ORIGINAL ARTICLE EXPERIMENTAL ALLERGY AND IMMUNOLOGY

Allergen cross reactions: a problem greater than ever thought?

P. Pfiffner[1], R. Truffer[1], P. Matsson[2], C. Rasi[3], A. Mari[3,4] & B. M. Stadler[1]

[1]University Institute of Immunology, Bern, Switzerland; [2]Phadia AB, Uppsala, Sweden; [3]Center for Clinical and Experimental Allergology, IDI-IRCCS, Rome; [4]Allergy Data Laboratories s.c., Latina, Italy

To cite this article: Pfiffner P, Truffer R, Matsson P, Rasi C, Mari A, Stadler BM. Allergen cross reactions: a problem greater than ever thought? *Allergy* 2010; **65**: 1536–1544.

Keywords
allergens; bioinformatics; sensitization; sequence motifs; specific IgE.

Correspondence
Pascal Pfiffner, University Institute of Immunology, Sahli Haus 2, Inselspital, 3010 Bern, Switzerland.
Tel.: +41 31 632 22 89
Fax: +41 31 632 35 00
E-mail: pascal.pfiffner@iib.unibe.ch

Accepted for publication 4 May 2010

DOI:10.1111/j.1398-9995.2010.02420.x

Edited by: Reto Crameri

Abstract

Background: Cross reactions are an often observed phenomenon in patients with allergy. Sensitization against some allergens may cause reactions against other seemingly unrelated allergens. Today, cross reactions are being investigated on a per-case basis, analyzing blood serum specific IgE (sIgE) levels and clinical features of patients suffering from cross reactions. In this study, we evaluated the level of sIgE compared to patients' total IgE assuming epitope specificity is a consequence of sequence similarity.

Methods: Our objective was to evaluate our recently published model of molecular sequence similarities underlying cross reactivity using serum-derived data from IgE determinations of standard laboratory tests.

We calculated the probabilities of protein cross reactivity based on conserved sequence motifs and compared these *in silico* predictions to a database consisting of 5362 sera with sIgE determinations.

Results: Cumulating sIgE values of a patient resulted in a median of 25–30% total IgE. Comparing motif cross reactivity predictions to sIgE levels showed that on average three times fewer motifs than extracts were recognized in a given serum (correlation coefficient: 0.967). Extracts belonging to the same motif group co-reacted in a high percentage of sera (up to 80% for some motifs).

Conclusions: Cumulated sIgE levels are exaggerated because of a high level of observed cross reactions. Thus, not only bioinformatic prediction of allergenic motifs, but also serological routine testing of allergic patients implies that the immune system may recognize only a small number of allergenic structures.

Determination of specific IgE in patient sera is a valuable test for allergologists (1). The number of potential allergens is steadily increasing (2) and suppliers of allergy tests are providing ever longer lists of allergenic preparations to be used for *in vitro* assays. In most instances, allergens are still relatively crude extracts of organisms or parts thereof (3). Recently, allergen diagnosis has improved by the use of highly purified natural or recombinant allergens and protein microarrays (3–8). This may improve allergy diagnostics in the future.

Cross reactions are allergic reactions against other allergens without prior sensitization. They have been extensively studied and a handful of well-defined cross reactivity syndromes are clinically highly important, e.g., the pollen-food syndromes (9). Cross reactions between recombinant allergens are also documented (3, 7, 10). Thus, the immune system might recognize common structures, allowing to predict allergic reactions that have not been tested physically but were derived by similarity.

We have previously shown that a bioinformatic approach is capable to define a much lower number of potentially allergenic structures, termed motifs, than the number of known protein sequences of allergens (11). These motifs represent a scaled profile over a window of 50 amino acids, derived from all currently known allergen protein sequences. They serve as an identifier for evolutionary conserved protein domains. Consequently, if protein sequences match a given motif, these proteins are predicted to fold into the same protein domain and therefore exhibit similar surface structures. We showed that

this method of cross reactivity prediction is superior to the FAO/WHO rule, which states that a protein is allergenic if it has either an identity of at least six continuous amino acids or more than 35% sequence similarity over a window of 80 amino acid residues. Especially in view of false positive matches (67.3% of all Swiss-Prot proteins were predicted to be allergenic by the FAO/WHO rule), the motif-based approach performed much better (2.6% predicted to be allergenic) (11).

Thus, the question remains whether the in silico prediction of allergenicity may be confirmed by wet lab data. For this purpose, we have analyzed 5362 sera corresponding to 203 283 specific IgE determinations. We could demonstrate that the degree of cross reaction was greater than ever thought.

Materials and methods

Serum samples

Data on 5456 serum samples were obtained by testing for IgE using Phadia's ImmunoCAP® (former UniCAP®, Phadia AB, Uppsala, Sweden) systems. These are sandwich immunoassay systems where serum IgE antibodies react with anti-IgE covalently coupled to the system in case of total IgE determination or with solid-phase bound allergen extracts to determine specific IgE. Bound antibodies are detected and quantified using enzyme-labeled anti-IgE-antibodies and fluorescence detection.

Tests were performed in the years from 1988 to 2006 in 17 different countries in different laboratories. Raw, anonymized IgE data (no age, sex, and other demographic and clinical information) were collected as quality assurance; therefore, no selection criteria were applied. Test results were collected in a clinical setting; most sera are presumably from patients with atopy.

All IgE levels are expressed in kilo units of antigen per liter serum (kUA/l). Specific IgE levels >0.35 kUA/l (Class I and higher) were regarded as a positive test result, levels >100 kUA/l were capped at 100 kUA/l, which affected 1578 values.

Included in the database were serum levels for 99 allergens as well as the total IgE level. According to the manufacturer, the 99 allergen extracts used to determine the specific IgE values are the 99 most tested allergens among a list of more than 700 allergens available in Phadia's catalog. Table 1 lists the extracts and groups them into major subsets.

Sera had to be tested for total IgE, against at least 10 different allergens and yield at least one positive specific IgE test result to be allowed for the final database. With a total of 203 283 specific IgE tests, 5362 sera met our criteria and were used for the analysis.

Databases and software

We created a MySQL database to hold the serum data (MySQL 5.0, obtained from http://www.mysql.com/). Allergen protein sequences were extracted from the Allergome database (http://www.allergome.org/ as of January 2009). MEME 3.5.7 (12) (obtained from http://meme.sdsc.edu/meme/) and pftools 2.3.4 (13) (obtained from http://www.isrec.isb-sib.ch/ftp-server/pftools/) were used for the iterative allergen motif discovery. Perl 5.8.8 (http://www.perl.org/), PHP 5.2+ (http://www.php.net/), and R 2.8 (14) (http://www.r-project.org/) scripts were created to extract the desired statistical calculations.

Allergen motifs

We performed the iterative allergen motif discovery according to Stadler and Stadler. (11) using 2189 protein sequences from Allergome. These sequences are known to encode allergenic proteins (2). We identified 97 motifs with a residue length of 50 amino acids. Three hundred and four of the sequences used for identification did not match an allergen motif, 96 thereof were shorter than 50 amino acids in length and therefore could not match a motif, and 26 additional proteins were known to only encode protein fragments.

To identify the motifs present in each allergen extract, the proteins used in the motif identification process had to be linked to extracts in which they occur. This linking was achieved by first matching the allergen extract to the corresponding allergen source within Allergome and then assigning the proteins to the extract as defined by Allergome (15).

Our database now made it possible to computationally determine which motifs occur in which extract(s), and therefore which extracts are likely to cross react due to their structural similarity.

Results

The serum database

We used the data of specific and total IgE determinations comprising a total of 5362 sera. Figure 1A shows the distribution of IgE levels in this serum collection. As expected for a serum collection of allergic individuals, total IgE levels peaked above the threshold value of 100 kUA/l (16), namely between 200 and 400 kUA/l. For our study purpose, it was not necessary to define a lower cut-off of total IgE levels.

Most sera were from the US (2264) and Sweden (2119), from Western European countries (635) and Russia (230) while the residual sera were either obtained from Japan, Southern Africa or Canada (81) or its origin was unspecified (33).

All sera were tested against a subset of a panel consisting of 99 allergens, as described in Table 1. Among these sera, 1471 sera have been tested against 90 or more of the allergens and 3448 sera against 10–30 allergens.

We used the traditional cutoff of 0.35 kUA/l (Class I or higher) to test positively against a given allergen extract. This yielded 99 276 positive values representing IgE from classes I–IV from a total of 203 283 allergen-specific IgE determinations.

Despite the similar number of tests that were performed (Fig. 1B), Fig. 1C shows that the percentage of positive specific IgE tests steadily increased with increasing total IgE, resulting in more than 90% positive tests for sera with very high total IgE (>3200 kUA/l).

Table 1 The allergen extracts used to determine specific IgE levels. Identifier for extracts is Allergome's accession number and Phadia's product code. The number of known proteins (excluding isoforms, within brackets the total number of proteins including isoforms) is estimated based on available sequence data. The number of distinct motifs present within the extract is derived from contained proteins

Group	Extract	Allergome accession number	Phadia product code	# known proteins in extract	# motifs associated with extract
Epidermals	Cat epithelium and dander	1819	e1	6 (8)	3
	Dog dander	1756	e5	5 (7)	4
	Guinea pig epithelium	1765	e6	2 (2)	0
	Horse dander	1813	e3	5 (6)	3
	Mouse epithelium; serum proteins and urine proteins	1881	e88	2 (2)	2
	Rat epithelium; serum and urine proteins	1957	e87	2 (2)	2
Foods of animal origin	Beef	2019	f27	5 (5)	3
	Blue mussel	1413	f37	1 (1)	1
	Chicken	2037	f83	0	0
	Egg white	1832	f1	6 (8)	4
	Fish (cod)	1831	f3	1 (2)	1
	Milk	1747	f2	7 (11)	6
	Pork	2088	f26	0	0
	Scallop	2012	f338	3 (3)	2
	Shrimp	1893	f24	0	0
	Tuna	2375	f40	0	0
Foods of plant origin	Almond	1948	f20	4 (5)	2
	Apple	1871	f49	4 (71)	4
	Banana	1882	f92	1 (1)	1
	Barley	2040	f6	7 (13)	4
	Brazil nut	1738	f18	2 (2)	2
	Buckwheat	1816	f11	4 (9)	3
	Cacao	1369	f93	0	0
	Carrot	1799	f31	2 (8)	2
	Celery	1721	f85	4 (5)	2
	Cherry	1946	f242	4 (7)	3
	Coconut	3559	f36	1 (1)	0
	Garlic	1706	f47	1 (2)	1
	Gluten	651	f79	0	0
	Hazel nut	2028	f17	7 (17)	5
	Kiwi	1697	f84	8 (21)	3
	Maize; Corn	2092	f8	4 (5)	3
	Oat	2018	f7	0	0
	Onion	1704	f48	1 (1)	1
	Orange	1774	f33	3 (6)	2
	Pea	1931	f12	2 (4)	1
	Peach	1949	f95	3 (6)	2
	Peanut	1723	f13	13 (36)	9
	Potato	1977	f35	6 (19)	4
	Rice	2058	f9	2 (2)	1
	Rye	2076	f5	1 (2)	0
	Sesame seed	1971	f10	7 (7)	3
	Soya bean	1834	f14	13 (32)	7
	Strawberry	2251	f44	3 (14)	3
	Tomato	1870	f25	5 (12)	5
	Wheat	1993	f4	13 (29)	8
	White bean	1923	f15	0	0
	Yeast	1960	f45	2 (2)	1

Table 1 (Continued)

Group	Extract	Allergome accession number	Phadia product code	# known proteins in extract	# motifs associated with extract
Grass pollens	Bermuda grass	1796	g2	9 (22)	6
	Cocksfoot	1798	g3	5 (16)	2
	Common reed	1927	g7	0	0
	Johnson grass	1979	g10	1 (1)	0
	Sweet vernal grass	1718	g1	1 (2)	0
	Timothy	1924	g6	10 (36)	8
Insects	Cockroach; German	1742	i6	11 (115)	10
Microorganisms	*Alternaria alternata (tenuis)*	1708	m6	12 (16)	10
	Aspergillus fumigatus	1730	m3	24 (27)	14
	Aureobasidium pullulans	2197	m12	0	0
	Botrytis cinerea	1630	m7	0	0
	Candida albicans	1757	m5	3 (3)	2
	Cladosporium herbarum (Hormodendrum)	1775	m2	12 (13)	7
	Epicoccum purpurascens	1810	m14	2 (2)	1
	Fusarium moniliforme	1554	m9	0	0
	Helminthosporium halodes	1125	m8	0	0
	Mucor racemosus	2291	m4	0	0
	Penicillium notatum	1912	m1	5 (8)	2
	Phoma betae	2303	m13	0	0
	Rhizopus nigricans	1622	m11	0	0
	Stemphylium botryosum	2637	m10	1 (1)	1
Mites	*Dermatophagoides pteronyssinus*	1803	d1	16 (32)	12
	Dermatophagoides farinae	1801	d2	20 (72)	14
Tree pollens	American beech	2249	t5	0	0
	Box-elder	2136	t1	0	0
	Common silver birch	1741	t3	6 (67)	6
	Cottonwood	2324	t14	0	0
	Elm	2385	t8	0	0
	Japanese cedar	1784	t17	5 (31)	6
	Maple leaf sycamore; London plane	1932	t11	3 (4)	3
	Mountain juniper	1851	t6	3 (4)	3
	Oak	1955	t7	1 (5)	1
	Olive	1888	t9	11 (97)	6
	Walnut	2044	t10	0	0
	White ash	2253	t15	0	0
	White pine	2312	t16	0	0
	Willow	2355	t12	0	0
Venoms	Common wasp (Yellow jacket)	2008	i3	4 (5)	4
	Honey bee	1722	i1	9 (13)	8
Weed pollens	Cocklebur	2401	w13	0	0
	Common ragweed	1710	w1	9 (17)	5
	Dandelion	6146	w8	1 (1)	1
	Goosefoot; Lamb's quarters	1768	w10	3 (3)	3
	Marguerite; Ox-eye daisy	1567	w7	0	0
	Mugwort	1728	w6	6 (25)	7
	Nettle	2390	w20	0	0
	Plantain (English); Ribwort	1933	w9	0	0
	Saltwort (prickly); Russian thistle	1961	w11	2 (3)	1
	Scale; Lenscale	2193	w15	0	0
	Sheep sorrel	2353	w18	0	0
	Wall pellitory (*Parietaria officinalis*)	1906	w19	1 (8)	0
	Wall pellitory (*Parietaria judaica*)	1904	w21	4 (9)	3
	Western ragweed	1711	w2	1 (2)	1

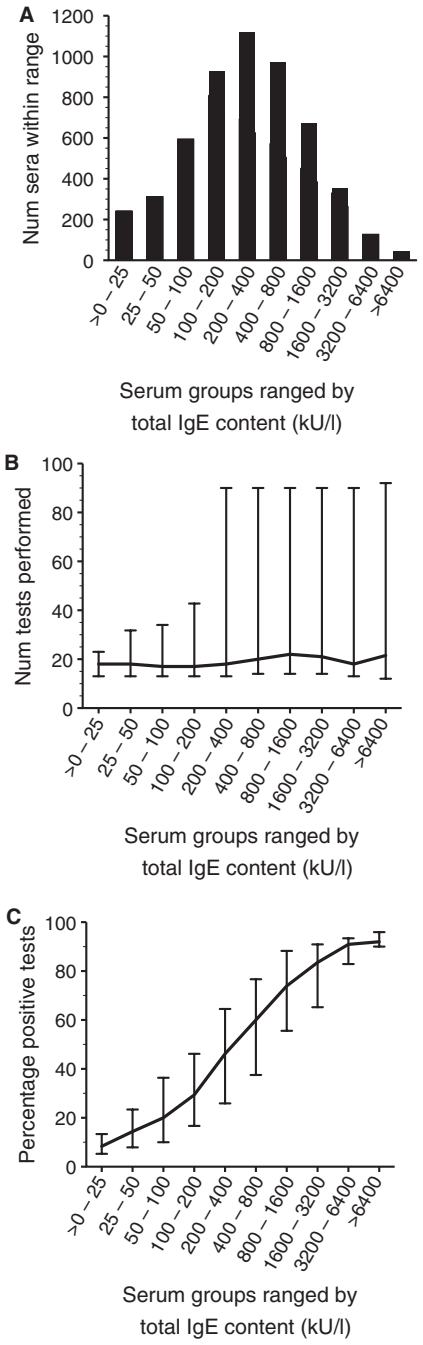

Figure 1 Grouping the sera by their total IgE content resulted in the distribution shown in (A). The amount of total IgE had no influence on the number of tests being performed (B); however, it affected the outcome of the tests (C). The lines in B and C indicate the median, the error bars represent the 25th and 75th percentile, respectively.

Cumulated specific IgE

Next, we cumulated all individual specific IgE levels within the 5362 sera and plotted them against the total serum IgE (Fig. 2A). The data show that in 88.95% of the sera, total serum IgE exceeds the cumulated specific IgE. Figure 2B shows the percentage of cumulated specific IgE if sera were grouped according to total IgE levels. Of all data, 90.2% lay within a range between >0 and 1600 kUA/l and in this range, specific IgE was between 25% and 30% of total IgE.

Interestingly, the residual 10% of sera with more than 1600 kUA/l total IgE showed a decreasing percentage of specific IgE. Thus, even though high total IgE sera at a greater number result in positive tests (Fig. 1C), their specific cumulated IgE fraction was lower (Fig. 2B). However, sIgE determinations were capped at 100 kUA/l and we found that with increasing total IgE sera were more likely to contain sIgE values affected by this capping (Fig. 2C), which would underestimate the percentage of cumulated specific IgE in these sera.

Cross reactivity by motifs

Based on the most recent sequence database, we found 64 of the 99 tested allergens to contain proteins containing our defined motifs. Some extracts contain more than one motif resulting in 86 motifs to be associated with the tested allergen extracts. In 13 extracts, we found only one motif (Table 1). On the other hand, extract d2 (House Dust Mite, *Dermatophagoides farinae*) contained 14 motifs, m3 (*Aspergillus fumigatus*) also 14, d1 (House Dust Mite, *Dermatophagoides pteronyssinus*) 12, i6 (German Cockroach) 10, and so on. Excluding those extracts without a known motif, we found a median of three motifs per extract.

Figure 3A shows that also the number of recognized motifs steadily increased with higher IgE levels. On the other hand, Fig. 3B shows that there are approximately three times less motifs recognized than extracts from different allergenic sources (correlation coefficient: 0.967).

The question remained whether this relation between motifs and extracts is linear and is directly due to cross reactions. Our next assumption was simple: If different extracts from different sources theoretically contain the same motif, one would expect a cross reaction. For this purpose, we analyzed all motifs that occurred in three or more different extracts. Table 2 shows all motifs that fulfill this criterion. For example, it is seen that motif 1 (corresponding to the group of the Bet v 1 allergen) occurs in 25 extracts of our allergen panel.

Figure 4 depicts two examples from the list in Table 2. We have chosen motif 4 occurring in 10 different extracts and motif 8, occurring in nine different extracts, as examples. Motif 4 was chosen because it is an example for relatively high cross reaction, while motif 8 is an example for a 'low' cross reactive motif. Our results, depicted in a spider form, show how closely related the different allergen motif containing extracts actually are. The graphical depiction of the relationship between the motif defined cross reactions also allows to create a ranking. For example, motif 4, extract f85 (Celery), shows cross reaction at an average of 92.7% with all other allergen extracts in the group, while the lowest extract (t3, Common silver birch) cross reacts at 69.6%. For illustra-

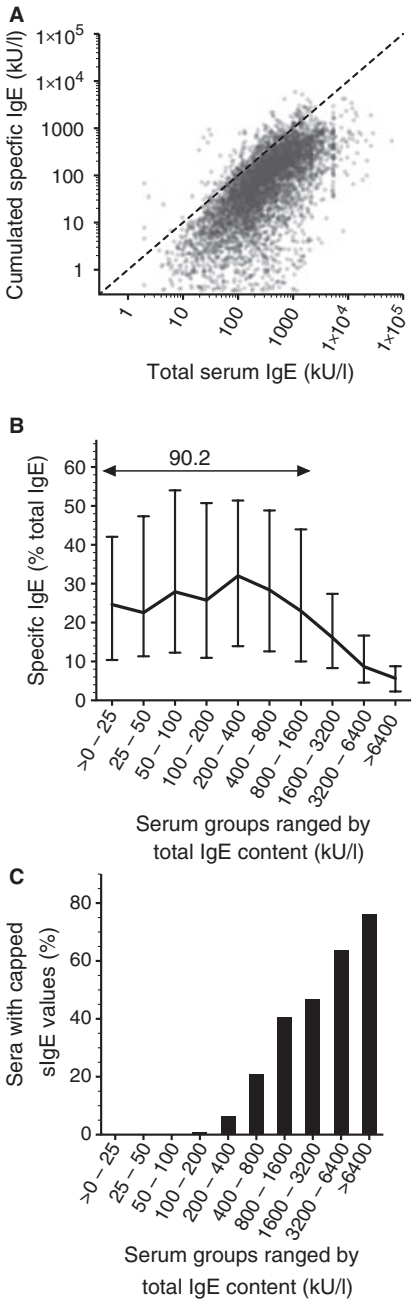

Figure 2 Relation of cumulated specific IgE to total IgE. The scatterplot (A) plots the cumulated specific IgE per serum to the total IgE. The dashed line in A indicates total IgE = cumulated specific IgE. For 90% of the sera, the cumulated specific IgE represented about 25–35% of the total IgE (B). The line in B represents the median, the error bars the 25th and 75th percentile, respectively. The histogram (C) shows the percentage of sera for which at least one extract value has been capped at 100 kUA/l.

tion, we have chosen this high and low cross reactive level from Table 2 that shows at the same time the absolute number of sera falling within this group.

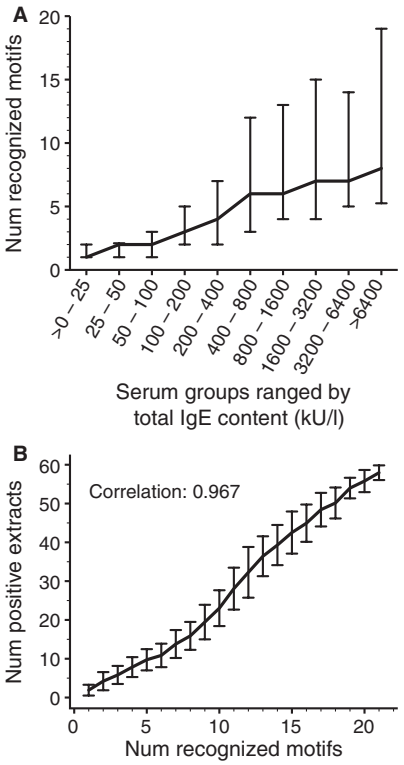

Figure 3 The number of recognized motifs within a serum increases with the total IgE content (A). In A, the line represents the median, the error bars the 25th and 75th percentile, respectively. Comparing the number of recognized motifs to the number of positive extracts shows that the number of recognized motifs is about three times lower than the number of positive extracts (B). In B, the line represents the mean while the error bars represent the standard deviation.

As cross reactions seemed to be very frequent, we analyzed how many sera recognized extracts without cross reaction. Three hundred and eighty-five sera (7.2%) were positive against one extract only. However, 12 single positive sera could not be considered as there are yet no protein sequences containing a motif within the positive tested extract [2x guinea pig epithelium (e6); 2x elm (t8); 2x wall pellitory (w19); 1x rye (f5); 1x oat (f7); 1x shrimp (f24); 1x chicken (f83); 1x mucor racemosus (m4); 1x walnut (t10)].

To truly exclude cross reactions, for each of the motifs occurring in the positive extract, the single positive sera would have to be tested against at least one other extract theoretically containing the motif. Otherwise, one might argue no cross reactions were observed because not all other extracts containing the potentially cross reactive motifs had been tested.

Only in 24 of the residual 373 sera, we observed a 'true' single positive reaction as all motifs theoretically contained within the positive extract were analyzed. In seven of the 24 sera, the specific IgE level against the single positive extract

Table 2 Range of cross reactivity by allergen motif. The percentage of sera reacting with another extract containing the same motif. For an illustration of the degree of cross reaction, see Fig. 4

		Most frequent cross reaction		Least frequent cross reaction	
Motif id	Number of extracts containing motif	% (extract)	Number of positive sera	% (extract)	Number of positive sera
1	25	91.3 (f33)	739	60.0 (g6)	3132
2	13	84.0 (t17)	887	64.1 (g6)	3132
7	13	94.5 (f33)	739	72.1 (w1)	2102
4	10	92.7 (f85)	151	69.6% (t3)	2916
8	9	73.1 (f27)	272	45.2 (d1)	2586
22	8	82.2 (f45)	656	52.7 (g6)	3132
28	8	77.5 (f27)	272	39.3 (e5)	3044
60	8	77.2 (m3)	1184	62.4 (t3)	2916
11	7	93.7 (f12)	777	74.8 (f13)	1975
10	6	91.6 (f18)	777	78.0 (f13)	1975
19	6	79.6 (f242)	39	57.0 (f49)	1179
20	6	79.6 (f27)	272	50.0 (e5)	3044
32	5	58.9 (i1)	604	47.0 (m6)	1526
33	5	88.3 (f6)	1051	83.3 (f17)	1588
52	5	94.3 (e87)	153	53.6 (e5)	3044
3	4	94.9 (f37)	358	58.9 (d1)	2586
18	4	84.8 (f35)	907	68.1 (t9)	1667
23	4	83.4 (i6)	1102	66.0 (f13)	1975
30	4	84.6 (f92)	230	63.5 (f4)	1472
37	4	73.8 (m2)	1201	47.4 (f20)	858
45	4	78.8 (m3)	1184	63.7 (f4)	1472
54	4	92.4 (f11)	919	78.4 (f13)	1975
57	4	94.5 (w11)	1043	70.1 (g6)	3132
63	4	61.3 (m2)	1201	54.4 (i3)	641
13	3	98.3 (g2)	1415	92.8 (g3)	1950
15	3	86.7 (i6)	1102	74.6 (d1)	2586
62	3	83.5 (m2)	1201	44.6 (d2)	1718
65	3	69.9 (m14)	668	53.0 (i1)	604

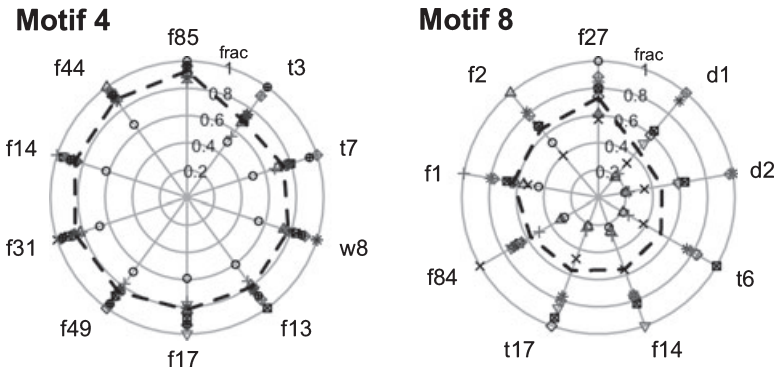

Figure 4 Degree of cross reaction. Spider plots for the extracts containing motif 4 or motif 8 as listed in Table 2. The number of sera positive against two extracts is expressed as a fraction of the number of sera positive against other extracts containing the same motif. The dashed line represents the mean fraction for each extract.

was only barely positive (≤0.5 kUA/l), but in general, the amount of specific IgE was not significantly different from the amount of specific IgE against these extracts in other sera (data not shown).

Additionally, we found that with an increasing number of extracts being tested, the fraction of single positive sera decreased. While 10.1% of the sera tested against 20 or less extracts were positive against one extract only (324 out of

3213 sera), this number dropped to 1.6% for sera tested against 90 or more extracts (15 out of 956 sera).

Discussion

We recently described a bioinformatic approach to predict allergenicity (11). Motifs in size of 50 amino acids were generated using bioinformatic algorithms using the most comprehensive database of allergen sequences, Allergome (2). We are presently at approximately 100 theoretical structures (motifs); whether such a general denomination of an allergen may truly predict allergen cross reactivity may be disputed.

Thus, we have used a large database of specific IgE determinations. These data seem to be normally distributed around an elevated IgE level as expected. The literature does not provide any assessments on how much of the specific IgE is actually directed against allergens causing symptoms. Our database was tested against a panel of 99 allergens and 1471 sera were tested against 90 or more of the allergens. During the rather long period over which our data were collected, the quality of allergen extracts has been changed by the manufacturer. Nevertheless, our analysis showed no statistically significant specific IgE level changes over time, neither in general nor grouped by country (data not shown).

Because it has been shown that ImmunoCAP sIgE determinations are accurate on a mass unit level (17), IgE levels of different extracts can be compared. As total and specific IgE determinations are based on a different measuring principle, one might argue their respective levels are not comparable, representing a systematic error. As quantification of human IgG and IgM has been transitioned from international units to absolute (g/l) values almost 30 years ago (18), this may also apply to IgE. The scientific literature has been converting 1 unit to 2.4 ng for IgE determinations performed on ImmunoCAP systems, and recent work comparing a new total IgE quantification method has shown that the absolute results are comparable to ImmunoCAP results (19). Thus, we assumed that total IgE determinations are accurate on a mass unit level and therefore may be compared to sIgE levels.

One may argue that the allergens of our panel were the true sensitizers for the generation of IgE (20, 21). We speculated that by using such a great number of allergens, most of the specific IgE might represent the total IgE level. Interestingly, this simple hypothesis delivered a value of 20–30% specific IgE in patient sera. While this percentage dropped in sera with total IgE higher than 1600 kUA/l, it is likely that we underestimate the amount of cumulated sIgE in these sera because specific IgE levels have been capped at 100 kUA/l, especially in these high total IgE sera, as Fig. 2C shows.

Another interesting finding is the percentage of positive test results related to total IgE levels. For the 169 sera with very high total IgE levels (higher than 3200 kUA/l), on average 85.8% of specific IgE tests yielded positive results. This suggested a broader reactivity against allergens in relation to higher total serum IgE.

The question remains whether the residual IgE, not to be attributed to one of these frequent allergens, is directed toward unknown allergens or represents IgE that cannot be assessed by using commercial allergen preparations. Of course, it may also be speculated that 'non specific' IgE may be a result of somatic mutation; in other words, IgE without a known specificity.

Here, we postulate that the estimation of 20–30% specific IgE may be exaggerated as a great proportion of this IgE is cross reacting. By comparing the specific IgE determinations against the various allergen extracts with the theoretically postulated motif, one may speculate that the accumulation of specific IgE using different extracts results in an overestimation. Indeed, as Fig. 3B shows, there seems to be an almost strict linear correlation between the number of recognized extracts and motifs. However, there seems to be about three times fewer motifs recognized than allergen extracts. Based on this result, one would have to assume that the actual percentage of specific IgE in patient sera is actually below 10% of the total IgE.

Figure 3B also addresses the question whether co-sensitization might be mistaken as cross reactivity. We observed a direct linear relationship of 3 : 1 between the recognition of allergens containing defined motifs and the number of motifs that were recognized. Co-sensitization would occur independently of structural similarities between proteins, resulting in a ratio closer to 1 : 1 as each recognized extract might be recognized by a different epitope. This is also exemplified by our finding for Bermuda grass (g2), an allergen that is not naturally occurring in Scandinavia. 61.6% of sera derived from Sweden were positive against Bermuda grass, arguably because of cross reaction rather than co-sensitization.

As motifs can easily be associated with known allergen sequences, we have grouped the cross reactions according to the theoretical presence of the motif. This approach resulted in an astonishing high frequency of cross reactions. It was also possible to generate a hierarchy which may or may not be of clinical relevance. Within one motif, it was always possible to rank cross reactions by their mean percentage of cross reactions with other allergen extracts containing the same motif. This, of course, may be a bias as the allergen extract with the highest degree of cross reaction may be an allergen extract where the epitopes that can be associated to motifs are well preserved, well presented or most prominently present in a given industrial allergen preparation coupled to a solid phase (22). However, our data strongly suggest that patients and doctors should realize that sensitization and reaction to a given allergen extract may not be the sensitizing organism or part thereof, like birch pollen, but that it is a molecular entity occurring in different organisms.

We were also interested whether there were patient sera without cross reactions. In our 5362 sera, we found 385 that were positive against one extract only. Twenty-four of these sera had been tested against at least one extract for every motif present in the positive extract, yet showed no cross reactions. Twelve sera were positive against an extract containing no currently known motif, and in the remaining 349 sera, a sensitization against one of the motifs that had not been tested would be possible. One might hypothesize that these patients developed an IgE response against an epitope not shared among proteins bearing the same motif, or even

simpler that not all cross reactive motifs have been found yet. Furthermore, the fact that with an increasing number of extracts tested, the percentage of single positive sera decreased from approx. 10% for sera tested against 20 or less extracts to 1.6% for those tested against at least 90 extracts suggests that the true number of single positive sera must be low.

For allergy diagnosis, the way out of this uncertainty whether a patient reacts to a certain part of a plant or a common allergenic structure may be diagnostic procedures using recombinant allergens. Thus, in the future, we intend to analyze whether a small number of recombinant proteins covering most of the known motifs would suffice to diagnose other sensitizations by predicting the allergen via the motif.

Acknowledgments

This work was supported by grant no 8803.1 from the Commission of Technology and Innovation (CTI), Switzerland.

References

1. Sanz ML, Prieto I, García BE, Oehling A. Diagnostic reliability considerations of specific IgE determination. *J Investig Allergol Clin Immunol* 1996;**6**:152–161.
2. Mari A, Scala E, Palazzo P, Ridolfi S, Zennaro D, Carabella G. Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cell Immunol* 2006; **244**: 97–100.
3. Steckelbroeck S, Ballmer-Weber BK, Vieths S. Potential, pitfalls, and prospects of food allergy diagnostics with recombinant allergens or synthetic sequential epitopes. *J Allergy Clin Immunol* 2008;**121**: 1323–1330.
4. Chapman MD, Smith AM, Vailes LD, Arruda LK, Dhanaraj V, Pomés A. Recombinant allergens for diagnosis and therapy of allergic disease. *J Allergy Clin Immunol* 2000;**106**:409–418.
5. Kazemi-Shirazi L, Niederberger V, Linhart B, Lidholm J, Kraft D, Valenta R. Recombinant marker allergens: diagnostic gatekeepers for the treatment of allergy. *Int Arch Allergy Immunol* 2002;**127**:259–268.
6. Harwanegg C, Hiller R. Protein microarrays for the diagnosis of allergic diseases: state-of-the-art and future development. *Clin Chem Lab Med* 2005;**43**:1321–1326.
7. Vrtala S. From allergen genes to new forms of allergy diagnosis and treatment. *Allergy* 2008;**63**:299–309.
8. Ferrer M, Sanz ML, Sastre J, Bartra J, del Cuvillo A, Montoro J et al. Molecular diagnosis in allergology: application of the microarray technique. *J Investig Allergol Clin Immunol* 2009;**19**(Suppl. 1):19–24.
9. Ferreira F, Hawranek T, Gruber P, Wopfner N, Mari A. Allergic cross-reactivity: from gene to the clinic. *Allergy* 2004;**59**:243–267.
10. Petersen A, Dresselhaus T, Grobe K, Becker WM. Proteome analysis of maize pollen for allergy-relevant components. *Proteomics* 2006;**6**:6317–6325.
11. Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. *FASEB J* 2003;**17**:1141–1143.
12. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994;**2**:28–36.
13. Bucher P, Karplus K, Moeri N, Hofmann K. A flexible motif search technique based on generalized profiles. *Comput Chem* 1996; **20**:3–23.
14. Team RDC. R: a language and environment for statistical computing. 2009.
15. Mari A, Rasi C, Palazzo P, Scala E. Allergen databases: current status and perspectives. *Curr Allergy Asthma Rep* 2009;**9**:376–383.
16. Zetterström O, Johansson SG. IgE concentrations measured by PRIST in serum of healthy adults and in patients with respiratory allergy. A diagnostic approach. *Allergy* 1981;**36**:537–547.
17. Wood RA, Segall N, Ahlstedt S, Williams PB. Accuracy of IgE antibody laboratory results. *Ann Allergy Asthma Immunol* 2007;**99**:34–41.
18. Bruyn AMD, Klein F, Neumann H, Sandkuyl LA, Vermeeren R, Blansch GL. The absolute quantification of human IgM and IgG: standardization and normal values. *J Immunol Methods* 1982;**48**:339–348.
19. Chen HX, Busnel JM, Peltre G, Zhang XX, Girault HH. Magnetic beads based immunoaffinity capillary electrophoresis of total serum IgE with laser-induced fluorescence detection. *Anal Chem* 2008;**80**:9583–9588.
20. Sutton BJ, Gould HJ. The human IgE network. *Nature* 1993;**366**:421–428.
21. Gould HJ, Sutton BJ, Beavil AJ, Beavil RL, McCloskey N, Coker HA et al. The biology of IGE and the basis of allergic disease. *Annu Rev Immunol* 2003;**21**:579–628.
22. Mari A. Multiple pollen sensitization: a molecular approach to the diagnosis. *Int Arch Allergy Immunol* 2001;**125**: 57–65.