Bioinformatics, 33(7), 2017, 1014–1020
doi: 10.1093/bioinformatics/btw767
Advance Access Publication Date: 30 December 2016
Original Paper

OXFORD

Structural bioinformatics

Cross-React: a new structural bioinformatics method for predicting allergen cross-reactivity

Surendra S. Negi* and Werner Braun

Sealy Center for Structural Biology and Molecular Biophysics, Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX 77555-0304, USA

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on July 20, 2016; revised on November 3, 2016; editorial decision on November 25, 2016; accepted on December 1, 2016

Abstract

The phenomenon of cross-reactivity between allergenic proteins plays an important role to understand how the immune system recognizes different antigen proteins. Allergen proteins are known to cross-react if their sequence comparison shows a high sequence identity which also implies that the proteins have a similar 3D fold. In such cases, linear sequence alignment methods are frequently used to predict cross-reactivity between allergenic proteins. However, the prediction of cross-reactivity between distantly related allergens continues to be a challenging task. To overcome this problem, we developed a new structure-based computational method, Cross-React, to predict cross-reactivity between allergenic proteins available in the Structural Database of Allergens (SDAP). Our method is based on the hypothesis that we can find surface patches on 3D structures of potential allergens with amino acid compositions similar to an epitope in a known allergen. We applied the Cross-React method to a diverse set of seven allergens, and successfully identified several cross-reactive allergens with high to moderate sequence identity which have also been experimentally shown to cross-react. Based on these findings, we suggest that Cross-React can be used as a predictive tool to assess protein allergenicity and cross-reactivity.

Availability and Implementation: Cross-React is available at: http://curie.utmb.edu/Cross-React.html

Contact: ssnegi@utmb.edu

1 Introduction

The immune system induces an IgE response to certain antigens that resemble primary sensitizing antigens. This phenomenon is known as cross-reactivity (CR). Considerable amount of experimental and theoretical work has been done to characterize the sequences and substructures of allergens that account for their cross-reactivity (Aalberse, 2000; Aalberse *et al.*, 2001; Breiteneder and Mills, 2006; Fedorov *et al.*, 1997; Ivanciuc *et al.*, 2009a; Ivanciuc *et al.*, 2009b). Earlier studies have shown that an antibody (Ab) recognizes and binds to a specific region on an antigen (Ag) surface known as epitope. An epitope can be a linear contiguous sequence of amino acids (known as linear epitope) or a group of sequentially separated amino acids in a protein sequence brought together by protein folding (known as conformational epitope). Compared to a linear epitope, a conformational epitope provides a correct scaffold for an antigenic determinant which is crucial for Ag–Ab interaction, identification of cross-reactive allergens, and development of new vaccines and therapeutics to treat allergenic diseases (Jutel *et al.*, 2016; Metcalfe, 2005; Tilles, 2016). It is believed that more than 90% of the clinically important epitopes recognized by antibodies are conformational in nature (Barlow *et al.*, 1986; Haste Andersen *et al.*, 2006; Van Regenmortel, 1996). The precise location of conformational epitope can be determined by X-ray crystallography, but solving co-crystal structure of an Ag–Ab complex is challenging, expensive and time-consuming. To overcome this problem, phage display technology (Smith, 1985) has been successfully used to characterize Ag–Ab interactions (Mittag *et al.*, 2006; Untersmayr *et al.*, 2006). It has been shown that peptides selected by phage

display mimic conformational epitopes on an antigen protein. But due to lack of consensus in the phage displayed peptides, mapping IgE binding sites (epitope) onto three-dimensional (3D) allergen structure is a challenging task (Chen *et al.*, 2016; Mayrose *et al.*, 2007; Negi and Braun, 2009; Tiwari *et al.*, 2012).

In most cases, the specificity of an antibody depends on a unique composition of amino acids at the epitope site. Thus, an antibody raised against one allergen may cross-react to similar allergens from different sources if they have high sequence or structural identity (Aalberse, 2000; Breiteneder and Mills, 2006; Dall'antonia *et al.*, 2014; Garcia and Lizaso, 2011). In such cases, potential cross-reactive allergens can be identified using global sequence alignment methods such as PSI-BLAST (Altschul *et al.*, 1997) or FASTA (Pearson, 1994). Alternatively, local peptide sequence similarity method such as property distance (PD) can be applied (Ivanciuc *et al.*, 2009a). However, these methods fail to detect cross-reactivity between distantly related allergens (Aalberse, 2000; Aalberse *et al.*, 2001; Breiteneder and Mills, 2006; Garcia and Lizaso, 2011; Guhsl *et al.*, 2014; Metcalfe, 2005; Vieths *et al.*, 2002; Weber, 2007). The definition of *in vitro* cross-reactivity of allergens should be distinguished from multi-sensitization in the clinical observation when a patient is exposed and sensitized to different allergen sources (Aalberse and Aalberse, 2015). In that case cross-reactivity might be due to binding of allergens to different IgE antibodies.

In the past, several computational methods have been developed to predict epitope sites on protein antigens (El-Manzalawy *et al.*, 2008; Singh *et al.*, 2013; Yao *et al.*, 2012). Most of these methods use a small linear segment of 6–10 amino acids to search for the epitopes. These methods are very useful to predict linear epitopes, but are limited to predict the location of conformational epitopes. Epitope prediction methods using structural properties of proteins have previously been reviewed by Dall'antonia *et al.* (2014). DiscoTope (Haste Andersen *et al.*, 2006; Larsen *et al.*, 2006), ElliPro (Ponomarenko *et al.*, 2008) and SPADE (Dall'Antonia *et al.*, 2011) uses 3D structure of an antigen, while PepSurf (Mayrose *et al.*, 2007) and EpiSearch (Negi and Braun, 2009) uses both 3D structure and phage display data, to predict conformational epitopes. However, all the methods developed so far are not designed for rapid large-scale screening of allergen cross-reactivity using epitope information.

Our objective was to find structural features of allergens responsible for their allergenicity and cross-reactivity with other allergens available in SDAP, a database that contains over a thousand 3D structures of allergens determined by X-ray crystallography or homology modeling (Ivanciuc *et al.*, 2003; Oezguen *et al.*, 2008; Power *et al.*, 2013). All modeled structures in the SDAP database have been energy minimized by FANTOM using ECEPP/2 force field (Ivanciuc *et al.*, 2003; Nemethy *et al.*, 1983; Oezguen *et al.*, 2008; Schaumann *et al.*, 1990) and validated by comparison with new allergen structures released in the PDB (Power *et al.*, 2013). In addition, due to large number of allergens in SDAP, we wanted the predictions to be rapid and efficient to perform such a large-scale structural bioinformatics analysis.

With this intent, we developed a new structure-based method, Cross-React, that enables prediction of cross-reactivity between distantly related allergenic proteins using their X-ray or homology modeled structures in combination with epitope analysis. The method uses a patch analysis, solvent accessible surface area (SASA) of amino acids, and structural similarity between amino acids in the epitope region of a query allergen and allergens in the SDAP database (target allergens). The search results are ranked based on the calculated Pearson correlation coefficient (PCC) between the amino acid composition in the query epitope and the accessible surface patches on the target allergens. We tested the performance of Cross-React for seven different allergenic proteins which have also been experimentally shown to cross-react (Ferreira *et al.*, 2004; Ivanciuc *et al.*, 2003; Konstantinou and Grattan, 2008; Mari *et al.*, 2006). Our analysis successfully predicted cross-reactive allergens which have similar tertiary structure, high PCC value and >30% sequence identity with the query allergen.

2 Methods

2.1 Selection of query allergens for Cross-React analysis

The query data set of seven allergenic proteins was divided into two types: Type I and Type II, depending on the nature of their experimentally determined epitopes. Type I included the allergenic proteins from birch pollen (Bet v 1) (Gajhede *et al.*, 1996), grass pollen (Phl p 2) (Padavattan *et al.*, 2009), honey bee venom (Api m 2) (Padavattan *et al.*, 2007), and bovine milk (Bos d 5) (Niemi *et al.*, 2007). For each of these allergens, location of the conformational epitope has previously been determined by X-ray crystallography of the allergen-Ab complex. Type II included the allergenic proteins from peach (Pru p 3) and peanut (Ara h 1 and Ara h 2), where co-crystal structures of allergen-Ab complexes are not available. Hence, we first mapped the potential conformational epitopes by using mimotopes from phage display for Pru p 3 (Pacios *et al.*, 2008), and linear epitopes for Ara h 1 and Ara h 2 (Ara h 1/2) (Barre *et al.*, 2005a; Barre *et al.*, 2005b) onto 3D structures determined by X-ray crystallography (Pru p 3 and Ara h 1) or homology modeling (Ara h 2).

2.2 Selection of amino acids in IgE binding site

Conformational epitope in a Type I allergen was defined as a change in the SASA of amino acids in the allergen upon formation of an allergen-Ab complex. The SASA values were calculated using GETAREA (Fraczkiewicz and Braun, 1998) with a probe of radius 1.4 Å. The amino acids on the antigen surface were considered to be part of an epitope if they lost more than 10 $Å^2$ of SASA upon allergen-Ab complex formation (Negi and Braun, 2007; Negi *et al.*, 2007). For the Type II allergen Pru p 3, we used peptides selected from phage display (Pacios *et al.*, 2008) in combination with EpiSearch (Negi and Braun, 2009) and PepSurf (Mayrose *et al.*, 2007) to map conformational epitopes. For Ara h 1/2, conformational epitopes were defined as a group of amino acids in a spherical patch around a surface exposed residue located at the center of a linear epitope.

2.3 Connectivity matrix

We introduced a connectivity matrix C_1 to describe the composition of surface exposed residues in a conformational epitope on a query allergen. The connectivity matrix is a 20x20 matrix which counts how often C_β atoms of the amino acids in the epitope region are in contact with other residues. Two amino acids were considered to be connected if the distance between their C_β atoms (C_α for Gly residue) was $\leq$8 Å. Similarly, a connectivity matrix C_2 was constructed for all surface exposed residues on a target allergen. The matrix counts the connections between amino acids present in a patch of radius 10 Å centered on a surface exposed residue in the target allergen. Thus, the total number of surface patches on the target allergen is equal to the number of surface exposed residues.

2.4 Correlation between connectivity matrices

The cross-reactivity between query and target allergen was assessed by calculating the correlation coefficient between connectivity matrices of query epitope and a surface patch on the target allergen. This process was repeated for all the surface patches on the target allergen. For the patch analysis, the SASA of amino acids in all the allergens were calculated using GETAREA, and amino acids with SASA greater than 10 Å^2 were selected and replaced by their C_β atoms (C_α in case of Gly residue). Next, we defined a spherical test patch of radius 10 Å around each surface exposed residue using their C_β atom as center. To improve search efficiency, we precomputed the connectivity matrices of all test patches to generate over one hundred thousand connectivity matrices. The similarities between epitope site on the query allergen and test patches on the target allergen were measured by calculating the Pearson correlation coefficient (PCC) between the connectivity matrices C_1 and C_2 as defined in equation 1

$$\mathrm{PCC}\ (C_1, C_2) = \frac{\sum C_1 C_2 - \frac{\sum C_1 \sum C_2}{N}}{\sqrt{\left[\sum {C_1}^2 - \frac{(\sum C_1)^2}{N}\right]\left[\sum {C_2}^2 - \frac{(\sum C_2)^2}{N}\right]}} \quad (1)$$

where N is the total number of elements in the connectivity matrix.

2.5 Prediction of cross-reactivity

Potential cross-reacting allergens were predicted by comparing: 3D structure of query allergen with the allergens in SDAP database, and the PCC value between query and predicted epitope. The structural alignment between query and target allergen was calculated using the combinatorial extension (CE) method (Shindyalov and Bourne, 1998). Based on the CE sequence identity, all the predicted allergens in SDAP were divided into two groups. The structural identity between query and predicted allergens was >60% for Group I, and 30–60% for Group II. Finally, the allergens predicted in each group were sorted based on the PCC value of surface patches.

3 Results and discussion

We investigated the performance of Cross-React using a diverse set of seven query allergens. As shown in Tables 1 and 2, the predicted allergens in Group I share high PCC value, high sequence and structural identity with the query allergen; and therefore represent strong candidates for cross-reactivity. The allergens predicted in Group II have moderate sequence identity, but high PCC value and similar 3D fold as the query allergen; and therefore represent probable candidates for cross-reactivity. In the following sections, we document our findings with Cross-React.

3.1 Type I: crystal structure of query allergen in complex with antibody is available

Before performing large scale SDAP database search, we first confirmed the ability of Cross-React method to correctly predict the experimentally known epitope site in Type I query allergens. The method identified all epitopes with PCC value: 0.99 for Bet v 1, 0.93 for Phl p 2 and 0.93 for Api m 2. In our cross-reactivity analysis a slightly lower PCC value is expected because computationally predicted epitopes are spherical patches on the allergen surface while natural epitope sites are not. Further, we tested the performance of Cross-React to find similar or distantly related allergens for Bet v 1, Phl p 2, Api m 2 and Bos d 5 using experimentally known epitope as input, and is described in the following sections.

Bet v 1 is a major birch pollen allergen that accounts for health problems related to pollen allergies in more than 20% of the European population (Ganglberger *et al.*, 2000). It is one of the most studied allergen and is known to cross-react with several other allergens (Vieths *et al.*, 2002). Patients sensitized to Bet v 1 can develop allergic reactions towards other plant derived foods such as apple (Mal d 1), cherry (Pru av 1), carrots (Dau c 1) and celery (Api g 1) (Aalberse, 2000; Breiteneder and Mills, 2006; Ganglberger *et al.*, 2000; Jensen-jarolim *et al.*, 1998; Klinglmayr *et al.*, 2009; Thomas *et al.*, 2005). In order to search for allergens with distribution of amino acids similar to Bet v 1 epitope, we first created a connectivity matrix of amino acids present in the Bet v 1 epitope region using the crystal structure of Bet v 1-mAb BV16 complex (Gajhede *et al.*, 1996; Mirza *et al.*, 2000). This information was then used to search the SDAP database using Cross-React. A closer examination of the results revealed that the method predicted the epitope similar to Bet v 1 in Group I (Aln g 1, Car b 1, Cor a 1, Cas s 1 and Mal d 1) and Group II (Gly m 4, Pru av 1, Dau c 1 and Api g 1) allergens (Table 1, Fig. 1a–c). The predicted epitopes in Group I allergens are surface patches with high PCC values and are spatially located in the same position relative to the experimentally known epitope of Bet v 1. The predicted allergens in Group II have moderate sequence identity of 30–60% with Bet v 1.

Next, we identified potential cross-reactive allergens for Phl p 2, Api m 2 and Bos d 5 using the methodology described for Bet v 1. The epitope of grass pollen allergen Phl p 2 consists of 15 surface exposed residues from four-stranded antiparallel beta strands (Padavattan *et al.*, 2009). In this case, Cross-React predicted several allergens with epitopes structurally similar to Phl p 2 (Table 1). Of particular interest is the experimentally known cross-reactive allergen Phl p 1 from timothy grass which has ~43% sequence identity with Phl p 2. Api m 2 is a 350 residue allergen protein found in honey bee venom. The epitope of Api m 2 is composed of a continuous group of nine surface exposed amino acids (Padavattan *et al.*, 2007). Using Api m 2 epitope information as input, Cross-React efficiently predicted Ves v 2 (wasp venom) and Dol m 2 (white face hornet) as cross-reactive allergens of Api m 2, despite moderate PCC value and moderate sequence identity (Table 1, Fig. 1d–f).

Finally, for the bovine milk allergen Bos d 5 (Niemi *et al.*, 2007), we predicted Asp f 3 and Bla g 4 as cross-reactive allergens which have less than 30% structural identity with Bos d 5. It is interesting to note that Bla g 4 and Bos d 5 belong to lipocalin group of proteins (Tan *et al.*, 2009), however, it is not known whether these allergens cross-react under experimental conditions. Based on our findings so far, we suggest that the prediction accuracy of Cross-React analysis is higher if the sequence or structural identity between a query allergen and target allergens in SDAP is more than 30%, and the PCC value is greater than 0.70.

3.2 Type II: crystal structure of query allergen in complex with antibody is not available

In this case, we investigated three allergens Pru p 3, Ara h 1/2. In the absence of crystal structures, we mapped the epitopes for Pru p 3 and Ara h 1/2 using phage display peptides and linear epitopes, respectively.

Peach allergen Pru p 3 is a common plant food allergen in Europe and Mediterranean (Borges *et al.*, 2007; Chen *et al.*, 2008; Pacios *et al.*, 2008). In an earlier attempt to map the conformational epitopes of Pru p 3, Pacios et al. used a consensus sequence obtained from phage display and identified two overlapping groups of amino acids: (i) between helix 2 and helix 3 and (ii) in the C-terminal

Table 1. Cross-reactive allergens predicted for Type I allergens Bet v 1, Phl p 2 and Api m 2

Query allergen[a]	Group	Predicted allergen[b]	% Seq ID[c]	% CE ID[d]	RMSD[e]	PCC
Bet v 1	I	**Aln g 1**	77.7	81.1	1.16	0.97
		Car b 1	69.9	75.5	1.09	0.97
		Cor a 1	69.3	66.7	1.12	0.97
		Cas s 1	65.1	67.9	0.73	0.93
		Mal d 1	62.7	64.2	0.74	0.92
	II	Gly m 4	45.8	46.8	1.19	0.97
		Pha v 1	42.2	44.2	0.77	0.94
		Ara h 8	44.0	46.8	1.96	0.94
		Pyr c 1	54.8	57.0	1.09	0.93
		Pru av 1	57.2	58.3	1.89	0.89
		Pru ar 1	57.2	59.7	0.72	0.87
		Api g 2	38.6	38.0	0.75	0.85
		Dau c 1	34.9	39.2	0.72	0.85
		Pet c PR10	36.1	39.0	0.69	0.85
		Tar o rap	33.7	35.7	1.75	0.85
		Api g 1	38.6	41.2	1.29	0.84
Phl p 2	I	Lol p 2	65.2	63.0	0.99	0.90
		Cyn d 2	98.9	98.9	0.63	0.90
		Dac g 2	92.9	100.0	0.65	0.89
		Tri a 3	52.2	51.6	0.78	0.88
		Hol l 1	43.3	40.7	1.43	0.85
		Dac g 3	54.6	56.7	1.46	0.84
		Lol p 3	55.6	57.8	1.63	0.82
		Zea m 1	33.4	33.0	1.52	0.82
		Phl p 1	43.4	40.7	1.42	0.82
		Ory s 1	44.4	37.4	1.49	0.82
		Phl p 3	53.5	56.0	1.20	0.80
Api m 2	I	**Ves v 2**	54.2	61.8	0.66	0.71
	II	Dol m 2	56.4	56.4	0.92	0.84
		Pol a 2	53.3	56.4	0.89	0.69

[a]Query allergen sequence was obtained from X-ray or model structure.
[b]Experimentally known cross-reactive allergens are shown in bold font.
[c]Predicted allergen sequence identity (Seq ID) was calculated using FASTA search in SDAP.
[d]Structure alignment identity was calculated using CE structure alignment method (CE ID).
[e]RMSD was calculated using CE method.

Table 2. Cross-reactive allergens predicted for Type II allergens Pru p 3, Ara h 1 and Ara h 2

Query allergen[a]	Group	Predicted allergen[b]	% Seq ID[c]	% CE ID[d]	RMSD[e]	PCC
Pru p 3	I	**Pru av 3**	81.6	87.0	1.28	0.91
		Hev b 12	60.2	62.0	1.35	0.88
		Pru ar 3	85.7	92.3	1.54	0.84
		Mal d 3	75.5	80.4	1.33	0.81
		Pru d 3	83.7	90.1	1.58	0.78
		Pyr c 3	72.5	78.0	1.62	0.74
	II	Zea m 14	62.0	58.7	1.25	0.86
		Vit v 1	18.5	40.0	1.22	0.84
		Hor v 1	44.5	44.0	1.01	0.81
		Cor a 8	58.7	58.7	1.31	0.67
		Amb a 6	34.8	36.0	1.44	0.61
Ara h 1	I	–	–	–	–	–
	II	Gly m conglycinin	53.6	54.2	1.12	0.86
		Jug r 2	41.9	36.1	1.11	0.83
		Jug n 2	41.7	35.5	1.09	0.81
		Ana o 1	33.6	30.8	1.18	0.73
		Cor a 11	37.9	36.5	1.15	0.64
Ara h 2	I	–	–	–	–	–
	II	**Ara h 6**	58.7	44.1	1.58	0.90

See Table 1 footnote for definitions of a–e.

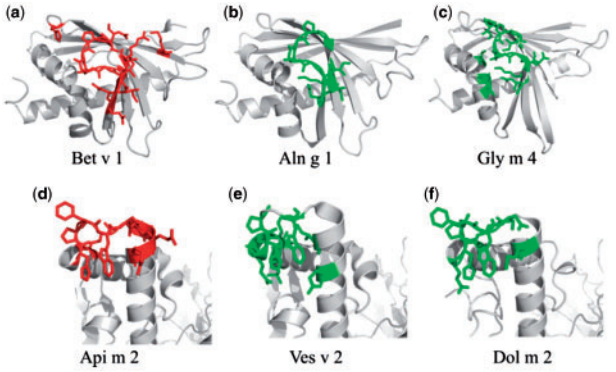

Fig. 1. Comparison of epitope sites in Type I query allergens and predicted cross-reactive allergens. All allergens are shown as ribbon models (grey). The epitope site residues are shown as sticks in Type I query allergens (red), and cross-reactive allergens (green) predicted using Cross-React. Only epitope site is shown for Api m 2, Ves v 2 and Dol m 2 (Color version of this figure is available at *Bioinformatics* online.)

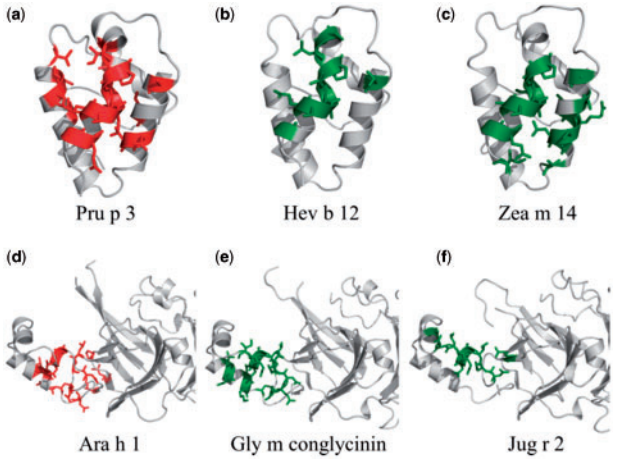

Fig. 2. Comparison of epitope sites in Type II query allergens and predicted cross-reactive allergens. All allergens are shown as ribbon models (grey). The epitope site residues are shown as sticks in Type II query allergens (red), and cross-reactive allergens (green) predicted using Cross-React. Only epitope site is shown for Ara h 1, Gly m conglycinin and Jug r 2 (Color version of this figure is available at *Bioinformatics* online.)

region. In another study, Borges et al. used hydropathy profile of Pru p 3 and predicted four linear IgE epitopes. A comparison between the above two methods revealed that the amino acids in the second linear epitope predicted by the hydropathy profile method are conserved in proteins from Pru p 3 family, and are similar to those in the epitope region predicted using phage display. Further, we analyzed the mimotope data for Pru p 3 using EpiSearch and PepSurf. Both methods predicted a similar group of amino acids that coincided with the previously known epitope in Pru p 3. The epitope information for Pru p 3 obtained using the computational approach was used to search the SDAP database for allergens with similar epitopes.

Our Cross-React method predicted potential cross-reactive allergens Pru av 3, Heb b 12, Pru ar 3 and Zea m 14 with high PCC values (Table 2). In all these allergens, the potential epitope patches were found within similar 3D location as observed in Pru p 3 (Fig. 2a–c). We also observed that the predicted epitopes on these allergens are surface exposed and hence could account for their cross-reactivity (Breiteneder and Mills, 2006; Garcia-Casado *et al.*, 2003; Pastorello *et al.*, 2000).

To confirm the feasibility of our method for predicted epitopes, the Cross-React analysis for Bet v 1 was repeated using the methodology similar to that described above for Pru p 3. We selected five phage displayed peptides that mimic Bet v 1 conformational epitope (Ganglberger *et al.*, 2000) and mapped them onto Bet v 1 crystal structure using EpiSearch. The method predicted a high scoring patch centered on V67 that overlapped with the Bet v 1 epitope previously identified by X-ray crystallography and site directed mutagenesis (Mirza *et al.*, 2000; Spangfort *et al.*, 2003). Further, Cross-React analysis of the predicted Bet v 1 epitope identified cross-reactive allergens same as those shown in Table 1.

Next, we analyzed two major peanut allergens Ara h 1/2 (Barre *et al.*, 2005b; de Leon *et al.*, 2007; Koppelman *et al.*, 2005; Naganawa *et al.*, 2008). Ara h 1 is a two domain allergen protein with similar fold but different sequences and forms a homotrimer through hydrophobic interactions. Earlier experimental studies using individual patient sera have shown that antibodies recognize twenty three linear epitopes in Ara h 1 (Shin *et al.*, 1998), and ten linear epitopes in Ara h 2 (Barre *et al.*, 2005a). For the cross-reactivity analysis of Ara h 1, we selected one major linear epitope (epitope 18, HRIFLAGDKD) and searched for its 3D homologs in SDAP using a patch size of 10 Å (as described in the method section). Cross-React predicted similar conformational epitope in: Gly m conglycinin, Jug r 2, Jug n 2, Ana o 1 and Cor a 11 (Table 2, Fig. 2d–f). In case of Ara h 2, we selected a high binding IgE epitope (epitope 2, DRRCQSQLER) (Barre *et al.*, 2005b) for the Cross-React analysis, and predicted similar conformational epitope in Ara h 6, Ara h 7, Sin a 7 and Hor v 1. The alignment results obtained from CE and FASTA show that Ara h 2 has high sequence and structural homology with Ara h 6, but low homology with Sin a 7, Ara h 7 and Hor v 1. Thus, by performing Cross-React analysis using a linear epitope, we have demonstrated the effectiveness of our method to predict potential cross-reactive allergens in the absence of known conformational epitopes.

3.3 Special case

To further evaluate our method, we performed Cross-React analysis for Bla g 2, a major cockroach allergen for which cross-reactive allergens are still unknown. We extracted Bla g 2 epitope information from the crystal structure of Bla g 2 in complex with mAb 7C11 (Li *et al.*, 2008). Our method did not find allergen surface patches with amino acids composition similar to those observed in the Bla g 2 epitope. In order to confirm our findings, we searched the SDAP database for Bla g 2 homologs using FASTA, and identified only one allergen Asp f 10 which has ~19% sequence identity with Bla g 2.

3.4 Clinical implications of cross-reactivity

In a clinical setting, serum of patients can be analyzed by component-resolved diagnostics (CRD) to determine if the serum IgE antibodies react to different purified allergens (Aalberse, 2007; Gadermaier *et al.*, 2011). A positive reaction of a patient serum in those experiments can have two different reasons: a) the patient has been previously exposed to different allergen sources and different IgE antibodies were induced, or b) a single IgE antibody in the serum binds to different allergens. In this study, we used the term cross-reactivity only for the second case which can be confirmed by in-vitro binding studies of isolated monoclonal IgE antibodies. Another reason for the observation of clinical cross-reactivity can be attributed to cross-reactive carbohydrate determinants (CCD) (Foetisch *et al.*, 2003; Mari *et al.*, 1999). It has been shown that some of the CCDs are highly cross-reactive and responsible for IgE reactivity

against plant and animal derived foods. Although clinical relevance of IgE specific for CCDs is still a matter of debate, they may represent a special case of panallergenic structures responsible for IgE cross-reactivity (Ferreira *et al.*, 2004; Hauser *et al.*, 2010). On the other hand, sequence and structural analysis of allergenic proteins suggests that the allergens can be grouped into panallergen families which share conserved motifs, 3D structures, function and cross-reactive epitopes. Hence, better understanding of amino acids in the epitope site and 3D structure of panallergens will improve diagnosis and treatment of allergenic diseases.

3.5 Implementation

A user friendly and fully automated version of the Cross-React method presented in this work is available online (http://curie.utmb.edu/Cross-React.html). To perform cross-reactivity analysis, users can upload X-ray or homology model structure of a query allergen as a PDB-formatted file, a list of amino acids in the allergen epitope site, and enter values for SASA and patch size threshold. The predicted results are displayed in a tabular form with name of the cross-reactive allergen, PCC value, its sequence identity with the query allergen and location of epitope on the cross-reactive allergen. In addition, user can also view a comparison between epitope sites on the query and the predicted cross-reactive allergens using JMOL molecular viewer (JMOL, 2007). The average time to process a query allergen may vary from 1 to 2 min.

4 Conclusion

Sensitization by an allergen induces IgE antibodies which can be cross-reactive to other allergen proteins with similar sequences or 3D folds in their tertiary structure. In the current work, we have developed and tested a new method, Cross-React, to predict cross-reactivity between allergenic proteins by comparing structural similarity between amino acids present in the epitope site. Using the concept of connectivity between amino acids within the epitope site in combination with data mining techniques, we have correctly predicted allergens which have also been experimentally shown to cross-react. Further, we suggest that the Cross-React method will help to understand why only selected proteins of the same family are cross-reactive, e.g. homologous of Bet v 1 in vegetables cause severe while those in fruit cause mild allergic reactions (Vieths *et al.*, 2002). We conclude that the structure-based computational method presented here will improve our understanding of allergen cross-reactivity, and help in development of new standardized approaches to immunotherapy and formulation practices.

Funding

This work was supported by grants from the National Institutes of Health (R21 AI1090909), and the U.S. Environmental Protection Agency under a STAR Research Assistance Agreement (No. RD834823). The article has not been formally reviewed by the EPA, and the views expressed in this document are solely those of the authors.

Conflict of Interest: none declared.

References

Aalberse,R.C. (2000) Structural biology of allergens. *J. Allergy Clin. Immunol.*, **106**, 228–238.

Aalberse,R.C. (2007) Assessment of allergen cross-reactivity. *Clin. Mol. Allergy*, **5**, 2.

Aalberse,R.C. and Aalberse,J.A. (2015) Molecular allergen-specific IgE assays as a complement to allergen extract-based sensitization assessment. *J. Allergy Clin. Immunol. Pract.*, **3**, 863–869. quiz 870.

Aalberse,R.C. *et al.* (2001) Cross-reactivity of IgE antibodies to allergens. *Allergy*, **56**, 478–490.

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Barlow,D.J. *et al.* (1986) Continuous and discontinuous protein antigenic determinants. *Nature*, **322**, 747–748.

Barre,A. *et al.* (2005a) Homology modelling of the major peanut allergen Ara h 2 and surface mapping of IgE-binding epitopes. *Immunol. Lett.*, **100**, 153–158.

Barre,A. *et al.* (2005b) Molecular modelling of the major peanut allergen Ara h 1 and other homotrimeric allergens of the cupin superfamily: a structural basis for their IgE-binding cross-reactivity. *Biochimie*, **87**, 499–506.

Borges,J.P. *et al.* (2007) How reliable is the structural prediction of IgE-binding epitopes of allergens? The case study of plant lipid transfer proteins. *Biochimie*, **89**, 83–91.

Breiteneder,H. and Mills,C. (2006) Structural bioinformatic approaches to understand cross-reactivity. *Mol. Nutr. Food Res.*, **50**, 628–632.

Chen,L. *et al.* (2008) Genomic characterization of putative allergen genes in peach/almond and their synteny with apple. *BMC Genomics*, **9**, 543.

Chen,X. *et al.* (2016) Conformational IgE epitopes of peanut allergens Ara h 2 and Ara h 6. *Clin. Exp. Allergy*, **46**, 1120–1128.

Dall'Antonia,F. *et al.* (2011) Prediction of IgE-binding epitopes by means of allergen surface comparison and correlation to cross-reactivity. *J. Allergy Clin. Immunol.*, **128**, 872–879. e878.

Dall'antonia,F. *et al.* (2014) Structure of allergens and structure based epitope predictions. *Methods*, **66**, 3–21.

de Leon,M.P. *et al.* (2007) IgE cross-reactivity between the major peanut allergen Ara h 2 and tree nut allergens. *Mol. Immunol.*, **44**, 463–471.

El-Manzalawy,Y. *et al.* (2008) Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.*, **21**, 243–255.

Fedorov,A.A. *et al.* (1997) The molecular basis for allergen cross-reactivity: crystal structure and IgE-epitope mapping of birch pollen profilin. *Structure*, **5**, 33–45.

Ferreira,F. *et al.* (2004) Allergic cross-reactivity: from gene to the clinic. *Allergy*, **59**, 243–267.

Foetisch,K. *et al.* (2003) Biological activity of IgE specific for cross-reactive carbohydrate determinants. *J. Allergy Clin. Immunol.*, **111**, 889–896.

Fraczkiewicz,R. and Braun,W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comp. Chem.*, **19**, 319.

Gadermaier,G. *et al.* (2011) Sensitization prevalence, antibody cross-reactivity and immunogenic peptide profile of Api g 2, the non-specific lipid transfer protein 1 of celery. *PloS One*, **6**, e24150.

Gajhede,M. *et al.* (1996) X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nat. Struct. Biol.*, **3**, 1040–1045.

Ganglberger,E. *et al.* (2000) Allergen mimotopes for 3-dimensional epitope search and induction of antibodies inhibiting human IgE. *FASEB J.*, **14**, 2177–2184.

Garcia-Casado,G. *et al.* (2003) Identification of IgE-binding epitopes of the major peach allergen Pru p 3. *J. Allergy Clin. Immunol.*, **112**, 599–605.

Garcia,B.E. and Lizaso,M.T. (2011) Cross-reactivity syndromes in food allergy. *J. Investig. Allergol. Clin. Immunol.*, **21**, 162–170. quiz 162 p following 170.

Guhsl,E.E. *et al.* (2014) Vig r 6, the cytokinin-specific binding protein from mung bean (*Vigna radiata*) sprouts, cross-reacts with Bet v 1-related allergens and binds IgE from birch pollen allergic patients' sera. *Mol. Nutr. Food Res.*, **58**, 625–634.

Haste Andersen,P. *et al.* (2006) Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.*, **15**, 2558–2567.

Hauser,M. *et al.* (2010) Panallergens and their impact on the allergic patient. *Allergy Asthma Clin. Immunol.*, **6**, 1–14.

Ivanciuc,O. *et al.* (2009a) The property distance index PD predicts peptides that cross-react with IgE antibodies. *Mol. Immunol.*, **46**, 873–883.

Ivanciuc,O. *et al.* (2003) SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res.*, **31**, 359–362.

Ivanciuc,O. *et al*. (2009b) Structural analysis of linear and conformational epitopes of allergens. *Regul. Toxicol. Pharmacol*., **54**, S11–S19.

Jensen-Jarolim,E. *et al*. (1998) Peptide mimotopes displayed by phage inhibit antibody binding to Bet v 1, the major birch pollen allergen, and induce specific IgG response in mice. *FASEB J*., **12**, 1635–1642.

JMOL (2007) An open-source Java viewer for three-dimensional molecular structures. http://www.jmol.org/.

Jutel,M. *et al*. (2016) Allergen immunotherapy: past, present, and future. *Allergy Asthma Immunol. Res*., **8**, 191–197.

Klinglmayr,E. *et al*. (2009) Identification of B-cell epitopes of Bet v 1 involved in cross-reactivity with food allergens. *Allergy*, **64**, 647–651.

Konstantinou,G.N. and Grattan,C.E. (2008) Food contact hypersensitivity syndrome: the mucosal contact urticaria paradigm. *Clin. Exp. Dermatol*., **33**, 383–389.

Koppelman,S.J. *et al*. (2005) Purification and immunoglobulin E-binding properties of peanut allergen Ara h 6: evidence for cross-reactivity with Ara h 2. *Clin. Exp. Allergy*, **35**, 490–497.

Larsen,J.E. *et al*. (2006) Improved method for predicting linear B-cell epitopes. *Immunome Res*., **2**, 2.

Li,M. *et al*. (2008) Crystal structure of a dimerized cockroach allergen Bla g 2 complexed with a monoclonal antibody. *J. Biol. Chem*., **283**, 22806–22814.

Mari,A. *et al*. (1999) Specific IgE to cross-reactive carbohydrate determinants strongly affect the in vitro diagnosis of allergic diseases. *J. Allergy Clin. Immunol*., **103**, 1005–1011.

Mari,A. *et al*. (2006) Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cell. Immunol*., **244**, 97–100.

Mayrose,I. *et al*. (2007) Pepitope: epitope mapping from affinity-selected peptides. *Bioinformatics*, **23**, 3244–3246.

Metcalfe,D.D. (2005) Genetically modified crops and allergenicity. *Nat. Immunol*., **6**, 857–860.

Mirza,O. *et al*. (2000) Dominant epitopes and allergic cross-reactivity: complex formation between a fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. *J. Immunol*., **165**, 331–338.

Mittag,D. *et al*. (2006) A novel approach for investigation of specific and cross-reactive IgE epitopes on Bet v 1 and homologous food allergens in individual patients. *Mol. Immunol*., **43**, 268–278.

Naganawa,Y. *et al*. (2008) Epitope analysis of peanut allergen Ara h1 with oligoclonal IgM antibody from human B-lymphoblastoid cells. *Cytotechnology*, **57**, 177–180.

Negi,S.S. and Braun,W. (2007) Statistical analysis of physical-chemical properties and prediction of protein-protein interfaces. *J. Mol. Model*., **13**, 1157–1167.

Negi,S.S. and Braun,W. (2009) Automated detection of conformational epitopes using phage display Peptide sequences. *Bioinform. Biol. Insights*, **3**, 71–81.

Negi,S.S. *et al*. (2007) InterProSurf: a web server for predicting interacting sites on protein surfaces. *Bioinformatics*, **23**, 3397–3399.

Nemethy,G. *et al*. (1983) Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J. Phys. Chem*., **87**, 1883–1887.

Niemi,M. *et al*. (2007) Molecular interactions between a recombinant IgE antibody and the beta-lactoglobulin allergen. *Structure*, **15**, 1413–1421.

Oezguen,N. *et al*. (2008) Comprehensive 3D-modeling of allergenic proteins and amino acid composition of potential conformational IgE epitopes. *Mol. Immunol*., **45**, 3740–3747.

Pacios,L.F. *et al*. (2008) Mimotope mapping as a complementary strategy to define allergen IgE-epitopes: Peach Pru p 3 allergen as a model. *Mol. Immunol*., **45**, 2269–2276.

Padavattan,S. *et al*. (2009) High-affinity IgE recognition of a conformational epitope of the major respiratory allergen Phl p 2 as revealed by X-ray crystallography. *J. Immunol*., **182**, 2141–2151.

Padavattan,S. *et al*. (2007) Identification of a B-cell epitope of hyaluronidase, a major bee venom allergen, from its crystal structure in complex with a specific Fab. *J. Mol. Biol*., **368**, 742–752.

Pastorello,E.A. *et al*. (2000) The maize major allergen, which is responsible for food-induced allergic reactions, is a lipid transfer protein. *J. Allergy Clin. Immunol*., **106**, 744–751.

Pearson,W.R. (1994) Using the FASTA program to search protein and DNA sequence databases. *Methods Mol. Biol*., **25**, 365–389.

Ponomarenko,J. *et al*. (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, **9**, 514.

Power,T.D. *et al*. (2013) Assessment of 3D models for allergen research. *Proteins*, **81**, 545–554.

Schaumann,T. *et al*. (1990) The program fantom for energy refinement of polypeptides and proteins using a Newton-Raphson minimizer in Torsion Angle Space. *Biopolymers*, **29**, 679–694.

Shin,D.S. *et al*. (1998) Biochemical and structural analysis of the IgE binding sites on ara h1, an abundant and highly allergenic peanut protein. *J. Biol. Chem*., **273**, 13753–13759.

Shindyalov,I.N. and Bourne,P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng*., **11**, 739–747.

Singh,H. *et al*. (2013) Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PloS One*, **8**, e62216.

Smith,G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, **228**, 1315–1317.

Spangfort,M.D. *et al*. (2003) Dominating IgE-binding epitope of Bet v 1, the major allergen of birch pollen, characterized by X-ray crystallography and site-directed mutagenesis. *J. Immunol*., **171**, 3084–3090.

Tan,Y.W. *et al*. (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.

Thomas,W.R. *et al*. (2005) Structural biology of allergens. *Curr. Allergy Asthma Rep*., **5**, 388–393.

Tilles,S.A. (2016) Allergen immunotherapy – an old friend coming of age. *Immunol. Allergy Clin. N. Am*., **36**, xiii.

Tiwari,R. *et al*. (2012) Validation of a phage display and computational algorithm by mapping a conformational epitope of Bla g 2. *Int. Arch. Allergy Immunol*., **157**, 323–330.

Untersmayr,E. *et al*. (2006) Mimotopes identify conformational epitopes on parvalbumin, the major fish allergen. *Mol. Immunol*., **43**, 1454–1461.

Van Regenmortel,M.H.V. (1996) Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity. *Methods*, **9**, 465–472.

Vieths,S. *et al*. (2002) Current understanding of cross-reactivity of food allergens and pollen. *Ann. N. Y. Acad. Sci*., **964**, 47–68.

Weber,R.W. (2007) Cross-reactivity of pollen allergens: impact on allergen immunotherapy. *Ann. Allergy Asthma Immunol*., **99**, 203–211. quiz 212-203, 231.

Yao,B. *et al*. (2012) SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PloS One*, **7**, e45152.