencyjną między homologicznymi alergenami może w przyszłości zostać wykorzystana do celów farmaceutycznych. Wyniki tych badań pokazują, że zastosowanie połączenia różnych technik umożliwia otrzymanie optymalnych rezultatów.

2. Abstract

Human kind has been troubled by allergies since the beginning of written history. Allergic reactions to certain substances have been known even in the ancient Egypt, Mesopotamia or Greece. What we call an allergy is a state when organism reacts to otherwise unharmful substance in the violent and inappropriate manner to real danger posed by the given substance. An allergic reaction can occur under the influence of different factors and different forms. Asthma, rhinitis, rash, digestive problems are the main ailments related to allergies. Asthma is one of the most common (Masoli et al., 2004) and one of the most serious diseases of the airways caused by inhaled allergens (Busse and Lemanske, 2001). The course of asthma may be acute or chronic, and asthma attacks may be caused by exposure to pollen, animal saliva, animal dander, feces of the house dust mites, cockroach particles, as well as certain non-allergenic. The elucidation of pneumoallergen structures and an attempt to understand the molecular basis of the interactions between analyzed allergens and antibodies may contribute to the development of the proper immunotherapy and thus reduce asthma symptoms in people affected by it.

The main part of this project was to analyze the molecular basis of the interaction between Group 1 major allergens from house dust mites – Der p 1 coming from European house dust mite (*Dermatophagoides pteronyssinus*) and Der f 1 coming from American house dust mite (*Dermatophagoides farinae*) with 4C1, 5H8 and 10B9 monoclonal antibodies. A detailed analysis of the interaction surface between the epitopes and the paratopes of the obtained complexes has been conducted.

The analysis of the Bla g 4 allergen coming from German cockroach (*Blatella germanica*) was another part of the project concerning pneumoallergens. This allergen belongs to lipocalin protein family, which usually bind small ligands. The lipocalins are characterized

by their relatively low sequence conservation, but strong structural similarity. The function of Bla g 4 as well as the ligand it binds were previously unknown, but the structure determination of Bla g 4 presented herein allowed for the identification of the ligand. The analysis of the structure and sequence of the closest homolog – Per a 4 as well as other homologous allergens allow recognition of interrelationships.

The elucidation of the structure of Alt a 1 was the last part of the project. The Alt a 1 allergen comes from black mold (*Alternaria alternata*), which is common in the outdoor environment in the mild climate zones and is a major health hazard for humans when . Both the structure and the function of the Alt a 1 allergen are unknown. The structure solution together with the analysis of its sequence is the first step for the future research.

The Der p 1 allergen complexed with monoclonal antibodies 10B9 and 5H8, as well as 10B9 antibody in its uncomplexed form provided a unique opportunity to study the mechanics of the binding of the complementarity determining regions to the epitopes. A very close homology between the Der p 1 and Der f 1 allergens together with the variety of the obtained structures allowed for the comparison of the changes undergoing upon the binding of the antibodies, as well as the molecular determinants involved in this process. This includes the changes in the conformation of the 10B9 antibody and the comparison with the results of previous study on the binding of the cross-reactive antibody 4C1 by Der p 1 and Der f 1. The obtained results show that both “lock and key” and “induced fit” binding models can coexist even in the same area of the epitopes. The identification of the amino acid residues having important role in the allergen-antibody interactions and the understanding of the molecular basis of the complementarity between them can be used in the design of allergens with the epitopes of lesser affinity to the given antibodies that may be beneficial in immunotherapy.

Thanks to x-ray crystallography, it was possible to study the details of the Bla g 4 allergen with tyramine, and as a result of the analysis of the structures together with the sequences of its homologs, it was possible to determine the conservation level of the binding site. This, in turn, provided clues to the ligand binding among homologs related to Bla g 4. It turned out that even the closest homolog to Bla g 4 – the Per a 4 allergen from the American cockroach (*Periplaneta americana*) does not have conserved key amino acids responsible for the binding of tyramine or octopamine, thus most likely it binds to other ligands and serves a different purpose.

Revealing the Alt a 1 allergen structure as a unique dimeric β-barrel protein, as well as solving it as a first structure of the whole protein family of unknown function and exclusive to fungi, is the first step for further research and identifying structure-function relationship, which can lead to the development of the new forms of immunotherapy for Alt a 1 sensitive patients.

The knowledge gained by the elucidation of the crystal structures with x-ray crystallography methods, together with the result of the analysis of the molecular basis of allergen-antibody interactions as well as the structure and sequence analysis between the homologous allergens, may be used in the future for the pharmaceutical purposes. The outcome of the experimental and theoretical approach presented herein shows that the combination of different techniques provides more information than just the sum of the individual results.

3. Abbreviations

4-HTP – 4-hydroxy-2,5,6-triaminopyrimidine

8-ACA – 8-aminocaprylic acid

Alt a 1 – *Alternaria alternata* major allergen 1

Bla g 4 – *Blatella germanica* major allergen 4

Der f 1 – *Dermatophagoides farinae* major allergen 1

Der p 1 – Dermatophagoides pteronyssinus major allergen 1

DLS – Dynamic Light Scattering

HDM – house dust mites

PDB – Protein Data Bank

NCBI – National Center for Biotechnology Information

VL – light chain of variable region of Ig

VH – heavy chain of variable region of Ig

LV – low viscosity

PEG – polyethylene glycol

PNMT – phenylethanolamine N-methyltransfrase

RMSD – root mean square deviation

4. Introduction

4.1. Allergy

An allergy (or allergic hypersensivity) and allergic diseases have been described in various sources since ancient Mesopotamia, Egypt, Asia or America (Bergmann and Ring, 2014). In short, an allergy is a destructive, hypersensitive inflammatory response of immune system to normally harmless substances known as allergens. An allergic response of an organism may cause multiple diseases or maladies associated with allergies such as asthma, allergic rhinitis (hay fever), atopic dermatitis, and food allergy. An atopy can defined as the genetic tendency to develop an hypersensivity to common allergens (Coca and Cooke, 1923). However, the immune system does not always mediate a hypersensivity to specific substance, but rather other mechanisms are involved (Szczeklik and Stevenson, 2003; Szczeklik, 1997). The terms allergy and atopy are often interchanged, but they are different. Atopy is an IgE-mediated immune response, but allergy is any exaggerated immune system response regardless of mechanism.

The allergen-specific lymphocytes play a dominant role in IgE-mediated allergic inflammation. In the response to allergic inflammation, mucous membranes present an increased sensitivity and may provoke or exacerbate other allergic symptoms by non-immunological factors like infection (Jackson et al., 2008; Kelly and Busse, 2008; Newcomb and Peebles, 2009; Wu et al., 2008), irritants, or physical exertion (Weiler et al., 2007). Normal antigen exposure causes IgG antibodies to be produced without visible symptoms, a process called lymphocyte sensitization (Hart and Mitchell, 1971). High concentrations of IgG, as well as IgE, antibodies have been shown to be of importance in allergic bronchopulmonary aspergillosis (Patterson et al., 1986); therefore, medium levels of

IgG antibodies to an antigen is not necessarily a sign of allergic disease. However, the presence of IgE antibodies may be a result of an allergic inflammation.

Certain allergic diseases are not mediated by IgE antibodies. Serum sickness (Jackson, 2000) was the disease that made the term allergy popular. Similarly to anaphylaxis, which is caused by complement activated by immunocomplexes containing dextran (Hedin et al., 1976), serum sickness is caused by the antibodies of the IgG isotype; however, specific lymphocytes mediate the disease in allergic contact dermatitis (Kimber et al., 2002).

4.2. Allergic asthma

Asthma is one of the most common diseases in the world. It is estimated that around 300 million people in the world currently have asthma (Masoli et al., 2004). Asthma is becoming more common in the United States and the number of people with asthma increased by 2.9% every year from 20.3 million in 2001 to 25.7 million in 2010 (Moorman et al., 2012) and accounts for ~1.6 million visits to emergency departments each year (Schappert and Rechtsteiner, 2008). In most cases asthma is initiated by IgE antibodies and the full name of this malady is IgE-mediated allergic asthma. Asthma can be described as a common, life-long, and the most serious, allergic disease that affects today's society. Asthma is a multiple form disorder that affects the freedom of airflow to the lungs by contraction of the bronchial tubes caused by hyper responsiveness to inflammation or as a response to the environment, a pathogen or drug treatment (Lodge et al., 2011).

The beginning of this disease can be traced to the infancy or early childhood (Jackson et al., 2008). However, adults were also found to be susceptible (Knutsen et al., 2012). Social background, age and ethnicity are among the factors that contribute to the development of asthma (Moorman et al., 2012). The stages of asthma may be acute or chronic. The rates of asthma related deaths have decreased in developed countries in the last two decades (Bartolomei-Díaz et al., 2011; Garne et al., 2014; Goldacre et al., 2012; Hasegawa et al., 2013;

Lotufo and Bensenor, 2012; Pesut et al., 2011; Prietsch et al., 2012; Roberts et al., 2013), perhaps as the result of asthma education and proper treatment (Chawla et al., 2012); however, asthma related deaths increased in Taiwan (Hsiao et al., 2013).

Inhaled allergens can exacerbate asthma symptoms. Fungal spores, cockroach particles, feces of the house dust mites, pollen, animal dander and dust are among the top ten most common allergens in the United States (Arbes et al., 2007) and they have been intensively studied. However, the amount of information that can be obtained from analysis of allergen structures exceeds the limitations of other methods.

4.3. Exacerbating factors

There are several factors that exacerbate the course of asthma. Exercise (Weiler et al., 2007), respiratory tract infections caused by viruses (Jackson et al., 2008; Kelly and Busse, 2008; Wu et al., 2008), *Chlamydia* or *Mycoplasma* species (Newcomb and Peebles, 2009), nonsteroidal anti-inflammatory drugs (Szczeklik and Stevenson, 2003), gastroesophageal reflux disease (Coughlan et al., 2001) or allergens. Persistent exposure to small amounts of indoor allergens, mainly coming from house dust mites and cockroaches, takes part in both early asthma stages and subsequent symptom escalation (Platts-Mills, 2008).

There are multiple sources of the inhaled allergens that can provoke asthma symptoms, but sensitization to house dust mites (Celedón et al., 2007), cockroach particles (Gruchalla et al., 2005), various species of mold from the *Alternaria* genus (Bush and Prochnau, 2004), and cat (Arbes et al., 2007) is significant in the asthma development. The exposure to a number of pets (Ownby et al., 2002) and frequent contact with them during infancy may reduce subsequent risk of the future development of allergic sensitization (Bufford et al., 2008). However, ownership of a cat does not seem to have this effect (Popp et al., 1990).

The level of complexity of interactions found in association with these findings is significant and it shows involvement of the genotype-environment interactions (Tabery, 2007).

Allergen specific immunotherapy, with the use of grass and birch pollen, can prevent the development of asthma and reduce risk hyperresponsiveness and asthma in children with seasonal allergic rhinitis (Möller et al., 2002) and it has long-term clinical effect (Niggemann et al., 2006).

4.4. Antibody structure and Fab

Antibodies comprise two identical heavy chains and two identical light chains and have a Y-shaped structure; therefore, the symmetry of the antibody can be described as pseudo-twofold. The stem contains the constant region, named after the fact, that its structure is strongly conserved among antibodies. This region triggers various mechanisms as a response to recognized antigens (Edelman, 1973; Porter, 1973). There is a group of antibodies, however, that has a quite different overall structure that comprises only single variable part of the heavy chain.

The average antibody (Figure 1) consists of two regions, a fragment crystallizable region (Fc region) and fragment antigen binding (Fab fragment). The Fc region interacts with the surface receptors and the complement proteins and therefore transmits signals from antibodies. The Fc region is usually created by the second and third constant domain of two heavy chains. The Fab is the part of an antibody responsible for binding of an antibody to antigens. The Fab fragment is built by one variable and one constant domain of each of the chain (Al-Lazikani et al., 1997). Papain can be used to cleave an antibody into one Fc fragment and two Fab fragments (Newkirk et al., 1987).

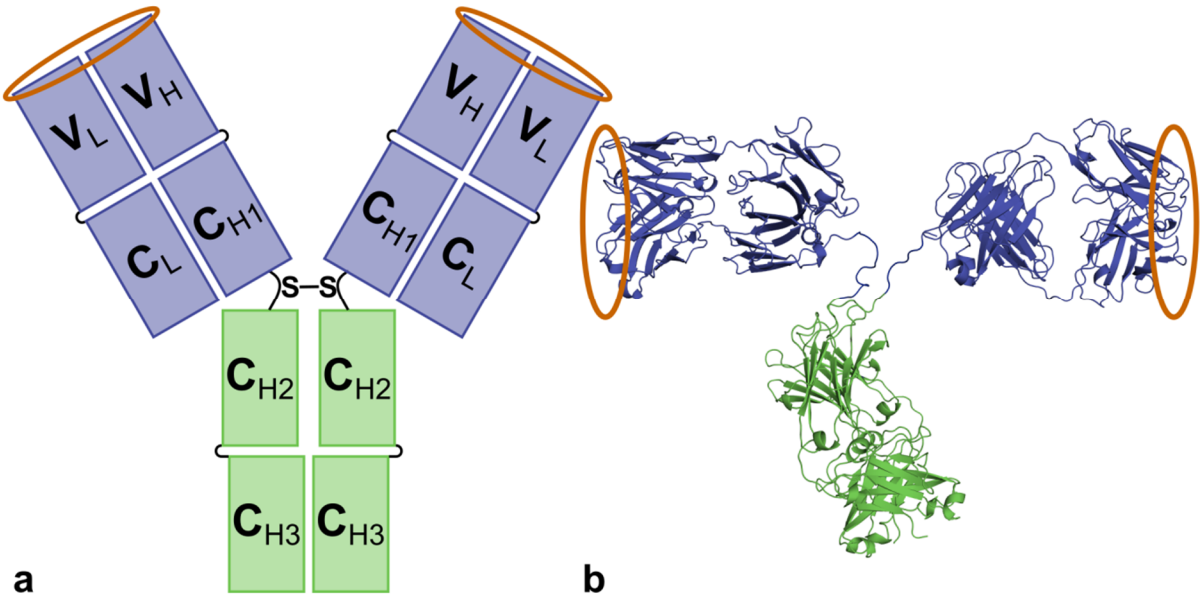

Figure 1 Schematic model of an IgG (a) and structure of an IgG (b).

Two Fab fragments are colored blue and one green colored Fc fragment. The compatibility determining regions (CDRs) are in orange circles. The PDB code of the structure used: 1IGT.

4.5. Paratopes and CDRs

Antibodies are produced with variety of specificities toward many antigens, yet the number of genes coding antibodies is limited. Genes coding the variable region of antibodies are segmented and are joined by recombination (Hozumi and Tonegawa, 1976; Matthyssens et al.; Tonegawa, 1976). Successful recombination prevents further somatic gene segments rearrangements, essentially locking each cell and its progenitors with one type of receptor specificity. The order of gene segments is random; therefore, genes coding variable regions of light and heavy chains can be unique and expressed by one type of lymphocytes. Thanks to this system, it is possible to create an enormous number of different antibodies from a limited number of genes.

The recognition of an antigen is carried by three loops from the light chain (L1, L2 and L3) and three loops from the heavy chain (H1, H2 and H3) forming the paratope (Figure 2). These loops are called complementarity determining region (CDR). The overall structure of the light chain and the heavy chain variable domains is highly conserved with the exception for

CDRs (Wu and Kabat, 1970), thus they were rendered a very interesting target for classification. There were numerous attempts to classify CDRs and almost all of them focused on identifying a small discrete set of main chain conformations – "canonical structures" – that could be used to group conformational classes of CDRs (Al-Lazikani et al., 1997; Chothia and Lesk, 1987; Martin and Thornton, 1996; Morea et al., 1998; Shirai et al., 1999).

The CDR H3 loop was hard to classify due to large conformational variety, but in recent studies (Nikoloudis et al., 2014; North et al., 2011) much larger datasets allowed for better classification and increased the potential for CDR predictability from sequence.

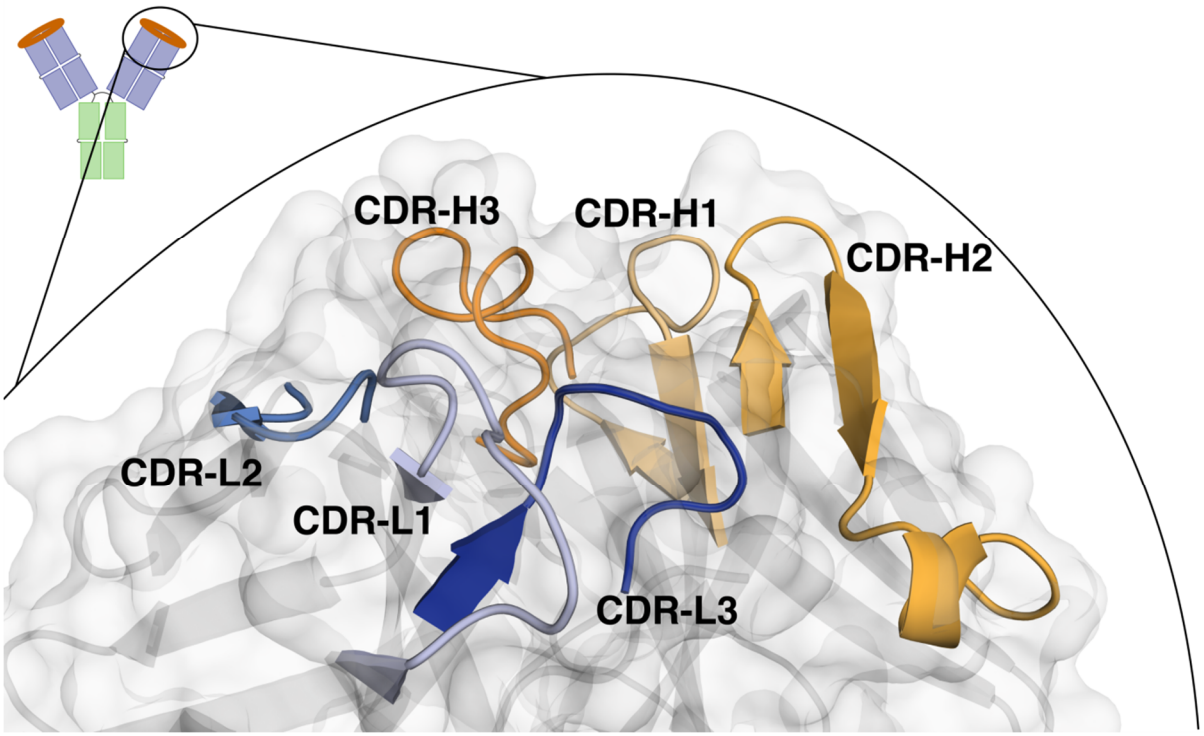

Figure 2 An example of the CDR loops (10B9 antibody).

The CDR loops from light chain are colored in different shades of blue color, the CDR loops from heavy chain are colored in different shades of orange. (PDB code: 4POZ)

4.6. House Dust Mites

House dust mites (HDM) are a common source of indoor allergens in households around the world and a major cause of perennial asthma worldwide (Platts-Mills et al., 1992; Thomas et al., 2010). House dust mites can be found in beds, carpets and soft furniture. The members of the *Dermatophagoides* genus feed on dander and small particles of shed skin, which is common in their habitat. The digestive enzymes of the house dust mites are potent proteases that are abundant in their feces, and are highly allergenic. Der p 1 is a major allergen and a cysteine protease (Chapman and Platts-Mills, 1980), that catalyzes the cleavage of the amide linkages in a variety of substrates, including α1-antitrypsin, the CD23 receptor on human B cells, the IL-2 receptor (CD25) on human T cells and the Der p 1 pro-polypeptide sequence (Shakib et al., 2008). Strong evidence suggests that the allergenicity of Der p 1 is exacerbated by Der p 1-related cleavage of these receptors (Schulz et al., 1998; Shakib et al., 1998). The exposure to Der p 1 occurs by inhalation of HDM fecal matter and if it is persistent may cause the production of IgE antibodies in susceptible individuals. The WHO/IUIS nomenclature (Radauer et al., 2014) contains seventeen registered allergens originating in European house dust mite (*Dermatophagoides pteronysinnus*) and twenty seven allergens from American house dust mite (*Dermatophagoides farinae*).

The structures of proenzyme and mature forms of recombinant Der p 1 were determined previously (Chruszcz et al., 2009; De Halleux et al., 2006; Meno et al., 2005). The structure of natural Der f 1, which has 81% sequence identity to Der p 1 (Figure 3), was determined as well (Chruszcz et al., 2009). In addition, structures of natural Der f 1 and natural Der p 1 in complex with the Fab fragment of a cross-reactive monoclonal antibody (mAb) 4C1 were also elucidated (Chruszcz et al., 2012b).

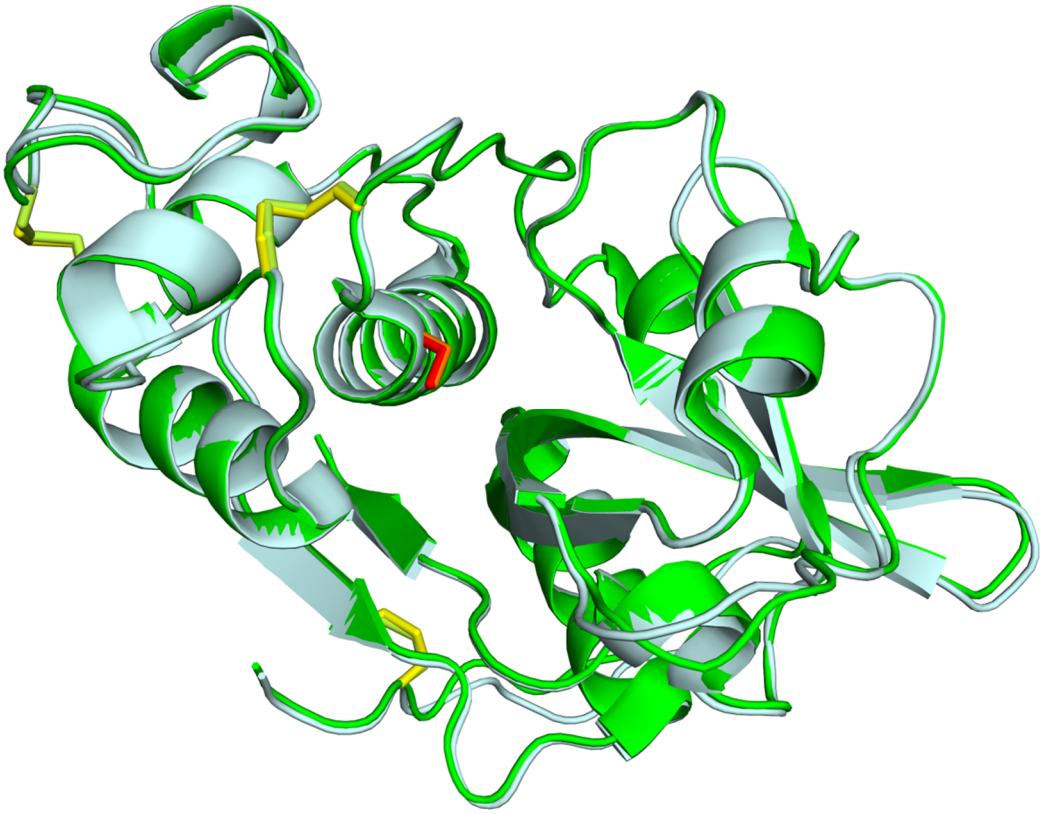

Figure 3 Der p 1 (green) and Der f 1 (cyan) superimposition.

Disulfide bridges are yellow (Der f 1 - light green). Cys34 in the active site is colored red (Cys35 from Der f 1 is orange).

Herein, I present the crystal structures of Der p 1, isolated from its natural source, complexed with the Fab fragment of 5H8 (Der p 1-5H8), Der p 1 complexed with the Fab fragment of 10B9 (Der p 1-10B9), and the Fab fragment of mAb 10B9 alone. Both 10B9 and 5H8 are species specific, whereas the 4C1 antibody is cross-reactive between Der p 1 from *D. pteronyssinus* and Der f 1 from *D. farinae*. This enabled the Der p 1 epitopes for mAbs 10B9, 5H8 and 4C1 to be compared with the corresponding surface on Der f 1 (Chruszcz et al., 2012b, 2009). It was discovered that the Der p 1 epitopes, which bind 4C1 and 10B9 antibodies, partially overlap and these two antibodies compete for the same binding site (Chapman et al., 1987). The 5H8 antibody, however binds to the epitope located on a different side of Der p 1, and does not compete with 4C1 or 10B9 for binding (Chapman et al., 1987). The binding interfaces of Der p 1 with mAbs 4C1, 5H8 and 10B9 with the binding interfaces of all currently

known structures of complexes of proteins or peptides with monoclonal antibodies were also compared.

4.7. German cockroach

One of the first reports associating cockroaches with allergy comes from 1964 (Bernton and Brown, 1964). Since then a variety of cockroach allergens have been reported. Most of them come from two cockroach species belonging to two cockroach families – *Blatella germanica* (German cockroach; *Blatellidae* family) and *Periplaneta americana* (American cockroach; *Blattidae* family)(Arruda et al., 1997; L. K. Arruda et al., 1995; L. Karla Arruda et al., 1995; Helm et al., 1996; Santos et al., 1999). The German cockroach is most commonly found in Europe and the US, while the American cockroach is more characteristic to South America and some regions in Asia.

The cockroach particles are among the allergens responsible for vast majority of asthma (Call et al., 1992; Cohn et al., 2006; Eggleston et al., 1998; Gelber et al., 1993; Matsui et al., 2003). The inner city households occupants are the most often subject to exposure and sensitization to cockroaches (Call et al., 1992; Cohn et al., 2006; Eggleston et al., 1998), but also the inhabitants of suburban areas (Matsui et al., 2003) and even the hospital patients (Gelber et al., 1993).

The allergen nomenclature maintained by WHO/IUIS (Radauer et al., 2014) lists nine allergens originating in German cockroach and seven allergens originating in American cockroach. Four allergen structures from *Blattella germanica* - Bla g 1 (Mueller et al., 2015), Bla g 2 (Gustchina et al., 2005), Bla g 4 (Yih et al., 2009), Bla g 5 (Mueller et al., 2013) and one allergen structure from *Periplaneta americana* - Per a 4 (Yih et al., 2009) have been characterized as of 2015. The Bla g 2 allergen is an aspartic protease (Gustchina et al., 2005; Pomés et al., 2002) with an unusual hydrogen bonded network. The Bla g 4 allergen belongs to the lipocalin protein family (L. Karla Arruda et al., 1995), but the ligand it binds, as well as its

function, is unknown. Lipocalins are a family of proteins that bind small extra-cellular ligands and the overall structure is usually highly conserved despite low sequence similarity among the family (Flower, 1996; Flower et al., 2000). The prevalence of serum IgE antibody to recombinant Bla g 4 in 73 cockroach allergic patients with asthma ranged from 40% (antigen binding radioimmunoassay) to 60% (plaque immunoassay) (L. K. Arruda et al., 1995).

Bla g 4 is produced in the male reproductive system by both utricles and the conglobate gland and is passed within a spermatophore from the male to the female during copulation and immediately after mating its level dramatically decreases (Fan et al., 2005; Gore and Schal, 2007). The fate of Bla g 4 protein inside the female is unknown, but it could serve as a structural component of the spermatophore or as part of the seminal fluid for sperm protection, storage and activation, sperm competition or as a modulator of female reproductive physiology (fecundity, ovulation, oviposition) and behavior or other functions (Gillott, 2003).

4.8. Mold

The genus *Alternaria* contains molds occurring naturally on outdoor vegetation. The most studied and representative species of *Alternaria* is *Alternaria alternata*, a member of the imperfect fungi and one of the most important allergenic fungi (Bush and Portnoy, 2001; Durham, 1937). *Alternaria alternata* is listed by WHO/IUIS (Radauer et al., 2014) nomenclature as containing twelve allergens, and *Alternaria* is one of the best-studied molds. Sensitization to *Alternaria* and its correlation to allergic asthma, rhinitis or airway hyperresponsiveness have been described in the United States, Europe, Canada and Australia (Chan-Yeung et al., 2010; Downs et al., 2001; Fernández Rodríguez et al., 2011; Marks and Bush, 2007; Perzanowski et al., 1998). The effects of *Alternaria* sensitivity were described primarily in reports across the United States – from California, through Arizona to Mid-West (Bush and Portnoy, 2001; Delfino et al., 1997; O'Hollaren et al., 1991; Stern et al., 2008) and around 12.9% among US citizens have positive skin prick test response to *Alternaria* species

(Arbes et al., 2005a). Sensitization and exposure to *Alternaria* that has been associated with asthma in various studies (Arbes et al., 2005b; O'Hollaren et al., 1991), and in the recent reports it has been noted that exposure to outdoor than rather indoor *Alternaria* spores plays a role in the sensitization (Pongracic et al., 2010). The spores of *Alternaria* may be the cause of thunderstorm-related outbreak of asthma, caused by increased exposure to the allergen (Nasser and Pulimood, 2009; Pulimood et al., 2007)

Alt a 1 is a major allergen produced by *Alternaria alternata* is responsible for IgE antibody responses in around 80% of *Alternaria*-allergic patients (Bush and Portnoy, 2001; Lopez and Salvaggio, 1985). The closest homolog of Alt a 1 is an isoallergen of 90% sequence identity produced by *Alternaria brassicicola,* that is responsible for fungal pathogenesis (brassica dark leaf spot) in *Brassicaceae* crops (Cramer and Lawrence, 2003; Hong et al., 2005; Nowicki et al., 2012). The small number of known Alt a 1 homologs and little structural data does not reflect its importance as an allergen. An immunotherapy with standardized *Alternaria alternata* extract significantly reduced a combined symptom and medication score in trial involving children with allergic rhinoconjuctivitis and asthma (Kuna et al., 2011).

Alt a 1 forms a disulfide bond linked dimer of 30 kDa mass and pI of 4.2 as its natural form. Under reducing conditions of SDS-PAGE it migrates as two bands of 16.4 and 15.3 kDa. Alt a 1 has been successfully cloned and the expressed recombinant allergen has been used to measure IgE and IgG antibody responses in *Alternaria*-sensitive patients (Achatz et al., 1995; De Vouge et al., 1996; Deards and Montague, 1991; Unger et al., 1999; Vailes et al., 2001).

5. Aims of the work

Inhaled allergens are exacerbating factors in asthma. The cooperation between Wladek Minor's Laboratory at the University of Virginia with Indoor Biotechnologies created a unique opportunity to thoroughly analyze various interactions of the selected allergens with antibodies, ligands and themselves to create the background rendering some aspects of further research possible and providing information to help better understand inhaled allergens. The whole project consists of three subprojects focusing on allergens having their origin in house dust mites (Der p 1), cockroach (Bla g 4) and fungus (Alt a 1). All of the subprojects involved many people; therefore, parts of the subprojects not performed by me will be appropriately marked.

In the subproject where the Der p 1 allergen was analyzed, the main goal was to find out the molecular mechanisms of specific antibody binding in complexes of Der p 1 with three monoclonal antibodies (4C1, 10B9 and 5H8) and compare these complexes with the homologous Der f 1 allergen and its complex with the cross-reactive 4C1 monoclonal antibody. The properties of the interface between complexes of these allergens with antibodies were compared with the interfaces of all available complexes of monoclonal antibodies with proteins or peptides found in the Protein Data Bank. The results can be used in the development of immunotherapy by design of less potent allergens.

The project involving the Bla g 4 allergen had the research focused at identification of the ligand by solving its structure with a bound ligand. The exploration of structural conservancy of homologous allergens and finding its implications on ligands that bind to them was second part of this project and greatly increased our understanding about this allergen.

The goal of project involving Alt a 1 allergen was to obtain a high resolution structure and to investigate structural relationships among its homologs.

6. Materials and Methods

6.1. Der p 1

6.1.1. Der p 1 – Data Collection and Structure Determination

Der p 1 was purified from *D. pteronyssinus* mite culture as described previously for Der f 1 (Chruszcz et al., 2012b). Data were collected at the Structural Biology Center Collaborative Access Team 19-BM and 19-ID beamlines (Rosenbaum et al., 2006), and at the 21-ID-D beamline of the Life Sciences Collaborative Access Team respectively at the Advanced Photon Source, Argonne National Laboratory. Data reduction and structure determination by molecular replacement were performed using HKL-3000 (Minor et al., 2006; Otwinowski and Minor, 1997) which incorporates MOLREP (Vagin and Teplyakov, 1997) and some of the programs included in the CCP4 package (Collaborative Computational Project, 1994). The structure of the Fab fragment of mAb 10B9 was determined using the Fab fragment of mAb 4C1 (PDB code: 3RVT) as a starting model. Structures of Der p 1 in complex with the Fab fragments of mAb 5H8 and 10B9 were determined utilizing the Der p 1 allergen complexed with the Fab fragment of mAb 4C1 (PDB code: 3RVW) (Chruszcz et al., 2012b) as the start model. The sequences of the mAbs were obtained by sequencing reverse-transcribed mRNA isolated from hybridomas producing the mAb (Chruszcz et al., 2009). The models were refined with COOT (Emsley and Cowtan, 2004) and REFMAC5 (Murshudov et al., 2011). TLS groups used in the refinement of Der p 1 with the Fab fragment of mAb 10B9 were generated using the TLSMD web server (Painter and Merritt, 2006). All three structures were validated with MOLPROBITY (Davis et al., 2007) and ADIT (Yang et al., 2004).

Previous results showed that natural Der p 1 binds a calcium ion (Chruszcz et al., 2012b); therefore, the metal bound by Der p 1-10B9 and Der p 1-5H8 complexes was determined to be calcium. Structures and structure factors were deposited to the PDB (Berman

et al., 2000) with accession code 4POZ, 4PP1 and 4PP2 for the Fab fragment of mAb 10B9, the Der p 1-5H8 complex and the Der p 1-10B9 complex, respectively. Data collection and refinement statistics can be found in Table I.

Structure (PDB code)	10B9 (4POZ)	Der p 1-10B9 (4PP2)	Der p 1-5H8 (4PP1)
Data collection			
Space group	$P2_12_12$	$P2_1$	$P2_1$
Cell dimensions: *a*,*b*,*c* (Å)	67.0, 121.8, 55.0	50.7, 74.9, 184.1	47.7, 73.3, 200.3
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 97.4, 90.0	90.0, 91.1, 90.0
Resolution (Å)	1.75 (1.75 - 1.78)	2.74 (2.74 - 2.79)	3.00 (3.00 - 3.05)
R_{sym}	0.069 (0.633)	0.139 (0.684)	0.165 (0.653)
I/σI	27.0 (2.2)	14.5 (2.3)	10.0 (2.3)
Completeness (%)	99.0 (96.9)	99.2 (97.9)	99.9 (100.0)
Redundancy	5.4 (4.7)	4.3 (4.2)	4.1 (4.2)
Refinement			
Resolution (Å)	1.75	2.74	3.00
No. Reflections	43428	34176	26581
R_{work}/R_{free} (%)	17.2 / 20.7	20.0 / 26.1	21.9 / 26.5
No. Atoms			
Protein	3364	9866	9748
Ligand/ion		30	48
Water	449	22	74
B-factors (Å^2)			
Protein	25	47	50
Ligand/ion		30	78
Water	32	34	27
r.m.s. deviations			
Bond lengths (Å)	0.018	0.007	0.009
Bond angles (°)	1.8	1.2	1.2

Table I Data collection and refinement statistics of 10B9, Der p 1-10B9 and Der p 1-5H8. Numbers in parentheses refer to the highest resolution shell.

6.1.2. Der p 1 – Structure analysis of Der p 1 with 10B9 and Der p 1 with 5H8

Sequences of proteins whose structures are similar to the heavy and light chains of the Fab fragments of mAb 10B9, 5H8 and 4C1, were obtained by a BLAST (Altschul et al., 1997,

1990; Camacho et al., 2009) search with an expectation value (e-value) of $1e^{-10}$ against the pdbaa BLAST sequence database (NCBI Resource Coordinators, 2013). The resulting list of the structures of the antibodies were downloaded, and STRIDE (Frishman and Argos, 1995) was used to determine their secondary structure. The antibodies containing an α-helix in the CDR H3 region were used to prepare the dataset for the analysis of the Fab fragment of mAb 10B9 and superposed in PYMOL (Schrodinger LLC, 2010). The CDR regions of these structures were compared and analyzed.

The list of sequences and structures of Fab fragments of monoclonal antibodies complexed with various macromolecules were obtained by running a BLAST (Altschul et al., 1997, 1990; Camacho et al., 2009) search against the pdbaa database (NCBI Resource Coordinators, 2013) using the sequences of the Fab fragments of mAbs 5H8, 10B9 and 4C1 as queries. The result of these searches was combined and redundancy was removed. The list of obtained structures was used for the preparation of two datasets – the first one, which contains structures solved at a resolution of 3.0Å or better and the second one, which contains structures solved at a resolution of 3.5Å or better. Preliminary analysis of the epitope and paratope content and hydrogen bonds between them were virtually identical for both datasets. All of the analyses were performed simultaneously on both datasets, however, the results from the first dataset, which contains 314 structures of Fab fragments in complex with protein or peptides solved at 3.0Å resolution or better, were chosen for this study and are referred to as the dataset, unless otherwise noted. PISA (Krissinel and Henrick, 2007) was used to estimate interfaces and hydrogen bonds formed by structures in the dataset. Amino acid residues were considered to be part of the interface if the calculated buried surface area for a particular residue was greater than 10 $Å^2$ or was involved in forming hydrogen bonds.

The amino acid residues involved in the formation of hydrogen bonds between proteins were identified with PISA (Krissinel and Henrick, 2007) and a cutoff distance of 3.3 Å for

hydrogen bonds was chosen. The salt bridges between Der p 1 and the Fabs of 10B9 and 5H8 were identified by PISA (Gibbs free energy based) and VMD (distance based) (Humphrey et al., 1996). A salt bridge was considered to be formed if the distance between any of the oxygen atoms of acidic residues (carboxylate ions; acceptors) and the nitrogen atoms of basic residues (ammonium or guanidinum ions; donors) were within the cut-off distance of 4.0 Å (Barlow and Thornton, 1983; Kumar and Nussinov, 2002a, 2002b)

The VMD (Humphrey et al., 1996) program was used for structural conservation analysis between Der p 1 and Der f 1. Der p 1 (3F5V), Der f 1 (3D6S), Der f 1-4C1 (3RVV), Der p 1-4C1 (3RVW), Der p 1-10B9 (4PP2), and Der p 1-5H8 (4PP1) were the structures used for comparison.

RING (Martin et al., 2011) was used as a tool for analyzing residue interaction networks (RINs) in order to describe the protein three-dimensional structure and the nature of interactions (e.g. hydrogen bonds, van der Waals contacts or π-cation interactions and π-π stacking interactions).

6.1.3. Der p 1 – Other techniques

Residues on the protein's surface were identified with PYMOL. The pairwise protein sequence identity and similarity was calculated using the EMBOSS (Rice et al., 2000) package. The overall frequency of amino acid residues are shown per 100 surface-exposed amino acids found in all protein chains of Fab-protein or Fab-peptide complexes reported in the PDB database (as of 2013). Amino acid residue frequencies calculated this way were used to estimate the expected amino acid content of the epitopes and paratopes in order to compare it with the observed amino acid content of the interfaces.

A program was written in Python with the use of numpy (Oliphant, 2007), matplotlib (Hunter, 2007) and mmlib (Painter and Merritt, 2004) libraries to generate three types of plots:

(I) bar plots for comparison between observed and expected distribution of the amount of amino acids among epitopes and paratopes, (II) rainbow colored bar plots showing the relative amount of surface area contributed to the antigen-antibody interface by each type of the amino acid, and (III) array plots for displaying the number of hydrogen bonds between epitopes and paratopes as a greyscale. The values of the expected number of amino acid residues in the interfaces for the bar plot (I) were derived from the frequencies of amino acid residues occurrences in all protein sequences found in BLAST nr database. The dataset obtained in the first step of the structural analysis was used for plots (II) and (III). The area contributed by a particular amino acid residue to the interface in the rainbow colored plot (II) was calculated with PISA. The number of hydrogen bonds in antigen-antibody complexes shown on plot (III) was also calculated with PISA. Some of the code was developed and used in a previous work (Chruszcz et al., 2012b).

Molecular Operating Environment (MOE) (Chemical Computing Group Inc., 2013) was used to verify the potential hydrogen bonds involved in forming the interfaces between Der p 1 and both Fab fragments of mAb 5H8 and 10B9. Swiss-Pdb Viewer (Guex and Peitsch, 1997) was used to display φ and ψ angles of the CDR H3 loop in the Fab fragment of mAb 10B9.

6.2. Bla g 4

6.2.1. Bla g 4 – Structure Determination

Protein production, crystallization, and data collection have been described previously (Yih et al., 2009). The reinterpretation of diffraction data using a new methodology included in the HKL-3000 package (Minor et al., 2006) is presented herein. During these studies, data obtained from both Se-Met labeled and native Bla g 4 crystals (Yih et al., 2009) were reinvestigated . The Se-Met and native structures were determined using the multi-wavelength anomalous diffraction (MAD) technique and Molecular Replacement (MR), respectively, by

HKL-3000 coupled with SHELXD/C/E (Sheldrick, 2008), MLPHARE (Otwinowski, 1991), DM (Cowtan and Main, 1993), ARP/wARP (Perrakis et al., 1999), MOLREP (Vagin and Teplyakov, 1997), SOLVE/RESOLVE (Terwilliger, 2004), and selected programs from the CCP4 package (Collaborative Computational Project, 1994). Both the Se-Met derivative and the native crystal structures were re-examined in the $P4_12_12$ space group. The models were later refined and validated as described in the section 6.1.1. Structures and structure factors were deposited to the PDB (Berman et al., 2000) with accession code 4N7D and 4N7C for Se-Met derivative and native Bla g 4, respectively. Refinement statistics are summarized in Table II.

Structure (PDB code)	Se-Met (4N7D)	Native (4N7C)
Data collection		
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions: *a,b,c* (Å)	60.17, 60.17, 124.93	60.35, 60.35, 125.41
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	2.1 (2.10 – 2.18)	1.75 (1.75 - 1.81)
R_{sym}	0.069 (0.38)	0.064 (0.5)
I/σI	23.71 (4.0)	43.2 (4.75)
Completeness (%)	100.0 (100.0)	99.9 (100.0)
Redundancy	5.1 (4.9)	15.0 (14.9)
Refinement		
Resolution (Å)	2.1	1.75
No. Reflections	13356	22948
R_{work}/R_{free} (%)	17.6 / 21.7	18.6 / 21.3
No. Atoms		
Protein	1361	1370
Ligand/ion	25	35
Water	117	127
r.m.s. deviations		
Bond lengths (Å)	0.018	0.024
Bond angles (°)	2.0	2.23

Table II Data collection and refinement statistics of Bla g 4.

Numbers in parentheses refer to the highest resolution shell.

6.2.2. Bla g 4 – Sequence analysis

The dataset used for analysis was prepared by running PSI-BLAST (Altschul et al., 1997) against the UniProt database (Uniprot version: 2013_2) (Consortium and The Uniprot Consortium, 2012) using the sequences of Bla g 4 and Per a 4 (a close homologous protein to Bla g 4 from *P. americana*) as the queries. The searches were executed in two steps. The first step allowed for preparation of position-specific scoring matrix (PSSM) profiles that were used as queries in the second step. The PSSM profiles were prepared by performing two independent searches with Bla g 4 and Per a 4 as queries with an expectation value (e-value) of 10^{-5} for three cycles. In the second step, the PSSM profiles prepared in the previous step were used to perform searches against the database used in the first search with an e-value of 10^{-3} until convergence

was achieved. The protein structures, homologous to Bla g 4 and Per a 4 were obtained by running PSI-BLAST searches with an expectation value of 10-3 against the pdbaa database (NCBI Resource Coordinators, 2013) from NCBI. The obtained sequences of the structures were then added to the sequence dataset. Identification of a particular sequence membership in specific protein family of PFAM (Punta et al., 2012) database was achieved by preparing a BLAST database from the PFAM protein families PF00061 (Lipocalin), PF08212 (Lipocalin-like), PG03973 (Triabin). PFAM database version 26.0 was used. Two allergen families (AF015—Lipocalin, AF119—Triabin family) included in AllFam (Radauer et al., 2008) allergen families database were also added to the dataset. AllFam database version 2011–09-12 was used. Sequences obtained from searches against UniProt were subjected to CD-HIT (Fu et al., 2012), where sequences with 80% identity or higher were removed. The prepared dataset was merged with results from searches against PFAM and pdbaa (Sayers et al., 2012) databases, ultimately returning 1561 non-redundant protein sequences listed in Lipocalin, Lipocalin-like and Triabin PFAM protein families and sequences obtained from other sources. CLANS (Frickey and Lupas, 2004) was used to prepare a 2D visualization of sequences pairwise similarity by using the Fruchterman–Reingold graph layout algorithm. Clustering was performed with an e-value of 10^{-6} until convergence was achieved. Allergens found in identified clusters were aligned with MAFFT (Katoh and Standley, 2013) with the L-INS-I option and later adjusted manually in Jalview (Waterhouse et al., 2009) according to the 2D projection of the structural alignment of representative allergens found in the sequence dataset—Bla g 4, Per a 4, Can f 2 and Equ c 1 (PDB codes: 3EBK, 3EBW, 3L4R and 1EW3) prepared in Swiss-PdbViewer (Guex and Peitsch, 1997).

6.2.3. Bla g 4 – Evolutionary analysis

Sequences from AllFam families AF015 and AF119 were mapped on the dataset used for clustering, then were aligned by MAFFT (Katoh and Standley, 2013) with the L-INS-I

option to increase accuracy. The obtained sequence alignment was subjected to MEGA5 (Tamura et al., 2011). Phylogeny reconstruction was performed using the Maximum-Likelihood method with the WAG (Whelan and Goldman, 2001) amino acid substitution model with gamma-distributed rates among patterns. The bootstrap method with 1000 replications was used to test branch probabilities.

6.2.4. Bla g 4 – Structure analysis

The dataset used for structural conservancy of Bla g 4 was obtained after several steps. The first step was performing a PSI-BLAST search against the pdbaa BLAST database (as of January 2013). The second step was the manual selection of the "best" structures representing given blast database record. The "best" structure was defined as the one with the best resolution/R-factors values out of a group representing given blast database record. The dataset was formed from 39 structures with the following PDB codes: 1AVG, 1DZK, 1EPA, 1EW3, 1EXS, 1GKA, 1GM6, 1I06, 1N0S, 1PM1, 1QWD, 1S2P, 1SXX, 1XKI, 1YUP, 1Z24, 2A2U, 2CZT, 2HLV, 2HYS, 2HZQ, 2L5P, 2L9C, 2RA6, 2RD7, 2WWP, 2XST, 3BX7, 3CQR, 3DSZ, 3DTQ, 3EBW, 3FMZ, 3L4R, 3QKG, 3S26, 3SAO, 4GE1, and 4N7C. The structures were then retrieved from the PDB and chains homologous to Bla g 4 were used to prepare a structural alignment in the VMD (Humphrey et al., 1996) program. Structural conservancy between protein structures was measured by using the QH algorithm (O'Donoghue and Luthey-Schulten, 2003). Calculated Q_H values for given residues (Q_{RES}) were then applied to the Bla g 4 structure, instead of B-factor values, and displayed in Pymol (Schrodinger LLC, 2010).

6.2.5. Bla g 4 – Other techniques

Solvent-accessible surface areas were calculated with PDBePISA (Krissinel and Henrick, 2007). Figures were prepared with Pymol (Schrodinger LLC, 2010). Modeling of the binding of two enantiomers of octopamine was performed in Pymol by superposing them on

the tyramine molecule with the use of Least Squares (LSQ) algorithm (Kabsch, 1978, 1976) implemented by Jason Vertrees in an extension to Pymol. DALI (Holm and Rosenström, 2010) and FATCAT (Ye and Godzik, 2004) were used for search to identify similar proteins.

6.3. Alt a 1

6.3.1. Alt a 1 – Data Collection and Structure Determination

Protein production, crystallization, and data collection have been described previously in supplementary materials (Chruszcz et al., 2012a). Data collection was performed at the 19-BM Beamline of the Structural Biology Center (Rosenbaum et al., 2006) and the 21-ID-G Beamline of the Life Sciences Collaborative Access Team at the Advanced Photon Source (Argonne, Il). Data were collected at 100K and processed with HKL-2000 (Otwinowski and Minor, 1997). The initial model was obtained from a low-resolution dataset (2.8 Å) collected at 19-BM from a crystal soaked in a solution containing the Ta6Br122+ cluster. The single-wavelength anomalous diffraction technique was used for structure determination. Calculations were performed with HKL-3000 (Minor et al., 2006; Otwinowski and Minor, 1997), which integrates SHELXC/D/E (Sheldrick, 2008), MLPHARE (Otwinowski, 1991), DM (Cowtan and Main, 1993), PARROT (Zhang et al., 1997), RESOLVE (Terwilliger, 2004), ARP/wARP (Perrakis et al., 1999), and selected programs from the CCP4 package (Collaborative Computational Project, 1994). The partial model was obtained by using a combination of manual building and building with RESOLVE. This model was used as a starting model for building with ARP/wARP. For the ARP/wARP calculation, a higher-resolution native dataset (1.9 Å) collected at 21-ID-G was used. The model was later updated with COOT (Emsley and Cowtan, 2004) and refined and validated as described in the section paragraph 6.1.1. The coordinates and structure factor for Alt a 1 were deposited in the PDB (Berman et al., 2000) with accession code 3V0R. Statistics from data processing and

structure determination are reported in Table III.

Structure (PDB code)	3V0R
Data collection	
Space group	$I4_122$
Cell dimensions: *a*,*b*,*c* (Å)	70.0, 70.0, 179.3
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	1.90 (1.90 – 1.93)
R_{sym}	0.068 (0.777)
I/σI	59.4 (4.2)
Completeness (%)	99.9 (100.0)
Redundancy	13.8 (13.9)
Refinement	
Resolution (Å)	1.90
No. Reflections	18063
R_{work}/R_{free} (%)	16.7 / 19.0
No. Atoms	
Protein	1014
Ligand/ion	61
Water	161
B-factors (Å^2)	
Protein	42.7
Ligand/ion	57.4
Water	54.0
r.m.s. deviations	
Bond lengths (Å)	0.020
Bond angles (°)	1.9

Table III Data collection and refinement statistics of Alt a 1.
Numbers in parentheses refer to the highest resolution shell.

6.3.2. Alt a 1 – Sequence analysis

The dataset used in analysis was prepared from the sequences obtained by running PSI-BLAST (Altschul et al., 1997) against non-redundant (nr) BLAST database (Sayers et al., 2012) with Alt a 1 sequence (gi number: 14423645) as a query. Searches were performed with expectation value (e-value) of 0.001 until convergence was achieved. All the sequences retrieved in the first step were also used for reciprocal PSI-BLAST searches to identify distant homologs. Sequences were retrieved from GenBank. Sequences annotated as incomplete were

removed. The final sequence dataset was aligned using MAFFT (Katoh and Standley, 2013) with L-INS-I option and further manually refined in Jalview (Waterhouse et al., 2009) and Bioedit (Hall, 1999) with regard to secondary structure obtained from STRIDE (Frishman and Argos, 1995). CD-HIT (Fu et al., 2012) was used to reduce the size of the dataset by removing the sequences of 80 percent of higher pairwise identity.

6.3.3. Alt a 1 – Evolutionary analysis

Two datasets have been prepared. The first dataset was prepared from the results of a PSI-BLAST (Altschul et al., 1997) search with Alt a 1 allergen sequence (gi number: 14423645) against nr BLAST database (NCBI Resource Coordinators, 2013) and then reducing the number of sequences with CD-HIT (Fu et al., 2012) by removing the sequences of 80% or higher identity. The MEGA package (Tamura et al., 2011) was used for phylogenetic analysis. The tree was computed by using the Maximum Likelihood statistical method with the WAG (Whelan and Goldman, 2001) amino acid substitution model. The number of discrete gamma categories with Invariant sites (G+I) was set to 5. Gaps and missing data were treated as complete deletion. The Nearest-Neighbor-Interchange was used as a maximum likelihood heuristic method. The initial tree was prepared automatically. Phylogeny was tested with the bootstrap method by using 1000 replications. The multiple sequence alignment programs – CLUSTAL W (Thompson et al., 1994) and CLUSTAL X (Larkin et al., 2007) - were used to prepare initial version of the alignment that was used to analyze cysteine conservation within Alt a 1 homologs. The overall conservancy was mapped on the Alt a 1 structure with Consurf (Ashkenazy et al., 2010; Celniker et al., 2013).

6.3.4. Alt a 1 - Other computational methods

The STRIDE (Frishman and Argos, 1995) was used to determine the secondary structure of Alt a 1. The search for similar structures was performed with DALI (Holm and Rosenström,

2010) and FATCAT (Ye and Godzik, 2004) against the structures deposited in PDB (Berman et al., 2000). PISA (Krissinel and Henrick, 2007) was used for analysis of the oligomeric assembly and calculation of the dimer's interface area. Figures were prepared with PYMOL (Schrodinger LLC, 2010). Electrostatic surface was calculated in APBS (Baker et al., 2001) with model prepared by PDB2PQR (Dolinsky et al., 2007, 2004). The identity and similarity values were calculated with lalign program from the FASTA package (Pearson and Lipman, 1988).

7. Results

7.1. Der p 1

Der p 1 with the Fab fragment of mAb 5H8 and Der p 1 with the Fab fragment of mAb 10B9 were purified, crystallized, and their structures were determined. Space groups were identified as $P2_12_12$ for the Fab fragment of mAb 10B9 and $P2_1$ for the Der p 1-5H8 and Der p 1-10B9 complexes. Previously, the crystal structure of Der p 1 complexed with the Fab fragment of mAb 4C1 (Chruszcz et al., 2012b) was elucidated. Figure 4 shows the relative position of the Fab fragments of 4C1, 10B9 and 5H8 antibodies in complex with Der p 1. More than 70% of the 10B9 binding epitope overlaps with the epitope that binds the 4C1 antibody; the epitope binding 5H8 antibody is located on a different part of Der p 1.

The electron density found in close proximity to Asn52 in each of the Der p 1 chains was interpreted as *N*-acetylglucosamine in the structures of both complexes, additionally it was discovered that the same, or similar, residue binds *N*-acetylglucosamine in the structures of Group 1 dust mite allergens previously solved, where in case of Der f 1 N-acetylglucosamine was bound by Asn53 (Chruszcz et al., 2012b, 2009). An exception was found in a structure where Asn52 was replaced by Gln52 (De Halleux et al., 2006). The *N*-acetylglucosamine

binding site is far from the 4C1, 10B9 or 5H8 epitopes and does not interfere with the binding of these antibodies.

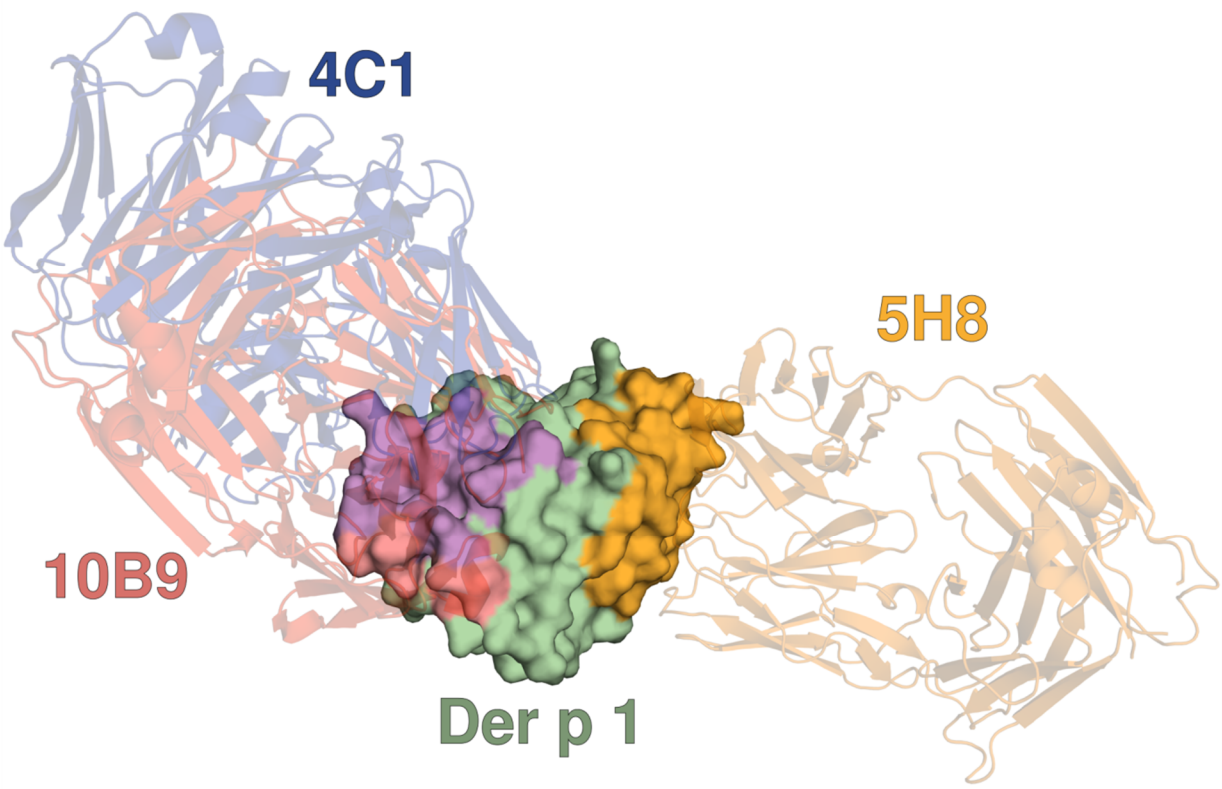

Figure 4 The relative position of Der p 1, Fab of 5H8, 4C1 and 10B9.

The Fab fragments of antibodies are colored as follows 4C1 blue, 10B9 red and 5H8 orange. Der p 1 is represented as green surface with epitopes binding each antibody colored in similar color to the corresponding antibody. Overlapping area of 4C1 and 10B9 binding epitopes is purple.

Three well-ordered complementarity-determining regions (CDRs) per antibody chain are involved in forming the interface areas (Figure 5). The conformations of the CDRs in the Fab fragments of mAb 10B9 and 5H8 classified according to the rules described by North et al. (North et al., 2011) can be categorized into the same categories as the CDRs of mAb 4C1 (Chruszcz et al., 2012b); CDR L1 – L1-11-2; CDR L2 – L2-8-1; CDR L3 – L3-9-cis7-1; CDR H1 – H1-14-1; CDR H2 – H2-9-1.

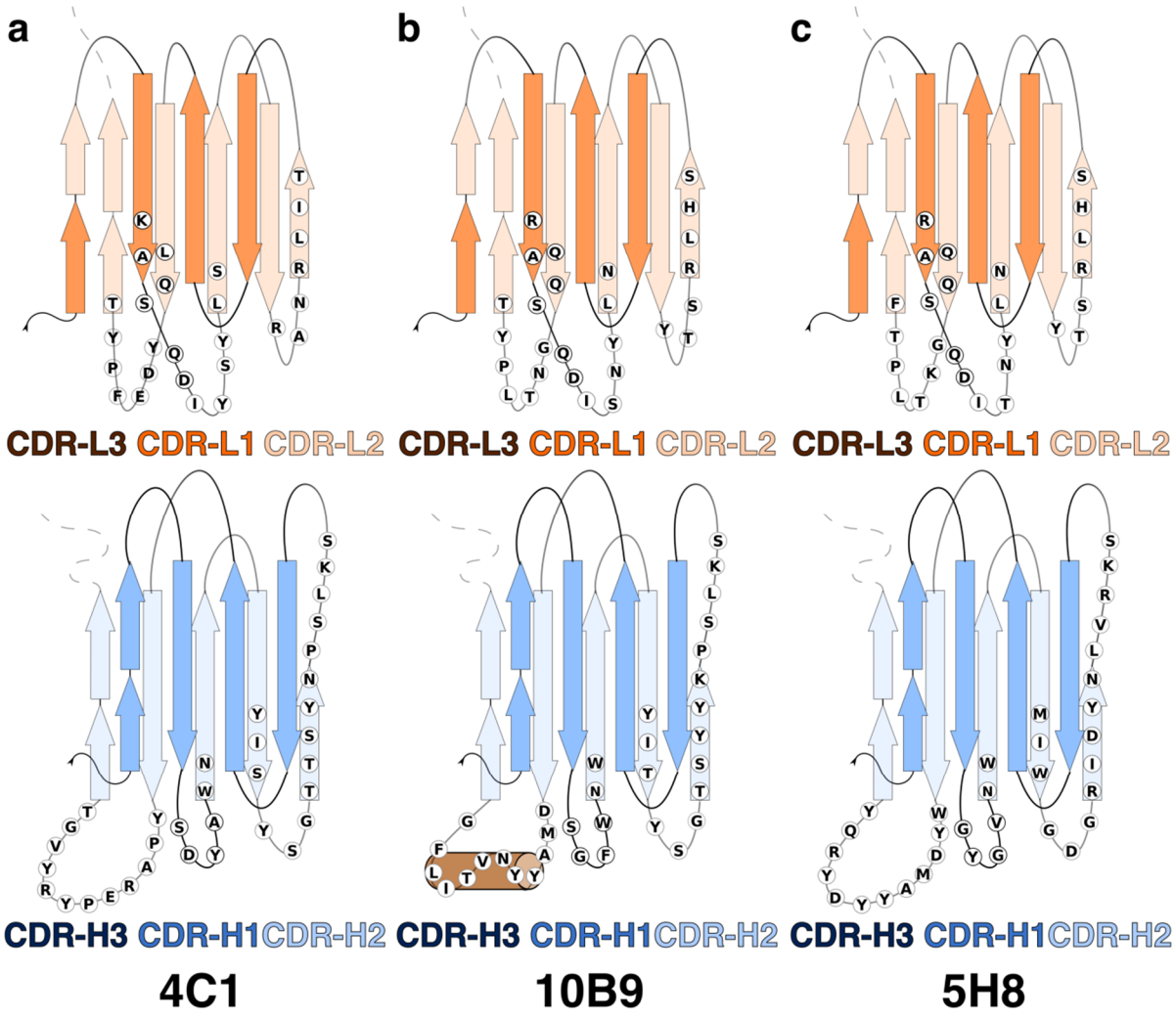

Figure 5 Comparison of CDRs of (a) 4C1, (b) 10B9 and (c) 5H8 antibodies.

Light chains are marked in the shades of orange, heavy chains are marked in shades of blue. Secondary structure is mapped as: arrow – beta-sheet, barrel – alpha-helix. The CDR loops are marked under each chain.

	Der f 1	Der p 1	ASA (Å^2)	BSA (Å^2)
		Chains A/B with V_H of 5H8		
	Tyr48	Tyr47	16.8/15.3	16.8/15.3
	Arg52	Arg51	142.3/139.5	65.2/66.9
*	Thr54	Gln53	98.2/97.0	79.7/77.2
	Ser55	Ser54	86.9/87.0	16.6/15.2
	Leu56	Leu55	18.2/18.4	13.4/13.5
	Gln85	Gln84	49.4/50.8	15.5/15.4
	Asn87	Asn86	79.2/77.7	11.0/11.0
*	Glu91	Gln90	46.6/47.4	39.5/38.7
*	Ser94	Tyr93	120.4/119.6	57.1/55.1
	Asn108	Asn107	152.1/153.2	73.4/75.6
*	Ser109	Ala108	24.5/24.0	24.4/24.1
	Gln110	Gln109	133.5/137.0	127.5/130.8
*	His111	Arg110	95.1/98.8	54.0/53.7
*	Tyr112	Phe111	37.6/39.4	37.6/39.4
	Gly113	Gly112	20.4/20.8	18.0/19.0
	Ser115	Ser114	83.3/78.0	10.4/10.8
		Chains A/B with V_L of 5H8		
	Tyr51	Tyr50	100.9/102.3	13.1/16.1
	Arg52	Arg51	142.3/139.5	77.1/72.1
*	Asn53	Gln53	98.2/97.0	10.5/7.3
	Ser115	Ser114	83.3/78.0	72.1/66.8
	Ile222	Ile221	77.7/78.8	21.3/21.7
*	Met223	Leu222	43.7/46.0	40.8/43.7

Table IV The 5H8-binding epitope on Der p 1 and the corresponding residues of Der f 1.

Different amino acids between Der p 1 and Der f 1 are indicated with an asterisk. The amino acids with buried surface area that were calculated as less than 10 Å^2 to the interface were omitted. ASA - Accessible Surface Area, BSA - Buried Surface Area. The values of ASA and BSA are presented for both the A and B chains, respectively.

5H8	ASA (Å^2)	BSA (Å^2)
V_H	Chains F/D	
Ser28	64.6/64.0	11.3/11.2
Thr30	83.3/84.0	25.1/25.0
Gly31	46.0/46.2	45.9/46.2
Tyr32	46.0/48.3	13.3/14.9
Gly33	10.9/10.8	10.0/10.4
Trp52	60.9/59.6	60.9/59.5
Gly35	17.3/17.9	11.2/11.2
Asp54	102.3/101.0	41.3/39.8
Arg56	135.8/133.9	58.7/58.8
Asp58	37.0/38.7	16.7/16.6
Gln100	118.6/115.8	77.8/77.7
Arg101	131.2/124.6	114.3/108.0
Tyr102	95.0/98.0	86.2/88.6
Asp103	93.9/92.7	59.4/59.9
Tyr105	89.5/90.9	10.2/9.0
V_L	Chains E/C	
Thr30	64.0/67.3	40.5/43.0
Tyr32	94.0/93.8	46.8/48.6
Tyr50	78.7/78.0	26.5/27.5
Gly91	35.6/35.3	12.1/11.9
Lys92	89.5/87.3	60.3/59.0
Thr93	63.1/70.8	19.3/23.4
Leu94	141.3/139.5	14.9/7.5

Table V The Der p 1-binding paratope on the 5H8 antibody.

The amino acids with buried surface area that were calculated as less than 10 Å^2 to the interface were omitted. ASA - Accessible Surface Area, BSA - Buried Surface Area.

7.1.1. The Der p 1 – 5H8 complex

Crystals of Fab 5H8 alone were not obtained. In the structure of the Der p 1-5H8 complex, the epitope binding 5H8 to Der p 1 is located far from the Der p 1's active site. The buried surface area at the interface of Der p 1 and 5H8 is around 910 $Å^2$. The interface areas contributed by the heavy and light chains are around 75% and 25%, respectively.

Der f 1	Der p 1	5H8		Distance (Å)
		V_L	V_H	Chains A/B
Arg52	Arg51		Asp103	3.0/2.9
		Gly91		2.7/3.0
Thr54	Gln53		Asp58	2.5/2.7
Ser55	Ser54		Arg56	2.9/3.0
Gln85	Gln84		Arg101	2.6/2.4
Ser94	Tyr93		Thr30	2.5/2.6
Ser109	Ala108		Gln100	3.2/3.1
Gln110	Gln109		Gly53	2.9/3.1
His111	Arg110		Gln100	2.7/2.7
			Tyr102	2.6/2.6
Gly113	Gly112		Asp103	3.2/3.1
Ile114	Ile113		Arg101	2.5/3.2
Ser115	Ser114	Tyr32		2.9/3.0
		Tyr50		3.1/3.0
Met223	Leu222	Lys92		2.4/2.8

Table VI The hydrogen bonds formed between Der p 1 and the Fab of mAb 5H8.

Der f 1 was added to highlight the differences in Der p 1 residues that form the 5H8 binding epitope. The hydrogen bonds were found by PISA and MOE. The residues forming hydrogen bonds are in rows.

The Der p 1 epitope binding 5H8 antibody is formed by nineteen amino acids, where sixteen residues from Der p 1 interact with fifteen residues from the heavy chain and six residues from Der p 1 interact with seven residues from the light chain of 5H8 antibody (Table IV and Table V).

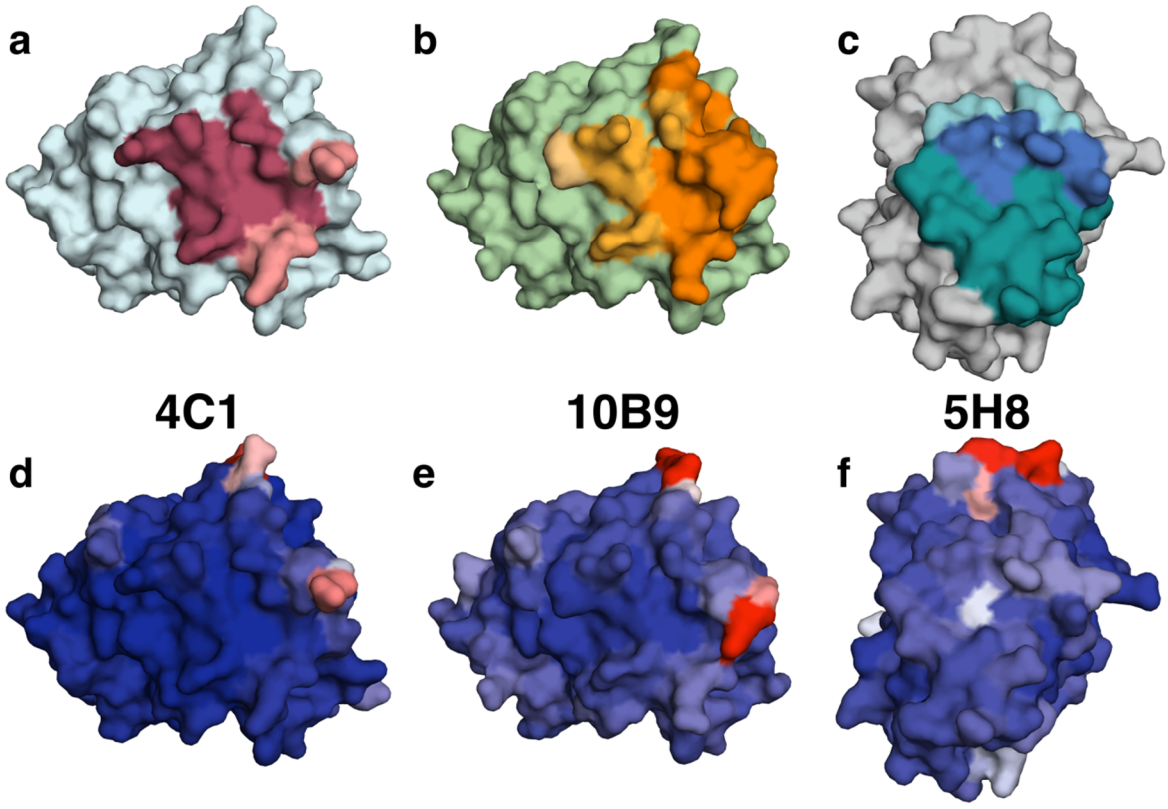

Figure 6 The surface of Der p 1 with various aspects of structure highlighted.

The surface of Der p 1 highlighting the epitope for the Fab fragments of mAb 4C1 (figures a and d), 10B9 (figures b and e), and 5H8 (figures c and f) are shown in surface representation. The colors of the epitopes are based on the chains of the antibodies that are being bound. The 4C1-binding epitope is colored as follows: pink (V_H), dark pink (V_H and V_L). The 10B9-binding epitope is colored as follows: orange (V_H), light orange (V_H and V_L) and yellow (V_L only). The 5H8-binding epitope is colored as follows: dark cyan (V_H), blue (V_H and V_L) and cyan (V_L only). Panels d, e and f represent the structural conservation between Der p 1 and Der f 1 calculated with VMD. The color scale starts from dark blue (high structural conservation), goes through white and ends in red (no structural conservation)

The amino acid residue composition of the epitope on the surface of Der p 1 (Figure 6a-f) where the Fab of mAb 5H8 is bound differs significantly from those epitopes where the Fab of mAb 4C1 and 10B9 are bound. Fifteen hydrogen bonds were identified between Der p 1 and 5H8 antibody. Residue Arg51 of Der p 1 forms hydrogen bonds with both the V_L and V_H chains

of mAb 5H8 (Gly91 and Asp103 respectively). Some of the residues, such as Ser114 with Tyr32 and Tyr50 and C-terminal Leu222 with Lys92, form hydrogen bonds only with the V_L chain, while residues Gln53, Ser54, Gln84, Tyr93, Ala108, Gln109, Arg110, Gly112, and Ile113 form hydrogen bonds only with the V_H chain of 5H8 antibody (Table VI). Figure 11 and Figure 12 demonstrate that hydrogen bond formation is quite specific, where a tyrosine residue from the antibody tends to form hydrogen bonds with an arginine residue from the antigen. This occurrence can be observed at the interfaces between Der p 1-4C1 and Der p 1-5H8, but not in Der p 1-10B9 complex (Figure 7 and previous work (Chruszcz et al., 2012b)). Arg51 forms cation-π interactions with Tyr32 of V_L and Tyr102 of V_H. The amino acids that make up mAb 5H8's paratope of the 5H8-Der p 1 interface are tyrosine, arginine, aspartate, glutamine, lysine and a relatively high content of glycine. Phe111 of Der p 1 forms π-π stacking interactions with Trp52 (edge-to-face orientation) and Tyr102 (face-to-edge orientation) of V_H as well as Tyr47 of Der p 1. The Fab fragment of mAb 5H8, which is similar to the Fab fragment of mAb 10B9, is also a species-specific antibody. It binds to a different fragment of Der p 1 and the epitope for this antibody does not overlap with the epitopes for the 10B9 and 4C1 antibodies.

.

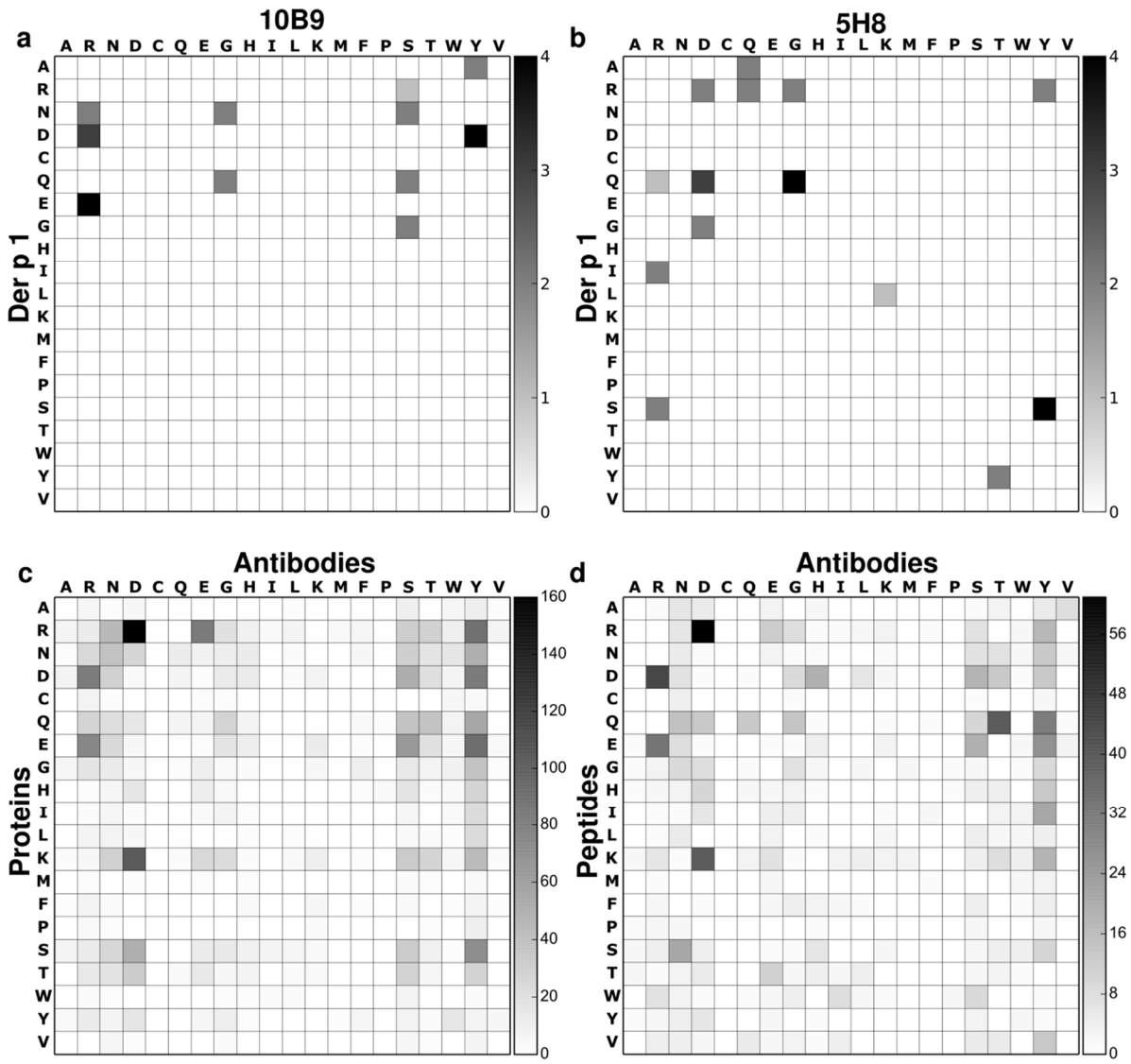

Figure 7 The hydrogen bonds formed between various complexes.

The hydrogen bonds formed between (a) Der p 1 and Fab fragment of mAb 10B9 (two copies from asymmetric unit), (b) Der p 1 and Fab fragment of mAb 5H8 (two copies from asymmetric unit), (c) all non-redundant structures of complexes of proteins with Fab fragments of monoclonal antibodies and (d) all non-redundant peptides with Fab fragments of monoclonal antibodies. The proteins and peptides complexes used for analyses were solved at resolution equal or higher than 3.0 Å. Number of hydrogen bonds between particular amino acids is shown in greyscale (on right) where the higher number of hydrogen bonds, the darker square is marked on figure. Antibodies are on X axis, antigens on Y axis of each plot.

7.1.2. The Der p 1 – 10B9 complex and 10B9 in uncomplexed form

The epitope for the Fab fragments of mAb 10B9 is also located far from the catalytic site of the enzyme (Cys34). The buried surface area at the interface of Der p 1 and 10B9 is 815

Å^2 and the interface areas contributed by the heavy chain and light chain are around 75% for the former and 25% for the latter (Table VII).

Der f 1	Der p 1	10B9		Distance (Å)
		V_L	V_H	Chains E/F
Ser13	Ala12		Tyr53	2.7/2.6
Glu14	Glu13		Arg98	2.9/3.0
			Arg98	2.8/2.7
Asp16	Asp15		Tyr107	2.8/2.7
Ser19	Gln18	Gly91		3.3/3.2
		Asn92		3.3/3.5
Arg20	Arg20	Ser30		2.9/NA
Ser180	Asn179		Ser31	2.6/2.7
			Gly32	3.0/2.8
Gln182	Gln181		Ser31	3.6/3.3
Gly183	Gly182		Ser31	2.8/2.8
Asp199	Asp198		Tyr106	2.7/2.8
		Arg53		2.7/3.0
		Arg53		3.0/3.2
Ser200	Asn199	Arg53		2.5/2.7

Table VII The hydrogen bonds formed between Der p 1 and the Fab of mAb 10B9.

Der f 1 was added to highlight the differences in Der p 1 residues that form the 10B9 binding epitope. The hydrogen bonds were estimated by PISA and MOE. The residues forming hydrogen bonds are shown in rows. Underlined amino acids also form the salt bridges.

The Der p 1 epitope that binds 10B9 is formed by seventeen amino acids, where fifteen residues from Der p 1 interact with eleven residues from the heavy chain and five residues from Der p 1 interact with five residues from the light chain of 10B9 (Table VIII and Table IX).

	Der f 1	Der p 1	ASA (Å^2)	BSA (Å^2)
		Chains F/E with V_H of 10B9		
*	Ser13	Ala12	94.7/90.2	86.2/79.3
	Glu14	Glu13	111.5/113.9	78.9/83.5
	Asp16	Asp15	29.5/29.8	24.3/26.1
	Arg18	Arg17	55.2/62.7	26.9/26.1
*	Ser19	Gln18	151.7/147.1	46.9/45.7
	Arg157	Arg156	199.1/197.0	14.9/12.2
*	Gly179	Ser178	17.1/16.6	13.6/11.8
*	Ser180	Asn179	82.5/83.3	80.2/79.2
*	Thr181	Ala180	40.4/45.4	40.4/45.1
	Gln182	Gln181	145.93/134.8	35.7/30.1
	Gly183	Gly182	77.5/79.2	18.5/19.9
	Asp185	Asp184	65.1/61.5	16.0/15.2
	Tyr186	Tyr185	31.6/34.9	30.7/34.2
	Asp199	Asp198	56.4/54.1	25.9/25.5
	Tyr204	Tyr203	30.0/32.5	29.3/30.8
		Chains F/E with V_L of 10B9		
	Arg18	Arg17	55.2/62.7	14.8/16.3
*	Ser19	Gln18	151.7/147.1	104.4/99.2
	Arg21	Arg20	153.8/166.1	41.2/43.5
	Asp199	Asp198	56.4/54.1	26.7/26.0
*	Ser200	Asn199	103.7/96.5	27.4/23.2

Table VIII The 10B9-binding epitope on Der p 1 and the corresponding residues of Der f 1.

The table presents the amino acids that form the 10B9 binding epitope on Der p 1 and the complimentary amino acids found on Der f 1. The differences are indicated with an asterisk. The amino acids with buried surface area that were calculated as less than 10 Å^2 to the interface were omitted. ASA - Accessible Surface Area, BSA - Buried Surface Area.

10B9	ASA (Å^2)	BSA (Å^2)
V_H	Chains A/C	
Ser31	69.3/65.4	58.1/57.4
Gly32	12.0/10.6	10.7/9.4
Phe33	33.8/36.4	33.8/36.1
Tyr53	115.0/112.1	85.0/81.2
Thr56	88.7/88.4	15.1/10.8
Gly99	25.4/26.1	18.7/17.2
Leu101	152.1/146.2	142.9/136.5
Val104	36.3/35.9	36.0/34.9
Asn105	82.1/80.5	76.7/76.1
Tyr106	172.8/173.6	50.4/49.7
Tyr107	164.3/161.9	83.0/82.2
V_L	Chains B/D	
Ser30	59.8/57.7	33.8/41.1
Tyr32	102.7/104.7	61.2/60.1
Arg53	136.2/134.0	56.5/59.0
Asn92	55.3/55.8	36.7/38.4
Tyr96	111.7/114.2	14.4/15.9

Table IX The Der p 1-binding paratope on the 10B9 antibody.

The table presents the amino acids that form the Der p 1 binding paratope on the Fab fragment of mAb 10B9. The amino acids with buried surface area that were calculated as less than 10 Å^2 to the interface were omitted. ASA - Accessible Surface Area, BSA - Buried Surface Area.

The uncomplexed Fab fragment of mAb 10B9 crystallized in the orthorhombic space group $P2_12_12$. The overall conformation is almost identical to the Fab fragment of mAb 10B9 that crystallized in complex with Der p 1, however there are conformational differences between the CDRs.

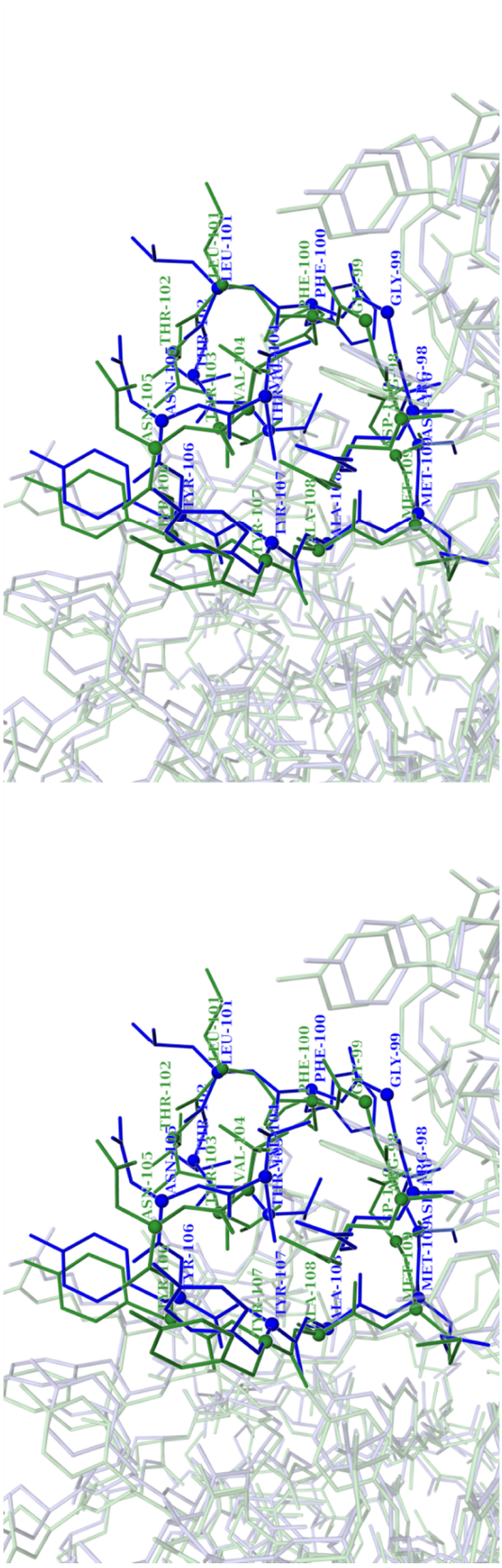

Figure 8 The stereogram of the CDR H3 loop of 10B9 (residues 98-110) in two forms.

The CDR H3 loop on V_L 10B9 between residues 98 and 110 in the complexed and uncomplexed form is shown in stereo. The uncomplexed form of the Fab fragment of mAb 10B9 is colored green; the complexed form with Der p 1 is colored blue.

A superposition of C_L, C_H, V_L, V_H from the 10B9 crystal and corresponding chains from 10B9 from the complex with Der p 1 have a root mean square deviation of 0.8 Å. CDRs L1 (residues 23 to 34), L2 (residues 49 to 56), L3 (residues 88 to 97), H1 (residues 23 to 36) and H2 (residues 50 to 67) have the same side chain conformations and almost identical placement of Cα atoms in both the non-bonded and bound antibody structures. However, CDR H3 (residues 98 to 110) has a very rare conformation in the binding area – an alpha helix. A comparison of uncomplexed 10B9 with Der p 1-10B9 complex revealed that the CDR H3 undergoes a few significant conformational changes upon binding to Der p 1.

The change is especially well visible for residues Phe100, Leu101, Thr102, Thr103 and Asn105 which adopt a well-defined α-helical conformation (one and a half turn alpha-helix as shown on Figure 9 and Table X).

	amino acid	uncomplexed		complexed	
		φ (°)	ψ (°)	φ (°)	ψ (°)
	Arg98	-120.0	87.6	-121.6	92.7
	Gly99	-131.1	-30.5	-136.9	-6.4
*	Phe100	-54.6	133.0	-77.6	154.3
*	Leu101	-58.8	-38.8	-54.7	-40.8
*	Thr102	-74.0	-28.0	-68.0	-57.0
*	Thr103	-107.7	-17.0	-64.5	-29.4
*	Val104	-55.1	-39.7	-70.2	-45.0
*	Asn105	-79.7	3.4	-66.2	-24.3
	Tyr106	-114.0	13.9	-123.1	23.8
	Tyr107	59.9	40.5	52.1	55.9
	Ala108	-101.4	145.7	-128.6	158.3
	Met109	-91.0	101.3	-104.0	118.7
	Asp110	-84.2	-10.8	-83.9	-57.3

Table X The φ and ψ angles found in CDR H3 in two forms of the Fab of mAb 10B9.

The phi and psi angles found in CDR H3 in the uncomplexed and complexed form of the Fab of mAb 10B9. Residues that form a short alpha helix upon binding are indicated with an asterisk.

The loop placement during the binding of Der p 1 is shown in Figure 8 and Figure 9. It also indicates that the overall conformation of the Fab fragment of mAb 10B9 is not significantly changed in comparison to the uncomplexed form, but rather all the CDR regions are shifted upon binding to Der p 1. There are two salt bridges formed by Der p 1 and 10B9. The first is formed by Asp198 and Arg53 of the V_L and the second is formed between Glu13 and Arg98 of the V_H. The asymmetric unit contains two copies of Der p 1-10B9 complex and fourteen hydrogen bonds have been identified in the first copy and thirteen in the second copy of the complex. The hydrogen bonds between Der p 1 and V_L of 10B9 involve the following residues: Gln18 of Der p 1 and both Gly91 and Asn92 of 10B9 (asparagine forms a hydrogen bond only in one of the copies of Der p 1-10B9 complex; Asp198 of Der p 1 and Arg53 (as well as Tyr106 of V_H) of 10B9; Asn199 of Der p 1 and Arg53 of 10B9. Arg20 was found to form a hydrogen bond with Ser30 of the V_L, but only in one of the copies of the complex. The hydrogen bonds between Der p 1 and V_H of 10B9 involve, respectively, Ala12 and Tyr53, Glu13 and Arg98, Asp15 and Tyr107, Asn179 and both Ser31 and Gly32, Gly182 and Ser31, Asp198 and Tyr106 (Table VII). Gln181 was found to form a hydrogen bond with Ser31 in only one of the copies of the Der p 1-10B9 complex. There is a possibility that two Der p 1 residues form cation-π interactions: Gln18 with Tyr107 and Arg17 with Tyr106 of the V_H of 10B9, but a higher resolution structure of Der p 1-10B9 complex would be required to verify these interactions.

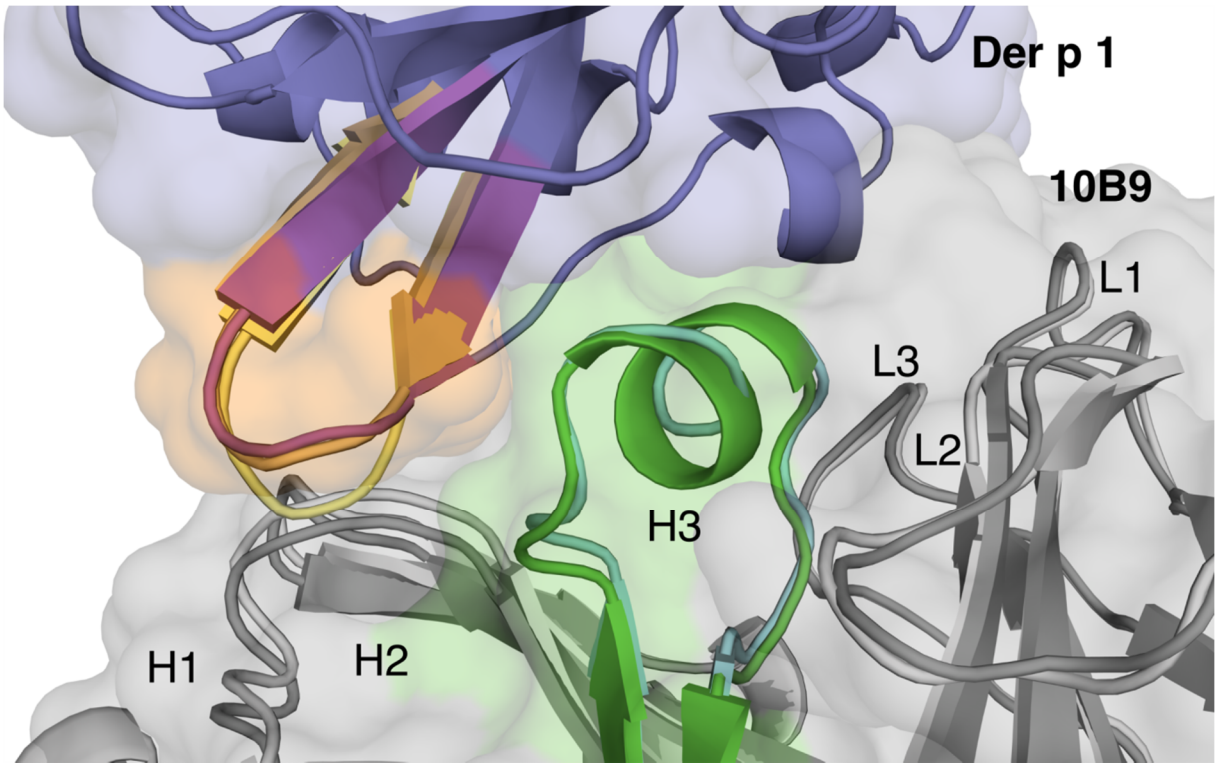

Figure 9 The Fab fragment of mAb 10B9 complexed with Der p 1.

Der p 1 is colored blue. Loop 179-185 of Der p 1 is colored according to the complexes it comes from: Der p 1-10B9 is colored magenta, Der p 1-4C1 is yellow, and Der p 1 in uncomplexed form is orange. CDR H3 of the Fab fragment of mAb 10B9 has a very rare helical form and is colored green and cyan in its complexed and uncomplexed form, respectively. The rest of 10B9 is colored grey.

7.1.3. Relative location of the epitopes for the mAbs 10B9 and 4C1 on Der p 1

Even though the epitopes for both 4C1 and 10B9 largely overlap (Table VIII and Table XI), the relative position of the antibodies is not shifted, but rather the 4C1 binds to its epitope in a position that is "rotated counter-clockwise" by a little more than 90 degrees with respect to the position assumed by 10B9 upon binding to its epitope on Der p 1 (Figure 6a and Figure 6b). The residues of Der p 1 that participate in hydrogen bonding with both 10B9 and 4C1 are Glu13,

Arg17, Gln18 and Asp198. Residues specific to hydrogen bond formation with 4C1 are Arg20 and Arg156, while Ala12, Asp15, Gln181, Gly182, Tyr185 and Asn199 participate in hydrogen bonding with 10B9. A surface exposed tyrosine residue of the Fab fragment of mAb 4C1 forms hydrogen bonds with an arginine residue on the surface of the antigen, which is observed at the Der p 1 and 4C1 interface (Figure 11a, Figure 11b, Figure 12a, and Figure 12b as well as previous work (Chruszcz et al., 2012b)). These bonds are not observed in the Der p 1-10B9 complex. However hydrogen bonds are present between a surface tyrosine on the antibody and an aspartate residue on the surface of Der p 1. Furthermore, such hydrogen bonds are not present at the interfaces formed by the Fab fragments of mAb 4C1 or 5H8 and Der p 1. The interface between 10B9 and Der p 1 comprises several hydrogen bonds that are formed between serine and arginine as well as asparagine/glutamine and glycine. In the 4C1 complex however, the interface is formed between arginine, aspartate, serine, or threonine from 4C1 and tyrosine, arginine, aspartate, glutamine or glutamate on Der p 1, with the most abundant linkage being between tyrosine from 4C1 and arginine from Der p 1 (Table XI, Table XII and Figure 3 of previous work (Chruszcz et al., 2012b)).

	Der f 1	Der p 1	ASA (Å^2)	BSA (Å^2)
		Chain A with V_H of 4C1		
	Glu14	Glu13	114.3/104.9	49.6/40.9
	Asp16	Asp15	34.8/27.0	19.1/13.6
	Arg18	Arg17	69.3/74.4	60.2/62.0
*	Ser19	Gln18	85.4/127.8	69.1/72.3
	Arg21	Arg20	179.5/193.6	59.3/92.7
	Arg157	Arg156	196.5/190.9	60.10/59.2
	Thr158	Thr157	90.2/89.7	8.5/10.2
	Ile159	Ile158	31.4/39.6	16.1/16.6
*	Thr181	Ala180	57.67/39.5	34.1/23.2
	Tyr186	Tyr185	39.4/35.9	35.8/33.8
	Asp199	Asp198	63.3/50.7	50.3/45.2
*	Ser200	Asn199	70.0/101.3	13.1/18.9
	Tyr202	Tyr201	18.5/18.2	13.7/10.5
	Tyr204	Tyr203	29.8/27.2	29.8/27.2
		Chain A with V_L of 4C1		
	Arg157	Arg156	196.5/190.9	134.4/127.6
*	Thr181	Ala180	57.6/39.5	20.2/10.6
	Gln182	Gln181	190.2/189.0	52.8/50.9

Table XI The 4C1-binding epitopes on Der p 1 and Der f 1.

The table presents the amino acids that form the 4C1 binding epitopes on Der p 1 and Der f 1. The differences are indicated with an asterisk. The amino acids with buried surface area that were calculated as less than 10 Å^2 to the interface were omitted. ASA - Accessible Surface Area, BSA - Buried Surface Area.

4C1	ASA (Å^2)	BSA (Å^2)
V_H	Chain D	
Thr30	62.4/62.2	8.2/16.4
Ser31	70.3/69.3	19.9/30.5
Asp32	44.3/36.0	23.4/16.4
Tyr54	66.0/64.7	53.6/53.4
Ser55	75.0/80.5	30.4/22.9
Tyr102	152.1/151.0	103.0/99.8
Arg103	210.3/210.9	89.1/79.1
Tyr104	160.1/155.1	134.4/131.0
Pro105	92.1/91.1	80.3/76.3
Glu106	126.1/119.1	25.1/21.7
Arg107	102.5/103.8	21.7/18.6
V_L	Chain C	
Tyr30	133.2/131.2	41.0/40.6
Tyr32	76.3/77.3	29.2/24.1
Arg50	70.1/75.4	10.4/17.8
Asn52	91.4/-	10.8/-
Arg53	115.1/124.7	56.4/59.8
Asp92	52.8/52.9	18.1/18.9

Table XII The Der p 1 and Der f 1 binding paratopes on the 4C1 antibody.

The table presents the amino acids that form the Der p 1 (left sides of the ASA and BSA columns) and Der f 1 (right sides of the ASA and BSA columns) binding paratopes on the Fab fragment of mAb 4C1. The amino acids with buried surface area that were calculated as less than 10 Å^2 to the interface were omitted. ASA - Accessible Surface Area, BSA - Buried Surface Area

There is a conformational change of the Der p 1 loop from residue 179 to 185 (rmsd value of 0.7 Å, where overall rmsd is 0.6 Å) calculated between Der p 1 complexed with 4C1 and Der p 1 complexed with 10B9. This conformational change however, is not observed between uncomplexed Der p 1 and the Der p 1-10B9 complex. This loop provides strong interactions with the paratope on the heavy chain of the Fab fragment of mAb 10B9, but it is on the edge of the epitope and does not form any substantial interactions with the light chain of

the Fab fragment of mAb 4C1 (Figure 9). Therefore, the conformational change might be induced upon binding of water molecules between the 4C1 antibody and Der p 1, which would therefore form bridging interactions between Asn179, Ser180 and Tyr186 with Glu106 of CDR H3 (as described in previous work (Chruszcz et al., 2012b)). However, this rearrangement is located on a peripheral part of the Der p 1-4C1 interface and other variables might contribute to the rearrangement.

7.1.4. Comparison between Der p 1 allergen epitopes for 10B9, 4C1 and 5H8 antibodies and the corresponding surface on Der f 1

The interfaces formed between Der p 1 and 10B9 or Der p 1 and 5H8 (both species-specific mAb) have larger contact areas than the interfaces between Der p 1 and 4C1 and Der f 1 and 4C1 (cross-reactive mAb) (~910 $Å^2$ and ~845 $Å^2$, versus ~745 $Å^2$ and ~760 $Å^2$, respectively; calculated by PISA). As mentioned previously, both 10B9 and 5H8 are species specific, whereas the 4C1 antibody is cross-reactive between Der p 1 from *D. pteronyssinus* and Der f 1 from *D. farinae*. The deletion of serine in the 8^{th} position of Der p 1 causes an offset, by one residue in the numbering of the amino acid residues between Der p 1 and Der f 1, thus Der f 1 numbering of the amino acid residues is used in comparisons of Der f 1 with Der p 1 for both allergens in this section.

The Q_H algorithm was used as the scoring function on structures superimposed in VMD. The calculated Q_{RES} values were assigned for every residue and the surface was marked in such a way that the amino acids with sidechains that had the most similar orientation and shape of the sidechains are colored blue through white to red (which indicates that the orientation and shape of the sidechains were different) (Figure 6d, Figure 6e and Figure 6f).

The 4C1 antibody is cross-reactive between two dust mite allergens, Der p 1 and Der f 1, while 10B9 is specific for Der p 1. In order to explain this observation I compared the residues of Der p 1 that form interactions with 10B9 to corresponding residues from Der f 1. The residues

that form the 10B9 binding epitope on Der p 1 are partially conserved in Der f 1. In addition to the surrounding amino acids not directly involved in the interaction with 10B9, the major differences between Der p 1 and Der f 1 in the 10B9 binding epitope include a substitution of residues Ser13, Ser19, Ser180, Thr181 and Ser200 in Der f 1 by Ala, Gln, Asn, Ala and Asn respectively in Der p 1 (Figure 6e and Figure 10c) which allows for hydrogen bond formation between Der p 1 and 10B9 (Table VII). However, a lack of these hydrogen bonds is not the major factor preventing the binding of 10B9 to Der f 1. A comparison of the available Der p 1 and Der f 1 structures reveals that there is a significant difference in conformation and composition of the 179-183 region (Asn-Ala-Gln-Gly-Val) of Der p 1 and corresponding 180-184 fragment (Ser-Thr-Gln-Gly-Asp) of Der f 1. The different conformations of these regions results in a significant shift of the central Gln position, with corresponding Cα atoms being shifted by ~3 Å. This shift results in Gln181 in Der p 1 being "parallel" to the allergen's surface, where in the case of Der f 1 this residue is "perpendicular" to the allergen’s surface. A superposition of the Der p 1-10B9 complex and Der f 1 shows that the "perpendicular" conformation of Gln182 in Der f 1 results in a steric clash with Ser31 from the heavy chain of the antibody. The density map, which is of good quality around Gln181 allows for the assumption that this shift may be a result of induced fit.

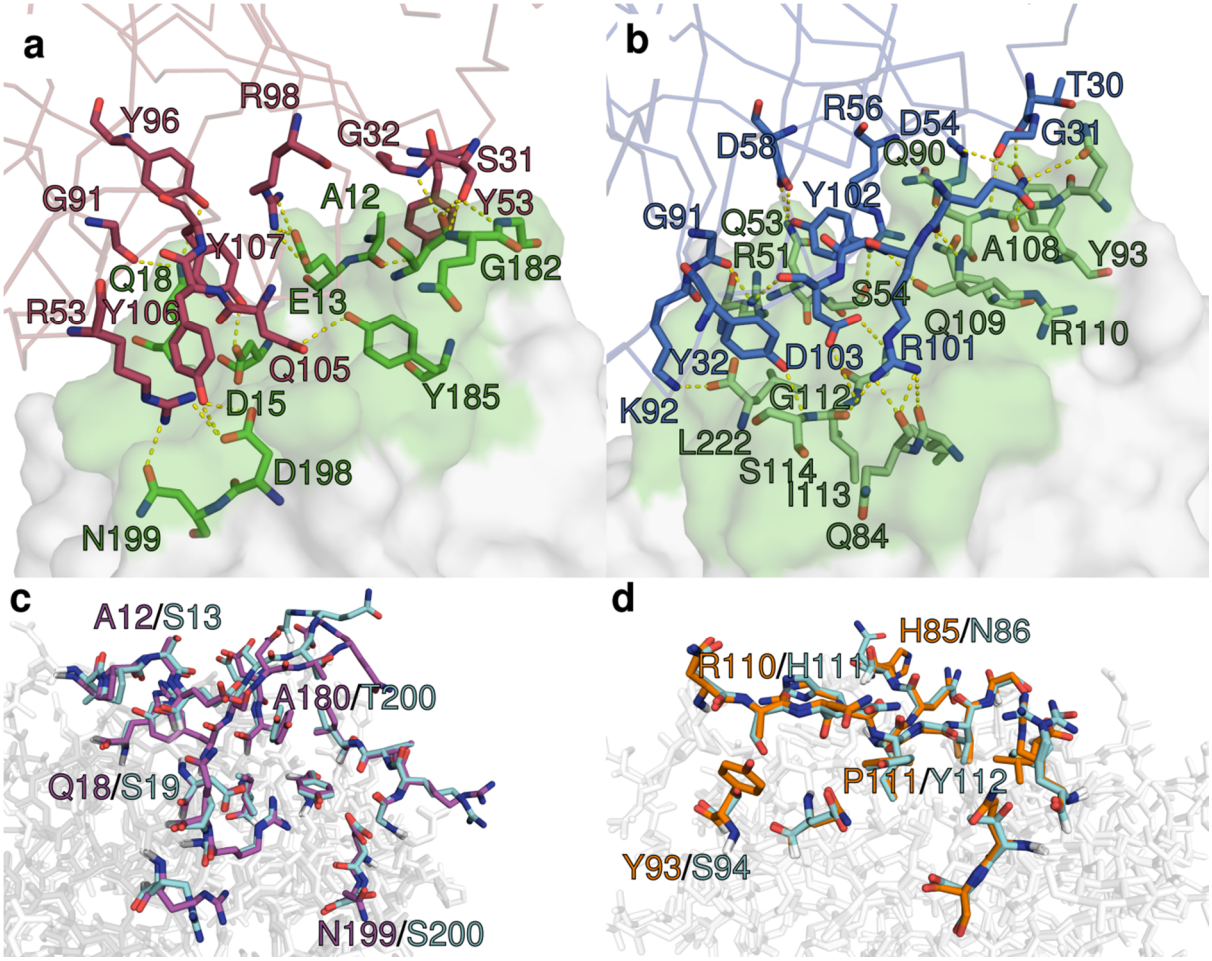

Figure 10 The 10B9 (a and c) and 5H8 (b and d) binding epitopes of Der p 1.

The 10B9 (a and c) and 5H8 (b and d) binding epitopes of Der p 1. The amino acids that form hydrogen bonds between Der p 1 and antibodies are colored as follows: Der p 1 – shades of green, 10B9 – purple, 5H8 – blue. The labels showing residue names and numbers are colored in the same way. The amino acid residues in Der f 1 corresponding to residues of Der p 1 epitopes binding 10B9 is shown in c) and colored purple and 5H8 is shown in d) and colored orange. The cyan colored residues shown in c and d come from Der f 1 and are corresponding to residues from Der p 1 labeled purple in c and orange in d.

The 5H8 binding epitope area on Der p 1 is mostly conserved in Der f 1 (fifteen out of twenty four amino acid residues), but there are several differences between them (Table VI). Residues Thr54 and Ser94 of Der f 1 are substituted by Gln and Tyr respectively in Der p 1, thus making it impossible for Der f 1 to form hydrogen bonds with the Fab fragment of mAb 5H8 (Figure 10d). Der f 1 amino acid residues Thr54, Glu91, Ser94, Ser109 and Tyr112 are substituted respectively by Gln, Gln, Tyr, Ala and Gly residues in Der p 1 respectively, which forms a non-conserved spot in the central part of the 5H8 epitope surface area (the white spot in the central part of Figure 6f) and therefore changes the shape of the surface of the 5H8 binding

epitope (Figure 10d). Some of the amino acids in Der f 1 (Gln86, His111 and Met223) that correspond to the 5H8 binding epitope on Der p 1 are substituted (His, Arg and Leu), but in the Der p 1-5H8 complex all the interactions are mediated by the main chain atoms for these residues and the side chains do not form steric clashes. Therefore these residues should not interfere with the possibility of 5H8 binding to Der f 1.

7.1.5. Analysis of antigen-antibody interactions in complexes reported in PDB

The expected frequency of amino acids on the surface of proteins was calculated as described in Materials and Methods section and is available in Table XIII.

Amino acid									
A	R	N	D	C	Q	E	G	H	I
4.9	3.8	5.4	5.1	0.5	4.4	3.5	8	1.6	3.5
Frequency (%)									

Amino acid									
L	K	M	F	P	S	T	W	Y	V
4.7	5.4	1.6	3.1	7.1	13.1	11.5	2.1	3.7	7.1
Frequency (%)									

Table XIII The frequency of amino acids on the surface of non-redundant Fab-protein and Fab-peptide complexes deposited in PDB (as of 2013).

The methods used to investigate Der p 1 complexes with antibodies made it possible to analyze not only these complexes, but all non-redundant complexes of Fab fragments of antibodies with proteins or peptides available in the PDB of resolution equal to, or better than, 3.0 Å (n = 314 as of 2013). The amount of particular amino acids found in the protein-antibody or peptide-antibody interfaces are shown in Figure 11g and Figure 11i. However, the observed frequency of the amino acids on the surface of the epitopes, in general, is very similar to the expected frequency of these amino acids on the surface of proteins from the dataset, with

charged or polar residues more frequent, and hydrophobic residues less frequent. Serine and threonine are the exception, however, with their occurrence being essentially two times less frequent in epitopes than in other parts of the surface of proteins from the dataset. Glutamine residues in the epitopes are approximately one third more frequent than expected. The observed frequency of the amino acids in the paratopes differ significantly from the expected frequency, with tyrosine being six times more frequent and tryptophan being three times more frequent in the paratope interfaces than on the surface of proteins from the dataset (Table XIV).

The observed value of serine frequency in the paratopes of Fab-protein complexes is almost equal to the expected serine frequency on the protein's surface (Table XIV). This frequency of serine is not observed in the paratopes of the antibodies interacting with peptides and is half of the expected value (Figure 11j and Table XIV). Among charged amino acids, only arginine and histidine are more frequent in the paratopes than expected on the surface. The frequency of lysine is lower than expected for protein-antibody complexes or almost equal to the expected value for peptide-antibody complexes.

	Antigens					Antibodies				
Interface	proteins	peptides	Der p 1			antibodies		10B9	5H8	4C1
amino acids	antibodies		10B9	5H8	4C1	proteins	peptides	Der p 1		
A	0.8	1.2	2.0	1.0	2.0	0.4	0.6	0.5	0.0	0.0
R	1.8	1.6	6.0	4.0	6.0	1.6	1.6	3.0	2.0	8.0
N	1.3	0.7	2.0	2.0	1.0	1.3	1.3	2.0	0.0	0.0
D	1.4	1.2	3.0	0.0	4.0	1.6	1.6	0.0	3.0	2.0
C	2.0	4.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Q	1.4	1.8	2.5	4.0	4.0	0.5	0.5	0.0	1.0	0.0
E	2.0	1.4	2.0	0.0	2.0	1.1	1.2	0.0	0.0	2.0
G	0.9	0.9	0.7	0.7	0.0	0.9	0.8	1.0	2.7	0.0
H	2.5	1.8	2.0	0.0	0.0	1.9	2.5	0.0	0.0	0.0
I	1.4	1.4	0.0	2.0	2.0	0.6	0.9	0.0	0.0	0.0
L	1.1	1.7	0.0	2.0	0.0	0.6	1.1	1.0	0.5	0.0
K	1.5	1.3	0.0	0.0	0.0	0.4	0.6	0.0	1.0	0.0
M	1.2	1.3	0.0	0.0	0.0	0.6	0.6	0.0	0.0	0.0
F	1.0	1.0	0.0	2.0	0.0	1.3	1.6	2.0	0.0	0.0
P	0.8	0.9	0.0	0.0	0.0	0.3	0.3	0.5	0.0	1.0
S	0.5	0.5	0.4	0.8	0.0	1.0	0.5	1.4	0.4	1.0
T	0.4	0.4	0.0	0.0	0.7	0.5	0.5	0.8	1.2	0.5
W	1.0	1.9	0.0	0.0	0.0	2.9	2.9	0.0	2.0	0.0
Y	1.1	0.8	5.0	6.0	5.0	5.4	5.2	10.0	4.5	10.0
V	0.6	0.7	0.0	0.0	0.0	0.3	0.4	1.0	0.0	0.0

Table XIV The ratio between the observed and the expected amino acid residues occurrences in the interfaces between the antigens and the antibodies.

Ratio was calculated from a frequency of occurrences of a particular amino acid residue in all surfaces of non-redundant Fab-protein or Fab-peptide complexes available in the pdbaa NCBI's BLAST database as of 2013. Stronger green intensity indicates higher ratio.

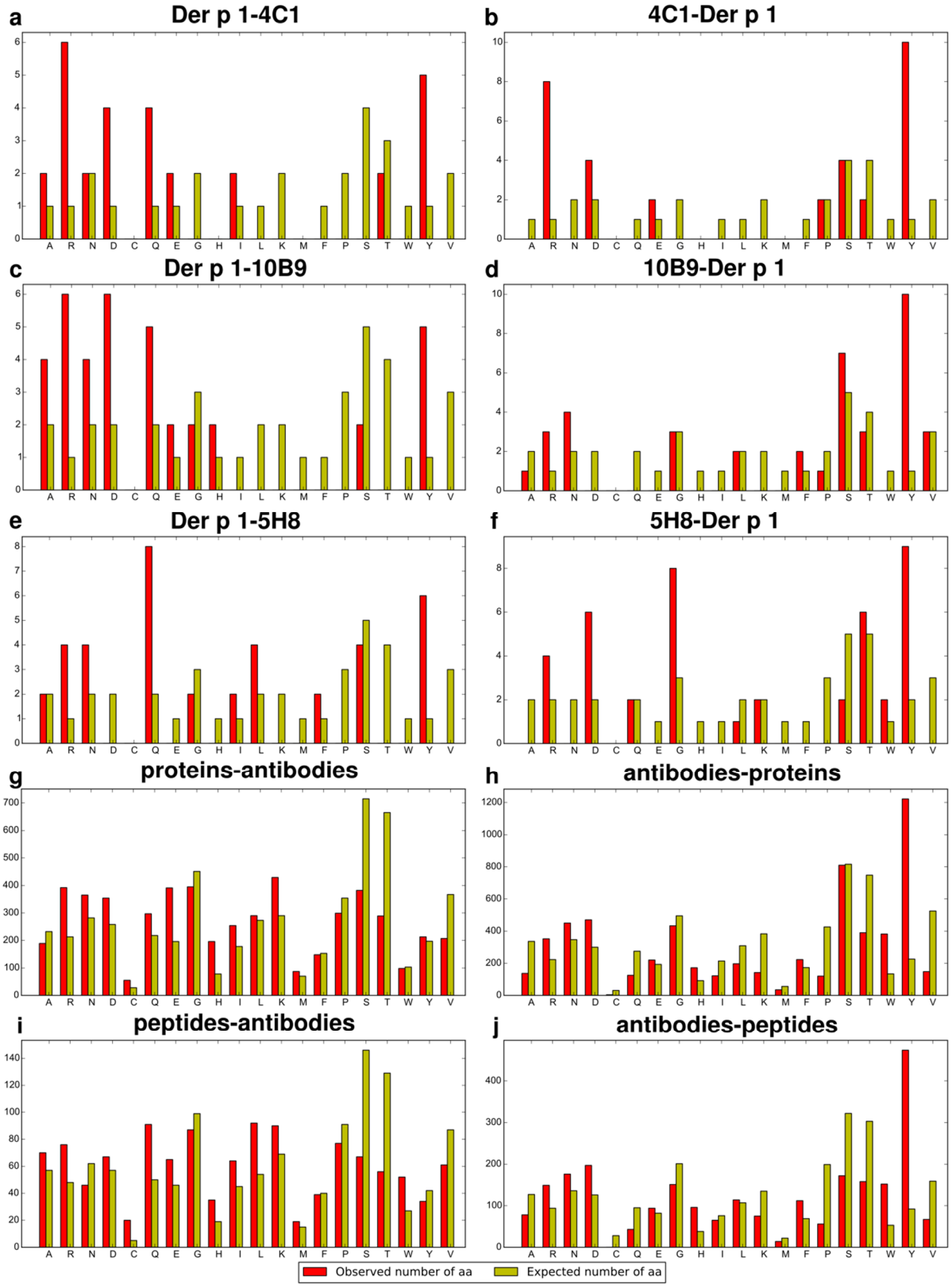

Figure 11 The amino acids found in the Der p 1-antibody interfaces.

Each plot shows the distribution of residues in the binding surface. The epitopes are shown in a), c), e), g) and i), the paratopes are shown in b), d), f), h) and j). The number of residues found in the given interface is colored red and expected number of residues (based on the frequency of the amino acids found on the surfaces of protein-Fab complexes from the dataset used) is colored yellow.

7.1.6. Amino acid composition of the interface of the complexes of Der p 1 with antibodies

The interface area of Der p 1 complexed with the Fab fragments of mAb 4C1, 10B9 or 5H8 (Figure 11 and Figure 12) is composed of almost every charged or polar amino acid residue (tyrosine, threonine, arginine, serine, glutamine, aspartate, glutamate and aspartate). However, hydrophobic amino acids like leucine and isoleucine are also present. The amino acids that form the interface and are common between all three antibodies include tyrosine, arginine, threonine and serine. Tyrosine is one of the main amino acids that contribute the most surface area to the interfaces in both the epitopes of Der p 1 and the paratopes of the antibodies. The number of tyrosines is approximately five (5H8) to ten times (10B9 and 4C1) higher in paratopes than the average tyrosine frequency on protein surfaces of Fab-protein and Fab-peptide complexes (Table XIV). Arginine is significantly more abundant in the epitopes of Der p 1 and the paratopes of all three antibodies. In the paratope of the Fab fragment of mAb 4C1, arginine is eight times more frequent than it is on protein surfaces Fab-protein and Fab-peptide complexes (Table XIV). The interface area on the Der p 1 epitopes contains a significantly higher frequency of arginine, aspartate, glutamate and glutamine than the general frequency of these amino acids on the surface of the proteins from the dataset. Arginine is four to six times more frequent, aspartate is three to four times more frequent, glutamate is two times more frequent and glutamine is two and a half to four times more frequent than what is expected. Asparagine, despite the physicochemical similarity to glutamine, is only present in the paratope of the 4C1 antibody in twice the expected frequency, and is absent in the 10B9 and 5H8 paratopes. A rather unusual observation, in comparison to other antibodies, is that the paratope of the 4C1 antibody contains eight times more arginine than the usual arginine frequency on protein surfaces, and it is much higher than the general arginine content in the paratopes, as well as in the paratopes of 10B9 and 5H8 (Table XIV). The hydrogen bonds formed between glycines of the Fab fragment of mAb 5H8 or the Fab fragment of mAb 10B9 and the Der p 1 allergen are quite common in

comparison to the hydrogen bonds formed between all antibodies and all proteins or peptides (Figure 7). The differences between amino acids involved in hydrogen bonds between the paratopes of the Fab fragments of monoclonal antibodies 4C1, 10B9 and 5H8 and Der p 1 allergen are significant. Most of the hydrogen bonds in the paratope of 10B9 are formed by serine with arginine, asparagine, glutamine or glycine and by tyrosine with alanine or aspartate on Der p 1 (Figure 7a). The hydrogen bonds are formed on the 5H8 paratope by aspartate with arginine, glutamine or glycine; by glutamine with arginine or asparagine and by tyrosine with serine or arginine on the surface of Der p 1 (Figure 7b). A quite rare hydrogen bonding interactions can be observed at the Der p 1 epitope and 5H8 paratope interface. Namely, the hydrogen bonds are formed between the mainchain atoms of Ile83 and Ile113 of Der p 1 and sidechain of Arg101 from 5H8. Additionally, Ile113 from Der p 1 and Asp103 from 5H8 also form hydrogen bonds. Surprisingly, there are also not common hydrogen bonds at the interface between Der p 1 and 5H8 - formed between glutamines from Der p 1 and arginines and aspartates from 5H8 antibody.

The amino acid composition of the Der p 1 epitopes that bind to the Fab of mAb 10B9, 5H8 and 4C1, does not contain all of the standard amino acids. The amino acids that are missing or underrepresented are mostly non-polar, however, some polar or charged amino acids may not be present in the every epitope. Aspartate and glutamate are not present in the 5H8 binding epitope, but very common in the 4C1 and 10B9 binding epitopes. Serine is absent in the 4C1 binding epitope, but present in other epitopes. The 4C1 binding epitope contains threonine which is absent in both 5H8 and 10B9 binding epitopes. Valine and lysine are absent in all of the three epitopes. Alanine is overrepresented in the 4C1 and 10B9 binding epitopes and is present in equal amount to the expected value in the 5H8 binding epitope on Der p 1. The content of isoleucine, despite being a non-polar amino acid, is significantly higher than expected in the 4C1 and 5H8 epitopes, but isoleucine is absent in the 10B9 epitope. Even though the

Der p 1 epitopes that bind 4C1 and 10B9 mostly overlap, their amino acids composition is different. The epitope that binds 10B9 contains glycine and serine in lower than expected quantities and histidine in higher than the expected amount. However, it does not contain either threonine or isoleucine, both of which are present in the 4C1 binding epitope.

7.1.7. Analysis of antigen-antibody interactions in complexes reported in PDB - areas contributed by particular amino acids

Figure 12 shows the area contributed to the interfaces by particular amino acids. The amino acids found in the epitopes of proteins and peptides generally contribute more area to the interfaces than amino acids found in the paratopes. Residues that contribute the most area to the protein-antibody interfaces found in the epitopes are arginine, histidine, lysine, methionine, phenylalanine, and tryptophan, while the residues that contribute the most area to the protein-antibody interfaces found in the paratopes are arginine, methionine, phenylalanine, and tryptophan. The amino acids found in the epitopes of the peptide-antibody interfaces contribute even more area to the interface as compared to the protein-antibody interface, but this observation can be explained by the small size of peptides. In the paratopes of the peptide-antibody interfaces, methionine contributes the most area. Curiously, and in contrast with the larger area contribution of methionine in all of these interfaces, the observed frequency of methionine found in both the epitopes and paratopes is half the expected frequency of methionine on the surface of proteins in the dataset, with the exception of methionine found in the peptide epitopes (Table XIV).

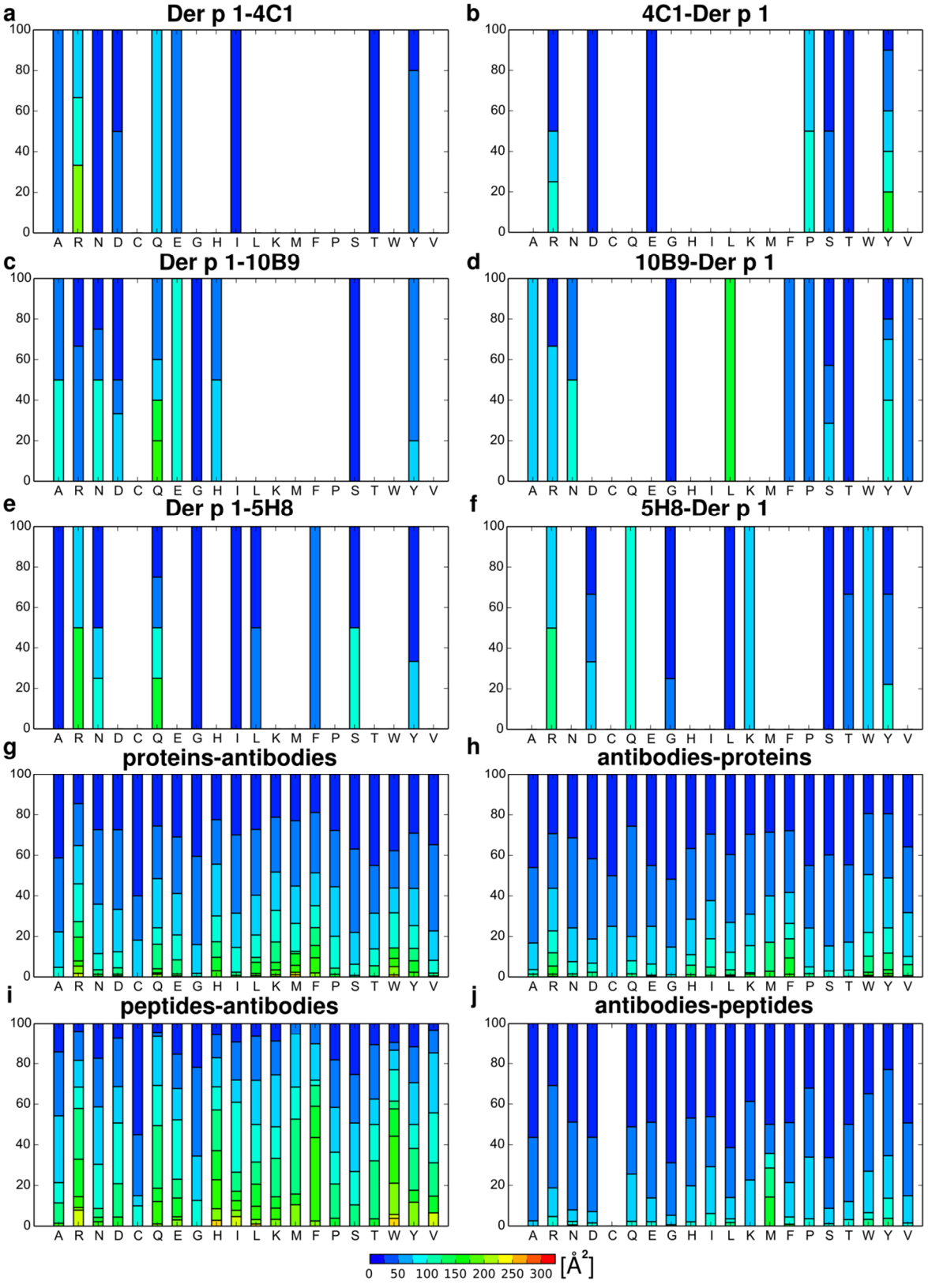

Figure 12 The area contributed to the Der p 1-antibody interface by each amino acid.

Each plot shows the distribution of the area contributed by the each amino acid residue forming an interface. The epitopes are shown in a), c), e), g) and i), the paratopes are shown in b), d), f), h) and j).

7.1.8. Analysis of the amino acid composition of the complexes of proteins or peptides with Fab fragments of monoclonal antibodies

The comparative analysis of interfaces formed by Der p 1 with the Fab fragments of mAb 4C1, 10B9 and 5H8 required an investigation of the content of the interfaces and hydrogen bonds formed by all Fab fragments of antibodies complexed with proteins or peptides. The analysis of some of the antigen-antibody interfaces properties showed (Figure 11 and Figure 12) that tyrosine is found over five times more often on the surface of the paratopes than on the surface of proteins in general (Table XIV). It strongly suggests the important structural role of tyrosine in forming interfaces and its presence in interface areas found in all complexes of Fab fragments of antibodies with proteins, peptides or nucleic acids (unpublished data) and it is consistent with earlier studies (Koide and Sidhu, 2009). Tryptophan, which shares some of the properties of tyrosine, is found to be three times more frequent on the surface of paratopes than on the surface of proteins in general. A higher concentration of aromatic amino acids at the binding interfaces is consistent with a previous report (Wu et al., 2010). The quantitative analysis of the binding interfaces showed that aromatic amino acids found on paratopes tend to contribute significantly more surface area to the interfaces than other amino acids found in paratopes, with the exception of methionine. Methionine, despite its very low frequency of occurrences in antigen-antibody interfaces, usually contributes significant surface area to the interface when present. Serine and threonine, despite having polar side chains, are found to be two times less frequent in the protein-antibody or peptide-antibody interfaces as in other parts of proteins. The observed frequency of serine in the paratopes, however, is almost equal to the expected value.

7.2. Bla g 4

7.2.1. Structural analysis of an apo and liganded forms of Bla g 4

The Bla g 4 allergen is a compact, globular protein that folds into a typical lipocalin fold (Yih et al., 2009), where the core of the protein contains a β-barrel formed by eight anti-parallel strands. A bent α-helix is placed adjacent and on the outside of the β-barrel (Figure 13).

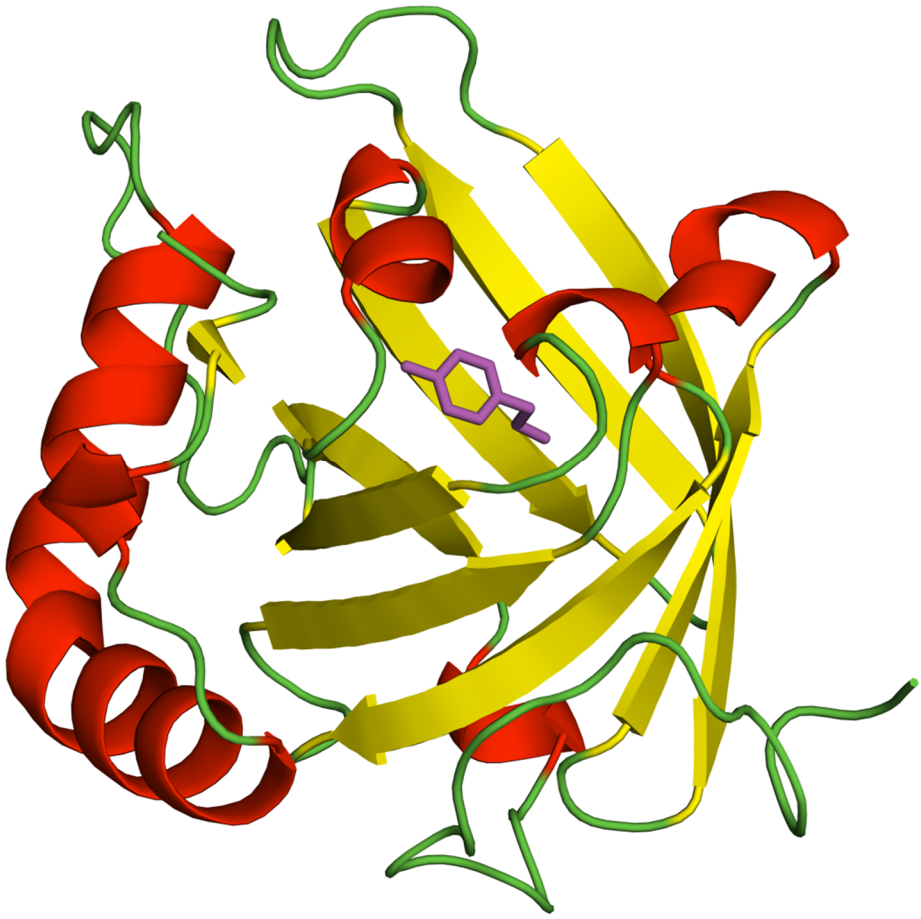

Figure 13 The Bla g 4 allergen with tyramine.
Cartoon representation of Bla g 4 with α-helices colored red, β-sheets colored yellow and coiled coils colored green.

Two long-range disulfide bonds (Cys10-Cys112 and Cys44-Cys175) that connect the central β-barrel with other structural elements have structural role by strengthening the compactness of the Bla g 4 structure. The Se-Met derivative and the native forms of the allergen crystallized in the same space group with very similar unit cell parameters. Additionally, the

Se-Met derivative is an apo form, while the native version of the protein is a ligand bound form of the protein. The superimposition of Cα of these structures yields a rmsd value of 0.3 Å. The most significant differences between apo and native form of Bla g 4 are observed in the mainchain conformations of the loop regions between residues ranging from 61 to 65 and from 134 to141. The analysis of the electron density maps for the structure of the native Bla g 4 revealed the presence of a large continuous section of electron density, which does not correspond to any proteinous part of the allergen. The shape of the difference map for this region suggests that tyramine, or a similar small molecule, is present in the structure (Figure 14).

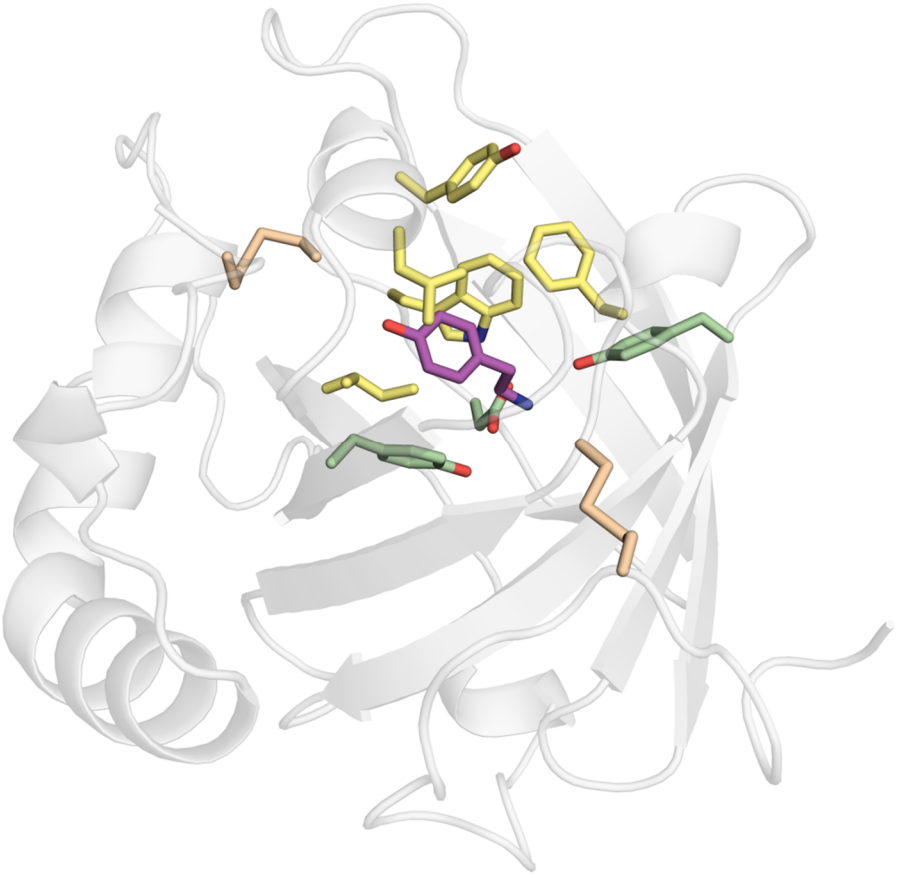

Figure 14 Bla g 4 ligand-binding site with bound tyramine shown in purple.

Tyrosines and aspartate forming the hydrogen bonds with the amine part of tyramine are green. The disulfide bonds are beige. The central part of the ligand is located in a hydrophobic cage which is closed by small cap hiding the ligand inside of the protein.

Although the sequence conservation among proteins from lipocalin protein family is quite low, their structure is highly conserved (Figure 15). The superimposition of structures of

close homologs of the Bla g 4 allergen revealed the structural conservation pattern. The β-barrel forming the core of the protein is very highly conserved. The conservation of the kinked α-helix adjacent to the protein core is divided into two parts and decreases gradually from the moderately high conservation to the lack of conservation along the succession of the α-helix. The short α-helix and small loop forming the hydrophobic cap, which covers tyramine binding site, were not found to be conserved. The N-terminal part of Bla g 4 is not conserved, but it can be attributed to the lack of constraints of its lability.

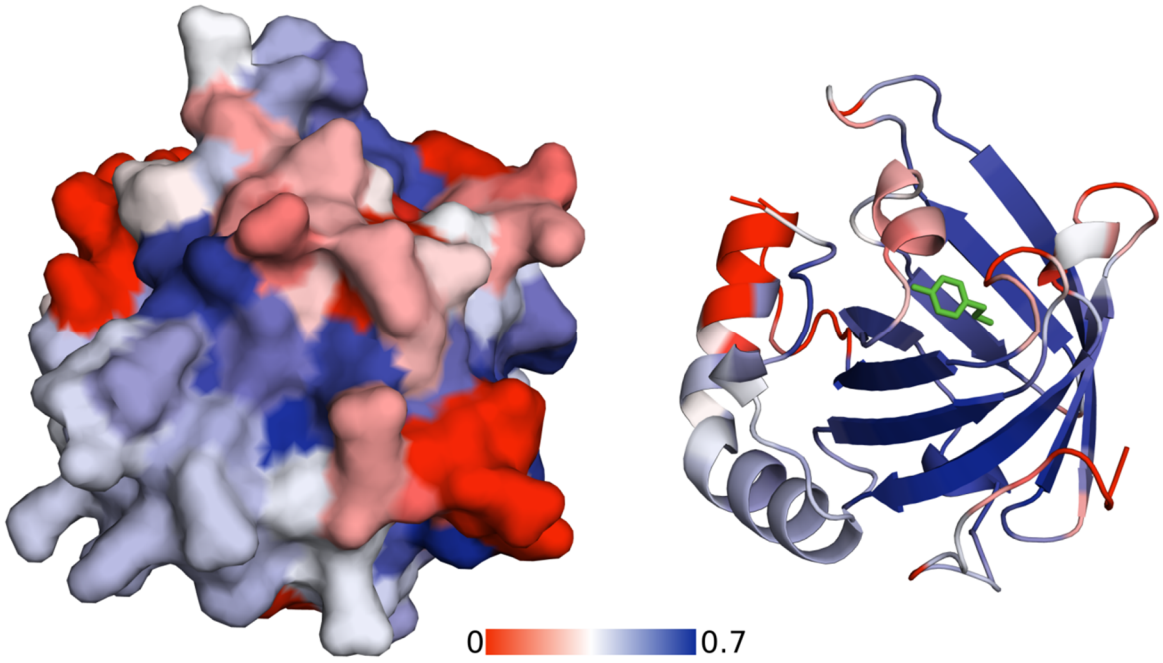

Figure 15 Structural conservation in Bla g 4.

Two representations of Bla g 4 (surface on the left and cartoon on the right side). The color scale represents mapping of Q_{RES} values onto the structure, where red represents no structural conservation and the blue represents conserved regions. Tyramine is colored in magenta. Q_{RES} can assume values from 0 to 1.

7.2.2. Sequence analysis

The sequence clustering was done using CLANS, based on a variant of the Fruchterman and Reingold graph layout algorithm representing sequence pairwise similarities and groups of the similar sequences are divided into the color-coded clusters (Figure 16). The sequences from the brown cluster almost exclusively belong to one of the two PFAM families (Lipocalin and

Lipocalin-like—PF00061 and PF08212) and are mixed with each other. The Bla g 4 allergen belongs to this cluster as well as its closest allergenic homolog – the Per a 4 allergen, both of which come from two different species of cockroaches. This cluster contains subclusters designated with roman numerals. Subcluster I contains chloroplastic lipocalins. Plasma retinol binding protein from various Vertebrata mainly from *Canis lupus familiaris*, *Mus musculus*, *Rattus norvegicus* and *Homo sapiens* are contained within cluster II. Several isoforms of the Bla g 4 allergen are found in the small cluster marked as 'III'. The Bla g 4 allergen, despite being classified by PFAM as part of the Triabin protein family, lies closer to Lipocalin protein family than to the Triabin protein family. Subcluster IV contains uncharacterized proteins from Insecta. Subcluster V is the largest subcluster and is the most diverse of all the other clusters. The known protein structures contained within this cluster provide a good representation of the structural diversity that can be used for structural studies. The most notable structures are human apolipoprotein D complexed with progesterone (PDB codes: 2HZQ, 2HZR), a subunit of crustacyanin-A2 (PDB code: 1GKA), apocrustacyanin C2 (PDB code: 1S2P), lipocalin complexed with fluorescein (PDB code: 1N0S), insecticyanin-A (PDB code: 1Z24), and Per a 4 allergen from cockroach (PDB code: 3EBW). The other proteins within this subcluster are classified as apolipoprotein D, chlorophyllide A binding protein, bilin binding protein, bombyrin, and biliverdin binding protein. The cluster marked with light green contains only one protein with a known structure - bacterial lipocalin (PDB code: 1QWD). Inside this cluster one can find outer membrane lipoproteins Blc, lipocalin-like and lipoprotein-like proteins from the Lipocalin-like PFAM protein family (PF08212) and almost all of those sequences are from various bacterial strains. The cyan cluster contains proteins similar to milk proteins from two cockroach species *Diploptera punctata* and *Rhyparobia maderae*. Various proteins from Drosophila sp. are located within the yellow cluster. The navy-blue cluster contains mammalian Alpha-1 acid glycoproteins. The cluster marked in green contains lipocalins, retinoic

acid-binding proteins, prostaglandin-D synthetases, siderocalins, prostaglandin isomerases, neutrophil gelatinases and Alpha-1 microglobulins. Several of the proteins from this cluster have their structures determined. These are human lipocalin 15 (PDB code: 2XST), human lipocalin 2 (PDB code: 3DSZ, 3BX7, 3BX8, 3DTQ), epidymal retinoic acid-binding protein (PDB codes: 1EPA, 1EPB), Human Complement Protein C8alpha (PDB code: 2RD7), siderocalin (PDB code: 3S26, 3SAO), lipocalin-type prostaglandin D synthase (PDB codes: 2CZT, 2CZU, 2RQ0, 2WWP), lipocalin 12 (PDB code:2L5P), and alpha-1-microglobulin (PDB code: 3QKG). The cluster marked in orange contains mammalian β-lactoglobulins, major urinary proteins, salivary lipocalins, and odorant-binding proteins. Furthermore, this cluster contains the largest number of determined protein structures (PDB codes: 1GM6, 2L9C, 1EXS, 2R73, 2R74, 2RA6, 1XKI, 3EYC, 2A2G, 2A2U, 1EW3, 1I04, 1I05, 1I06, 3L4R, 1DZJ, 1DZK, 1DZM, 1DZP, 1E00, 1E02, 1E06, 1HQP, 1YUP, and 2HLV). Almost all protein sequences from the green and orange cluster are from the Lipocalin PFAM protein family (PF00061). The small purple cluster in the top left corner of Figure 16 contains violaxanthin de-epoxidases from *Arabidopsis thaliana* (PDB codes: 3CQN and 3CQR) from plants and bacteria. The pink cluster contains sequences from the Triabin PFAM protein family (PF03973). This cluster comprises proteins like nitrophorin, thrombin inhibitor or amine binding protein (PDB codes: 1PM1, 2ALL, 2AMM, 1SXX, 1SY2, 1AVG, 1EUO, 1PEE, 1T68, 2A3F, 2ACP, 2AH7, 2AL0, 2HYS, 2GTF, 4GE1, and 4GET) and salivary lipocalins. Despite the PFAM classification, which includes Bla g 4 into the Triabin family, this allergen differs enough from this protein family to be clustered with the Lipocalin family which apparently have higher sequence similarity.

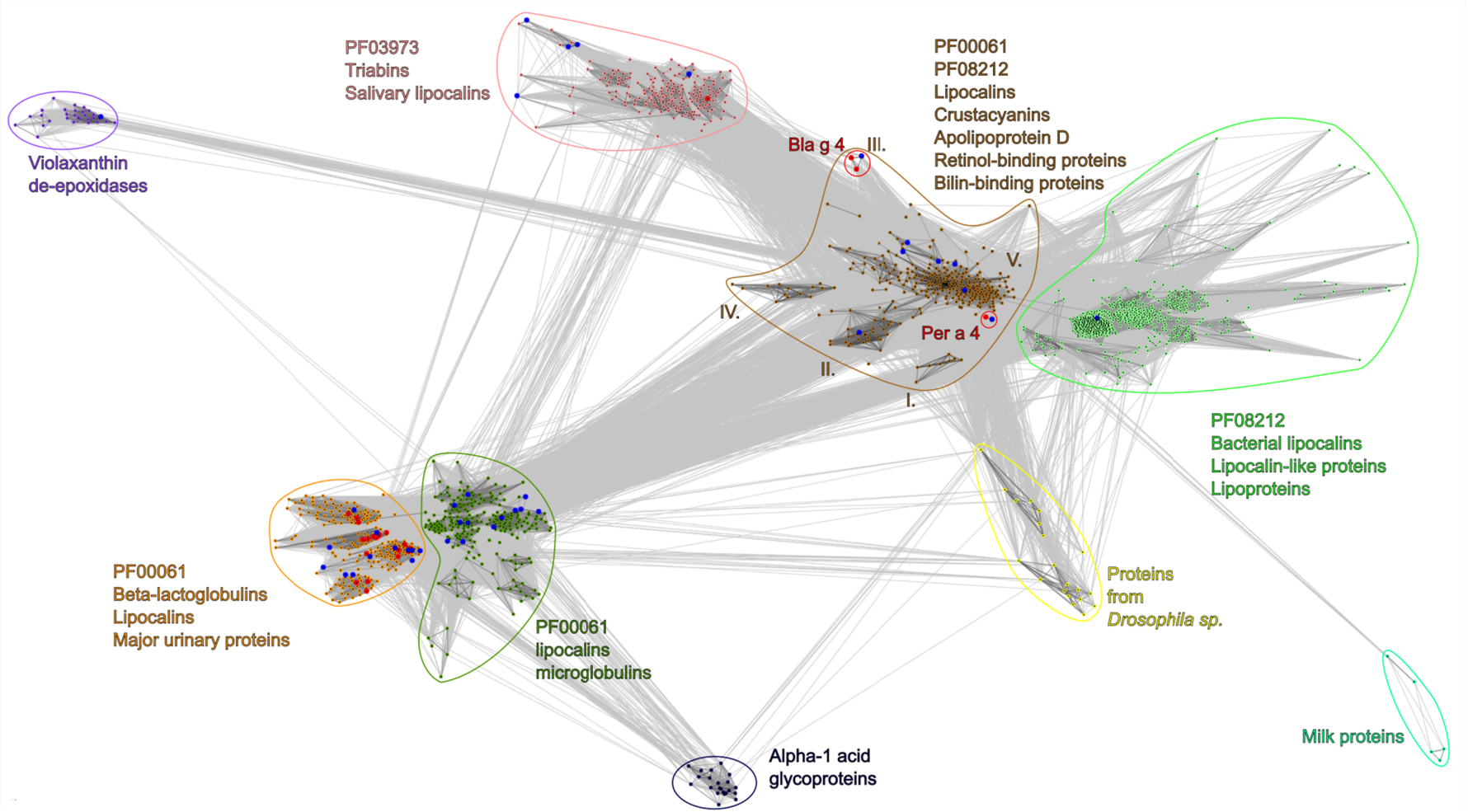

Figure 16 The two-dimensional projection of the sequence similarity based clustering.

The clustering was performed in CLANS (Frickey and Lupas, 2004). The proteins are indicated by dots. Lines indicate sequence similarity detectable with BLAST (Altschul et al., 1990) and are colored by a spectrum of grey shades according to the BLAST p value (black: p value <10−225, light grey: p value <10−5). Allergen sequences are marked as red dots. Sequences with known structure are marked as blue dots. Allergen sequences with known structure are marked as red dots with blue boarders. Bla g 4 and Per a 4 allergens protein sequences are in the red circles. (See the main text for a description of the colors.)

7.2.3. Evolutionary analysis

All allergen sequences were extracted from the dataset used for sequence analysis in this work. Redundancy was removed and extracted allergen sequences were used to compute a Maximum Likelihood phylogenetic tree (Figure 17). The similarity between the sequences is reflected by their branch length. The β-lactoglobulin protein family (Bos d 5, Bub b BLG, Ran t BLG, Cap h BLG, Ovi a BLG) is represented by the largest and the most closely related group. The level of similarity between the β-lactoglobulin proteins found in the dataset is so high that, it is almost impossible to distinguish between them on this tree and it is not necessary for this work. The horse β-lactoglobulins (Equ c BLG 1 and Equ c BLG 2), however, are placed further from the large β-lactoglobulin group. Two groups of allergens are equally distant from β-lactoglobulin group and from each other. The first one is the dog lipocalin allergen (Can f 2) and the second one consist of a small group containing lipocalins from human, dog, and cat (Hum s TL, Can f 1, Fel d 7). Another closely related and large group is formed by the major urinary proteins from rat (Rat n 1) and mouse (Mus m 1) and lipocalins from dog, cat, and horse (Can f 6, Fel d 4, Equ c 1). The group containing the major urinary protein group consists of the lipocalin allergens from guinea pig and bovine (Cav p 2, Cav p 3 and Bos d 2). The group with the shortest distance to the Bla g 4 is formed by the protein from the procalin protein family - Tria p 4 from California kissing bug (*Triatoma protracta*) and it is the closest homolog to both Bla g 4 and Per a 4.

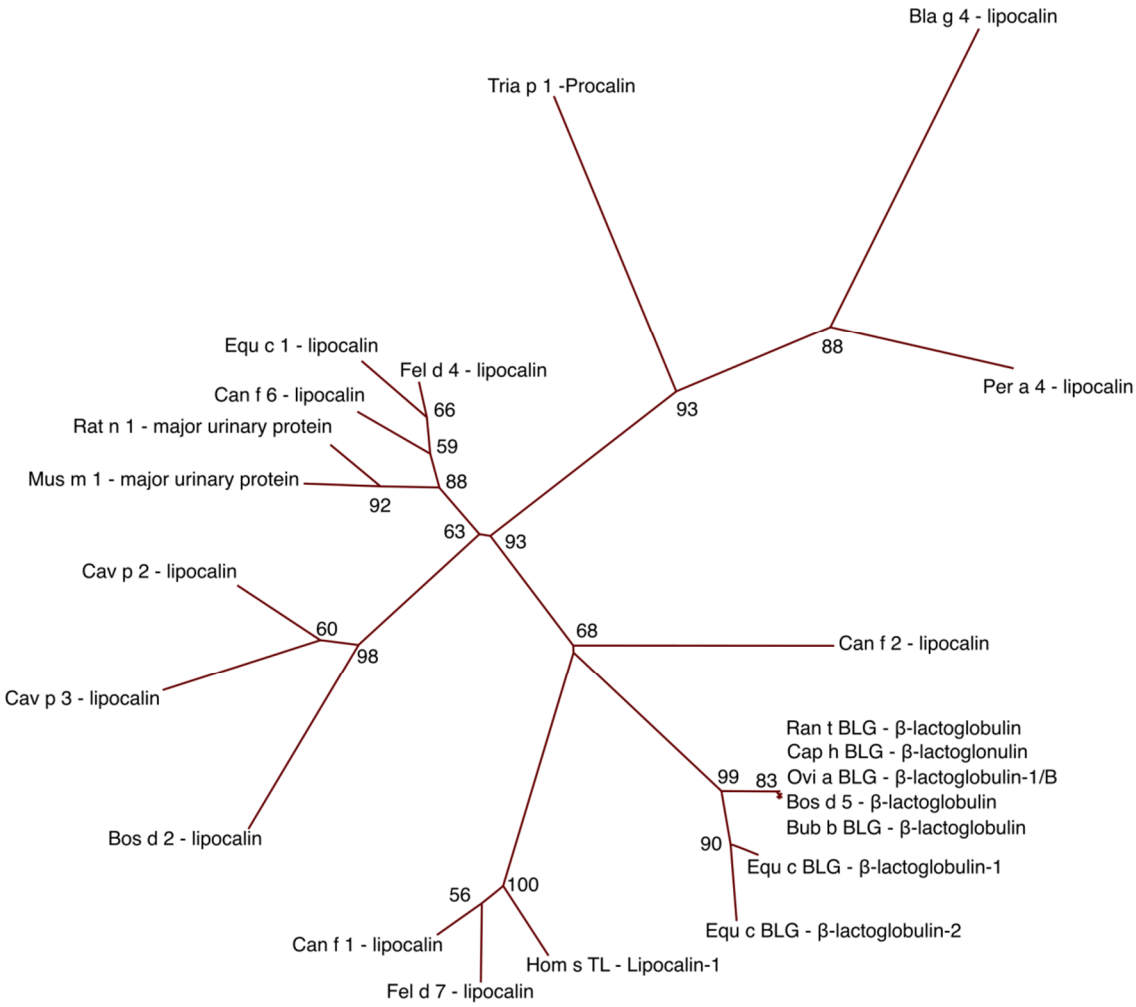

Figure 17 Maximum likelihood unrooted tree of Bla g 4 allergenic homologs.

The length of the branches (red) reflects the similarity between the sequences. The numbers near the nodes show the bootstrap values for that particular node. The bootstrap values below 50 are not shown.

The last group consist of the Bla g 4 and Per a 4 allergens, which are the closest homologs. However, there are several differences between them and Tria p 1, despite being the closest allergen to Bla g 4 and Per a 4. The sequence alignment of these allergens (Figure 18) shows not all amino acids involved in tyramine binding present in Bla g 4 are conserved, even in those that are the most closely related. Tria p 1 belongs to one of the lipocalin protein subfamily – procalins and Asp47 and Tyr85 (from Bla g 4) are substituted with Asn45 and Phe89, respectively. Since the data were not sufficient for preparing a reliable sequence alignment of the area surrounding Tyr122 in Bla g 4, it was omitted from this analysis.

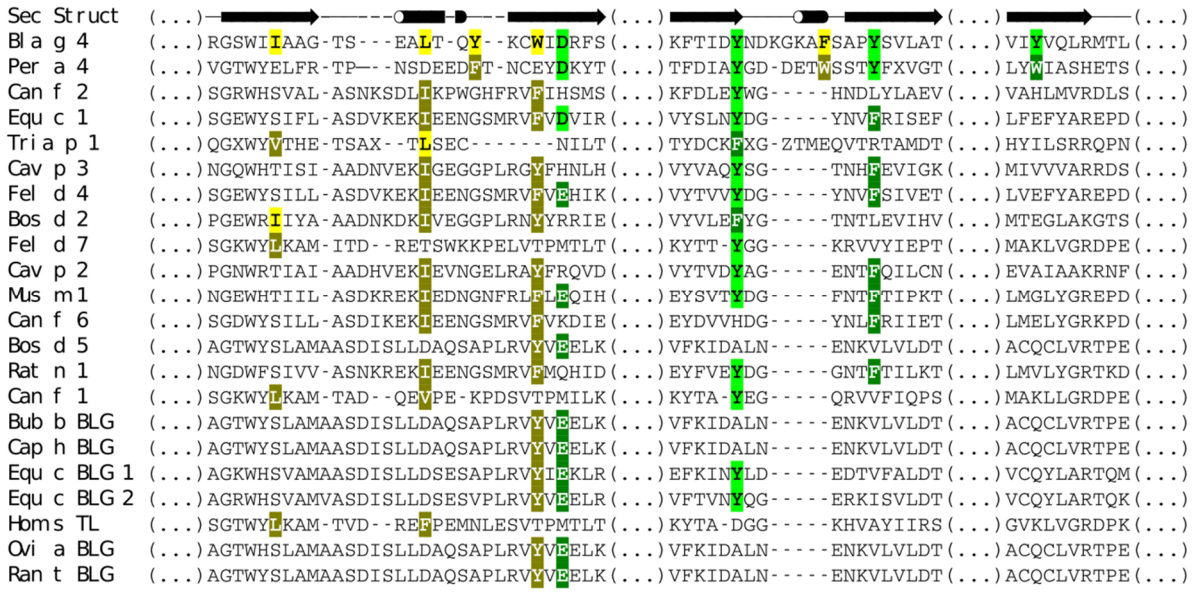

Figure 18 Alignment of allergens found in sequence dataset.

Amino acids involved in binding of tyramine are colored green (identical to amino acids in Bla g 4 are colored light green; similar are colored dark green). Amino acids involved in forming the hydrophobic pocket are colored yellow (identical to amino acids in Bla g 4 are colored yellow; similar are colored in dark yellow).

7.3. Alt a 1

7.3.1. Structural analysis

The Alt a 1 allergen crystallized in tetragonal system and $I4_122$ space group with one protein chain in the asymmetric unit. Residues 28-157 of the Alt a 1 sequence were fit into the electron density. The protein is formed by eleven β-strands that wrap around forming a β-barrel (Figure 19). Pro125 breaks the seventh strand into two fragments - β7a (residues 116-123) and β7b (residues 126-130). An eight strand is broken into three subunits by a β-bulge and Leu150 - β8a (residues 138-142), β8b (residues 146-149), and β8c (residues 151-153). Despite the fact that β8b and β8c are successive to each other, they belong to the different sheets forming the barrel (Chruszcz et al., 2012a).

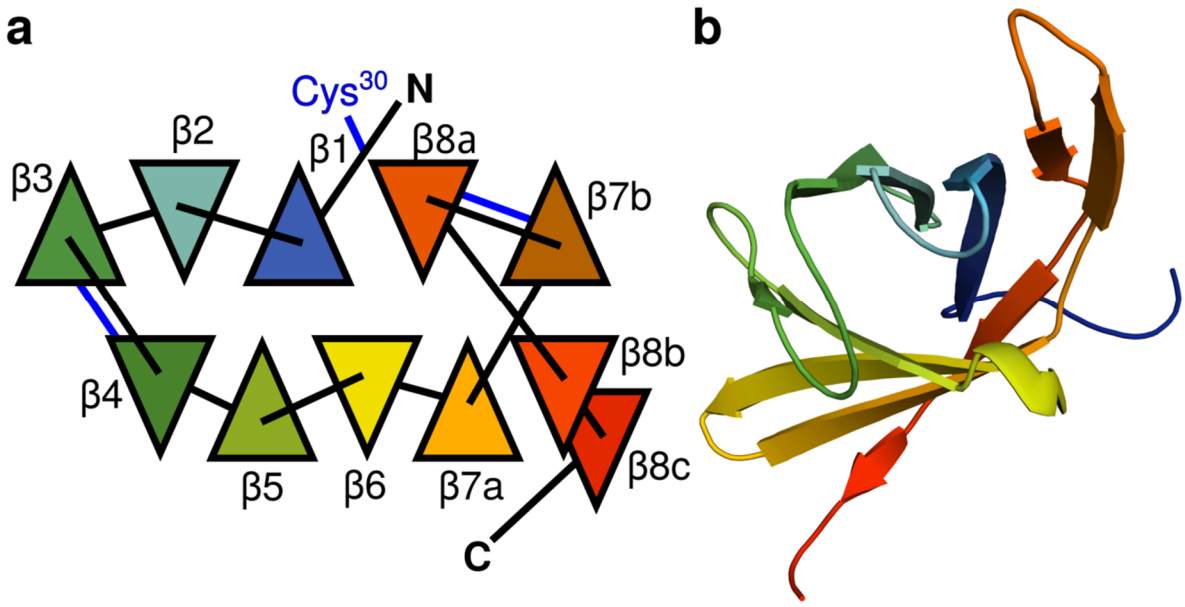

Figure 19 Overall structure of Alt a 1.

Two representations of Alt a 1: (a) the topology diagram and (b) cartoon. β-strands are labelled and colored in rainbow scale for both representations from N-terminus in blue to C-terminus in red. Disulfide bonds and Cys30 are represented as blue lines.

All of the five cysteine residues are involved in forming of the disulfide bridges. The bridge formed by Cys128-Cys140 joins two close β-strands - β7b and β8a. The strand β3 is linked to the β4 strand by the Cys74-Cys89 bridge. This bridge could be described as a clamp holding both barrel-forming β-sheets. Two symmetry related Alt a 1 chains are covalently linked with two Cys30, one coming from each chain. This stabilizes the dimer form of the protein (Figure 20), however, Cys30 is not conserved in all homologous sequences (Figure 23). The hydrophobic residues fill the center of the barrel and there is no inner cavity that could be used for ligand binding, as it is the case for lipocalins.

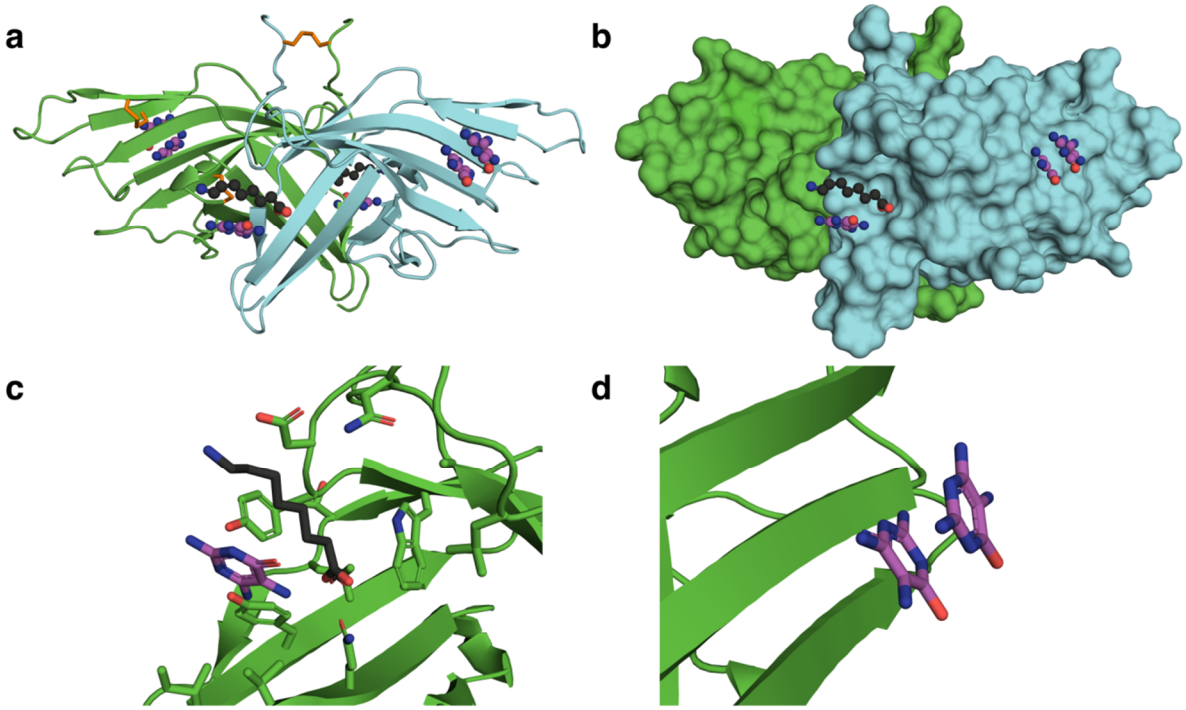

Figure 20 The Alt a 1 dimer and small molecular compounds.

The dimer is shown with molecular compounds (explanation in text) as a) cartoon and b) molecular surface. Alt a 1 molecules are colored green and cyan. Disulfide bridges are shown in orange. Disulfide bridge between Cys30 of two Alt a 1 monomers is orange. There are two locations of compounds in the Alt a 1 structure c) containing 4-HTP and 8-ACA and d) containing two copies of 4-HTP.

The dimer is stabilized by the disulfide bridge and by a mixture of hydrophobic and hydrophilic interactions. Both N-terminal and C-terminal regions of the protein participate in dimer formation (Figure 20). Several of the residues (Asp100, Arg103, Thr121, Thr123, Thr149, Leu150, and Thr152) form hydrogen bonds between the protein chains, forming the dimeric assembly. The dimer interface is large, having an area of 1345Å^2 (Chruszcz et al., 2012a).

The surface of Alt a 1 contains a substantial number of acidic residues (12 Asp and 5 Glu) relative to the number of basic residues (3 Arg, 9 Lys, and 1 His), resulting in a theoretical pI of 4.7 (for surface residues). The acidic amino acids are located on the surface of the Alt a 1 dimer (Figure 21), where they form three negatively charged patches. The sequence of the Alt a 1 allergen (theoretical pI = 4.75) contains more of basic amino acids (4 Arg, 9 Lys, and 1 His) than acidic residues (12 Asp and 6 Glu). Possibly the negatively charged patches take part in

the binding of the ligand and the dimerization process. However, there is nothing bound to the third patch, so its role is not clear. Moreover, the different structure of Alt a 1 deposited in PDB (PDB code: 4AUD) does not contain ligand.

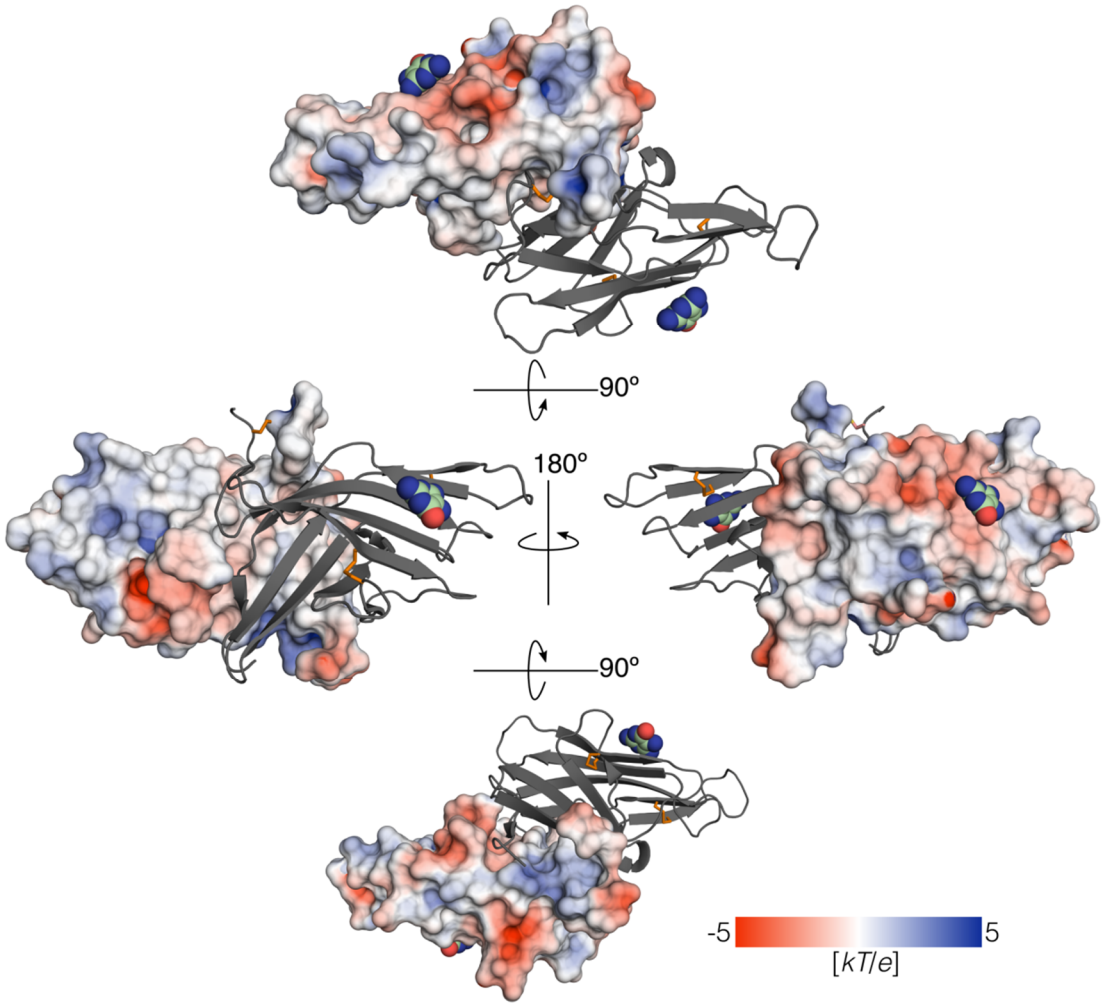

Figure 21 Surface of the Alt a 1 dimer with mapped electrostatic charges.

The dimer of Alt a 1 is shown from different angles with one of the monomers shown as semi-transparent cartoon to highlight the charges in the dimer interface area. The ligand bound (2,5,6-triaminopyrimidin-4-ol) is shown as spheres. The electrostatic charges are colored from red (negative charge), through white (no charge) to blue (positive charge).

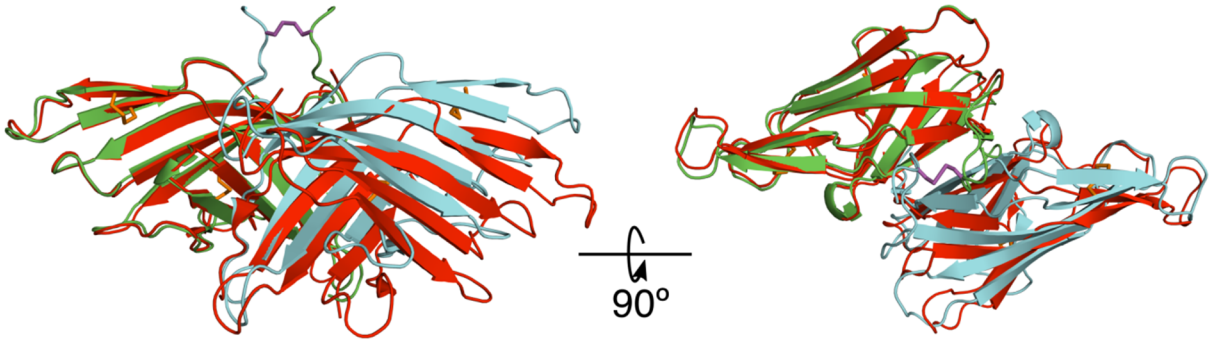

Figure 22 Superimposition of two dimers of Alt a 1.

First dimer is recreated from 3V0R (original chain is shown in green and recreated dimeric chain is shown in cyan), second dimer is 4AUD (shown in red). The dimers are superimposed with one of the Alt a 1 monomers from 3V0R structure (green).

Another structure of Alt a 1 (PDB code: 4AUD) is also dimeric (Figure 22). However, it does not have Cys^{30} in the structure, as it is missing from 4AUD sequence. Overall the dimer formed by Alt a 1 in 4AUD structure is almost identical to the other Alt a1 structure, with the exception of a few degrees difference of the angle formed between the two Alt a 1 chains in each structure. No other structures with a significant degree of similarity to Alt a 1 were identified by either DALI or FATCAT. The highest similarity structures identified by these programs yielded values of rmsd around 3 Å and mainly residues forming the β-barrel have been aligned. The aligned fragments had their sequence identity ranging from 5 to 12% in most cases.

7.3.2. Evolutionary and sequence analysis

All sequences obtained from exhaustive searches against the nr database (NCBI Resource Coordinators, 2013) contained only fungal proteins from Dothideomycetes (where *Alternaria alternata* species is classified) and Sordariomycetes classes that belong to Pezizomycotina subphylum. Both Dothideomycetes and Sordariomycetes are classified as Leotiomyceta (Schoch et al., 2009). The dataset was then prepared and narrowed to include several sequences of varying degree of similarity (Table XV) representing.

Organism	Identity (%)	Similarity (%)
Alternaria brassicicola (gi: 20279107)***	89.6	98.7
Crivellia papaveracea (gi: 307147764)***	82.1	95.9
Alternaria brassicae (gi: 49476533)***	81.8	95.6
Embellisia telluster (gi: 49476565)***	81.6	96.3
Pleospora tarda (gi: 74588233)***	81.0	94.2
Alternaria argyranthemi (gi: 49476475)***	80.6	93.5
Alternaria species CID76 (gi: 148792960)***	79.9	97.1
Embellisia proteae (gi: 209363449)***	70.1	92.7
Lewia species CID245 (gi: 148357931)***	66.9	84.5
Leptosphaeria maculans (gi: 312213528)***	63.2	85.8
Mycosphaerella graminicola (gi: 339470688)**	40.9	71.3
Pyrenophora teres f. teres (gi: 330921398)**	39.3	74.2
Myceliophthora thermophila (gi: 347012821)*	32.4	55.3
Verticillium dahliae (gi: 321271123)*	31.9	60.1
Phaeosphaeria nodorum (gi: 169594938)*	31.7	56.1
Mycosphaerella graminicola (gi: 339469016)**	31.2	58.1
Chaetomium globosum (gi: 116179392)*	31.1	60.5
Pyrenophora tritici-repentis (gi: 189202924)*	30.1	56.9
Magnaporthe oryzae (gi: 39970423)*	29.0	60.0
Glomerella graminicola (gi: 310797708)*	27.0	58.3
Glomerella graminicola (gi: 310798150)*	25.6	56.9
Gibberella zeae (gi: 46114386)*	24.7	55.1

Table XV Sequence identity and sequence similarity to Alt a 1 (gi: 14423645).

Sequences are sorted by identity. Names of the most similar sequences are marked with *** (they are shown in black on Figure 23 and Figure 24), sequences with lower degree of similarity are marked with ** (they are shown in grey on Figure 23 and Figure 24), and names of the least similar sequences are marked with * and are shown in brown on Figure 23 and Figure 24.

Cysteines corresponding to Cys30 of Alt a 1 is only conserved among the homologs closest to Alt a 1, while Cys74, Cys89, Cys128, and Cys140 are conserved among almost all sequences with the exception of one (gi: 307147764), which is of shorter length (Figure 23).

This is represented on a tree computed from the alignment of selected Alt a 1 homologs and a distinct division into three groups (Figure 24) can be observed.

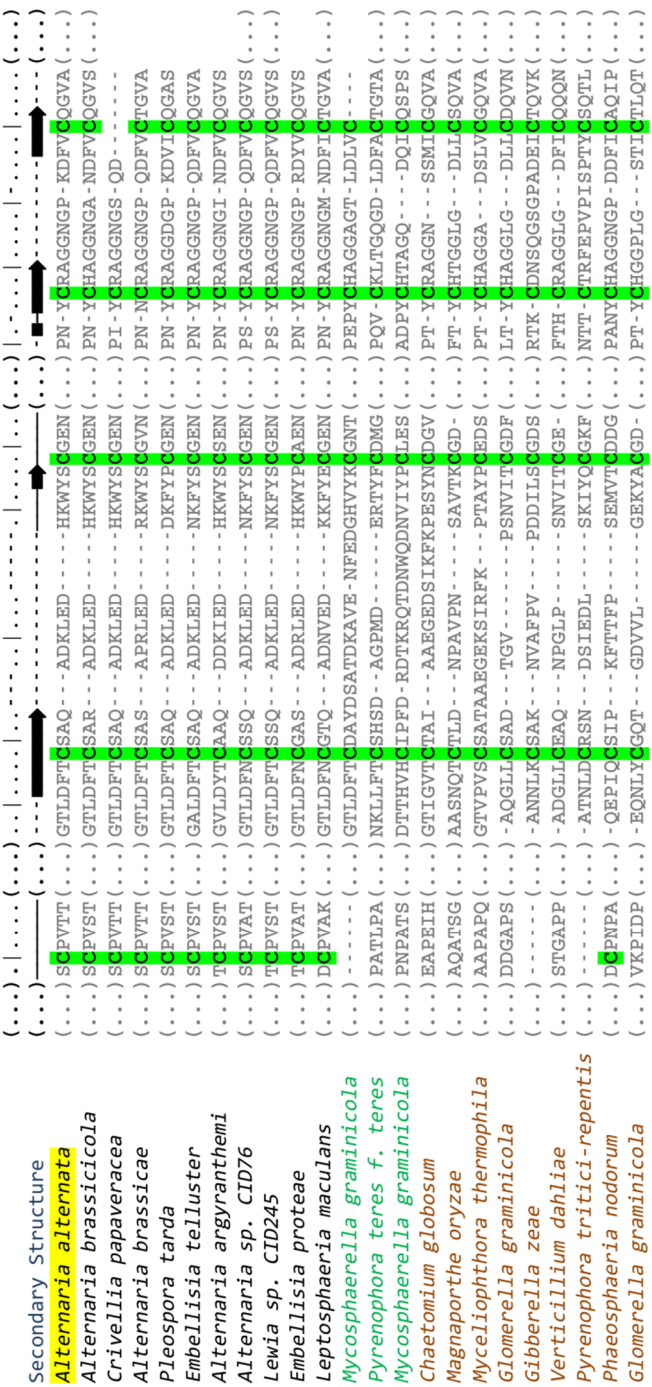

Figure 23 Sequence alignment of selected Alt a 1 homologs.

Secondary structure is on the top row and beta-sheets are marked with an arrow. The cysteine residues involved in disulfide bonds formation are highlighted in green. Names of the most similar sequences are shown in black, sequences with lower degree of similarity are shown in green, and names of the least similar sequences are shown in brown.

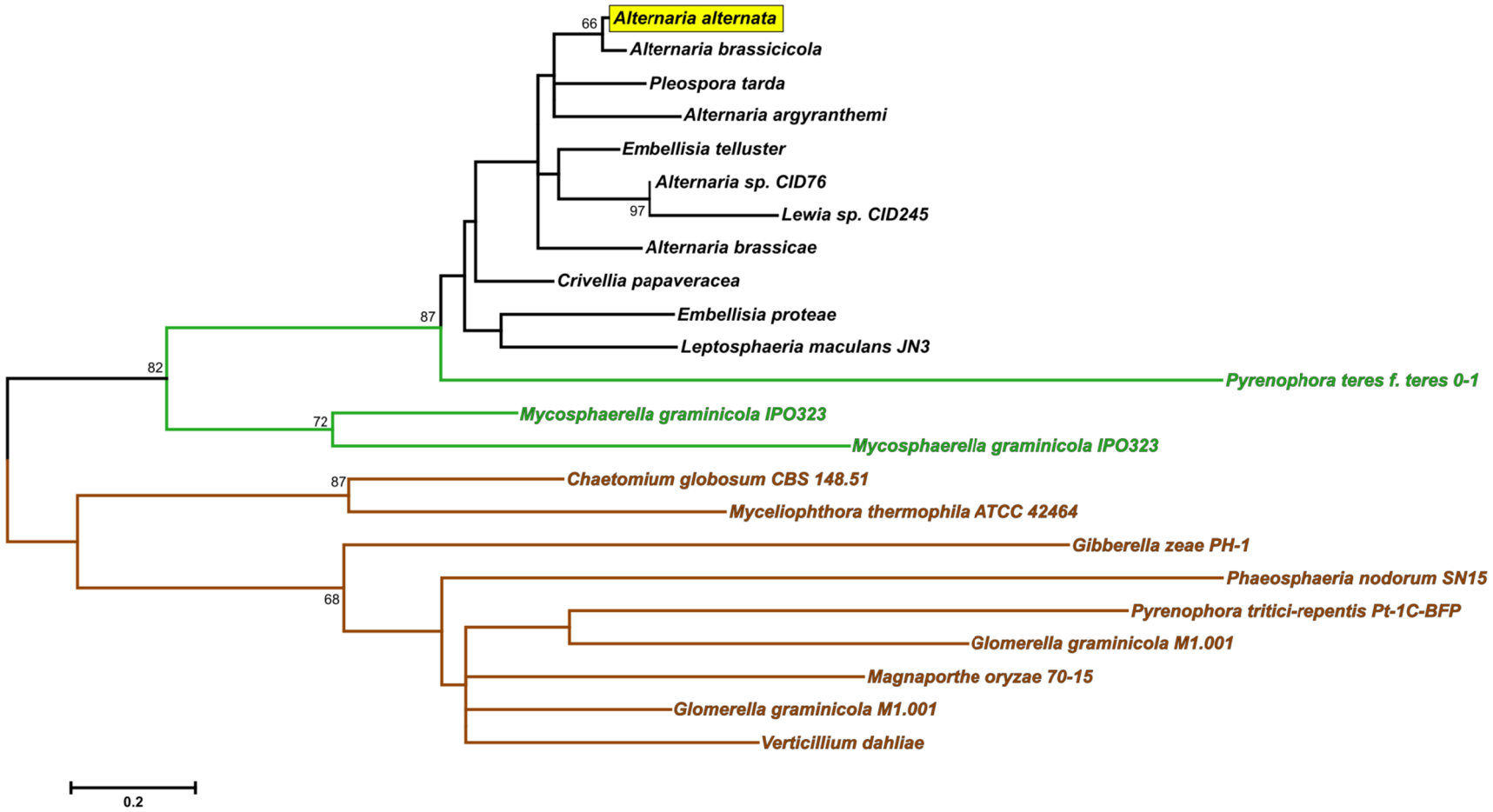

Figure 24 Maximum likelihood tree of *A. alternata* homologs.

The sequences are labeled with an organism name, a gi number, and a sequence length used for the analysis, respectively. The bootstrap values are near the nodes the values refer to. The bootstrap values below 50 are not shown. Branches are colored according to alignment shown in Figure 23. Branch lengths are drawn to scale with scale bar indicating the number of the substitutions of the amino acid residues per site.

The Pfam database (Punta et al., 2012) (version 28.0) classifies the Alt a 1 allergen in Pfam-A since June 2015 as the AltA1 family (PF16541). The AltA1 Pfam family contains 461 sequences from 301 species from the Fungi kingdom (mostly from *Alternaria, Ulocladium, Embellisia, Nymbya, Lewia* genera). The sequences of full length from AltA1 Pfam family were downloaded and aligned by MAFFT (Katoh and Standley, 2013) using the L-INS-I option. The alignment shows that four cysteines forming the two disulfide bridges (Cys74 with Cys89 and Cys128 with Cys140) are conserved in all sequences, but Cys30 is not included in this Pfam family and is not conserved.

The Allergome database (Mari et al., 2009) lists 174 Alt a 1-related allergens. Each of these comes from different fungal species and is available in UniProt database (The UniProt Consortium, 2014). The sequence alignment computed by MAFFT (Katoh and Standley, 2013) using the L-INS-I option to increase accuracy shows, that Cys30 taking part in dimerization is conserved in all but three of the sequences. In one case the sequence is missing the N-terminus (*Alternaria citri*; Alt it 1 allergen), in one case there is deletion in Cys30 area (*Alternaria nobilis*; Alt no 1 allergen) and in one case cysteine is substituted by arginine (*Ulocladium dauci*; Ulo da 1 allergen). Cysteines forming the first of two disulfide bridges, Cys74 and Cys89, are conserved in all sequences. The second disulfide bridge formed by Cys128 and Cys140 is conserved in almost all sequences, but eight, which do not have full length. The sequence from *Alternaria ascaloniae* (Alt as 1 allergen), does not have either Cys128 or Cys140, resulting from a truncation. The other seven sequences (*Alternaria iranica* – Alt ir 1; *Alternaria citri* – Alt it 1; *Alternaria porri* – Alt po 1; *Alternaria prasonis* – Alt ra 1; *Alternaria vanuatuensis* – Alt vn 1; *Alternaria penicillata* – Cri pa 1; *Sinomyces fusoideus* – Sin fu 1) has only Cys140.

The sequence conservation among the whole AltA1 protein family is relatively high with the structure mostly invariant with the exception of β7a, β8a and β8c β-strands and some loops (Thr33-Asp37, Gln77-Leu81, Ala130-Phe138, and Val112-Ile116) being not well

conserved. The conserved patches on the surface of Alt a 1 have been identified (Figure 25) and they are possible targets for the design of the cross-reactive antibodies directed against the many members of the protein family to which Alt a 1 belongs. Despite the conservation of Cys74, Cys89, Cys128 and Cys140 among the whole group, the amino acids involved in forming the dimerization interface (Ile148-Ser157) are not well conserved. The amino acids present in the previously described IgE epitopes (Kurup et al., 2003) are well conserved for the peptides (Lys41-Pro50 and Tyr54-Lys63) that caused strong reaction. The amino acids present in the peptides (Tyr87-Asp96 and Val119-Cys128) that showed weak IgE binding are less conserved among the AltA1 pfam protein family.

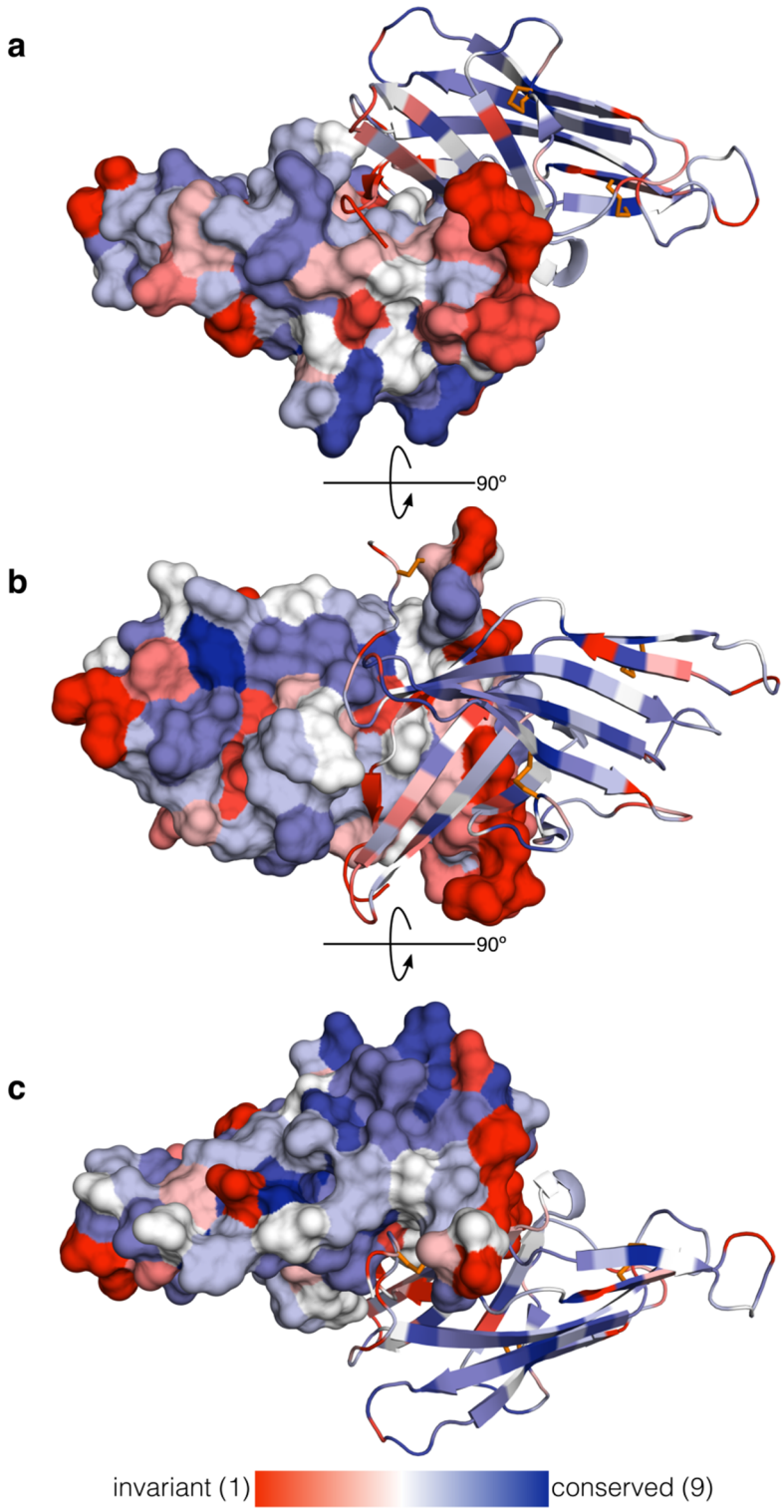

Figure 25 Sequence conservation among close Alt a 1 homologs mapped on Alt a 1.

Each monomer of Alt a 1 dimer is in different representation – first is shown as surface, second as cartoon. Color scale, based on Consurf (Ashkenazy et al., 2010; Celniker et al., 2013) scores, ranges from red (1) for residues with no sequence conservation to blue (9) for conserved residues.

8. Discussion

8.1. Der p 1

The cross-reacting epitopes of Der p 1 and Der f 1 are targeted by most human IgE antibodies (>80%) (Arruda et al., 1991; Heymann et al., 1986). Thus, most species-specific mAb developed against Der p 1 did not inhibit IgE binding to this allergen. The species-specific monoclonal antibodies 10B9 and 5H8, as well as the cross-reactive mAb 4C1, partially inhibited IgE antibody binding to Der p 1 (Chapman and Platts-Mills, 1980; Thomas et al., 2010). The three epitopes of these mAbs, identified on Der p 1 by X-ray crystallography, provide the basis for ongoing analysis of antigenic determinants involved in IgE antibody binding (Chapman et al., 1987, 1984; Heymann et al., 1986).

The unique alpha-helical loop in CDR-H3 of 10B9 antibody, which fits into a cavity on the epitope on Der p 1's surface, is formed by a conformational change that occurs upon binding to Der p 1. An alpha-helix is found in the CDR-H3 region of antibodies in less than twenty out of approximately one thousand of monoclonal antibody structures reported in the PDB as of 2014. This form of secondary structure in CDR-H3 is very rare, and its role has been described in only a few cases as inhibitory (it is mimicking the structure of other protein and competes for binding site) (Dong et al., 2010), as affecting antigen recognition (Blech et al., 2012; Bulek et al., 2012), or it can form a tentacle which provides a large interaction surface for antigen binding (Blech et al., 2012). In almost all other structures of Fab fragments of antibodies, the CDR-H3 forms a random coil loop without distinctive characteristics. The CDR-H3 loop of 10B9 can assume multiple functions that can be explained as affecting recognition of Der p 1; providing the structural backbone for the formation of the Der p 1-10B9 interface; stabilization of Der p 1-10B9 interface by increasing the size of the paratope area and participating in forming hydrogen bonds. The interaction between 10B9 and Der p 1 is an example of the "induced fit" model of protein interactions rather than "lock-and-key". Considering the

"lock-and-key" interaction of Der p 1 with 4C1 antibody described previously (Chruszcz et al., 2012b), this study shows that both types of antigen-antibody interactions can occur even in a similar area of the antigen surface.

The Der p 1 epitope that binds the 5H8 antibody may be affected by sequence polymorphisms in a few sites (Ala108 and Ile113) (Piboonpocanun et al., 2006). However, the structural analysis revealed that these amino acids are involved in the interface creation by their mainchain atoms and their variance most likely does not interfere with 5H8 binding. Both 10B9 and 4C1 binding epitopes on Der p 1 do not contain residues associated with sequence polymorphisms (Piboonpocanun et al., 2006). A comparison of the mature Der p 1-antibody complex structures with the structure of pro-rDer p 1 clearly indicates that the presence of the pro-peptide does not interfere with the binding of 4C1, 10B9 and 5H8 Fab fragments to the allergen.

The structural studies of Der p 1 and Der f 1 allergens allowed us to directly compare the epitopes that bind the Fab fragments of mAb 4C1, 10B9 or 5H8. Such a comparison provides the detailed insight, at the molecular level, about which of the characteristics of these antibodies surfaces are responsible for the fact that 10B9 and 5H8 antibodies are Der p 1 specific, while 4C1 is cross-reactive and binds both Der f 1 and Der p 1 (Chruszcz et al., 2012b). The results presented here indicate which of the Der p 1 amino acids involved in binding the 10B9 and 5H8 antibodies are different between Der p 1 and Der f 1, rendering the latter impossible to bind 10B9 and 5H8. The residues that were different between the 10B9 and 5H8 binding epitopes on Der p 1, and the corresponding surface area of Der f 1, are mostly charged and polar. However, there are a few substitutions of the non-polar amino acids like alanine to small polar amino acids like threonine and serine, which may affect their attractions with the amino acids found in the paratopes of the 10B9 and 5H8 antibodies, and therefore contribute to the species specificity. In Der f 1, the number of amino acids that differ from the corresponding residues

that form hydrogen bonds between 10B9 or 5H8 and Der p 1 is significant. These substitutions apparently contribute to the lack of binding of these two antibodies to the Der f 1 allergen, despite a high sequence identity (~75%) and similarity (~83%) between these two allergens at surface level. These results indicate that the relatively small differences in the allergen surface may impair the antibody binding. For example, such changes may involve the presence of a few non-conserved residues, which were observed in the case of the Der f 1 region that corresponds to the 5H8 epitope in Der p 1. However, these differences may be more subtle as it occurs on the surface of Der f 1, in the region corresponding to the 10B9 epitope in Der p 1. In this case, Der f 1 has amino acids with smaller sidechains than the corresponding amino acids in Der p 1, and therefore it is unable to form as many binding interactions with 10B9 as Der p 1. In addition, to the "missing" interactions, the conformation of the 180-184 fragment (Ser-Thr-Gln-Gly-Asp) in Der f 1 is not compatible with the CDR on 10B9, as the main chain of Gln182 generates a steric clash between the allergen and the antibody.

The amino acid composition of the Der p 1 epitopes for the 4C1, 10B9 and 5H8 antibodies demonstrates that in general there is little preference of the particular amino acids; however, all the polar or charged amino acids are much more frequent than the non-polar ones. The surface of the paratope is built from the polar amino acids like tyrosine or serine, possibly to increase the steric compatibility and to compensate for the epitope's charge. In the contrast to 10B9 and 5H8, the paratope of the 4C1 antibody has a very high arginine content which, by adding a positive charge to the Der p 1 binding paratope, may partially explain its cross-reactivity.

The analysis of the Der p 1 and Der f 1 allergen-antibody complexes prompted to investigate all the protein-antibody and the peptide-antibody complexes that have their structures determined. A methodology that allows analysis of the properties of the epitopes and the paratopes with regards to the relative frequency of the amino acids, the areas of the

interactions for the particular amino acids, and the hydrogen bonding interactions has been developed. This methodology allows for the generation of several types of the plots highlighting the diverse aspects of the epitopes and the paratopes characteristics. The analysis of the amino acid composition of all of the antigen-antibody interfaces in the dataset used versus interfaces formed by Der p 1 with 10B9, 5H8 and 4C1 (Table IV, Table V, Table VIII, Table IX, Table XI, Table XII, Table XIV and Figure 11) indicated that the amino acid composition of the Der p 1 epitopes is similar to them. None of the amino acids are significantly overrepresented in the epitopes. However, the amino acid composition of the epitopes shows some bias towards the polar (with the exception of serine and threonine) and the charged amino acids. Such bias is observed for both the proteins and the short peptides. The observed frequency of serine and threonine is half of the expected value, which may be caused by the size of their side chains and a significant amount of tyrosines in the paratopes. One could expect such results for the proteins, taking into account that the residues forming the epitopes are on their surface, but this trend is not so obvious for the peptide-antibodies complexes. The amino acid composition of the paratopes is highly biased toward the polar (except of serine and threonine) and the charged amino acids with a significant absence of the non-polar amino acids. It is also striking that tryptophan and especially tyrosine are significantly overrepresented in the CDRs of the antibodies. The special role of tyrosine was already noticed (Koide and Sidhu, 2009) and most likely could be explained by the chemical nature of this amino acid residue that allows it to participate in many types of the bonding interactions.

The analysis of the antigen-antibody interfaces revealed a preference for the certain amino acids that are involved in forming the hydrogen bonds between the antibodies and the antigens. The amino acids that are the most involved in the hydrogen bond formation on the surface of the paratopes are tyrosine (22%), serine (14%), aspartate (13%), arginine (10%), and asparagine (9%). On the surface of the epitopes, the most hydrogen bonds are formed by

arginine (15%), glutamate (11%), aspartate (10%), lysine (10%), serine (9%), and asparagine or glutamine (both are 8% each). A closer look at the surface area of the antigen-antibody interfaces indicates that the number of amino acids is lower for the antigens and higher for the antibodies. The relative area contributed to the interfaces by particular amino acids, however, is quite opposite, higher for the former and lower for the latter. This is consistent with the fact that, in general, antibodies must have the possibility to evolve and quickly adjust to recognize a wide variety of antigens (clonal selection and affinity maturation). Having several residues in the binding site provides a greater flexibility to precisely recognize the different epitopes. It is a nature of the immune response and such variability is not needed at the level of the epitopes as it is not their role to try and bind to an antibody, but it is the role of the antibody to bind to an antigen. Although this might not have the desired response in the case of allergens, the antibody is doing what it is supposed to. It may not even be desired in the design of the molecules that no longer bind the antibody, as the mutation of just a few residues that contribute a large surface area, might be enough to prevent an undesired recognition by the antibodies and lead to the design of a molecule usable for immunotherapy.

There are various methods that currently exist for the prediction of the linear (Hopp and Woods, 1981; Parker et al., 1986; Welling et al., 1985) and the conformational (Kulkarni-Kale et al., 2005; Lo et al., 2013; Negi and Braun, 2009; Zhang et al., 2012) epitopes. The linear epitope prediction algorithms do not provide enough value because most of the epitopes in the globular proteins are discontinuous or non-linear and brought together during protein folding (Barlow et al., 1986; Thornton et al., 1986). The algorithms used in these methods consider various properties of the epitopes like the hydrophilicity scale, the protein secondary structure, the protein flexibility, but their accuracy was claimed by the authors at usually around 80%. There is no the epitope prediction algorithm that takes into account the surface contributed to the interface area by the particular amino acids. The method describing the contribution of the

surface area and the preferences in forming hydrogen bonds between the antigens and the antibodies may be used alongside the sequences and the structures as an element of creating a new epitope prediction tools with the very high prediction accuracy.

8.2. Bla g 4

The Bla g 4 allergen is a member of a lipocalin protein family, and similar to other members of this family binds small molecule ligands. A re-examination of two crystal forms of Bla g 4 resulted in the opening of new possibilities to analyze the function of this allergen. The crystallization of Bla g 4 in the apo and native forms can be attributed to different protein production protocols. The high resolution of the structure allowed the examination of the electron density and indicated that it was tyramine that was bound to Bla g 4 as a ligand. The tyramine binding was later verified through both NMR titration and isothermal titration calorimetry experiments (Offermann et al., 2014). In addition, both these techniques indicated that tyramine and octopamine (Figure 26) are bound by the allergen in a similar manner (Offermann et al., 2014).

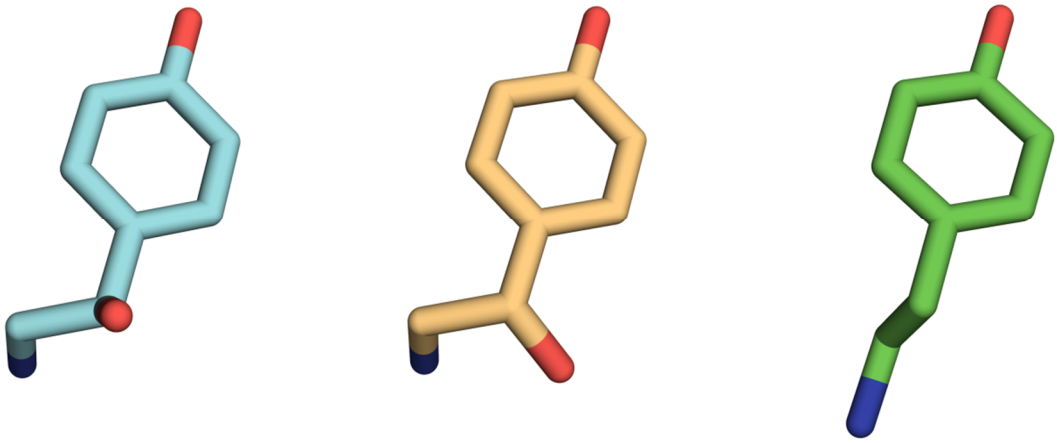

Figure 26 R-Octopamine (cyan), S-Octopamine (orange) and Tyramine (green). Hydroxyl groups are red, amine groups are blue.

The interaction of Bla g 4 and ligand is of dual nature and consist of both hydrogen bonds and hydrophobic interactions. The hydroxyl group of Tyr85 and carboxylate of Asp47 form hydrogen bond with the amine group of tyramine. The third hydrogen bond links the amine group to a water molecule bound to Tyr122. Bla g 4 does not form any direct hydrogen bonds with the hydroxyl group of tyramine, but does interact with the water molecules that fill part of the ligand binding cavity. All water molecules that are in the vicinity of the tyramine's hydroxyl group form hydrogen bonds with other waters or main chain atoms of the protein. The sidechains of Ile31, Leu39, Trp45, and Tyr122 form a hydrophobic cage that surrounds central part of the ligand. The binding site is closed by a cap formed by Tyr42 and Phe92. The binding site in the structure of Bla g 4 without the ligand is filled with approximately 12 water molecules, some of which are displaced upon the ligand binding. The positions of the water molecules that act as bridges between Bla g 4 and tyramine are not substantially changed. The conformational change of the residues that forming the hydrophobic cage is minimal with Ile31 and Asp47 having the small changes in the conformation of the side chains (Offermann et al., 2014)

Identification of structures similar to Bla g 4 was performed with DALI (Holm and Rosenström, 2010) and FATCAT (Ye and Godzik, 2004). The structure of the Per a 4 allergen was found to be the most similar (~2 Å rmsd) (Offermann et al., 2014). At the value of around 2 Å rmsd searches with Bla g 4 structure as a query, most of the identified proteins belong to the lipocalin family, thus a substantial number of them have structures complexed with small molecules, like for example biogenic amine-binding protein from saliva of *Rhodnius prolixus* (PDB code: 4GE1) (Xu et al., 2013). None of the ligands complexed with the identified proteins are bound in the same manner as tyramine in Bla g 4. Tyramine is identified as a ligand in only two structures in PDB (as of July 2015) with Bla g 4 being one (PDB code 4N7C) and Alzheimer's disease drug target BACE-1 (PDB code: 3BRA) being the other (Kuglstatter et al.,

2008). BACE-1 protein, however, binds tyramine in a vastly different mode than Bla g 4. Octopamine can be found in three structures in the PDB (July 2015). Goat lactoperoxidase is among the structures in PDB that bind octopamine binds the R-octopamine enantiomer (PDB code: 4QJQ) and the binding mode resembles Bla g 4, with octopamine buried inside the binding cavity and interacting via hydrogen bonds, hydrophobic forces, and some of the interactions are being carried by water molecules. Another protein that binds octopamine is human phenylethanolamine N-methyltransfrase (hPNMT) (PDB codes: 3HCA, 3HCE and 2AN4). PNMT is an adrenaline-synthesizing enzyme and it also binds octopamine in similar mode to Bla g 4 – the ligand is buried in the protein, and it interacts with the macromolecule through both hydrogen bonds and hydrophobic interactions. Some interactions also involve bridging water molecules. Human tyrosyl-DNA phosphodiesterase (Tdp1) is the remaining protein that binds octopamine. Both enantiomers of octopamine (PDB codes: 1RG2 and 1RGT) or S-octopamine only (PDB codes: 1RGU and 1RH0). Octopamine binding, however, is bridged by vanadate in the Tdp1 structures and it is different than in Bla g 4 (Offermann et al., 2014).

A comparison of the Bla g 4 and Per a 4 binding sites revealed partial conservation of the residues that form the hydrophobic cage (Figure 27). Asp47 and Tyr85, which participate in hydrogen bonding with the amide moiety of tyramine are among those conserved residues (Figure 14). However, some of the residues in Bla g 4 that directly interact with the ligand are not conserved in Per a 4 (Offermann et al., 2014).

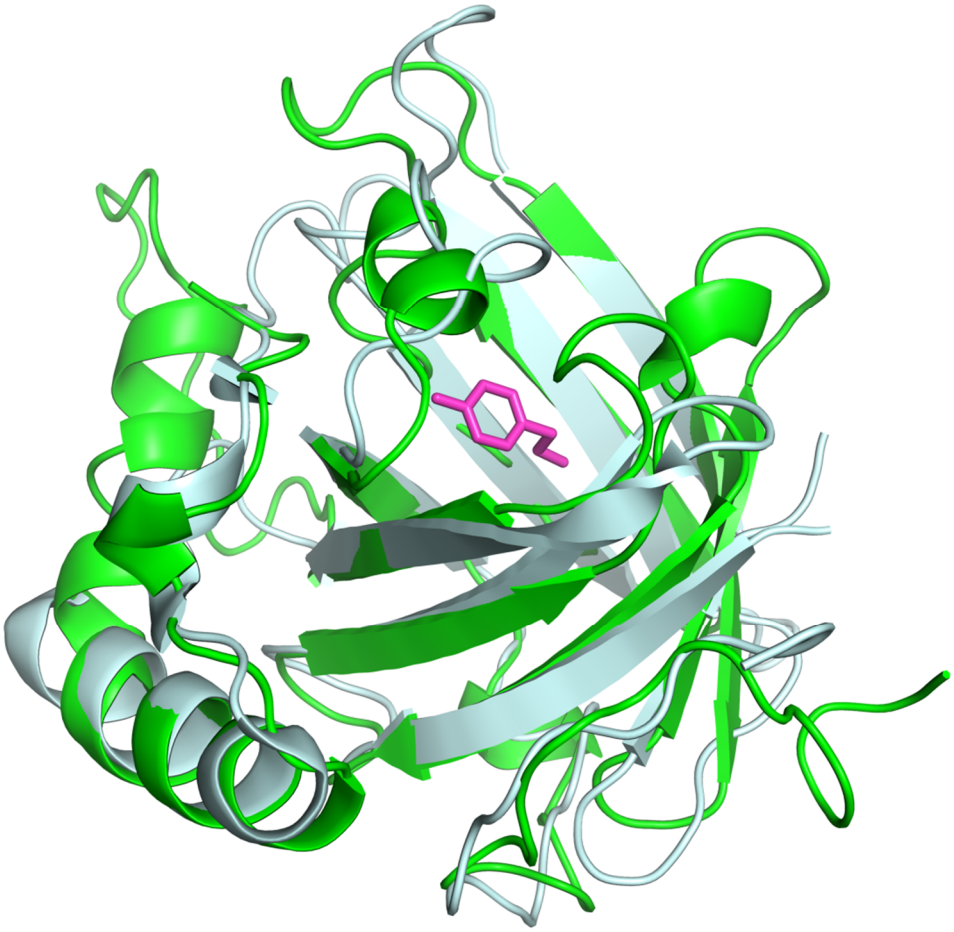

Figure 27 Superimposition of Bla g 4 (green) and Per a 4 (cyan) with0 tyramine (purple). Both structures are in the cartoon representation. PDB codes: 4N7C (Bla g 4), 3EBW (Per a 4).

The simple modelling of binding different ligands is shown in Figure 28 and the results (Offermann et al., 2014) suggest that both enantiomers of octopamine can fit into the binding cavity in Bla g 4. The results of NMR and ITC experiments support this proposition and suggest that Bla g 4 binds octopamine in the same manner as tyramine with the exception is Tyr96, which may form a hydrogen bond with the hydroxyl group of the octopamine. The difference between binding of the enantiomers of octopamine is that the hydrogen bond formed between O7 of R-octopamine and hydroxyl group of Tyr96 may be stronger that analogous bond formed by S-octopamine, as the length of the bond is shorter for R-octopamine (2.7 Å) than for S-octopamine (3.3 Å). The distance from N8 of both enantiomers of octopamine to the hydroxyl group of Tyr85 is 2.9 Å (Offermann et al., 2014).

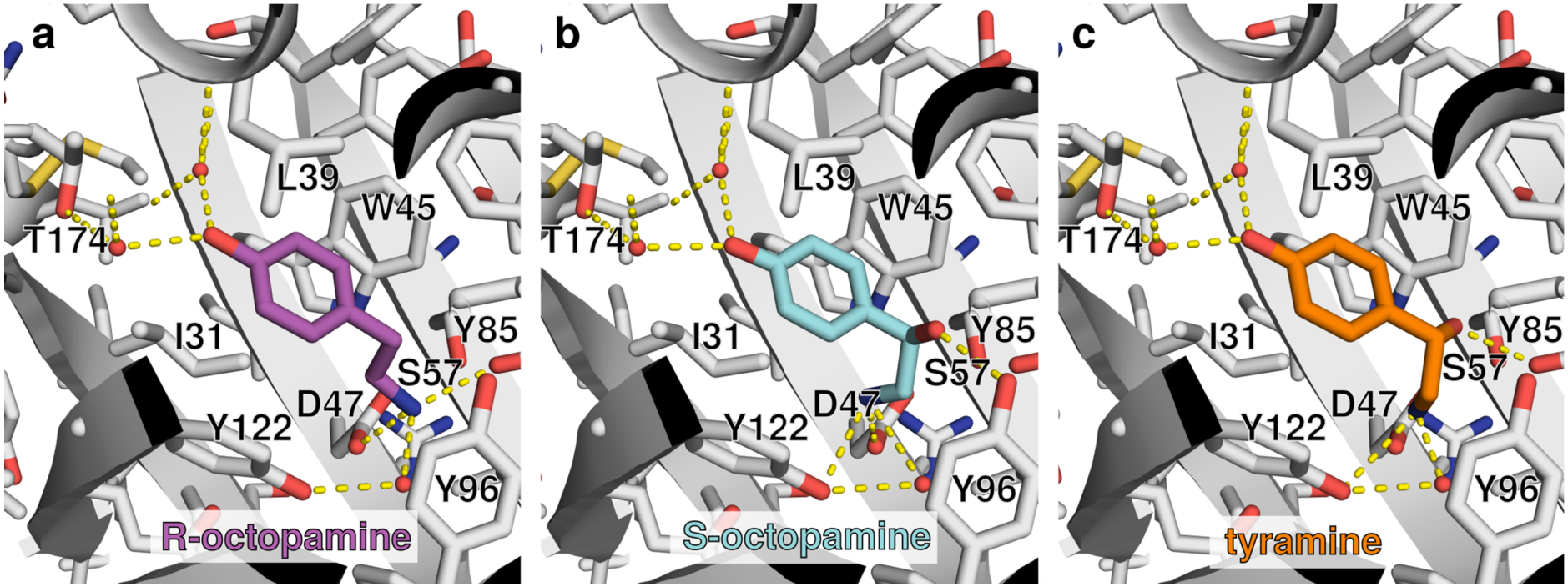

Figure 28 Ligands bound by the Bla g 4 allergen.

Tyramine (a) and two enantiomers - R-octopamine (b; shown in cyan) and S-octopamine (c; shown in orange). The results described in this figure were taken from Offermann et. al. 2014 (Offermann et al., 2014) which I am coauthoring.

Bla g 4 and Per a 4 are cross-reactive despite the fact that their sequence identity is very low (21%). One explanation for this phenomenon assumes that a common IgE epitope exists for both proteins. In order to investigate the existence of such an epitope, structures of both allergens were superposed and continuous surface patches containing identical and similar amino acids were identified (Figure 29). In addition the previously reported studies of Bla g 4 and Per a 4 mutants were used to determine whether some of the conserved regions may at least partially overlap with IgE epitopes (Yih et al., 2009).

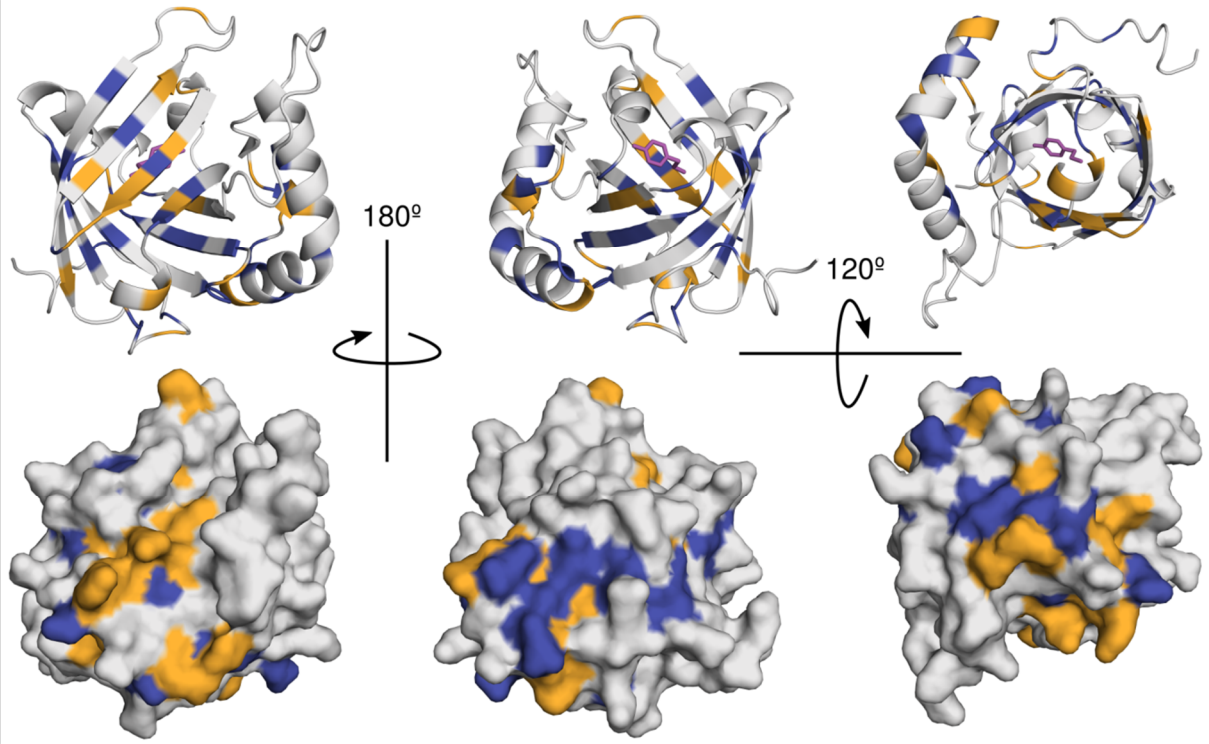

Figure 29 Similarity between the Bla g 4 and Per a 4 allergens.

Bla g 4 position is identical in columns. The upper row show Bla g 4 in cartoon representation, in the lower row Bla g 4 is shown as surface. Residues that are identical in Bla g 4 and Per a 4 are highlighted in blue while similar residues are highlighted in orange.

The comparison of the Bla g 4 and Per a 4 structures reveal the presence of three surface patches that contain residues that are conserved in terms of sequence and structure. The first of these patches (P1) is composed of Gly27, Ser28, Arg48, Phe49, Leu55, Val56, and Gly72 (sequence numbering for Bla g 4; identical residues are in bold). The second patch (P2) contains

residues Asp21, Arg24, Thr101, Asp102, Tyr103, Glu104, and Tyr106, while the third patch (P3) is composed of Thr35, Lys154, Glu158, Glu159, Leu166, Glu169, and Asp170. Solvent-accessible surface areas for these patches are 270 Å2, 360 Å2, and 530 Å2, respectively. These three regions are potential candidates for being conformational IgE epitopes for both allergens. A large portion of the second region (P2) is composed of residues identical or similar between Per a 4 and Bla g 4. The patch P2 can be considered as a potential cross-reactive antibody binding site or a part of a larger IgE-binding epitope. The sequence polymorphism identified in Bla g 4 (Jeong et al., 2009, 2008) in patches P1 (Ser28) and P2 (Arg24) contributed to increased similarity between Bla g 4 and Per a 4 in corresponding areas. Three linear regions containing antibody binding sites were identified by studying IgE binding with five overlapping peptide fragments (E1: aa1-100, E2: aa 34-77, E3: aa 74-117, E4: aa 114-156, E5: aa 153-182) (Shin et al., 2009). Two of the peptides (E1 and E2) correspond to patches P1 and P2. The major IgE-binding region is located at residues 118-152 and it is not overlapping with either of patches similar between Bla g 4 and Per a 4. The result of analysis of IgE binding the residues forming the patch P2 suggests in some cases their role in cross-reactivity. The role of the residues forming patch P3 in cross-reactivity between Bla g 4 and Per a 4 have not yet been analyzed.

The comparison of Bla g 4 to its closest documented allergenic homolog, Per a 4, suggests that the ligands bound by them are most likely different ligands or have different modes for tyramine or octopamine binding. The amine group of tyramine forms hydrogen bonds with two of Bla g 4 residues - Asp47 and Tyr85. Moreover, these residues are also conserved in Per a 4. This fact suggests the possibility of the presence of an amine group in the physiological ligand bound by Per a 4 and its overall tyramine similarity. The binding of octopamine by Bla g 4 most probably involves Tyr96, which is also conserved in Per a 4.

Several studies regarding the role of tyramine or octopamine in reproduction of insects have been conducted (Lange, 2009). Tyramine was found to be present in oviducts (Donini and Lange, 2004) and spermatheca of locust (da Silva and Lange, 2008; Lange, 2009). Tyramine also suppresses pheromone production after mating in *Bombyx mori* (Hirashima et al., 2007), and octopamine had no effect on this process. It is highly possible and in consensus with previous findings that tyramine or octopamine or one of them is the physiological ligand of Bla g 4.

8.3. Alt a 1

Alt a 1 and its homologs are unique to Fungi and characteristic for mostly Dothideomycetes and Sordariomycetes classes. The uniqueness of Alt a 1 is not only limited to its sequence, but the structure it forms does not have homologs in the PDB (July 2015) and Pfam protein family was named after it. The intramolecular pattern of the disulfides bonds are observed in all known homologs of Alt a 1 and presented in the dataset of selected representatives in the Figure 23.

It was observed that the Alt a 1 allergen is localized exclusively in the cell wall of *Alternaria* spores, but not in the cytoplasm (Twaroch et al., 2012). The airborne fungal spores are extremely resistant against antagonistic conditions. The IgE epitopes of Alt a 1 can resist heat treatment of 95°C (Twaroch et al., 2012). This observation is consistent with the previous report speculating that Alt a 1 is involved in spore germination (Mitakakis et al., 2001). The presence of disulfide bridges increase the stability of Alt a 1 and its homologs. The biological function of Alt a 1 is still unknown, but elucidation of its structure is a great step towards identifying it. The conservation of Cys30 among the closest Alt a 1 homologs strongly suggests that they also form dimers which are linked by a covalent bond. The second structure of Alt a 1 deposited in the PDB (PDB code: 4AUD) does not have Cys30 in its sequence, but it still

forms a dimer Figure 22 strongly suggesting that it is possible that homologs in which Cys30 is not conserved may still form dimeric structures, in which dimers are stabilized by hydrophobic and hydrogen bonding interactions. There is a possibility is that the negatively charged surface of the allergen may be involved in the binding of a ligand and stabilization of the homodimer.

The result of an experiment performed with the Dynamic Light Scattering method has shown that the dimer of recombinant Alt a 1 may have a less compact structure than in the crystal or that dimers form a higher order oligomeric assemblies, such as a tetramer (Chruszcz et al., 2012a).

Small molecular compounds which are present in the crystal structure of Alt a 1 - 4-hydroxy-2,5,6-triaminopyrimidine (4-HTP) and 8-aminocaprylic acid (8-ACA) originate from the crystallization solution and most probably their binding is not physiological. 4-HTP is located on symmetry elements and 8-ACA is located close to one of the 4-HTP molecules, and not on a symmetry element (Figure 20). The orientation of 8-ACA points the amino group towards the solution and enables the carboxylic group to form two hydrogen bonds with an amide group of Gln110. The residues Asp37 and Tyr38 provide additional non-specific interaction for the 8-aminocaprylate binding (Chruszcz et al., 2012a). Moreover, Alt a 1 does not have a binding cavity that is characteristic for lipocalins, streptavidin and other protein with the β-barrel fold. However, it can be speculated that the binding of a physiological Alt a 1 ligand may induce significant conformational changes resulting in the opening of the hydrophobic β-barrel core or that the closed conformation is forced in a crystal (Chruszcz et al., 2012a).

A search for three-dimensional functional fragments with PROFUNC (Laskowski et al., 2005) did not return any significant hits (Chruszcz et al., 2012a). However, structure of streptavidin (PDB codes: 2JGS, 1SWG, 1SWF, and 4GD9) was found to have somewhat matching fold, albeit with high rmsd value (~3 Å). Although some similarity of the Alt a 1 to

other proteins with a β-barrel can be determined, the structural comparison does not provide any direct hint of a possible function of the allergen. The results obtained from the searches with DALI (Holm and Rosenström, 2010) and FATCAT (Ye and Godzik, 2004) showed the structure (PDB code: 2Q03) of an uncharacterized protein from *Shewanella denitrificans* (deposited in GenPept under locus YP_563039 which was discontinued and replaced with WP_011496472.1) as the top hit (Chruszcz et al., 2012a). Only the proteins β-barrels were found as the structures somewhat similar to Alt a 1. The results contained membrane proteins, streptavidin, odorant binding proteins, plant transcriptional regulators, RNA binding proteins, lipid binding proteins and a few lipocalin antibodies (Bos d 2, Can f 2 and Equ c 1), but these allergens contain an α-helix besides β-barrels (Chruszcz et al., 2012a).

Previous studies of IgE binding to the synthetic peptides of Alt a 1 (Kurup et al., 2003) are confirmed with the structural data. The study consist of four peptides derived from the sequence of Alt a 1, two of which (K41-P50 and Y54-K63) were shown to bind strongly IgE from the sera of *Alternaria alternata* allergic individuals. The two remaining peptides (Y87-D96 and V119-C128) showed weak IgE binding and in the case of the Y87-D96 the results were inconsistent. The peptides that were characterized as strong IgE binding are surface exposed and easily accessible for the interaction with antibodies (Figure 30). These two peptides correspond to strands β1 and β2 of the reported by us structure. The peptides that showed weak IgE binding are partially inside of the structure and this may clarify lower affinity towards IgE. The localization of the IgE binding epitopes on the surface of the dimer also provides an explanation why it is possible to use a single monoclonal antibody for “sandwich” ELISA measurements of Alt a 1. In this method the same antibody is used as both capture and biotinylated secondary antibody.

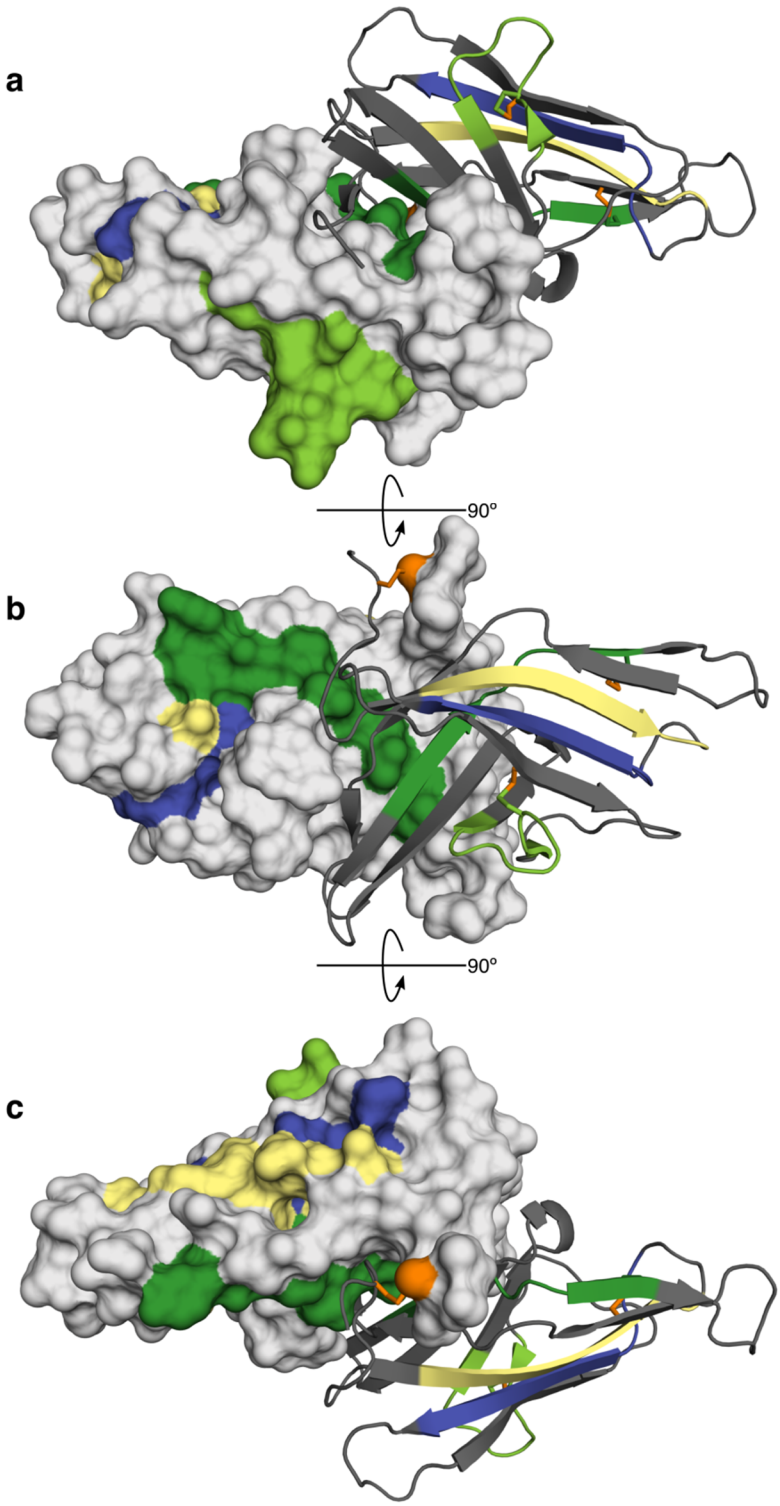

Figure 30 Dimer of Alt a 1 shown as a split representation of molecular surface and cartoon.

The IgE antibody–binding peptides are marked in yellow (residues Lys41-Pro50) and blue (residues Tyr54-Lys63). Peptides Tyr87-Asp96 and Val119-Cys128, which showed weak IgE antibody binding, are shown in light green and dark green, respectively. Each part of the figure is showing a different orientation of the dimer around the x-axis with **a** being at -90°, **b** at 0° and **c** at 90°.

8.4. Summary

Allergy or hypersensivity to various substances affect almost every citizen of industrialized countries. The everyday contact with allergic individuals is commonplace for not only allergists, but pediatricians or primary care physicians, who deal with allergic patients every day. This is quite different than it was a half century ago, when the diagnosed prevalence of allergy was minimal. Allergy care has become a subject for politicians, the media, and, the most importantly, is a problem for the allergic individuals, their everyday life and their families.

The elucidation of the three-dimensional structures of the allergens complexed with antibodies reveals the conformational epitopes of allergens such as Der p 1 (Lombardero et al., 1990), that are the main structural features responsible for the allergic response. This approach for mapping the antibody epitopes represents a major advantage, compared to the previously used strategies that are based on the identification of the antibody-binding peptides that only contain the linear epitopes. Therefore, the results of the described structural studies of Group 1 mite allergens may be used not only to design the experiments allowing for the identification of the IgE binding epitopes, but also to design the recombinant versions of the allergens for the application in the immunotherapy (Thomas et al., 2010).

Bla g 4 is a male cockroach specific protein that is a member of the lipocalin family of proteins, and binds small molecule ligands. It binds both tyramine and octopamine, and most likely the function of the protein is to deliver a neurotransmitter to a female. The structural analysis shows that tyramine binding does not cause any significant changes in the ligand binding site. The results allowed to speculate that the delivery of tyramine causes post-mating inactivation of pheromone production in a female. However, it cannot be excluded that Blag 4 works as a carrier for both tyramine and octopamine, which is consistent with the NMR and ITC results indicating that octopamine binds to Bla g 4 in the same manner as tyramine

(Offermann et al., 2014). Comparison of Bla g 4 and Per a 4 structures and sequences suggests that these proteins bind different ligands, and therefore have different physiological roles. Currently lack of experimental data does not allow for the identification of Bla g 4 and Per a 4 fragments that are responsible for cross-reactivity.

Alt a 1 is a protein produced by common black mold – *Alternaria alternata*. It is the first protein from the entirely new protein family with no previously known structures. The structure of Alt a 1 forms a dimer which helps explain the differences observed in previous studies (Kurup et al., 2003). The disulfide bridge, formed by Cys30 between the monomers, is not necessary to form a dimer, as observed in another Alt a 1 structure (PDB code: 4AUD), but it may be involved in the stabilization of the dimer. The other cysteines found in Alt a 1 are highly conserved among its homologs which can help explaining their stability. Further studies may identify the function of Alt a 1, which can lead to increased knowledge about this unique protein family and will provide insights into their biology, fungal allergenicity and the immune response to fungal allergens.

The unique opportunity to analyze different aspects of three inhaled allergens was an amazing challenge that allowed for not only the recognition of structures of three selected allergens, but showed more detailed knowledge on interfaces in antigen-antibody complexes. All of the aims of this work have been met.

9. References

Achatz, G., Oberkofler, H., Lechenauer, E., Simon, B., Unger, A., Kandler, D., Ebner, C., Prillinger, H., Kraft, D., Breitenbach, M., 1995. Molecular cloning of major and minor allergens of Alternaria alternata and Cladosporium herbarum. Mol. Immunol. 32, 213–227. doi:10.1016/0161-5890(94)00108-D

Al-Lazikani, B., Lesk, A.M., Chothia, C., 1997. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol 273, 927–948. doi:10.1006/jmbi.1997.1354

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402.

Arbes, S.J., Gergen, P.J., Elliott, L., Zeldin, D.C., 2005a. Prevalences of positive skin test responses to 10 common allergens in the US population: results from the third National Health and Nutrition Examination Survey. J. Allergy Clin. Immunol. 116, 377–383. doi:10.1016/j.jaci.2005.05.017

Arbes, S.J., Gergen, P.J., Vaughn, B., Zeldin, D.C., 2007. Asthma cases attributable to atopy: results from the Third National Health and Nutrition Examination Survey. J. Allergy Clin. Immunol. 120, 1139–45. doi:10.1016/j.jaci.2007.07.056

Arbes, S.J., Sever, M., Mehta, J., Collette, N., Thomas, B., Zeldin, D.C., 2005b. Exposure to indoor allergens in day-care facilities: Results from 2 North Carolina counties. J. Allergy Clin. Immunol. 116, 133–139. doi:10.1016/j.jaci.2005.04.022

Arruda, L.K., Rizzo, M.C., Chapman, M.D., Fernandez-Caldas, E., Baggio, D., Platts-Mills, T.A., Naspitz, C.K., 1991. Exposure and sensitization to dust mite allergens among asthmatic children in São Paulo, Brazil. Clin. Exp. Allergy 21, 433–439.

Arruda, L.K., Vailes, L.D., Hayden, M.L., Benjamin, D.C., Chapman, M.D., 1995. Cloning of cockroach allergen, Bla g 4, identifies ligand binding proteins (or calycins) as a cause of IgE antibody responses. J. Biol. Chem. 270, 31196–31201. doi:10.1074/jbc.270.52.31196

Arruda, L.K., Vailes, L.D., Mann, B.J., Shannon, J., Fox, J.W., Vedvick, T.S., Hayden, M.L., Chapman, M.D., 1995. Molecular cloning of a major cockroach (Blattella germanica) allergen, Bla g 2. Sequence homology to the aspartic proteases. J. Biol. Chem. 270, 19563–19568. doi:10.1074/jbc.270.33.19563

Arruda, L.K., Vailes, L.D., Platts-Mills, T.A., Hayden, M.L., Chapman, M.D., 1997. Induction of IgE antibody responses by glutathione S-transferase from the German cockroach (Blattella germanica). J. Biol. Chem. 272, 20907–20912. doi:10.1074/jbc.272.33.20907

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N., 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38, W529–33. doi:10.1093/nar/gkq399

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, 10037–10041. doi:10.1073/pnas.181342398

Barlow, D.J., Edwards, M.S., Thornton, J.M., 1986. Continuous and discontinuous protein antigenic determinants. Nature 322, 747–8. doi:10.1038/322747a0

Barlow, D.J., Thornton, J.M., 1983. Ion-pairs in proteins. J. Mol. Biol. 168, 867–885. doi:10.1016/S0022-2836(83)80079-5

Bartolomei-Díaz, J.A., Amill-Rosario, A., Claudio, L., Hernández, W., 2011. Asthma mortality in Puerto Rico: 1980-2007. J. Asthma 48, 202–9. doi:10.3109/02770903.2010.528498

Bergmann, K.-C., Ring, J., 2014. History of Allergy. Karger Medical and Scientific Publishers. doi:10.1159/isbn.978-3-318-02195-0

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Res. 28, 235–242. doi:10.1093/nar/28.1.235

Bernton, H.S., Brown, H., 1964. Insect allergy—Preliminary studies of the cockroach. J. Allergy 35, 506–513. doi:10.1016/0021-8707(64)90082-6

Blech, M., Seeliger, D., Kistler, B., Bauer, M.M.T., Hafner, M., Horer, S., Zeeb, M., Nar, H., Park, J.E., Hörer, S., 2012. Molecular structure of human GM-CSF in complex with a disease-associated anti-human GM-CSF autoantibody and its potential biological implications. Biochem J 447, 205–215. doi:10.1042/BJ20120884

Bufford, J.D., Reardon, C.L., Li, Z., Roberg, K. a., DaSilva, D., Eggleston, P. a., Liu, a. H., Milton, D., Alwis, U., Gangnon, R., Lemanske, R.F., Gern, J.E., 2008. Effects of dog ownership in early childhood on immune development and atopic diseases. Clin. Exp. Allergy 38, 1635–1643. doi:10.1111/j.1365-2222.2008.03018.x

Bulek, A.M., Cole, D.K., Skowera, A., Dolton, G., Gras, S., Madura, F., Fuller, A., Miles, J.J., Gostick, E., Price, D. a, Drijfhout, J.W., Knight, R.R., Huang, G.C., Lissin, N., Molloy, P.E., Wooldridge, L., Jakobsen, B.K., Rossjohn, J., Peakman, M., Rizkallah, P.J., Sewell, A.K., 2012. Structural basis for the killing of human beta cells by CD8(+) T cells in type 1 diabetes. Nat. Immunol. 13, 283–289. doi:10.1038/ni.2206

Bush, R.K., Portnoy, J.M., 2001. The role and abatement of fungal allergens in allergic diseases. J. Allergy Clin. Immunol. 107, S430–S440. doi:10.1067/mai.2001.113669

Bush, R.K., Prochnau, J.J., 2004. Alternaria-induced asthma. J. Allergy Clin. Immunol. 113, 227–34. doi:10.1016/j.jaci.2003.11.023

Busse, W.W., Lemanske, R.F., 2001. Asthma. N. Engl. J. Med. 344, 350–62. doi:10.1056/NEJM200102013440507

Call, R.S., Smith, T.F., Morris, E., Chapman, M.D., Platts-Mills, T.A., 1992. Risk factors for asthma in inner city children. J. Pediatr. 121, 862–866.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. doi:10.1186/1471-2105-10-421

Celedón, J.C., Milton, D.K., Ramsey, C.D., Litonjua, A.A., Ryan, L., Platts-Mills, T.A.E., Gold, D.R., 2007. Exposure to dust mite allergen and endotoxin in early life and asthma and atopy in childhood. J. Allergy Clin. Immunol. 120, 144–9. doi:10.1016/j.jaci.2007.03.037

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., Ben-Tal, N., 2013. ConSurf: Using evolutionary data to raise testable hypotheses about protein function. Isr. J. Chem. 53, 199–206. doi:10.1002/ijch.201200096

Chan-Yeung, M., Anthonisen, N.R., Becklake, M.R., Bowie, D., Sonia Buist, a., Dimich-Ward, H., Ernst, P., Sears, M.R., Siersted, H.C., Sweet, L., Van Til, L., Manfreda, J., 2010. Geographical variations in the prevalence of atopic sensitization in six study sites across Canada. Allergy Eur. J. Allergy Clin. Immunol. 65, 1404–1413. doi:10.1111/j.1398-9995.2010.02399.x

Chapman, M.D., Heymann, P.W., Platts-Mills, T. a, 1987. Epitope mapping of two major inhalant allergens, Der p I and Der f I, from mites of the genus Dermatophagoides. J. Immunol. 139, 1479–84.

Chapman, M.D., Platts-Mills, T. a, 1980. Purification and characterization of the major allergen from Dermatophagoides pteronyssinus-antigen P1. J. Immunol. 125, 587–592.

Chapman, M.D., Sutherland, W.M., Platts-Mills, T. a, 1984. Recognition of two Dermatophagoides pteronyssinus-specific epitopes on antigen P1 by using monoclonal antibodies: binding to each epitope can be inhibited by serum from dust mite-allergic patients. J. Immunol. 133, 2488–95.

Chawla, J., Seear, M., Zhang, T., Smith, A., Carleton, B., 2012. Fifty years of pediatric asthma in developed countries: How reliable are the basic data sources? Pediatr. Pulmonol. 47, 211–219. doi:10.1002/ppul.21537

Chemical Computing Group Inc., 2013. MOE : Molecular Operating Environment.

Chothia, C., Lesk, A.M., 1987. Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. 196, 901–917. doi:10.1016/0022-2836(87)90412-8

Chruszcz, M., Chapman, M.D., Osinski, T., Solberg, R., Demas, M., Porebski, P.J., Majorek, K.A., Pomés, A., Minor, W., Pomes, A., 2012a. Alternaria alternata allergen Alt a 1: a

unique β-barrel protein dimer found exclusively in fungi. J. Allergy Clin. Immunol. 130, 241–7 e9. doi:10.1016/j.jaci.2012.03.047

Chruszcz, M., Chapman, M.D., Vailes, L.D., Stura, E.A., Saint-Remy, J.M., Minor, W., Pomes, A., 2009. Crystal structures of mite allergens Der f 1 and Der p 1 reveal differences in surface-exposed residues that may influence antibody binding. J Mol Biol 386, 520–530. doi:10.1016/j.jmb.2008.12.049

Chruszcz, M., Pomes, A., Glesner, J., Vailes, L.D., Osinski, T., Porebski, P.J., Majorek, K.A., Heymann, P.W., Platts-Mills, T.A.E., Minor, W., Chapman, M.D., Pomés, A., 2012b. Molecular determinants for antibody binding on group 1 house dust mite allergens. J Biol Chem 287, 7388–7398. doi:10.1074/jbc.M111.311159

Coca, A.F., Cooke, R.A., 1923. On the Classification of the Phenomena of Hypersensitiveness. J. Immunol. 8, 163–182.

Cohn, R.D., Arbes, S.J., Jaramillo, R., Reid, L.H., Zeldin, D.C., 2006. National prevalence and exposure risk for cockroach allergen in U.S. households. Environ. Health Perspect. 114, 522–526. doi:10.1289/ehp.8561

Collaborative Computational Project, N., 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50, 760–763. doi:10.1107/S0907444994003112

Consortium, T.U., The Uniprot Consortium, 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res. 40, D71–5. doi:10.1093/nar/gkr981

Coughlan, J.L., Gibson, P.G., Henry, R.L., 2001. Medical treatment for reflux oesophagitis does not consistently improve asthma control: a systematic review. Thorax 56, 198–204. doi:10.1136/thorax.56.3.198

Cowtan, K.D., Main, P., 1993. Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. Acta Crystallogr D Biol Crystallogr 49, 148–157. doi:10.1107/S0907444992007698

Cramer, R.A., Lawrence, C.B., 2003. Cloning of a gene encoding an Alt a 1 isoallergen differentially expressed by the necrotrophic fungus Alternaria brassicicola during Arabidopsis infection. Appl. Environ. Microbiol. 69, 2361–2364. doi:10.1128/AEM.69.4.2361-2364.2003

Da Silva, R., Lange, A.B., 2008. Tyramine as a possible neurotransmitter/neuromodulator at the spermatheca of the African migratory locust, Locusta migratoria. J. Insect Physiol. 54, 1306–13. doi:10.1016/j.jinsphys.2008.07.001

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall 3rd, W.B., Snoeyink, J., Richardson, J.S., Richardson, D.C., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35, W375–83. doi:10.1093/nar/gkm216

De Halleux, S., Stura, E., VanderElst, L., Carlier, V., Jacquemin, M., Saint-Remy, J.-M.M., 2006. Three-dimensional structure and IgE-binding properties of mature fully active Der p 1, a clinically relevant major allergen. J. Allergy Clin. Immunol. 117, 571–576. doi:10.1016/j.jaci.2005.11.032

De Vouge, M.W., Thaker, A.J., Curran, I.H., Zhang, L., Muradia, G., Rode, H., Vijay, H.M., 1996. Isolation and expression of a cDNA clone encoding an Alternaria alternata Alt a 1 subunit. Int. Arch. Allergy Immunol. 111, 385–395.

Deards, M.J., Montague, a E., 1991. Purification and characterisation of a major allergen of Alternaria alternata. Mol. Immunol. 28, 409–15.

Delfino, R.J., Zeiger, R.S., Seltzer, J.M., Street, D.H., Matteucci, R.M., Anderson, P.R., Koutrakis, P., 1997. The effect of outdoor fungal spore concentrations on daily asthma severity. Environ. Health Perspect. 105, 622–635. doi:10.1289/ehp.97105622

Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., Baker, N. a., 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 35, 522–525. doi:10.1093/nar/gkm276

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N. a., 2004. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32, 665–667. doi:10.1093/nar/gkh381

Dong, J., Thompson, A. a, Fan, Y., Lou, J., Conrad, F., Ho, M., Pires-Alves, M., Wilson, B. a, Stevens, R.C., Marks, J.D., 2010. A single-domain llama antibody potently inhibits the enzymatic activity of botulinum neurotoxin by binding to the non-catalytic alpha-exosite binding region. J Mol Biol 397, 1106–1118. doi:10.1016/j.jmb.2010.01.070

Donini, A., Lange, A.B., 2004. Evidence for a possible neurotransmitter/neuromodulator role of tyramine on the locust oviducts. J. Insect Physiol. 50, 351–61. doi:10.1016/j.jinsphys.2004.02.005

Downs, S.H., Mitakakis, T.Z., Marks, G.B., Car, N.G., Belousova, E.G., Leüppi, J.D., Xuan, W., Downie, S.R., Tobias, A., Peat, J.K., 2001. Clinical importance of Alternaria exposure in children. Am. J. Respir. Crit. Care Med. 164, 455–459. doi:10.1164/ajrccm.164.3.2008042

Durham, O.C., 1937. Incidence of air-borne fungus spores: I. Alternaria. J. Allergy 8, 480–490. doi:10.1016/S0021-8707(37)90405-2

Edelman, G.M., 1973. Antibody structure and molecular immunology. Science 180, 830–40.

Eggleston, P. a., Rosenstreich, D., Lynn, H., Gergen, P., Baker, D., Kattan, M., Mortimer, K.M., Mitchell, H., Ownby, D., Slavin, R., Malveaux, F., 1998. Relationship of indoor allergen exposure to skin test sensitivity in inner-city children with asthma. J. Allergy Clin. Immunol. 102, 563–570. doi:10.1016/S0091-6749(98)70272-6

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126–2132. doi:10.1107/S0907444904019158

Fan, Y., Gore, J.C., Redding, K.O., Vailes, L.D., Chapman, M.D., Schal, C., 2005. Tissue localization and regulation by juvenile hormone of human allergen Bla g 4 from the German cockroach, Blattella germanica (L.). Insect Mol. Biol. 14, 45–53. doi:10.1111/j.1365-2583.2004.00530.x

Fernández Rodríguez, C., Bevilacqua, E., Fernández, N., Gajate, P., de la Cámara, a. G., Garcimartín, M., Vives, R., Rodríguez, J., 2011. Asthma related to Alternaria sensitization: An analysis of skin-test and serum-specific IgE efficiency based on the bronchial provocation test. Clin. Exp. Allergy 41, 649–656. doi:10.1111/j.1365-2222.2010.03645.x

Flower, D.R., 1996. The lipocalin protein family: structure and function. Biochem. J. 318 (Pt 1, 1–14.

Flower, D.R., North, A.C.T., Sansom, C.E., 2000. The lipocalin protein family: structural and sequence overview. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1482, 9–24. doi:10.1016/S0167-4838(00)00148-5

Frickey, T., Lupas, A., 2004. CLANS: A Java application for visualizing protein families based on pairwise similarity. Bioinformatics 20, 3702–3704. doi:10.1093/bioinformatics/bth444

Frishman, D., Argos, P., 1995. Knowledge-based protein secondary structure assignment. Proteins 23, 566–579. doi:10.1002/prot.340230412

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics 28, 3150–3152. doi:10.1093/bioinformatics/bts565

Garne, E., Hansen, A.V., Birkelund, A.S., Marie, A., Andersen, N., Andersen, A.M.N., 2014. Major congenital anomalies in a Danish region. Dan. Med. J. 61, A4825.

Gelber, L.E., Seltzer, L.H., Bouzoukis, J.K., Pollart, S.M., Chapman, M.D., Platts-Mills, T. a, 1993. Sensitization and exposure to indoor allergens as risk factors for asthma among patients presenting to hospital. Am. Rev. Respir. Dis. 147, 573–578. doi:10.1164/ajrccm/147.3.573

Gillott, C., 2003. Male accessory gland secretions: modulators of female reproductive physiology and behavior. Annu. Rev. Entomol. 48, 163–84. doi:10.1146/annurev.ento.48.091801.112657

Goldacre, M.J., Duncan, M.E., Griffith, M., 2012. Death rates for asthma in English populations 1979-2007: comparison of underlying cause and all certified causes. Public Health 126, 386–93. doi:10.1016/j.puhe.2012.01.022

Gore, J.C., Schal, C., 2007. Cockroach allergen biology and mitigation in the indoor environment. Annu. Rev. Entomol. 52, 439–463. doi:10.1146/annurev.ento.52.110405.091313

Gruchalla, R.S., Pongracic, J., Plaut, M., Evans, R., Visness, C.M., Walter, M., Crain, E.F., Kattan, M., Morgan, W.J., Steinbach, S., Stout, J., Malindzak, G., Smartt, E., Mitchell, H., 2005. Inner City Asthma Study: relationships among sensitivity, allergen exposure, and asthma morbidity. J. Allergy Clin. Immunol. 115, 478–85. doi:10.1016/j.jaci.2004.12.006

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18, 2714–2723. doi:10.1002/elps.1150181505

Gustchina, A., Li, M., Wünschmann, S., Chapman, M.D., Pomés, A., Wlodawer, A., 2005. Crystal structure of cockroach allergen Bla g 2, an unusual zinc binding aspartic protease with a novel mode of self-inhibition. J. Mol. Biol. 348, 433–444. doi:10.1016/j.jmb.2005.02.062

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98. doi:citeulike-article-id:691774

Hart, P.D., Mitchell, D.N., 1971. Lymphocyte sensitization. Br. Med. J. 3, 246.

Hasegawa, K., Tsugawa, Y., Brown, D.F.M., Camargo, C. a., 2013. Childhood asthma hospitalizations in the United States, 2000-2009. J. Pediatr. 163, 1127–33.e3. doi:10.1016/j.jpeds.2013.05.002

Hedin, H., Richter, W., Ring, J., 1976. Dextran-induced anaphylactoid reactions in man: role of dextran reactive antibodies. Int. Arch. Allergy Appl. Immunol. 52, 145–59.

Helm, R., Cockrell, G., Stanley, J.S., Brenner, R.J., Burks, W., Bannon, G. a, 1996. Isolation and characterization of a clone encoding a major allergen (Bla g Bd90K) involved in IgE-mediated cockroach hypersensitivity. J. Allergy Clin. Immunol. 98, 172–180. doi:10.1016/S0091-6749(96)70240-3

Heymann, P.W., Chapman, M.D., Platts-Mills, T. a, 1986. Antigen Der f I from the dust mite Dermatophagoides farinae: structural comparison with Der p I from Dermatophagoides pteronyssinus and epitope specificity of murine IgG and human IgE antibodies. J Immunol 137, 2841–2847.

Hirashima, A., Yamaji, H., Yoshizawa, T., Kuwano, E., Eto, M., 2007. Effect of tyramine and stress on sex-pheromone production in the pre- and post-mating silkworm moth, Bombyx mori. J. Insect Physiol. 53, 1242–9. doi:10.1016/j.jinsphys.2007.06.018

Holm, L., Rosenström, P., 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545–9. doi:10.1093/nar/gkq366

Hong, S.G., Cramer, R. a, Lawrence, C.B., Pryor, B.M., 2005. Alt a 1 allergen homologs from Alternaria and related taxa: analysis of phylogenetic content and secondary structure. Fungal Genet. Biol. 42, 119–29. doi:10.1016/j.fgb.2004.10.009

Hopp, T.P., Woods, K.R., 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U. S. A. 78, 3824–8.

Hozumi, N., Tonegawa, S., 1976. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc. Natl. Acad. Sci. U. S. A. 73, 3628–32.

Hsiao, H.-J., Wang, L.-C., Yang, Y.-H., Lee, J.-H., Yu, H.-H., Lin, Y.-T., Chiang, B.-L., 2013. A nationwide survey of the severity, comorbidity, and mortality of hospitalized patients with asthma in Taiwan. Pediatr. Neonatol. 54, 254–60. doi:10.1016/j.pedneo.2013.01.017

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J Mol Graph 14, 27–28,33–38.

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95.

Jackson, D.J., Gangnon, R.E., Evans, M.D., Roberg, K. a., Anderson, E.L., Pappas, T.E., Printz, M.C., Lee, W.M., Shult, P. a., Reisdorf, E., Carlson-Dakes, K.T., Salazar, L.P., DaSilva, D.F., Tisler, C.J., Gern, J.E., Lemanske, R.F., 2008. Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. Am. J. Respir. Crit. Care Med. 178, 667–672. doi:10.1164/rccm.200802-309OC

Jackson, R., 2000. Serum sickness. J. Cutan. Med. Surg. 4, 223–5.

Jeong, K.Y., Lee, H., Shin, K.H., Yi, M., Jeong, K.Y., Hong, C.-S., Yong, T.-S., 2008. Sequence polymorphisms of major German cockroach allergens Bla g 1, Bla g 2, Bla g 4, and Bla g 5. Int. Arch. Allergy Immunol. 145, 1–8. doi:10.1159/000107460

Jeong, K.Y., Yi, M.-H., Jeong, K., Lee, H., Hong, C.-S., Yong, T.-S., 2009. Sequence diversity of the Bla g 4 cockroach allergen, homologous to lipocalins, from Blattella germanica. Int. Arch. Allergy Immunol. 148, 339–45. doi:10.1159/000170388

Kabsch, W., 1978. A discussion of the solution for the best rotation to relate two sets of vectors. Acta Crystallogr. Sect. A 34, 827–828. doi:10.1107/S0567739478001680

Kabsch, W., 1976. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr. Sect. A 32, 922–923. doi:10.1107/S0567739476001873

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30, 772–780. doi:10.1093/molbev/mst010

Kelly, J.T., Busse, W.W., 2008. Host immune responses to rhinovirus: Mechanisms in asthma. J. Allergy Clin. Immunol. 122, 671–682. doi:10.1016/j.jaci.2008.08.013

Kimber, I., Basketter, D.A., Gerberick, G.F., Dearman, R.J., 2002. Allergic contact dermatitis. Int. Immunopharmacol. 2, 201–211. doi:10.1016/S1567-5769(01)00173-4

Knutsen, A.P., Bush, R.K., Demain, J.G., Denning, D.W., Dixit, A., Fairs, A., Greenberger, P.A., Kariuki, B., Kita, H., Kurup, V.P., Moss, R.B., Niven, R.M., Pashley, C.H., Slavin, R.G., Vijay, H.M., Wardlaw, A.J., 2012. Fungi and allergic lower respiratory tract diseases. J. Allergy Clin. Immunol. 129, 280–91; quiz 292–3. doi:10.1016/j.jaci.2011.12.970

Koide, S., Sidhu, S.S., 2009. The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins. ACS Chem Biol 4, 325–334. doi:10.1021/cb800314v

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–97. doi:10.1016/j.jmb.2007.05.022

Kuglstatter, A., Stahl, M., Peters, J.-U., Huber, W., Stihle, M., Schlatter, D., Benz, J., Ruf, A., Roth, D., Enderle, T., Hennig, M., 2008. Tyramine fragment binding to BACE-1. Bioorg. Med. Chem. Lett. 18, 1304–7. doi:10.1016/j.bmcl.2008.01.032

Kulkarni-Kale, U., Bhosle, S., Kolaskar, A.S., 2005. CEP: A conformational epitope prediction server. Nucleic Acids Res. 33. doi:10.1093/nar/gki460

Kumar, S., Nussinov, R., 2002a. Relationship between ion pair geometries and electrostatic strengths in proteins. Biophys. J. 83, 1595–612. doi:10.1016/S0006-3495(02)73929-5

Kumar, S., Nussinov, R., 2002b. Close-range electrostatic interactions in proteins. Chembiochem 3, 604–17. doi:10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X

Kuna, P., Kaczmarek, J., Kupczyk, M., 2011. Efficacy and safety of immunotherapy for allergies to Alternaria alternata in children. J. Allergy Clin. Immunol. 127, 502–508.e6. doi:10.1016/j.jaci.2010.11.036

Kurup, V.P., Vijay, H.M., Kumar, V., Castillo, L., Elms, N., 2003. IgE binding synthetic peptides of Alt a 1, a major allergen of Alternaria alternata. Peptides 24, 179–185. doi:10.1016/S0196-9781(03)00024-X

Lange, A.B., 2009. Tyramine: From octopamine precursor to neuroactive chemical in insects. Gen. Comp. Endocrinol. 162, 18–26. doi:10.1016/j.ygcen.2008.05.021

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–8. doi:10.1093/bioinformatics/btm404

Laskowski, R.A., Watson, J.D., Thornton, J.M., 2005. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res. 33, W89–W93. doi:10.1093/nar/gki414

Lo, Y.-T., Pai, T.-W., Wu, W.-K., Chang, H.-T., 2013. Prediction of conformational epitopes with the use of a knowledge-based energy function and geometrically related neighboring residue characteristics. BMC Bioinformatics 14 Suppl 4, S3. doi:10.1186/1471-2105-14-S4-S3

Lodge, C.J., Lowe, A.J., Gurrin, L.C., Hill, D.J., Hosking, C.S., Khalafzai, R.U., Hopper, J.L., Matheson, M.C., Abramson, M.J., Allen, K.J., Dharmage, S.C., 2011. House dust mite sensitization in toddlers predicts current wheeze at age 12 years. J. Allergy Clin. Immunol. 128, 782–788.e9. doi:10.1016/j.jaci.2011.06.038

Lombardero, M., Heymann, P.W., Platts-Mills, T.A., Fox, J.W., Chapman, M.D., 1990. Conformational stability of B cell epitopes on group I and group II Dermatophagoides spp. allergens. Effect of thermal and chemical denaturation on the binding of murine IgG and human IgE antibodies. J. Immunol. 144, 1353–60.

Lopez, M., Salvaggio, J.E., 1985. Mold-sensitive asthma. Clin. Rev. Allergy 3, 183–96.

Lotufo, P.A., Bensenor, I.M., 2012. Temporal trends of asthma mortality rates in Brazil from 1980 to 2010. J. Asthma 49, 779–84. doi:10.3109/02770903.2012.693237

Mari, A., Rasi, C., Palazzo, P., Scala, E., 2009. Allergen databases: current status and perspectives. Curr. Allergy Asthma Rep. 9, 376–83.

Marks, G.B., Bush, R.K., 2007. It's blowing in the wind: New insights into thunderstorm-related asthma. J. Allergy Clin. Immunol. 120, 530–532. doi:10.1016/j.jaci.2007.07.012

Martin, A.C., Thornton, J.M., 1996. Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies. J. Mol. Biol. 263, 800–15. doi:10.1006/jmbi.1996.0617

Martin, A.J.M., Vidotto, M., Boscariol, F., Di Domenico, T., Walsh, I., Tosatto, S.C.E., 2011. RING: networking interacting residues, evolutionary information and energetics in protein structures. Bioinformatics 27, 2003–2005. doi:10.1093/bioinformatics/btr191

Masoli, M., Fabian, D., Holt, S., Beasley, R., 2004. The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy Eur. J. Allergy Clin. Immunol. 59, 469–78. doi:10.1111/j.1398-9995.2004.00526.x

Matsui, E.C., Wood, R.A., Rand, C., Kanchanaraksa, S., Swartz, L., Curtin-Brosnan, J., Eggleston, P.A., 2003. Cockroach allergen exposure and sensitization in suburban middle-class children with asthma. J. Allergy Clin. Immunol. 112, 87–92. doi:10.1067/mai.2003.1588

Matthyssens, G., Hozumi, N., Tonegawa, S.,. Somatic generation of antibody diversity. Ann. Immunol. (Paris). 127, 439–48.

Meno, K., Thorsted, P.B., Ipsen, H., Kristensen, O., Larsen, J.N., Spangfort, M.D., Gajhede, M., Lund, K., 2005. The crystal structure of recombinant proDer p 1, a major house dust mite proteolytic allergen. J. Immunol. 175, 3835–3845. doi:175/6/3835 [pii]

Minor, W., Cymborowski, M., Otwinowski, Z., Chruszcz, M., 2006. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallogr. D. Biol. Crystallogr. 62, 859–66. doi:10.1107/S0907444906019949

Mitakakis, T.Z., Barnes, C., Tovey, E.R., 2001. Spore germination increases allergen release from Alternaria. J. Allergy Clin. Immunol. 107, 388–90. doi:10.1067/mai.2001.112602

Möller, C., Dreborg, S., Ferdousi, H.A., Halken, S., Høst, A., Jacobsen, L., Koivikko, A., Koller, D.Y., Niggemann, B., Norberg, L.A., Urbanek, R., Valovirta, E., Wahn, U., 2002. Pollen immunotherapy reduces the development of asthma in children with seasonal rhinoconjunctivitis (the PAT-study). J. Allergy Clin. Immunol. 109, 251–256. doi:10.1067/mai.2002.121317

Moorman, J.E., Akinbami, L.J., Bailey, C.M., Zahran, H.S., King, M.E., Johnson, C. a, Liu, X., 2012. National surveillance of asthma: United States, 2001-2010., Vital & health statistics. Series 3, Analytical and epidemiological studies / [U.S. Dept. of Health and Human Services, Public Health Service, National Center for Health Statistics].

Morea, V., Tramontano, A., Rustici, M., Chothia, C., Lesk, A.M., 1998. Conformations of the third hypervariable region in the VH domain of immunoglobulins. J Mol Biol 275, 269–294. doi:10.1006/jmbi.1997.1442

Mueller, G.A., Pedersen, L.C., Glesner, J., Edwards, L.L., Zakzuk, J., London, R.E., Arruda, L.K., Chapman, M.D., Caraballo, L., Pomés, A., 2015. Analysis of glutathione S-transferase allergen cross-reactivity in a North American population: Relevance for molecular diagnosis. J. Allergy Clin. Immunol. doi:10.1016/j.jaci.2015.03.015

Mueller, G.A., Pedersen, L.C., Lih, F.B., Glesner, J., Moon, A.F., Chapman, M.D., Tomer, K.B., London, R.E., Pomés, A., 2013. The novel structure of the cockroach allergen Bla g 1 has implications for allergenicity and exposure assessment. J. Allergy Clin. Immunol. 132, 1420–6. doi:10.1016/j.jaci.2013.06.014

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, 355–367. doi:10.1107/S0907444911001314

Nasser, S.M., Pulimood, T.B., 2009. Allergens and thunderstorm asthma. Curr. Allergy Asthma Rep. 9, 384–90. doi:10.1007/s11882-009-0056-8

NCBI Resource Coordinators, 2013. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 41, D8–D20. doi:10.1093/nar/gks1189

Negi, S.S., Braun, W., 2009. Automated detection of conformational epitopes using phage display peptide sequences. Bioinform. Biol. Insights 2009, 71–81.

Newcomb, D.C., Peebles, R.S., 2009. Bugs and asthma: a different disease? Proc. Am. Thorac. Soc. 6, 266–71. doi:10.1513/pats.200806-056RM

Newkirk, M.M., Edmundson, A., Wistar, R., Klapper, D.G., Capra, J.D., 1987. A New Protocol to Digest Human IgM with Papain that Results in Homogeneous Fab Preparations that Can Be Routinely Crystallized. Hybridoma 6, 453–460. doi:10.1089/hyb.1987.6.453

Niggemann, B., Jacobsen, L., Dreborg, S., Ferdousi, H. a., Halken, S., Høst, a., Koivikko, a., Koller, D., Norberg, L. a., Urbanek, R., Valovirta, E., Wahn, U., Möller, C., 2006. Five-year follow-up on the PAT study: Specific immunotherapy and long-term prevention of asthma in children. Allergy Eur. J. Allergy Clin. Immunol. 61, 855–859. doi:10.1111/j.1398-9995.2006.01068.x

Nikoloudis, D., Pitts, J.E., Saldanha, J.W., 2014. A complete, multi-level conformational clustering of antibody complementarity-determining regions. PeerJ 2, e456. doi:10.7717/peerj.456

North, B., Lehmann, A., Dunbrack, R.L., Dunbrack Jr., R.L., Dunbrack, R.L., 2011. A new clustering of antibody CDR loop conformations. J Mol Biol 406, 228–56. doi:10.1016/j.jmb.2010.10.030

Nowicki, M., Nowakowska, M., Niezgoda, A., Kozik, E., 2012. Alternaria Black Spot of Crucifers: Symptoms, Importance of Disease, and Perspectives of Resistance Breeding. Veg. Crop. Res. Bull. 76, 5–19. doi:10.2478/v10032-012-0001-6

O'Donoghue, P., Luthey-Schulten, Z., 2003. On the evolution of structure in aminoacyl-tRNA synthetases. Microbiol. Mol. Biol. Rev. 67, 550–573. doi:10.1128/MMBR.67.4.550-573.2003

O'Hollaren, M.T., Yunginger, J.W., Offord, K.P., Somers, M.J., O'Connell, E.J., Ballard, D.J., Sachs, M.I., 1991. Exposure to an aeroallergen as a possible precipitating factor in respiratory arrest in young patients with asthma. N. Engl. J. Med. 324, 359–363. doi:10.1056/NEJM199102073240602

Offermann, L.R., Chan, S.L., Osinski, T., Tan, Y.W., Chew, F.T., Sivaraman, J., Mok, Y.-K.K., Minor, W., Chruszcz, M., 2014. The major cockroach allergen Bla g 4 binds tyramine and octopamine. Mol. Immunol. 60, 86–94. doi:10.1016/j.molimm.2014.03.016

Oliphant, T.E., 2007. Python for Scientific Computing. Comput. Sci. & Eng. 9, 10–20. doi:doi:http://dx.doi.org/10.1109/MCSE.2007.58

Otwinowski, Z., 1991. Isomorphous Replacement and Anomalous Scattering, in: Daresbury Study Weekend Proceedings. SERC Daresbury Laboratory, Warrington, U.K.

Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. doi:10.1016/S0076-6879(97)76066-X

Ownby, D.R., Johnson, C.C., Peterson, E.L., 2002. Exposure to dogs and cats in the first year of life and risk of allergic sensitization at 6 to 7 years of age. JAMA 288, 963–972. doi:10.1001/jama.288.8.963

Painter, J., Merritt, E.A., 2006. TLSMD web server for the generation of multi-group TLS models. J. Appl. Crystallogr. 39, 109–111. doi:doi:10.1107/S0021889805038987

Painter, J., Merritt, E.A., 2004. mmLib Python toolkit for manipulating annotated structural models of biological macromolecules. J. Appl. Crystallogr. 37, 174–178. doi:doi:10.1107/S0021889803025639

Parker, J.M., Guo, D., Hodges, R.S., 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 25, 5425–5432. doi:http://dx.doi.org/10.1021/bi00367a013

Patterson, R., Greenberger, P.A., Halwig, J.M., Liotta, J.L., Roberts, M., 1986. Allergic bronchopulmonary aspergillosis. Natural history and classification of early disease by serologic and roentgenographic studies. Arch. Intern. Med. 146, 916–8.

Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U. S. A. 85, 2444–2448. doi:10.1073/pnas.85.8.2444

Perrakis, A., Morris, R., Lamzin, V.S., 1999. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 6, 458–463. doi:10.1038/8263

Perzanowski, M.S., Sporik, R., Squillace, S.P., Gelber, L.E., Call, R., Carter, M., Platts-Mills, T. a E., 1998. Association of sensitization to Alternaria allergens with asthma among school-age children. J. Allergy Clin. Immunol. 101, 626–632. doi:10.1016/S0091-6749(98)70170-8

Pesut, D.P., Bulajic, M. V, Nagomi-Obradovic, L.M., Grgurevic, A.D., Gledovic, Z.B., Ponomarev, D.R., Blanka, A.Z., 2011. Asthma mortality in Serbia: a 30-year analysis. Respir. Med. 105 Suppl , S50–3. doi:10.1016/S0954-6111(11)70011-7

Piboonpocanun, S., Malainual, N., Jirapongsananuruk, O., Vichyanond, P., Thomas, W.R., 2006. Genetic polymorphisms of major house dust mite allergens. Clin. Exp. Allergy 36, 510–516. doi:10.1111/j.1365-2222.2006.02464.x

Platts-Mills, T.A., Thomas, W.R., Aalberse, R.C., Vervloet, D., Champman, M.D., 1992. Dust mite allergens and asthma: report of a second international workshop. J Allergy Clin Immunol 89, 1046–1060.

Platts-Mills, T.A.E., 2008. Allergen avoidance in the treatment of asthma: Problems with the meta-analyses. J. Allergy Clin. Immunol. 122, 694–696. doi:10.1016/j.jaci.2008.07.045

Pomés, A., Chapman, M.D., Vailes, L.D., Blundell, T.L., Dhanaraj, V., 2002. Cockroach allergen Bla g 2: structure, function, and implications for allergic sensitization. Am. J. Respir. Crit. Care Med. 165, 391–7. doi:10.1164/ajrccm.165.3.2104027

Pongracic, J.A., O'Connor, G.T., Muilenberg, M.L., Vaughn, B., Gold, D.R., Kattan, M., Morgan, W.J., Gruchalla, R.S., Smartt, E., Mitchell, H.E., 2010. Differential effects of outdoor versus indoor fungal spores on asthma morbidity in inner-city children. J. Allergy Clin. Immunol. 125, 593–9. doi:10.1016/j.jaci.2009.10.036

Popp, W., Rauscher, H., Sertl, K., Wanke, T., Zwick, H., 1990. Risk factors for sensitization to furred pets. Allergy 45, 75–79.

Porter, R.R., 1973. Structural Studies of Immunoglobulins. Science (80-.). 180, 713–716. doi:10.1126/science.180.4087.713

Prietsch, S.O.M., Zhang, L., Catharino, A.R., Vauchinski, L., Rodrigues, F.E., 2012. Asthma mortality among Brazilian children up to 19 years old between 1980 and 2007. J. Pediatr. (Rio. J). 88, 384–388. doi:10.2223/JPED.2215

Pulimood, T.B., Corden, J.M., Bryden, C., Sharples, L., Nasser, S.M., 2007. Epidemic asthma and the role of the fungal mold Alternaria alternata. J. Allergy Clin. Immunol. 120, 610–617. doi:10.1016/j.jaci.2007.04.045

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. Nucleic Acids Res. 40, D290–301. doi:10.1093/nar/gkr1065

Radauer, C., Bublin, M., Wagner, S., Mari, A., Breiteneder, H., 2008. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. J. Allergy Clin. Immunol. 121, 847–52.e7. doi:10.1016/j.jaci.2008.01.025

Radauer, C., Nandy, a., Ferreira, F., Goodman, R.E., Larsen, J.N., Lidholm, J., Pomés, a., Raulf-Heimsoth, M., Rozynek, P., Thomas, W.R., Breiteneder, H., 2014. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. Allergy Eur. J. Allergy Clin. Immunol. 69, 413–419. doi:10.1111/all.12348

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277. doi:10.1016/S0168-9525(00)02024-2

Roberts, N.J., Lewsey, J.D., Gillies, M., Briggs, A.H., Belozeroff, V., Globe, D.R., Chiou, C.-F., Lin, S.-L., Globe, G., 2013. Time trends in 30 day case-fatality following hospitalisation for asthma in adults in Scotland: a retrospective cohort study from 1981 to 2009. Respir. Med. 107, 1172–7. doi:10.1016/j.rmed.2013.04.004

Rosenbaum, G., Alkire, R.W., Evans, G., Rotella, F.J., Lazarski, K., Zhang, R.G., Ginell, S.L., Duke, N., Naday, I., Lazarz, J., Molitsky, M.J., Keefe, L., Gonczy, J., Rock, L., Sanishvili, R., Walsh, M.A., Westbrook, E., Joachimiak, A., 2006. The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. J Synchrotron Radiat 13, 30–45. doi:10.1107/S0909049505036721

Santos, A.B.R., Chapman, M.D., Aalberse, R.C., Vailes, L.D., Ferriani, V.P.R.L., Oliver, C., Candida Rizzo, M., Naspitz, C.K., Arruda, L.K., Rizzo, M.C., Naspitz, C.K., Arruda, L.K., 1999. Cockroach allergens and asthma in Brazil: Identification of tropomyosin as a major allergen with potential cross-reactivity with mite and shrimp allergens. J. Allergy Clin. Immunol. 104, 329–337. doi:10.1016/S0091-6749(99)70375-1

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S., Feolo, M., Fingerman, I.M., Geer, L.Y., Helmberg, W., Kapustin, Y., Krasnov, S., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Karsch-Mizrachi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y., Wilbur, W.J., Yaschenko, E., Ye, J., 2012. Database resources of

the National Center for Biotechnology Information. Nucleic Acids Res. 40, D13–25. doi:10.1093/nar/gkr1184

Schappert, S.M., Rechtsteiner, E. a, 2008. Ambulatory medical care utilization estimates for 2006. Natl. Health Stat. Report. 1–29.

Schoch, C.L., Wang, Z., Townsend, J.P., Spatafora, J.W., 2009. Geoglossomycetes cl. nov., Geoglossales ord. nov. and taxa above class rank in the Ascomycota Tree of Life. Persoonia Mol. Phylogeny Evol. Fungi 22, 129–138. doi:10.3767/003158509X461486

Schrodinger LLC, 2010. The PyMOL Molecular Graphics System, Version 1.3r1.

Schulz, O., Sewell, H.F., Shakib, F., 1998. Proteolytic cleavage of CD25, the alpha subunit of the human T cell interleukin 2 receptor, by Der p 1, a major mite allergen with cysteine protease activity. J Exp Med 187, 271–275.

Shakib, F., Ghaemmaghami, A.M., Sewell, H.F., 2008. The molecular basis of allergenicity. Trends Immunol. 29, 633–42. doi:10.1016/j.it.2008.08.007

Shakib, F., Schulz, O., Sewell, H., 1998. A mite subversive: cleavage of CD23 and CD25 by Der p 1 enhances allergenicity. Immunol. Today 19, 313–6.

Sheldrick, G.M., 2008. A short history of SHELX. Acta Crystallogr A 64, 112–122. doi:10.1107/S0108767307043930

Shin, K.H., Jeong, K.Y., Hong, C.-S., Yong, T.-S., 2009. IgE binding reactivity of peptide fragments of Bla g 4, a major German cockroach allergen. Korean J. Parasitol. 47, 31–6. doi:10.3347/kjp.2009.47.1.31

Shirai, H., Kidera, A., Nakamura, H., 1999. H3-rules: identification of CDR-H3 structures in antibodies. FEBS Lett. 455, 188–197. doi:10.1016/S0014-5793(99)00821-2

Stern, D. a., Morgan, W.J., Halonen, M., Wright, A.L., Martinez, F.D., 2008. Wheezing and bronchial hyper-responsiveness in early childhood as predictors of newly diagnosed asthma in early adulthood: a longitudinal birth-cohort study. Lancet 372, 1058–1064. doi:10.1016/S0140-6736(08)61447-6

Szczeklik, A., 1997. Mechanism of aspirin-induced asthma. Allergy 52, 613–619. doi:10.1111/j.1398-9995.1997.tb01039.x

Szczeklik, A., Stevenson, D.D., 2003. Aspirin-induced asthma: Advances in pathogenesis, diagnosis, and management. J. Allergy Clin. Immunol. 111, 913–921. doi:10.1067/mai.2003.1487

Tabery, J., 2007. Biometric and developmental gene-environment interactions: looking back, moving forward. Dev. Psychopathol. 19, 961–76. doi:10.1017/S0954579407000478

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–9. doi:10.1093/molbev/msr121

Terwilliger, T., 2004. SOLVE and RESOLVE: automated structure solution, density modification and model building. J Synchrotron Radiat 11, 49–52.

The UniProt Consortium, 2014. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–212. doi:10.1093/nar/gku989

Thomas, W.R., Hales, B.J., Smith, W.A., 2010. House dust mite allergens in asthma and allergy. Trends Mol Med 16, 321–328. doi:10.1016/j.molmed.2010.04.008

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–80.

Thornton, J.M., Edwards, M.S., Taylor, W.R., Barlow, D.J., 1986. Location of “continuous” antigenic determinants in the protruding regions of proteins. EMBO J. 5, 409–13.

Tonegawa, S., 1976. Proceedings: Determination of the number of antibody structural genes by DNA-RNA hybridization. Hoppe. Seylers. Z. Physiol. Chem. 357, 617.

Twaroch, T.E., Arcalís, E., Sterflinger, K., Stöger, E., Swoboda, I., Valenta, R., 2012. Predominant localization of the major Alternaria allergen Alt a 1 in the cell wall of airborne spores. J. Allergy Clin. Immunol. 129, 1148–9. doi:10.1016/j.jaci.2011.10.008

Unger, A., Stöger, P., Simon-Nobbe, B., Susani, M., Crameri, R., Ebner, C., Hintner, H., Breitenbach, M., 1999. Clinical testing of recombinant allergens of the mold Alternaria alternata. Int. Arch. Allergy Immunol. 118, 220–1. doi:24076

Vagin, A., Teplyakov, A., 1997. MOLREP: an Automated Program for Molecular Replacement. J. Appl. Crystallogr. 30, 1022–1025. doi:doi:10.1107/S0021889897006766

Vailes, L.D., Perzanowski, M.S., Wheatley, L.M., Platts-Mills, T.A., Chapman, M.D., 2001. IgE and IgG antibody responses to recombinant Alt a 1 as a marker of sensitization to Alternaria in asthma and atopic dermatitis. Clin. Exp. Allergy 31, 1891–5.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. doi:10.1093/bioinformatics/btp033

Weiler, J.M., Bonini, S., Coifman, R., Craig, T., Delgado, L., Capão-Filipe, M., Passali, D., Randolph, C., Storms, W., 2007. American Academy of Allergy, Asthma & Immunology Work Group report: exercise-induced asthma. J. Allergy Clin. Immunol. 119, 1349–58. doi:10.1016/j.jaci.2007.02.041

Welling, G.W., Weijer, W.J., van der Zee, R., Welling-Wester, S., 1985. Prediction of sequential antigenic regions in proteins. FEBS Lett. 188, 215–8.

Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–699. doi:10.1093/oxfordjournals.molbev.a003851

Wu, D., Sun, J., Xu, T., Wang, S., Li, G., Li, Y., Cao, Z., 2010. Stacking and energetic contribution of aromatic islands at the binding interface of antibody proteins. Immunome Res 6 Suppl 1, S1. doi:10.1186/1745-7580-6-S1-S1

Wu, P., Dupont, W.D., Griffin, M.R., Carroll, K.N., Mitchel, E.F., Gebretsadik, T., Hartert, T. V., 2008. Evidence of a causal role of winter virus infection during infancy in early childhood asthma. Am. J. Respir. Crit. Care Med. 178, 1123–1129. doi:10.1164/rccm.200804-579OC

Wu, T.T., Kabat, E.A., 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J. Exp. Med. 132, 211–50.

Xu, X., Chang, B.W., Mans, B.J., Ribeiro, J.M.C., Andersen, J.F., 2013. Structure and ligand-binding properties of the biogenic amine-binding protein from the saliva of a blood-feeding insect vector of Trypanosoma cruzi. Acta Crystallogr. D. Biol. Crystallogr. 69, 105–13. doi:10.1107/S0907444912043326

Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H.M., Westbrook, J.D., 2004. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. Acta Crystallogr D Biol Crystallogr 60, 1833–1839. doi:10.1107/S0907444904019419

Ye, Y., Godzik, A., 2004. FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res. 32, W582–5. doi:10.1093/nar/gkh430

Yih, W.T., Siew, L.C., Tan, C.O., Le, Y.Y., Yuen, S.T., Fook, T.C., Sivaraman, J., Yu, K.M., 2009. Structures of two major allergens, Bla g 4 and Per a 4, from cockroaches and their IgE binding epitopes. J. Biol. Chem. 284, 3148–3157. doi:10.1074/jbc.M807209200

Zhang, K.Y.J., Cowtan, K., Main, P., 1997. Macromolecular Crystallography Part B, Methods in Enzymology, Methods in Enzymology. Elsevier. doi:10.1016/S0076-6879(97)77006-X

Zhang, W., Niu, Y., Xiong, Y., Zhao, M., Yu, R., Liu, J., 2012. Computational prediction of conformational B-cell epitopes from antigen primary structures by ensemble learning. PLoS One 7. doi:10.1371/journal.pone.0043575

11. List of Figures

12. List of Tables