

Journal of Chromatography B, 756 (2001) 199-206

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Search for the determinants of allergenicity in proteins of the lipocalin family

Juha Rouvinen^{a,*}, Tuomas Virtanen^b, Rauno Mäntyjärvi^b

^aDepartment of Chemistry, University of Joensuu, P.O. Box 111, FIN-80101 Joensuu, Finland ^bDepartment of Clinical Microbiology, University of Kuopio, P.O. Box 1627, FIN-70211 Kuopio, Finland

Abstract

Three different lines of analysis have been applied to approach the problem of the allergenicity of certain proteins: biological functions, molecular structures and immunological properties. It is immediately obvious that these three are interdependent. The lipocalin family of proteins includes a significant number of allergens. A considerable amount of data is already available of lipocalins and some insights about allergenic determinants can now be presented. However, more information on the molecular structures and immunological parameters of lipocalin allergens is required. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Food allergy; Lipocalin proteins; Epitopes

1. Introduction

In recent years, the understanding of immune reactions during sensitization and in clinical manifestations of an allergy has increased rapidly. Still, the reasons why some particular proteins are able to sensitize and to trigger allergic reactions are not well understood. Biological function, proteolytic enzyme activity in particular, has been presented as properties associated with the allergenicity [1]. Another line of research has been to analyze the molecular characteristics of allergens for clues of allergenicity [2]. It is quite clear, however, that the reason for potential allergenicity of some particular proteins depends on several factors. Since atopic allergy is a misfunction of the immune system, the immunogenic

properties of proteins, in relation to the immune responsiveness of the genetically predisposed persons, are critical. Immunogenic properties are, in turn, defined by molecular characteristics ranging from the sequential level (T-cell epitopes) to the three-dimensional structure level (B-cell epitopes). In this respect, the recognition of linear sequences of allergens by T helper lymphocytes of the Th2 subset is known to be of special importance because Th2 cells are in a decisive role in favoring immunoglobulin E (IgE) synthesis [3], the hallmark of the immediate type of allergy. On the other hand, the immediate type of allergic reactions in a sensitized person are triggered by the binding of an allergen to IgE molecules on mast cells. At this stage, the three-dimensional structure of the protein plays a particularly crucial role. Furthermore, the binding of IgE antibodies to conformational B-cell epitopes and the subsequent internalization and processing by the

E-mail address: juha.rouvinen@joensuu.fi (J. Rouvinen).

 $0378\text{-}4347/01/\$-\text{see front matter}\quad \circledcirc \ 2001 \ Elsevier \ Science \ B.V. \ All \ rights \ reserved.$

PII: S0378-4347(01)00109-8

^{*}Corresponding author.

antigen presenting cells enhances the T-cell reaction by 100- to 1000-fold, as measured in vitro [4,5].

2. Allergenic proteins

Proteins able to induce allergy are present in a wide variety of biological materials. One salient characteristic of an allergen is the route of coming into contact with the host. Food allergens meet the host's immune system in the alimentary canal, respiratory or aeroallergens, e.g., pollen or animal dander, are present in the outdoor or indoor air and cause symptoms predominantly in the respiratory tract. Considerable effort has been made to identify the allergens among the proteins in crude extracts of allergenic materials, and a large pool of data on allergens has accumulated. The current list of known allergens contains about 300 different proteins (ftp:/ /biobase.dk/pub/who-iuis/allergen.list). lergenic capacity, measured as the prevalence of reactivity against an allergen among allergic patients, varies. An arbitrary distinction is being made between major and minor allergens [6]. Major allergens are proteins against which >50% of allergic patients have specific IgE, whereas IgE prevalence against minor allergens is <50%. While the allergenicity of these proteins in terms of IgE reactivity is established, the immune parameters of allergens are only partially known. T-Cell epitopes of several allergens have been characterized but information on B-cell epitopes is almost nil.

A large number of allergens have been cloned and their amino acid sequences are known. Still, the knowledge of the molecular details and especially the data concerning three-dimensional (3D) structures of allergenic proteins is quite limited. In fact, a majority of the papers describing structures of allergenic proteins have been published just in the last few years. The current list of allergens with a known 3D structure includes plant proteins: birch pollen Bet v 1 [7] and Bet v 2 [8], mouse ear cress Ara t [9] and timothy grass pollen Phl p 2 [10,11]; and proteins from animals: house dust mite Der f 2 [12] and Der p 2 [13], mouse urine Mus m 1 [14], and bovine dander Bos d 2 [2] and milk Bos d 5 (B-lactoglobulin) [15]. In addition, horse allergen Equ c 1 has been crystallized [16] and the crystal structure of rat Rat n 1 has been solved at medium resolution but coordinates are not available at the protein data bank [8]. With the exception of the food allergen Bos d 5, all the others cause allergic disorders of the respiratory tract, eyes and skin.

Taking a view on the allergen structures available at the PDB data bank reveals that they represent proteins which are relatively small (molecular mass $11\ 000-22\ 000$). The proteins are structurally quite diverse. They are composed mainly of β -structures, but this does not need to be a requirement of allergenicity. Four of the allergens belong to the same already well characterized family of lipocalins: Mus m 1 [17], Rat n 1 [18], bovine Bos d 2 [19], and Bos d 5 [20]. Other known allergens belonging to this family whose sequences are known but whose structures have not been reported are cockroach Bla g 4 [21], horse dander Equ c 1 [22], dog dander Can f 1 and Can f 2 [23].

3. The lipocalin protein family

Serum retinol binding protein [24] and β-lactoglobulin [25] were the first members of the lipocalin protein family to have their structures published, in 1984 and 1985, respectively. In spite of the weak homology, it was soon found that their 3D structures were similar [26] and the family name of lipocalin was proposed in 1987 [27,28]. Since then, the number of known proteins belonging to this family has increased considerably. The description of the family and the sequence signature are available in the Prosite data base (http://www.expasy.ch/cgibin/prosite-search-ac/ps00213). Lipocalin proteins are typically transport proteins able to bind small, principally hydrophobic ligands. Ligands which have been identified include retinoids and pheromones. Lipocalins present in saliva and nasal secretions probably bind ligands associated with taste or odor perception, respectively. As shown by in vitro experiments, the specificity of the ligand-binding does not need to be very high [29]. In addition to acting as carriers, the members of family have also other functions [30]. Allergenicity, a property not related to the biological function, has been attached to lipocalins in the last few years [19,21].

Lipocalins are quite small proteins. They exist as a monomer or an oligomer. The size of the monomeric unit is about 150–190 residues, thus the sub-unit

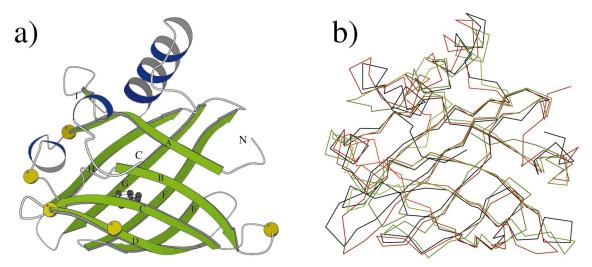


Fig. 1. (a) A three-dimensional structure of major bovine dander allergen Bos d 2. β -Strands are marked with green arrows and helices with blue ribbons. Known glycosylation sites of homologous allergenic lipocalin proteins are marked with yellow balls. (b) The superimposition of a $C\alpha$ -trace of three lipocalin allergens whose three-dimensional structures have been determined. Bos d 2 is in black, Bos d 5 in green and Mus m 1 in red.

molecular mass is approximately 17 000–22 000. If the protein is glycosylated the actual molecular mass can be significantly higher. A typical feature for the lipocalin protein family is the very low degree of amino acid homology. A consistent, definite similarity can be detected only in the beginning of the β-strand A where there is a motif GXW (Figs. 1a) and 2). Lipocalins consist of eight successive βstrands. Each \(\beta\)-strand is hydrogen-bonded to the previous and to the next β-strand until the last strand is hydrogen-bonded to the first strand and the barrel is closed. The central β -barrel is thus composed of eight antiparallel β -strands (A–H) with (+1)₈ topology. The structure may also contain a short Cterminal β -strand (I), short 3_{10} -helices and an α helix after strand H. Functionally, the essential feature is the ligand binding pocket inside the barrel. It is usually very hydrophobic, thus permitting the binding of a small hydrophobic molecule. Hence, lipocalins are able to adopt ligands to the protein's hydrophobic core.

4. Many major allergens are lipocalin proteins

Surprisingly, many major allergens are members of the lipocalin protein family (Fig. 2). Bos d 5 represents the food allergens among lipocalins. Bos d

5 or β-lactoglobulin of cow's milk, together with caseins and lactalbumin, are one of the predominant allergens of milk [31]. β-Lactoglobulins are also present in the milk of other ruminants but not in human milk. The rest of the known lipocalin allergens are respiratory allergens. Rodent lipocalin allergens Mus m 1 and Rat n 1 are urinary proteins acting as pheromone carriers. Can f 1, Can f 2 and Equ c 1 have been cloned from salivary glands or from the von Ebner's glands of the tongue, and the assumption is that they are associated with taste perception. Bos d 2 was cloned from the skin epithelium and the corresponding allergen is present in cow dander, but its biological function remains unknown. The same applies to Bla g 4, a lipocalin allergen of the German cockroach.

As described above, proteins of the lipocalin family are characterized by several common features. Therefore, it is enticing to investigate the factors which make so many lipocalins act as allergens. If these factors are found it could help to outline general rules for allergenicity and to understand the immunological phenomena of allergy. At the moment, it seems likely that allergenicity does not depend on the function of the protein, at least not directly, although the enzymatic activity of a protein may play role in some cases [32,33]. Furthermore, comparisons of primary or tertiary structures of

			=====	A	> =====	B	
Rat-Rat n 1		EEASFE	${\tt RGNLDVDKLN}$	GDWFSIVVAS	DKREKIEENG	-SMRVFVQHI	45
Mouse-Mus m 1		CVHAEEASST	${\tt GRNFNVEKIN}$	GEWHTIILAS	DKREKIEDNG	-NFRLFLEQI	49
Horse-Equ c 1		-QQEENSDVA	IRNFDISKIS	GEWYSIFLAS	DVKEKIEE N G	-SMRVFVDVI	48
Dog-Can f 2		EGNHE	EPQGGLEELS	GRWHSVALAS	$\mathbf{N} \texttt{KSDLIKPWG}$	-HFRVFIHSM	44
Bovine-Bos d 2		QET	PAEIDPSKIP	GEWRIIYAAA	DNKDKIVEGG	-PLRNYYRRI	42
Dog-Can f 1			DTVAVS	GKWYLKAMTA	DQEVPE	-KPDSVTPIL	31
Bovine-Bos d 5		LIVTQT	MKGLDIQKVA	${\tt GTWYSLAMAA}$	SDISLLDAQS	APLRVYVEEL	46
Cockroach Bla g	4	NEDCFRHESL	${\tt VPNLDYERFR}$	GSWIIAAGTS	EALT	-QYKCWIDRF	43
		>	-C> -	D>	E	>F-	
Rat-Rat n 1		DVLENSLG	FTFRIKENGV	CTEFSLVADK	$\mathtt{TAKDGEYF}$	VEYD-GENTF	90
Mouse-Mus m 1		HVLENSLV	LKFHTVRDEE	CSELSMVADK	$\mathtt{TEKAGEYS}$	${\tt VTYD-GFNTF}$	94
Horse-Equ c 1		RALD N -SSLY	${\tt AEYQTKVNGE}$	$\mathtt{CTEFPMVFDK}$	$\mathtt{TEEDGVYS}$	LNYD-GYNVF	94
Dog-Can f 2		SAKDGNLH	${\tt GDILIPQDGQ}$	CEKVSLTAFK	$\mathtt{TATSNKFD}$	LEYW-GHNDL	89
Bovine-Bos d 2		ECINDCESLS	${\tt ITFYLKDQGT}$	${\tt CLLLTEVAKR}$	QEGYVYV	${\tt LEFY-GTNTL}$	88
Dog-Can f 1		LKAQKGGNLE	${\tt AKITMLTNGQ}$	$\mathtt{CQN}\mathtt{ITVVLHK}$	TSEPGKYT	${\tt AYEG-QRVVF}$	78
Bovine-Bos d 5				CAQKKIIAEK			92
Cockroach-Bla g	4	SYDDALV	SKYTDSQGK N	RTTIRGRTKF	EGNKFTIDYN	DKGKAFSAPY	90
			_	H		======	
Rat-Rat n 1		TILKTDYDNY	VMFHLVNVNN	GETFQLME	LYGRTKDLS-	SDIKEKFA	
Mouse-Mus m 1		TILKTDYDNY TIPKTDYDNF	VMFHLVNVNN LMAHLINEKD	GETFQLME GETFQLMG	LYGRTKDLS- LYGREPDLS-	SDIKEKFA SDIKERFA	139
Mouse-Mus m 1 Horse-Equ c 1		TILKTDYDNY TIPKTDYDNF RISEFENDEH	VMFHLVNVNN LMAHLINEKD IILYLVNFDK	GETFQLME GETFQLMG DRPFQLFE	LYGRTKDLS- LYGREPDLS- FYAREPDVS-	SDIKEKFA SDIKERFA PEIKEEFV	139 139
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2		TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ-	SDIKEKFA SDIKERFA PEIKEEFV QDFLPAFE	139 139 134
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2		TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTE	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT-	SDIKEKFA SDIKERFA PEIKEEFV QDFLPAFE PEELEKYQ	139 139 134 132
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1		TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTE RQIRMAK	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS-	SDIKEKFA SDIKERFA PEIKEEFV QDFLPAFE PEELEKYQ QEALEDFR	139 139 134 132 122
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5		TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD-	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 139 134 132 122 137
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD-	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 139 134 132 122 137
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD-	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 139 134 132 122 137
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5 Cockroach-Bla g	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ NGHVIYVQIR	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD- FSVRRFHPKL	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 139 134 132 122 137 140
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5 Cockroach-Bla g	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY ===== KLCVAHGITR	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA I->==== DNIIDLTKT-	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ NGHVIYVQIRDRCLQA	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD- FSVRRFHPKL	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 139 134 132 122 137 140
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5 Cockroach-Bla g Rat-Rat n 1 Mouse-Mus m 1	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY ===== KLCVAHGITR QLCEEHGILR	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA I->=== DNIIDLTKT- ENIIDLSNA-	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ NGHVIYVQIRDRCLQA	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD- FSVRRFHPKL	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 139 134 132 122 137 140
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5 Cockroach-Bla g Rat-Rat n 1 Mouse-Mus m 1 Horse-Equ c 1	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY ===== KLCVAHGITR QLCEEHGILR KIVQKRGIVK	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA I->=== DNIIDLTKT- ENIIDLSNA- ENIIDLTKI-	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ NGHVIYVQIRDRCLQANRCLQAREDRCFQLRG	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD- FSVRRFHPKL	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 134 132 122 137 140 160 166 172
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5 Cockroach-Bla g Rat-Rat n 1 Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY ===== KLCVAHGITR QLCEEHGILR KIVQKRGIVK SVCEDIGLHK	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA I->=== DNIIDLTKT- ENIIDLSNA- ENIIDLTKI- DQIVVLSDD-	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ NGHVIYVQIRDRCLQANRCLQAREDRCFQLRGDRCQGSRD	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD- FSVRRFHPKL	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 134 132 122 137 140 160 166 172 161
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5 Cockroach-Bla g Rat-Rat n 1 Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY ===== KLCVAHGITR QLCEEHGILR KIVQKRGIVK SVCEDIGLHK QLNSERGVPN	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA I->=== DNIIDLTKT- ENIIDLSNA- ENIIDLTKI- DQIVVLSDD- ENIENLIKT-	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ NGHVIYVQIR DRCLQANRCLQAREDRCFQLRGDRCQGSRDDNCPP	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD- FSVRRFHPKL	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 134 132 122 137 140 160 166 172 161 156
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5 Cockroach-Bla g Rat-Rat n 1 Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1	4	TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY ===== KLCVAHGITR QLCEEHGILR KIVQKRGIVK SVCEDIGLHK QLNSERGVPN EFSRAKGL-N	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA I->=== DNIIDLTKT- ENIIDLSNA- ENIIDLTKI- DQIVVLSDD- ENIENLIKT- QEILELAQS-	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ NGHVIYVQIR DRCLQANRCLQAREDRCFQLRGDRCQGSRDDNCPPETCSPGGQ	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD- FSVRRFHPKL NGVAQA	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 139 134 132 122 137 140 160 166 172 161 156 148
Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2 Dog-Can f 1 Bovine-Bos d 5 Cockroach-Bla g Rat-Rat n 1 Mouse-Mus m 1 Horse-Equ c 1 Dog-Can f 2 Bovine-Bos d 2		TILKTDYDNY TIPKTDYDNF RISEFENDEH YLAEVDPKSY EVIHVS-ENM IQPSPVRDHY LVLDTDYKKY SVLATDYENY ===== KLCVAHGITR QLCEEHGILR KIVQKRGIVK SVCEDIGLHK QLNSERGVPN EFSRAKGL-N KALKALPMH-	VMFHLVNVNN LMAHLINEKD IILYLVNFDK LILYMINQYN LVTYVENYDG ILYCEGELHG LLFCMENSAE AIVEGCPAAA I->=== DNIIDLTKT- ENIIDLSNA- ENIIDLTKI- DQIVVLSDD- ENIENLIKT- QEILELAQSIRLSFNPTQ	GETFQLME GETFQLMG DRPFQLFE DDTSLVAH ERITKMTERQIRMAK PEQSLVCQ NGHVIYVQIR DRCLQANRCLQAREDRCFQLRGDRCQGSRDDNCPPETCSPGGQ LEEQCHI	LYGRTKDLS- LYGREPDLS- FYAREPDVS- LMVRDLSRQ- GLAKGTSFT- LLGRDPEQS- CLVRTPEVD- FSVRRFHPKL NGVAQA	SDIKEKFASDIKERFAPEIKEEFVQDFLPAFEPEELEKYQQEALEDFRDEALEKFD	139 134 132 122 137 140 160 166 172 161 156

Fig. 2. The alignment of lipocalin allergens whose amino acid sequences are available. Putative N-glycosylation sites are shown in boldface. The first row shows the secondary structure elements of Bos d 2: \rightarrow β -strand; == 3_{10} - or α -helix.

different allergens have not revealed any clear motives or features which could explain allergenicity. On the other hand, amino acid sequences and the tertiary structures determine elements recognized by immunocompetent cells and by antibodies. Sequence comparison of lipocalin allergens reveals a family where homology is very low, only residues G and W in the GXW triplet of the signature sequence, located in the β -strand A, are clearly conserved (Fig. 2). It is noteworthy that the sequence comparison does not reveal any long insertions or deletions, an indication of a very similar 3D structure and shape (Fig. 1b).

It is clear that the understanding of T-cell (sequential) epitopes or B-cell (conformational) epitopes and especially their mutual role in the formation of allergenicity is critical. At the moment, the information on the T-cell and B-cell epitopes of lipocalin allergens and of their relationship to the molecular structure is too limited to warrant meaningful attempts at defining allergenic determinants. In addition to identifying both sequential and conformational epitopes, it is necessary to assess the results of the immunological studies with reference to the molecular characteristics. Especially in the case of B-cell

epitopes, 3D structures are essential both for the localization of the IgE-binding regions and for the comparative structural analyses of lipocalin allergens.

5. T-Cell epitopes

The recognition of allergens as processed linear peptides by the receptor of T helper 2 (Th2) type cells is an essential prerequisite for allergenicity. Th2 cells recognize peptides in conjunction with major histocompatibility class II molecules on the surface of antigen-presenting cells. The MHC restriction of the presentation of several pollen allergens has been reported [34] but the information of lipocalin allergens is limited [35,36]. The information concerning T-cell epitopes of lipocalins is also scarce. T-Cell epitopes of Bos d 2 were mapped recently by using 16-mer peptides. Seven different epitopes were found. In the 3D structure, these epitopes are located mostly in the regions of secondary structural elements, although it can be pointed out that the lipocalin structure has a high proportion of secondary structural elements and less loops or irregular elements. Two of the epitope regions of Bos d 2 are preferentially recognized by T-cell lines and clones [35]. One corresponds to the region 13–26 (β-strand A) and the second one to the region 127-138 (Cterminal α-helix).

An analysis of human T-cell epitopes of \(\beta-lactoglobulin (Bos d 5) was made by Piastra et al. [36] by using the priming of cord mononuclear cells with β-lactoglobulin and subsequent stimulation with CNBr peptides. The highest number of positive reactions was observed with 145-161. The murine T-cell epitopes of Bos d 5 have been determined in three different mouse strains [37-39]. The three studies with BALB/c mice suggested that Bos d 5 contains at least 3-4 regions with T-cell stimulatory capacity, roughly corresponding to 42-47, 67-76, 107-117 and 134-148 [37-39]. These regions correspond again to secondary structure elements, namely, β-strands B, D, and H as well as C-terminal α-helix. The most vigorous response was observed against 67-76 but 134-148, which corresponds to the Cterminal α-helix (130–140) and is therefore equivalent with the epitope G of Bos d 2, also induced a good response [38,39].

6. B-Cell epitopes

Data on the B-cell epitopes are available only for a few lipocalin allergens. Because of its significance as a food allergen, the IgE binding of Bos d 5 has been studied extensively, mostly by using protein fragments. In a recent study, regions 41-60 (β -strands B and C), 102-124 (β -strands F and G) and 149-162 (C-terminus with a short β -strand and 3_{10} -helix) were identified as major epitopes [40]. In a more detailed analysis with synthetic peptides, region 97–108 (β -strand G) was recognized by all the patients [20]; a region roughly corresponding to that was identified as a minor epitope (92–100) by Sélo et al. [40].

Another approach to analyze the IgE binding of an allergen is to start with minor modifications of the complete molecule. If the modifications cause changes in the 3D structure, they can also be expected to affect the IgE binding. Intramolecular disulfide bridges are one of the characteristics of lipocalins. It has been reported of mite allergens which also have similar bridges that the disruption of the bridges interferes with the IgE binding [41–44]. A Cys(154)-Ala mutation in the recombinant Bos d 2 caused a significant decrease of the IgE binding when analyzed by enzyme-linked immunosorbent assay (ELISA) inhibition [45]. The loss of a disulfide bridge probably causes only a minor change in conformation, and therefore the IgE epitopes of Bos d 2 seem to be highly dependent on the 3D structure.

7. Glycosylation

Of the allergenic lipocalins, Bla g 4, Can f 1, Can f 2 have one and Equ c 1 two N-glycosylation sites. On the other hand, Bos d 2, Bos d 5, Mus m 1, and Rat n 1 are not glycosylated. This would suggest that glycosylation is not critical for allergenicity. On the other hand, it is obvious that the glycan moiety can affect the IgE binding of an allergen [46]. Our preliminary results have indicated that deglycosylation of Can f 2 would reduce IgE binding but more

studies are needed to confirm and explain these results [47]. As a theoretical consideration, it is possible that the glycan can also interfere with the IgE binding. The size of the N-glycan moiety can be large, especially when compared to a small protein, and thus it is quite possible that glycosylation may affect the binding of IgE. The N-glycosylation sites of the currently known lipocalin allergens are all located on one side of a molecule (Fig. 1a), especially in the β -strand D. If N-glycans in this area "shadow" the lipocalin protein surface, the epitopes available for antibodies would be located on the

other site of the protein, for example around the C-terminal α -helix.

8. Homology with endogenous lipocalins

Recently, a proposal was made that the molecular mimicry at the T-cell level between allergenic and endogenous lipocalins could explain why exposure to lipocalins results in allergy [47]. At least two human lipocalins which have a distinct homology with allergenic proteins can be found. Bos d 5 has a 43%

```
---A ---->
                                        === ==
Bos d 5
           LIVTQTMKGL DIQKVAGTWY SLAMAASDIS LLDAQSAPLR .VYVEELKPT
                                                                    49
paep human MDIPOTKODL ELPKLAGTWH SMAMATNNIS LMATLKAPLR .VHITSLLPT
                                                                    49
Can f 1
           ..... DTVAVSGKWY LKAMTADQ.. ..EVPEKPDS .VTPMILKAQ
                                                                    35
vegp human ..HHLLASDE EIODVSGTWY LKAMTVDR.. ..EFPEMNLE SVTPMTLTTL
                                                                    44
               ----C- ->
                           ----> D---->
                                             ---E-->
Bos d 5
           PEGDLEILLQ KWENGECAQK KIIAEKTKIP AVFKIDALNE NKVLVLDTDY
                                                                    99
paep human PEDNLEIVLH RWENNSCVEK KVLGEKTENP KKFKINYTVA NEATLLDTDY
                                                                    99
Can f 1
           KGGNLEAKIT MLTNGQCQNI TVVLHKTSEP GKYTAYEGQR VVFIQPSPVR
                                                                    85
vegp_human EGGNLEAKVT MLISGRCQEV KAVLEKTDEP GKYTADGGKH VAYIIRSHVK
                                                                    94
             ---G--->
                               - H-->
                                            Bos d 5
           KKYLLFCMEN SAEPEQSLAC QCLVRTPEVD DEALEKFDKA LKALPMH... 146
paep_human DNFLFLCLQD TTTPIQSMMC QYLARVLVED DEIMQGFIRA FRPLPRH... 146
           DHYILYCEGE LHGR-QIRMA KLLGRDPEQS QEALEDFREF SRAKGLNQE. 133
vegp_human DHYIFYCEGE LHGK-PVRGV KLVGRDPKNN LEALEDFEKA AGARGLSTES 143
           -I->
Bos d 5
                                                                   162
           IRLSFNPTQL EEQCHI...
paep_human LWYLLDLKQM EEPCRF...
                                                                   162
Can f 1
           ILELAQ.... SETCSPGGQ
                                                                   148
vegp_human ILIPRQ.... SETCSPGSD
                                                                   158
```

Fig. 3. The alignment of sequences of two lipocalin allergens (Bos d 5 and Can f 1) which are homologous with human endogenous lipocalins. Identical areas are marked with pink between pairs. In addition, if identical residues are found from the other (less homologues) allergen they are marked with blue. The secondary structure elements of Bos d 5 are on the first row and the same symbols as in Fig. 2 have been used.

identity with glycodelin or pregnancy-associated endometrial alpha-2-globulin (PAEP) and Can f 1 a 56% identity with tear lipocalin or von Ebner's gland protein (VEGP). We have performed a sequence comparison between both protein pairs (Fig. 3). As could be expected, there are clearly peptides several residues long which are identical between the allergens and human endogenous lipocalins. The mimicry model [47] assumes that autoreactive Th2 cells which present in low numbers in the periphery recognize, with a low avidity, self-resembling epitopes of allergens. The identical stretches are located in the secondary structure elements but also in some loops. This results in a clear similarity at the 3D structure level even though the level of the amino acid homology between lipocalins can be weak.

9. Concluding remarks

The fact that allergenicity is a property shared by several lipocalins offers a unique opportunity to approach the question of the determinants which make proteins allergenic. From the information available, it is already possible to say that allergenicity is not an immediately recognizable feature, even if the complete molecular structure of an allergen is known. Thus, there is no clear evidence for the factors which make so many lipocalins major inducers of allergy. The combination of immunological and biochemical data is essential for further analyses. A contributing factor may be that lipocalins are often present in secretions and therefore easily spread in the environment. However, the immunobiology of allergens is the key, and therefore the combination of immunological and biochemical data is essential for the further analysis. More detailed information about T-cell and B-cell epitopes are needed. When more 3D structures of allergenic and endogenous lipocalins are available more detailed structural comparisons will be possible in relation to allergenicity.

References

 T. Musu, C. Gregoire, B. David, J.-P. Dandeau, Clin. Rev. Allergy Immunol. 15 (1997) 485.

- [2] J. Rouvinen, J. Rautiainen, T. Virtanen, T. Zeiler, J. Kauppinen, A. Taivainen, R. Mäntyjärvi, J. Biol. Chem. 274 (1999) 2337.
- [3] S. Romagnani, Immunol. Today 18 (1997) 263.
- [4] D. Maurer, C. Ebner, B. Reininger, E. Fiebiger, D. Kraft, J.P. Kinet, G. Stingl, J. Immunol. 154 (1995) 6285.
- [5] F.L. van der Heijden, R.J. Joost van Neerven, M. van Katwijk, J.D. Bos, M.L. Kapsenberg, J. Immunol. 150 (1993) 3643.
- [6] J.N. Larsen, H. Lœwenstein, J. Allergy Clin. Immunol. 97 (1996) 577.
- [7] M. Gajhede, P. Osmark, F.M. Poulsen, H. Ipsen, J.N. Larsen, R.J. van Neerven, C. Schou, H. Løwenstein, M.D. Spangfort, Nat. Struct. Biol. 3 (1996) 1040.
- [8] A.A. Fedorov, T. Ball, N.M. Mahoney, R. Valenta, S.C. Almo, Structure 5 (1997) 33.
- [9] K.S. Thorn, H.E.M. Christensen, R. Shigeta Jr., D. Huddler Jr., L. Shalaby, U. Lindberg, N.H. Chua, C.E. Schutt, Structure 5 (1997) 19.
- [10] A.A. Fedorov, T. Ball, R. Valenta, S.C. Almo, Int. Arch. Allergy Immunol. 113 (1997) 109.
- [11] S. De Marino, M.A. Castiglione Morelli, F. Fraternali, E. Tamborini, G. Musco, S. Vrtala, C. Dolecek, P. Arosio, R. Valenta, A. Pastore, Structure 7 (1999) 943.
- [12] S. Ichikawa, H. Hatanaka, T. Yuuki, N. Iwamoto, S. Kojima, C. Nishiyama, K. Ogura, Y. Okumura, F. Inagaki, J. Biol. Chem. 273 (1998) 356.
- [13] G.A. Mueller, D.C. Benjamin, G.S. Rule, Biochemistry 37 (1998) 12707.
- [14] Z. Böcskei, C.R. Groom, D.R. Flower, C.E. Wright, S.E.V. Phillips, A. Cavaggioni, J.B.C. Findlay, A.C.T. North, Nature 360 (1992) 186.
- [15] S. Brownlow, J.H.M. Cabral, R. Cooper, D.R. Flower, S.J. Yewdall, I. Polikarpov, A.C.T. North, L. Sawyer, Structure 5 (1997) 481.
- [16] C. Grégoire, G.A. Tavares, H.K. Lorenzo, J.-P. Dandeu, B. David, P.M. Alzari, Acta Cryst. D55 (1999) 880.
- [17] J.R. Lorusso, S. Moffat, J.L. Ohman Jr., J. Allergy Clin. Immunol. 78 (1986) 928.
- [18] C. Bayard, L. Holmquist, O. Vesterberg, Biochem. Biophys. Acta 1290 (1996) 129.
- [19] R. Mäntyjärvi, S. Parkkinen, M. Rytkönen, J. Pentikäinen, J. Pelkonen, J. Rautiainen, T. Zeiler, T. Virtanen, J. Allergy Clin. Immunol. 97 (1996) 1297.
- [20] G. Ball, M.J. Shelton, B.J. Walsh, D.J. Hill, C.S. Hosking, M.E.H. Howden, Clin. Exp. Allergy 24 (1994) 758.
- [21] L.K. Arruda, L.D. Vailest, M.L. Hayden, D.C. Benjamin, M.D. Chapman, J. Biol. Chem. 270 (1995) 31196.
- [22] C. Gregoire, I. Rosinski-Chupin, J. Rabillon, P.M. Alzari, B. David, J.-P. Dandeu, J. Biol. Chem. 271 (1996) 32951.
- [23] A. Konieczny, J.P. Morgenstern, C.B. Bizinkauskas, C.H. Lilley, A.W. Brauer, J.F. Bond, R.C. Aalberse, B.P. Wallner, M.T. Kasaian, Immunology 92 (1997) 577.
- [24] M.E. Newcomer, T.A. Jones, J. Agvist, J. Sundelin, U. Eriksson, L. Rask, P.A. Peterson, EMBO J. 3 (1984) 1451.
- [25] L. Sawyer, M.Z. Papiz, A. North, E. Eliopoulos, Biochem. Soc. Trans. 13 (1985) 265.

- [26] M.Z. Papiz, L. Sawyer, E. Eliopoulos, A. North, J. Findlay, R. Sivaprasadarao, T.A. Jones, M.E. Newcomer, P. Kraulis, Nature 324 (1986) 383.
- [27] S. Pervaiz, K. Brew, FASEB J. 1 (1987) 209.
- [28] L. Sawyer, Nature 327 (1987) 659.
- [29] D.R. Flower, J. Mol. Recognit. 8 (1995) 185.
- [30] D.R. Flower, Biochem. J. 318 (1996) 1.
- [31] J.M. Wal, Allergy 53 (Suppl. 46) (1998) 114.
- [32] A. Bufe, Int. Arch. Allergy Immunol. 117 (1998) 215.
- [33] C.R.A. Hewitt, S. Foster, C. Phillips, H. Horton, R.M. Jones, A.P. Brown, B.J. Hart, D.I. Pritchard, Allergy 53 (1998) 60.
- [34] D.G. Marsh, in: D.G. Marsh, M.N. Blumenthal (Eds.), Genetic and Environmental Factors in Clinical Allergy, University of Minnesota Press, Minneapolis, MN, 1990, p. 97.
- [35] T. Zeiler, R. Mäntyjärvi, J. Rautiainen, M. Rytkönen-Nissinen, P. Vilja, A. Taivainen, J. Kauppinen, T. Virtanen, J. Immunol. 162 (1999) 1415.
- [36] M. Piastra, A. Stabile, G. Fioravanti, M. Castagnola, G. Pani, F. Ria, Int. Arch. Allergy Immunol. 104 (1994) 358.
- [37] T. Takahashi, S. Kaminogawa, T. Kuwata, O. Ando, K. Yamauchi, Agric. Biol. Chem. 52 (1988) 2485.

- [38] N.M. Tsuji, J. Kurisaki, K. Mizumachi, S. Kaminogawa, Immunol. Lett. 37 (1993) 215.
- [39] M. Totsuka, A. Ametani, S. Kaminogawa, Cytotechnology 25 (1997) 101.
- [40] I. Sélo, G. Clément, H. Bernard, J.-M. Chatel, C. Créminon, G. Peltre, J.-M. Wal, Clin. Exp. Allergy 29 (1999) 989.
- [41] S. Olsson, M. van Hage-Hamsten, P. Whitley, Mol. Immunol. 35 (1998) 1017.
- [42] G.A.J. Hakkaart, R.C. Aalberse, R. van Ree, Allergy 53 (1998) 165.
- [43] C. Nishiyama, M. Fukuda, Y. Usui, N. Iwamoto, T. Yuuki, Y. Okumura, H. Okudaira, Mol. Immunol. 32 (1995) 1021.
- [44] A.M. Smith, M.D. Chapman, Clin. Exp. Allergy 27 (1997) 593
- [45] J. Kauppinen, T. Zeiler, J. Rautiainen, M. Rytkönen-Nissinen, A. Taivainen, R. Mäntyjärvi, T. Virtanen, Clin. Exp. Allergy 29 (1999) 989.
- [46] R. van Ree, R.C. Aalberse, J. Allergy Clin. Immunol. 103 (1999) 1000.
- [47] T. Virtanen, T. Zeiler, J. Rautiainen, R. Mäntyjärvi, Immunol. Today 20 (1999) 398.