DOI 10.1002/mnfr.200500274 Mol. Nutr. Food Res. 2006, *50*, 628–632

Review

Structural bioinformatic approaches to understand cross-reactivity

Heimo Breiteneder[1] and Clare Mills[2]

[1] Department of Pathophysiology, Medical University of Vienna, Vienna, Austria
[2] Institute of Food Research, Norwich Research Park, Colney, Norwich, UK

Cross-reactivity of allergens results from the presence of antibody-accessible conserved surface structures. These can best be studied when allergens have been structurally defined by X-ray crystallography or another structure determination method. When this is not the case, mimotope technology provides a useful alternative for elucidating antibody-binding sites on allergens. Structural bioinformatic approaches have been used to study the cross-reactivity of inhalant allergens with labile food allergens (Bet v 1 family) as well as the cross-reactivity between stable food allergens such as members of the nonspecific lipid transfer protein family. It was found that the degree of similarity of the structures correlated with the observed IgE cross-reactivities. However, IgE cross-reactivity between structurally unrelated allergens has not been demonstrated to date.

Keywords: Allergen / Bioinformatics / Bet v 1 / IgE / nsLTP

Received: June 15, 2005; revised: February 21, 2006; accepted: April 8, 2006

1 The molecular basis of cross-reactivity: Antibody–antigen interactions

The molecular mechanics of antibody cross-reactivity are based on the physicochemical interactions between an antibody's-binding site and a target molecule. Since the binding site architecture of antibodies is independent of their isotype, studies to understand the mechanisms of antibody recognition, including cross-reactivity, using model antibody-target pairs, such as the hen egg white lysozyme (HEL) and a panel of anti-HEL antibodies [1] is highly relevant to understand cross-reactivity in relation to allergy. Like any molecular recognition event, antibody-antigen interactions are a dialogue between the antibody's-binding site and the region on the antigen to which it binds – the epitope.

Epitopes have been classified into two types [2]. The first are linear or continuous epitopes where recognition is based almost entirely on the amino acid sequence with very little effect of conformation. In general such antibodies can recognize unfolded proteins well, although antibodies deliberately raised against peptide sequences buried within the core of a protein will not bind to the native protein for steric reasons. The second type of epitope is conformational where the secondary, tertiary, and quaternary structural elements of a protein bring together sometimes quite distant regions of the polypeptide chain. Antibody binding to such epitopes is generally disrupted when proteins unfold. Structural studies have indicated that antibody binding to proteins involves a surface area of 650–900 Å [2, 3]. The majority of structural epitopes on a protein are regarded as conformational in nature [2]. By making such a clear distinction between continuous and discontinuous epitopes, this classification is somewhat of an oversimplification. In reality there is probably more of a continuum of epitopes running from strictly continuous to strictly conformational types. Thus, continuous epitopes may resemble conformational epitopes by adopting certain secondary structural elements (such as alpha-helix) especially when located within the context of a folded protein. In addition, antibodies may recognize elements of conformational epitopes that resemble linear epitopes, albeit with reduced affinity compared to the full epitope. Epitope mapping methods utilizing short (6–8 residues) synthetic peptides to determine antibody binding are biased toward detecting those antibodies that largely recognize linear epitopes, although they do have the potential to detect regions of a polypeptide that contribute to a conformational epitope.

Correspondence: Professor Heimo Breiteneder, Department of Pathophysiology, Medical University of Vienna, Waehringer Guertel 18–20, 1090 Vienna, Austria
Fax: +431-40400-5130
E-mail: Heimo.Breiteneder@meduniwien.ac.at

Abbreviations: nsLTP, nonspecific lipid transfer protein; **HEL,** hen egg white lysozyme

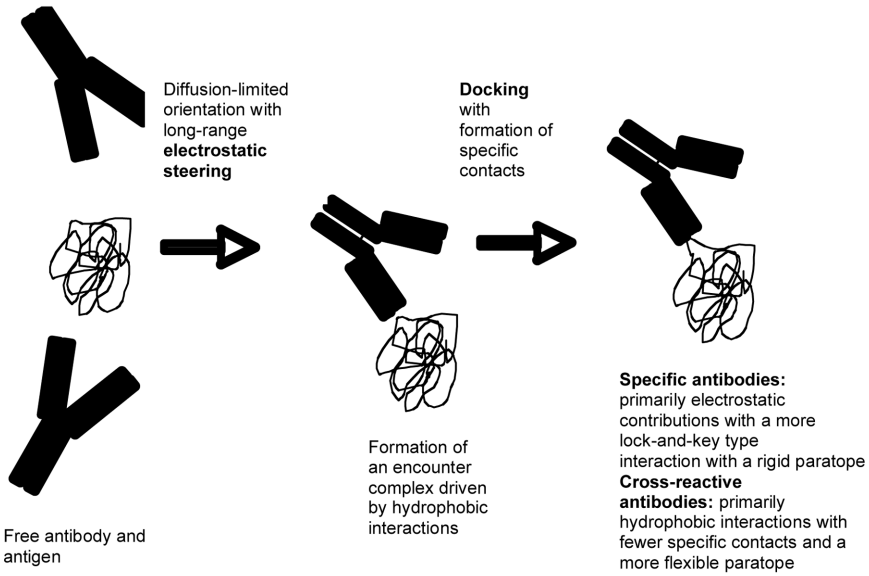

Figure 1. Schematic diagram of the steps involved in the association of an antibody with an antigen summarizing the types of interactions involved.

A peptide mimotope is defined as an amino acid sequence that differs from the one present in a protein but displays the same three-dimensional characteristics. Mimotopes account for the fact that unrelated amino acid sequences can be used to model equivalent three-dimensional structures. An antibody directed to the original epitope also recognizes the mimotope. Phage displayed peptides can be used to define mimotopes of conformational IgE epitopes on allergens. The mimotope technology was successfully used to characterize one IgE epitope of Bet v 1, the major birch pollen allergen. A computer-aided three-dimensional epitope search based on mimotope sequences followed by statistical analysis revealed an exposed area on the Bet v 1 molecule (located between residues 9 – 22 and 104 – 123) as the IgE-binding structure [4]. Combined use of phage display and molecular modeling also facilitated the localization of conformational IgE epitopes on the major grass pollen allergen Phl p 5 [5].

Figure 1 shows the stages involved in antibody-binding events. Initially the antibody and antigen loosely interact in an encounter complex, formation of which is governed by long-range electrostatic interactions with diffusion limited orientation of the molecules. The encounter complex is thought to be held together by hydrophobic associations which then undergo rearrangement (docking) to form the actual antibody-antigen complex. This rearrangement increases the number of specific contacts in a lock-and-key type interaction.

Three antibodies sharing > 90% sequence homology in their combining site recognize the same epitope in HEL, one of which (HH26) is highly specific while another (HH8) is cross-reactive [6, 7]. Structural studies indicate that the more specific antibodies have a more rigid (lock-like) binding site and the work of Sinha *et al.* [7] suggests that interactions with the epitope are dominated by electrostatic forces from salt bridges and hydrogen bonds. In contrast, the more cross-reactive antibodies have a more flexible binding site with fewer specific contacts and interactions appear to involve more hydrophobic interactions. Such a binding site will recognize an epitope in a manner more akin to a handshake rather than a lock-and-key.

For an antibody with a more rigid binding site to recognize a range of targets, those targets will clearly need to possess closely similar epitopes, with a high degree of molecular mimicry and similar structures (be it amino acid sequences or three-dimensional structures). This is probably the more common source of cross-reactivity.

2 Structural similarity in IgE cross-reactive proteins

Identification of potential cross-reactive proteins or epitopes is possible using sequence homology searches and bioinformatics approaches. Sequence homology searches identify potential cross-reactive continuous epitopes which are probably relevant in cases of sensitization to denatured or proteolyzed allergens, as may be the case for allergens proteolyzed in the gastrointestinal tract. The identification of sequential epitopes is well-served by current bioinformatics searching protocols, as long as the data on IgE epitopes are reliable.

Structural bioinformatics approaches, on the other hand, take into account both conservation of the sequence and the structure of allergens that sensitize in native or near-native forms. This is likely to be the case for inhalant allergens that cross-react with labile food allergens or proteolysis/heat-resistant food allergens. A methodology based on a structural bioinformatic approach [8] that was applied to plant foods allergens is summarized below. It could be equally well applied to structurally defined allergens from animal sources.

Common allergenic plant allergens [8] include the following: (i) members of the Bet v 1 family, found in foods that are thought to be allergenic as a result of sensitization to the major birch pollen allergen Bet v 1 [9], and (ii) members of the nonspecific lipid transfer protein (nsLTP) family, where cross-reactive allergies may arise from sensitization to the peach nsLTP allergen Pru p 3 [10], or to a pollen homolog [11]. In both cases, it is likely that the native, folded form of the protein is the sensitizing agent. Thus, Bet v 1 sensitization will occur to folded Bet v 1 that leaches from pollen grains lodged in the nasal mucosa or the lungs, while nsLTPs are highly resistant to digestion [12] and are likely to be present in their native form even in the gut.

IgE resulting from such primary sensitization is highly likely to preferentially recognize the native folded proteins, with many potential conformational epitopes present on the surface of the proteins. Certainly, the major Bet v 1 IgE epitopes are known to be conformational in nature. The fact that natural and recombinant Bet v 1 fragments fail to bind IgE revealed that IgE epitopes of Bet v 1 are discontinuous [13]. Moreover, recombinant Bet v 1 fragments had a >100-fold reduced capacity to induce basophil histamine release and immediate type skin reactions in patients allergic to birch pollen [14, 15]. Short of cocrystallizing an allergen in a specific complex with IgE, conformational epitopes are still difficult to map. However, conformational epitopes can be mapped by introducing point mutations in antibody-accessible molecular surfaces [16, 17] and studying binding of native and mutated allergens to monoclonal antibodies that block IgE [18], or by combining phage display with molecular modeling approaches [5].

An *in silico* approach was developed to quantify differences in surface residues accessible for antibody binding as well as changes in native-folded homologs of the Bet v 1 and nsLTP families [8]. In addition, information from comparisons of both the sequences and three-dimensional structures of selected allergens was combined [8].

2.1 Bet v 1 homologs

In order to assess variation in surface topography among proteins in the Bet v 1 family, the three-dimensional structures of Bet v 1 and several homologs were compared using Bet v 1a (PDB codes 1BV1 and 1B6F), Bet v 1l (1FM4), Pru av 1 from cherry (1E09) and Llpr10.1a and Llpr10.1b from lupin (1ICX and 1IFV). This showed that the Bet v 1a structure, 1BV1, could be aligned three-dimensionally with even the most distantly related structures; for example, the structures of Llpr10.1a (1ICX) and Llpr10.1b (1IFV) aligned with root mean square deviations of 1.514 and 1.728 Å, respectively, over Cα atoms of 135 and 149 residues.

The total surface area of Bet v 1a, including conserved residues and polypeptide backbone, was calculated using a 1.4 Å radius probe [8], and an approach akin to rolling a ball of water across the surface of the protein. The area calculated by this approach is larger than the area accessible to an antibody, since the 1.4 Å probe is small enough to enter the sterol-binding tunnel. In order to take this into account, a 3.5 Å radius probe was also used; this probe is too large to accurately measure the total surface area, but since it is excluded from the tunnel, it was useful to identify surface areas accessible to an antibody [8]. This procedure allowed the protein exterior surface area to be calculated, excluding the sterol-binding tunnel.

Antibody-binding reactions involve contacts between the binding site (paratope) and the epitope which can comprise main-chain atoms and amino acid side chain residues of the allergen, as demonstrated for HEL [19]. Thus, about 1000 $Å^2$ total surface area corresponded to the tunnel, and the exterior surface area of Bet v 1a was estimated to be approximately 8880 $Å^2$, 75% of which arose from side chains. This result is in agreement with another estimate of 990 $Å^2$ as the surface area of the tunnel, which was calculated using two deoxycholate molecules found in one structure of Bet v 1 [20].

The amino acid sequences of Bet v 1a homologs were aligned, including proteins from apple (Mal d 1b), soybean (Gly m 4), and celery (Api g 1.0101). This alignment did not require more than a single internal deletion between pairs of sequences. Conserved exterior surface residues were then identified by comparing structures of Bet v 1a, Mal d 1b, Gly m 4, and Api g 1.0201. This analysis illustrated the extent to which conserved surface areas form patches that might bind IgE. The most similar region of the surface for the most distantly related protein pair, Bet v 1a and Api g 1, contained a T-cell epitope identified by Bohle *et al.* [21] at residues 25–45. In contrast, Bet v 1a and Gly m 4 had a similar surface on the P-loop around amino acid residue 45 [8].

From this analysis it became evident that the level of conservation of amino acid side chains in the Bet v 1 homologs was considerably higher than indicated by overall similarity/identity levels. Thus, approximately 71% of the surface area of Mal d 1b is conserved, even though the overall sequence identity with Bet v 1 is 56%. Similarly, the overall

sequence identity between Gly m 4 and Bet v 1 is approximately 47%, but approximately 60% of the surface and main-chain residues are conserved. While the contribution from the main-chain residues is similar for all Bet v 1a homologs, the fraction of conserved external surface area from side chain residues decreased with decreasing sequence identity relative to Bet v 1a.

Thus, the sequence identity of Bet v 1a with Mal d 1 (56%) is lower than the conserved exterior surface area (71%) while these values for Gly m 4 are 47% *versus* 60%. However, for Api g 1.0101 the conserved surface area from the side chains (30%) is lower than the sequence identity (39%). The deletion in Api g 1.0101 (also seen in Api g 1.0201) relative to Bet v 1a will contribute to this and since it is located in the so-called P-loop around residue 45 where a major IgE epitope in Bet v 1 and Pru av 1 (originally known as Pru a 1) is located [22, 23] it is likely to modify IgE binding to a considerable extent. These structural biology calculations – although independent of antibody binding – form the basis for studies of the correlation of structural similarities of allergens with the actual *in vitro* and *in vivo* cross-reactivity.

2.2 nsLTP homologs

A similar analysis of the nsLTP family of the prolamin superfamily was performed comparing sequences from Pru av 3 from peach, Zea m 14 from maize, and wheat nsLTP1. These proteins were chosen because it is currently thought that peach nsLTP (Pru p 3) is the sensitizing agent and there is demonstrable cross-reactivity with nsLTP from maize (Zea m 14 [24]) with none reported for wheat nsLTP1. As shown for the Bet v 1 homologs, several "interior-surface" residues in the lipid-binding tunnel of nsLTPs are not expected to be accessible to IgE binding, as long as the protein retains its native conformation. In addition, for several residues in the loop regions between the helices, the side chains contribute to the lipid-binding cavity and the main-chain atoms contribute to both interior and exterior surfaces; these residues are potentially capable of interacting with IgE. Inclusion of these residues therefore results in overestimation of the calculated surface area. One of these loops is believed to be involved in IgE binding by peach nsLTP [25]. Peach and maize nsLTPs have a large proportion of their surface conserved, which could account for the observed cross-reactivity. While some of the surface of wheat nsLTP is conserved with peach, wheat nsLTP has a very different surface topography and hence would be unlikely to cross-react.

3 Antibody multispecificity and potentially unrelated cross-reactants

In addition to characterizing cross-reactivity to structurally similar homologous antigens, there is also the potential for cross-reactivity to unrelated antigens. In the past, IgE cross-reactivity has been wrongly ascribed to recognition of unrelated cross-reactants, as a consequence of our lack of understanding of structural relationships between proteins from different species. Thus, the availability of the three-dimensional structures of Bet v 1 homologs has revealed that cross-reactivity of allergens arises from their structural similarity. Nevertheless, there is good evidence that cross-reactive binding (or probably more correctly, polyspecificity) to unrelated ligands can result from plasticity in the antibody-binding site and the ability of another epitope to make appropriate contacts in the binding site.

The number of complementary ligands that can be identified for a given binding site is a function of the size of the library being screened and the affinity cut-off. Thus, alternative ligands can always be found for any protein provided a large enough library is screened. For example, 2000 compounds were screened for binding to an antidinitrophenol IgE antibody, SPE7 [26], and three structurally diverse compounds with similar affinities to the immunizing hapten were identified. One of these, Alizarin Red, is a large conjugated ring system that bound with the same K_d (0.04 μM) as 2,4-dinitrophenyl lysine. In this instance, promiscuous binding did not arise from hydrophobic interactions in the binding site which may be intrinsically "sticky", but it was instead due to topographical heterogeneity of the antibody-binding site and the ligand, which allowed specific contacts to be generated with little structural rearrangement.

Promiscuous binding to protein ligands would seem to be less likely than for haptens, given the surface area of the contacts involved. However, polyspecificity may also result from conformational diversity in the antibody-binding site [27]. For example, the model IgE antibody SPE7 adopts at least two pre-existing conformations in the antibody-binding site, each of which possesses different antigen-binding functions. These conformations resulted from changes in the disposition of the L3 and H3 loops, the different conformers being able to either bind haptens or cyclic peptides on the active site loop of thioredoxin (Trx-proteins). Haptens bound to the conformer presenting a deep narrow pocket, with an initial low-affinity interaction and subsequent induced-fit rearrangements to give the final hapten-antibody complex. In contrast, the Trx-proteins bound to a conformer with a shallow, more solvent-exposed binding site with a much lower affinity (~10 μM) than the haptens.

The concept that antibodies may bind promiscuously through flexibility in the antibody-binding site is also supported by studies on the anti-HEL antibody HH8 [6, 7]. For this antibody, plasticity in the binding site is achieved through rearrangement in the conformation of the H2 loop in the complementarity determining region.

4 Predicting cross-reactivity

The current allergy literature indicates that the majority of IgE-based cross-reactive allergy syndromes are caused by structural similarity between allergens. Molecular mimicry also plays a role in other immune-related disorders, notably autoimmune syndromes where microbial infections can give rise to immune (including antibody) responses which also recognize "self" molecules [28]. For those antibodies exhibiting this kind of behavior, potential cross-reactive allergens can be predicted. Confidence in the reliability of these predictions will undoubtedly increase in the future, if efforts on the reliability of IgE epitope identification improve and the database of allergen structures increases in size and completeness. However, predicting cross-reactivity between unrelated proteins is more difficult, because it could reflect promiscuity of the antibody. It is unclear at present what proportion of the antibody repertoire is likely to have sufficient flexibility in the binding site to allow promiscuous, yet specific, binding to take place. It is possible that such promiscuous antibodies would have the potential to manifest cross-reactivity between completely unrelated proteins. The significance of such interactions is unknown, and with a low-level of affinity, the clinical relevance of such interactions would also be questionable.

5 Conclusions

The study of surface structures of allergens – by structural bioinformatics or using mimotope approaches – is a more accurate way to assess and possibly predict IgE cross-reactivity of allergens and query proteins than simple sequence comparison. So far, IgE cross-reactivity that manifests with clinical symptoms related to food allergy has only been observed for members of the same protein family. In these cases, the degree of cross-reactivity is directly correlated to the degree of structural similarity of the allergens.

Author H. B. wishes to acknowledge the support of the Austrian Science Fund grant SFB F01802 and of the EU through QLK1-CT-2000-01394 (SAFE). Author CM acknowledges the support by the EU through InformAll (QLRT-2001-02284), QLK1-CT-2000-01394 (SAFE), and by the BBSRC through the competitive strategic grant to the Institute of Food Research.

6 References

[1] Lavoie, T. B., Mohan, S., Lipschultz, C. A., Grivel, J. C., *et al.*, *Mol. Immunol.* 1999, *36*, 1189–1205.

[2] Van Regenmortel, M. H. V., *Structure of Antigens*, Van Regenmortel MHV Edn., CRC Press, USA 1992, Vol. I, pp. 1–27.

[3] Laver, W. G., Air, G. M., Webster, R. G., Smith-Gill, S. J., *Cell* 1990, *61*, 553–556.

[4] Ganglberger, E., Grünberger, K., Sponer, B., Radauer, C., *et al.*, *FASEB J.* 2000, *14*, 2177–2184.

[5] Hantusch, B., Krieger, S., Untersmayr, E., Scholl, I., *et al.*, *J. Allergy Clin. Immunol.* 2004, *114*, 1294–1300.

[6] Mohan, S., Sinha, N., Smith-Gill, S. J., *Biophys. J.* 2003, *85*, 3221–3236.

[7] Sinha, N., Mohan, S., Lipschultz, C. A., Smith-Gill, S. J., *Biophys. J.* 2002, *83*, 2946–2968.

[8] Jenkins, J. A., Griffiths-Jones, S., Shewry, P. R., Breiteneder, H., *et al.*, *J. Allergy Clin. Immunol.* 2005, *115*, 164–171.

[9] Vieths, S., Scheurer, S., Balmer-Weber, B., *Ann. NY Acad. Sci.* 2002, *964*, 47–68.

[10] Asero, R., Mistrello, G., Roncarolo, D., Amato, S., *Ann. Allergy Asthma Immunol.* 2004, *92*, 268–272.

[11] San Miguel-Moncin, M., Krail, M., Scheurer, S., Enrique, E., *et al.*, *Allergy* 2003, *58*, 511–517.

[12] Asero, R., Mistrello, G., Roncarolo, D., de Vries, S. C., *et al.*, *Int. Arch. Allergy Immunol.* 2000, *122*, 20–32.

[13] Vrtala, S., Hirtenlehner, K., Vangelista, L., Pastore, A., *et al.*, *J. Clin. Invest.* 1997, *99*, 1673–1681.

[14] van Hage-Hamsten, M., Kronqvist, M., Zetterstrom, O., Johansson, E., *et al.*, *J. Allergy Clin. Immunol.* 1999, *104*, 969–977.

[15] Pauli, G., Purohit, A., Oster, J. P., De Blay, F., *et al.*, *Clin. Exp. Allergy* 2000, *30*, 1076–1084.

[16] Ferreira, F., Ebner, C., Kramer, B., Casari, G., *et al.*, *FASEB J.* 1998, *12*, 231–242.

[17] Holm, J., Gajhede, M., Ferreras, M., Henriksen, A., *et al.*, *J. Immunol.* 2004, *173*, 5258–5267.

[18] Wiche, R., Gubesch, M., Konig, H., Fotisch, K., *et al.*, *Biochem. J.* 2005, *385*, 319–327.

[19] Li, Y., Urrutia, M., Smith-Gill, S. J., Mariuzza, R. A., *Biochemistry* 2003, *42*, 11–22.

[20] Markovic-Housley, Z., Degano, M., Lamba, D., von Roepenack-Lahaye, E., *et al.*, *J. Mol. Biol.* 2003, *325*, 123–133.

[21] Bohle, B., Radakovics, A., Jahn-Schmid, B., Hoffmann-Sommergruber, K., *et al.*, *Eur. J. Immunol.* 2003, *33*, 3303–3310.

[22] Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., *et al.*, *Nat. Struct. Biol.* 1996, *3*, 1040–1045.

[23] Scheurer, S., Son, D. Y., Boehm, M., Karamloo, F., *et al.*, *Mol. Immunol.* 1999, *36*, 155–167.

[24] Pastorello, E. A., Farioli, L., Pravettoni, V., Ispano, M., *et al.*, *J. Allergy Clin. Immunol.* 2000, *106*, 744–751.

[25] Garcia-Casado, G., Pacios, L. F., Diaz-Perales, A., Sanchez-Monge, R., *et al.*, *J. Allergy Clin. Immunol.* 2003, *112*, 599–605.

[26] James, L. C., Tawfik, D. S., *Protein Sci.* 2003, *12*, 2183–2193.

[27] James, L. C., Roversi, P., Tawfik, D. S., *Science* 2003, *299*, 1362–1367.

[28] Oldstone, M. B., *FASEB J.* 1998, *12*, 1255–1265.

