Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Various cross-reactivity of the grass pollen group 4 allergens: crystallographic study of the Bermuda grass isoallergen Cyn d 4

Tse-Hao Huang,[a,b] Ho-Jen Peng,[c] Song-Nan Su[c] and Shwu-Huey Liaw[a,c]*

[a]Department of Life Sciences, National Yang-Ming University, Taipei 11221, Taiwan, [b]Institute of Biochemistry and Molecular Biology, National Yang-Ming University, Taipei 11221, Taiwan, and [c]Department of Medical Research and Education, Taipei Veterans General Hospital, Taipei 11217, Taiwan

Correspondence e-mail: shliaw@ym.edu.tw

Received 17 April 2012
Accepted 18 June 2012

PDB Reference: Cyn d 4, 4dns.

The structure of Cyn d 4, the group 4 allergen from Bermuda grass, is reported at 2.15 Å resolution and is the first crystal structure of a naturally isolated pollen allergen. A conserved N-terminal segment that is only present in the large isoallergens forms extensive interactions with surrounding residues and hence greatly enhances the structural stability of the protein. Cyn d 4 contains an FAD cofactor that is covalently linked to His88 and Cys152. To date, all identified bicovalent flavoproteins are oxidases and their substrates are either sugars or secondary metabolites. A deep large hydrophobic substrate-binding cleft is present. Thus, Cyn d 4 may be an oxidase that is involved in the biosynthesis of a pollen-specific metabolite. Cyn d 4 shares ∼70% sequence identity with the Pooideae group 4 allergens. Various cross-reactivities between grass pollen group 4 allergens have previously been demonstrated using sera from allergic patients. The protein surface displays an unusually large number of positively charged clusters, reflecting the high pI of ∼10. 38 decapeptides that cover the solvent-accessible sequences did not show any significant IgE-binding activity using sera with high Cyn d 4 reactivity from four patients, suggesting that the IgE epitopes of Cyn d 4 are predominantly conformational in nature. Several group 4 structures were then modelled and their potential cross-reactive and species-specific IgE epitopes were proposed.

1. Introduction

Grass pollens represent a major cause of allergic diseases that afflict up to 40% of the atopic population. In temperate climates, members of the Pooideae subfamily such as Timothy-grass (*Phleum pratense*) are clinically significant, while in warm temperate and subtropical areas Bermuda grass (*Cynodon dactylon*, subfamily Chloridoideae) becomes an important source of aeroallergens. The IgE responses of a large number of patients who are sensitized to various grass-pollen extracts show extensive cross-allergenicity between the Pooideae species, but Bermuda grass allergens display differences in immunological reactivity (Martin *et al.*, 1985; Schumacher *et al.*, 1985; Andersson & Lidholm, 2003; Johansen *et al.*, 2009). Consequently, an international standardized preparation of Bermuda grass pollen (BGP) extract has been established (Baer *et al.*, 1986). In these circumstances, characterization of the BGP allergens should provide insights that will help in the development of improved diagnostic approaches and better therapy for this type of allergic disease.

Grass pollen group 4 allergens have been identified as high-molecular-weight basic glycoproteins (Ekramoddoullah *et al.*, 1983; Haavik *et al.*, 1985; Su *et al.*, 1991, 1996; Fischer *et al.*, 1996; Leduc-Brodard *et al.*, 1996; Gavrović-Jankulović *et al.*, 2000) that are able to react with IgE antibodies from up to 80% of grass-allergic patients, and hence are regarded as major allergens (Rossi *et al.*, 2001; Stumvoll *et al.*, 2002). BG60 from Bermuda grass, with an apparent molecular weight of 56–60 kDa, has been purified as an isoallergen mixture over the pI range 9.8–10.5 (Su *et al.*, 1991, 1996). These isoallergens have been characterized as the first flavinylated allergens and also as glycoproteins that contain some unusual novel plant oligosaccharides (Liaw *et al.*, 2001; Ohsuga *et al.*, 1996). A full-length cDNA (AAS02108) has been cloned, with a predicted 25-residue signal peptide and a 497-residue mature protein. BG60 shares ∼70% protein-sequence identity with the Pooideae group 4 allergens, including Phl p 4, Lol p 4, Sec c 4, Hor v 4 and Tri a 4 (Nandy *et al.*, 2005). BG60 is referred to as Cyn d 4 throughout the rest of this paper.

Various cross-allergenicities between pollen group 4 allergens from different grass species have been demonstrated using polyclonal and monoclonal IgG and IgE antibodies of grass pollen-allergic patients. Rossi and coworkers found that the majority of 411 BGP-allergic patients were co-sensitive to native or recombinant Timothy-grass allergens, including the native Phl p 4 (Rossi *et al.*, 2008). In contrast, neither natural Lol p 4 nor Phl p 4 bound to polyclonal murine IgG antibodies that had been raised against Cyn d 4 (Su *et al.*, 1991). Nevertheless, Stumvoll and coworkers demonstrated that Phl p 4 possesses cross-reactive epitopes as well as grass species-specific epitopes by reacting sera from Phl p 4-reactive patients and rabbit antibodies with pollen extracts from trees, grasses and weeds (Stumvoll *et al.*, 2002). Here, we determine the crystal structure of native Cyn d 4 and then model several Pooideae group 4 allergen structures to provide a structural basis for the identification of the potential shared and unique epitopes of these allergens.

2. Materials and methods

2.1. Structural analysis

Cyn d 4 isoallergens were isolated from grass pollen (Su *et al.*, 1996) and green crystals were grown from the isoallergen mixture (Liaw *et al.*, 1999). X-ray diffraction data sets were collected and processed on beamlines BL13C at NSRRC, Hsinchu, Taiwan and BL-6B at the Photon Factory, Tsukuba, Japan. Since Cyn d 4 shares 42% sequence identity with the *Eschscholzia californica* berberine bridge enzyme (*Ec*BBE; PDB entry 3d2h; Winkler *et al.*, 2008), the structure was determined by the molecular-replacement method and refined using *CNS* (Brünger *et al.*, 1998). The statistics of data collection and structural refinement are summarized in Table 1. Each asymmetric unit contains two Cyn d 4 molecules without extensive contacts, which is consistent with the monomeric state in solution observed by size-exclusion chromatography (Su *et al.*, 1996). We modelled the three-dimensional structures of other grass pollen group 4 allergens using the *SWISS-MODEL* server (Arnold *et al.*, 2006) with the Cyn d 4 structure as a template. Solvent-accessible areas for each residue were calculated with *GETAREA* (Fraczkiewicz & Braun, 1998) and residues with a solvent accessibility of >50% were considered to be solvent-exposed. Figs. 1 and 3(*a*) were generated by *MolScript* (Esnouf, 1999) and Fig. 2 was generated by *BobScript* (Esnouf, 1999). Based on structural superposition, a multiple sequence alignment of several bicovalent flavoproteins was generated (Fig. 3*b*), from which the structures of several members were predicted using *Phyre* (Kelley & Sternberg, 2009). The electrostatic potentials in Fig. 4(*a*) were obtained using the *PyMOL* software (DeLano, 2002).

Table 1
Statistics of data collection and structural refinement.

Values in parentheses are for the highest resolution shell.

PDB code	4dns
Data collection	
Space group	$P4_32_12$
Unit-cell parameters (Å)	$a = b = 86.0$, $c = 309.4$
Resolution (Å)	50–2.15 (2.23–2.15)
Observed reflections	547591 (34592)
Unique reflections	63094 (6206)
Completeness (%)	98.2 (99.0)
$\langle I/\sigma(I)\rangle$	21.6 (6.1)
$R_{\rm merge}$† (%)	5.9 (31.1)
Refinement	
Resolution (Å)	50–2.15 (2.23–2.15)
Reflections [$F > 0\sigma(F)$]	61781 (5033)
$R_{\rm cryst}$‡ (%)	18.1 (20.9)
$R_{\rm free}$§ (%)	21.7 (25.5)
Root-mean-square deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.36
No. of protein atoms	7550
No. of FAD atoms	106
No. of water molecules	310
Mean B factor (Å^2)	
Protein atoms	27.3
FAD atoms	23.1
Water molecules	34.3
Ramachandran plot¶ (%)	
Favoured	96.2
Allowed	3.5
Disallowed	0.3

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the average intensity of equivalent reflections. ‡ $R_{\rm cryst} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}||/\sum_{hkl}|F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are observed and calculated structure factors, respectively. § $R_{\rm free}$ was calculated using 10% of the data that were randomly excluded from refinement. ¶ The Ramachandran analysis was performed using the *RAMPAGE* program (Lovell *et al.*, 2003).

2.2. IgE-binding assay

Designed decapeptides spanning the solvent-accessible sequences of Cyn d 4 were custom-synthesized. Serum samples were kindly provided by Su and coworkers (Su *et al.*, 1996). A dot-blot immunoassay of the IgE reactivity of the peptides was performed as described previously (Huang *et al.*, 2008). 10 µg of each peptide was applied onto a polyvinyl difluoride membrane in triplicate. Free sites were blocked with phosphate-buffered saline (PBS) containing 3% skimmed milk and 0.1% Tween-20 for 1 h. Four allergic patient sera

were diluted 1:10 in PBS and incubated with the dot blots overnight at 277 K. The membranes were then immersed in a 1:5000 dilution of horseradish-peroxidase-conjugated goat anti-human IgE (Amersham, Buckinghamshire, England) at room temperature for 5 h. The dot blots were extensively washed with PBS after each treatment. Finally, the signal from the reaction was developed with enhanced Amersham chemiluminescence reagent for 2 min and exposed to X-ray film for 5–30 s.

3. Results

3.1. The Cyn d 4 structure

N-terminal sequencing of the native protein did not reveal the start of Cyn d 4 (Liaw *et al.*, 2001). The protein residues of Cyn d 4 are numbered according to the predicted mature sequence. The current atomic model contains residues 12–497. A structural similarity search using *DALI* (Holm & Rosenström, 2010) revealed that Cyn d 4 belongs to the vanillyl alcohol oxidase (VAO) superfamily (Leferink *et al.*, 2008). As in other VAO members, the protein structure is composed of two distinct domains (Fig. 1*a*). The FAD-binding domain includes the N-terminal residues 12–224 and the C-terminal residues 442–497 and folds into two ($\alpha + \beta$) subdomains. One of the small subdomains (residues 12–111) consists of four central β-strands (β1–β4) sandwiched by three helices (αA, αA′ and αB), while the other consists of five β-strands (β5–β9) surrounded by five helices (αC, αD, αD′, αJ and αK). The substrate-binding domain (residues 225–441) is composed of seven β-strands (β10–β16) flanked by five helices (αE–αI). There are disulfide bonds between Cys16 and Cys73 and between Cys283 and Cys304. Two N-glycosylation sites were identified (Asn63 and Asn300). The electron density showed excellent correlation, with density corresponding to one or two *N*-acetyl-D-glucosamine rings linked to these residues.

3.2. A conserved N-terminal segment

SDS–PAGE analysis of the isolated Cyn d 4 isoallergens displayed two major protein bands with higher molecular weights (the L forms) and one band with a lower molecular weight (the S form) (Liaw *et al.*, 2001). The S isoform starts at residue Lys54, whereas the N-termini of the L isoforms could not be sequenced. The lack of the N-terminal segment in the S isoform results in higher protease susceptibility and a T_m that is ∼20 K lower (Liaw *et al.*, 2001). Crystals were only grown from the L isoforms (Liaw *et al.*, 1999). In the Cyn d 4 structure residues 12–53 form helices αA and αA′ and the β1 strand, which interact extensively with the surrounding residues (Fig. 1*b*). At least 20 residues are involved in the interaction network, including the Cys16–Cys73 disulfide bridge, hydrophobic patches and 20 direct hydrogen bonds between atoms within

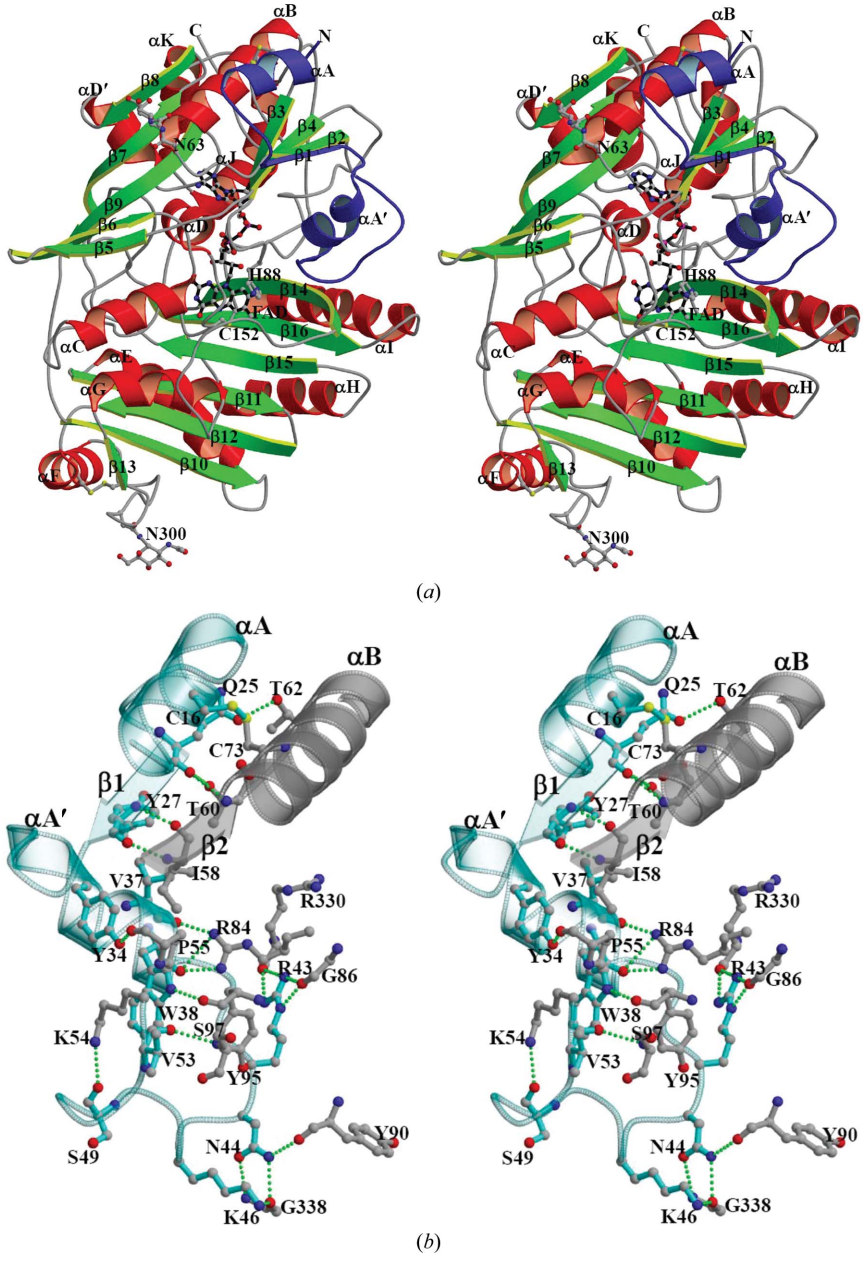

Figure 1
(*a*) Stereoview of the Cyn d 4 structure. The protein consists of a FAD-binding and a substrate-binding domain, with the N-terminal segment highlighted in blue. The cofactor FAD, the linking residues His88/Cys152 and the glycosylated Asn63/Asn300 are displayed in ball-and-stick representation. (*b*) Stereoview of the interaction networks between the N-terminal segment (cyan) and the surrounding residues (grey). Hydrogen bonds are shown as green dashed lines. These extensive interactions greatly enhance the structural stability of the N-terminal small subdomain and hence dramatically contribute to the structural integrity of Cyn d 4. The Cys16–Cys73 disulfide bond is conserved in the plant BBE-like family.

the protein. These extensive interactions greatly enhance the structural stability of the N-terminal small subdomain and hence dramatically contribute to the structural integrity of Cyn d 4. This segment is also conserved in other VAO members and is probably essential for the compact conformation.

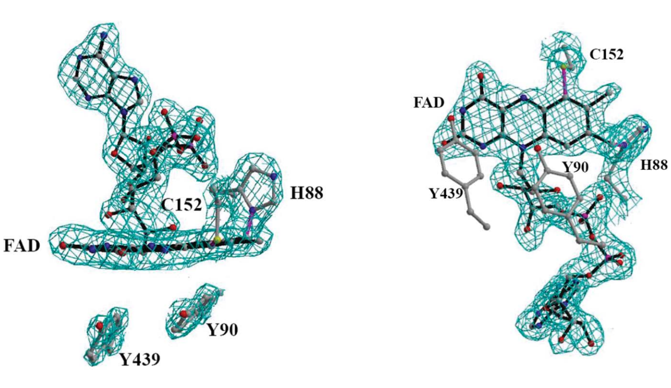

Figure 2
Two perpendicular views of the $2F_o - F_c$ electron-density map of the FAD cofactor, His88, Tyr90, Cys152 and Tyr439 contoured at the 2σ level. Tyr90 and Tyr439 are predicted to be involved in substrate binding and proton transfer during enzyme catalysis, respectively. The densities of Tyr90 and Tyr439 are not shown in the right-hand view for clarity.

3.3. The bicovalent flavoproteins

In Cyn d 4, the C6 atom and the 8α-methyl group of the isoalloxazine ring of the FAD cofactor are covalently linked to Cys152 S^{γ} and His88 $N^{\delta 1}$, respectively, to give 6-*S*-cysteinyl and 8α-*N*1-histidyl FAD (Fig. 2). As in other VAO members, most of the FAD-interacting residues in Cyn d 4 are located near to the two flavinylation sites within glycine-rich regions and utilize their amide backbones for cofactor binding (Huang *et al.*, 2005). A bicovalent FAD cofactor has also been observed in crystal structures of *Ec*BBE (Winkler *et al.*, 2008), glucooligosaccharide oxidase from *Acremonium strictum* (*At*GOOX; Huang *et al.*, 2005), aclacinomycin oxidase from *Streptomyces galilaeus* (*Sg*AknOX; Alexeev *et al.*, 2007), pregilvocarcin V oxidase from *S. galilaeus* (*Sg*GilR; Noinaj *et al.*, 2011) and the glycopeptide hexose oxidase Dbv29 from the filamentous actinomycete *Nonomurea* (Dbv29; Liu *et al.*, 2011). In addition, biochemical studies have revealed the same type of bicovalent linkage in *Chondrun crispus* hexose oxidase (*Cc*HEOX; Rand *et al.*, 2006) and *Fusarium graminearum* chitooligosaccharide oxidase (*Fg*ChitO; Heuts *et al.*, 2008).

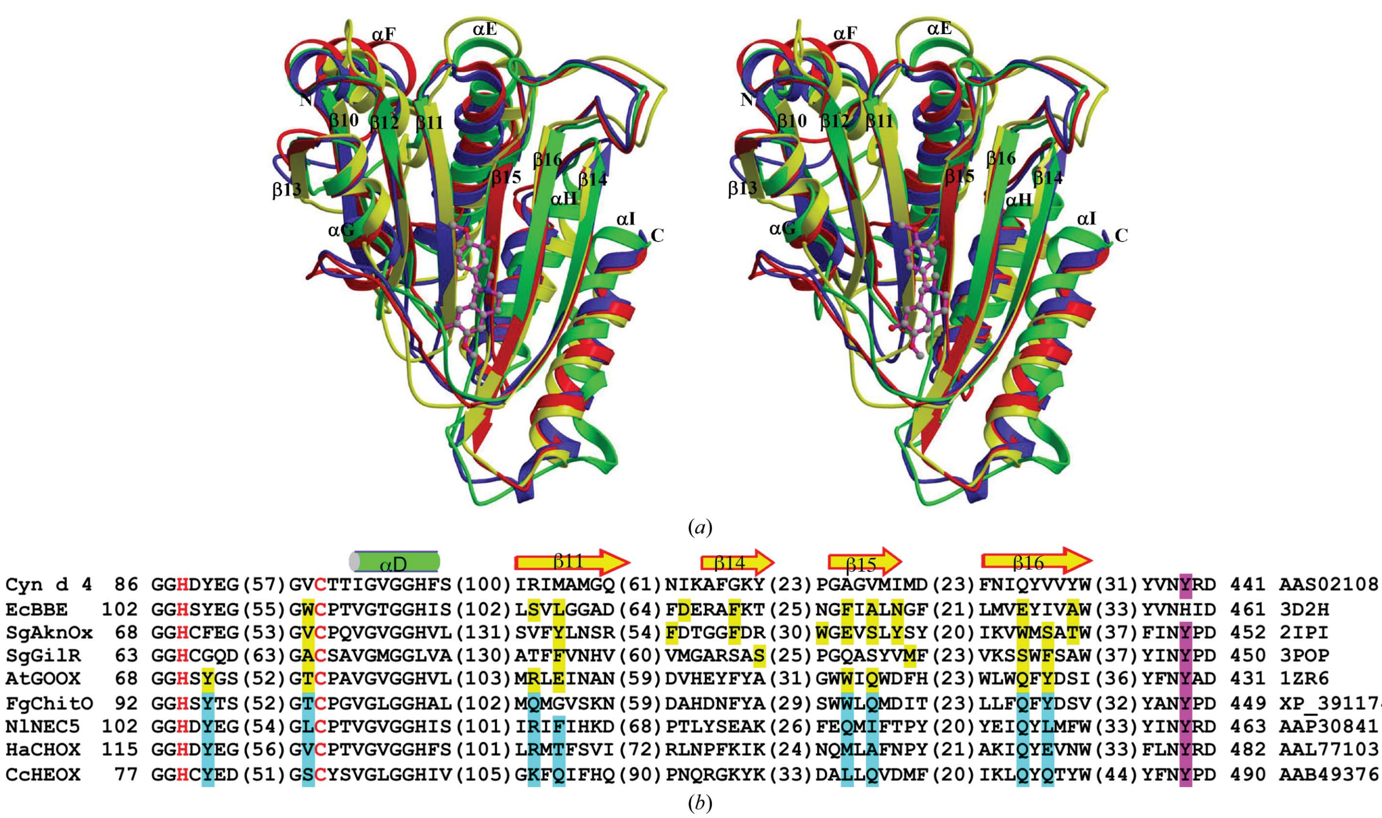

Figure 3
(*a*) Stereoview of structural superposition of Cyn d 4 (red), *Ec*BBE (blue), *At*GOOX (green) and *Sg*AknOx (yellow). The BBE substrate (*S*)-reticuline (magenta) is shown in ball-and-stick representation. These bicovalent flavoproteins share a structurally conserved substrate-binding domain. (*b*) Structure-based sequence alignment of some bicovalent flavoproteins. The substrate-binding residues in the known complex structures are shaded in yellow, while those that are predicted are shaded in cyan. The conserved tyrosine residue involved in proton transfer is shown in magenta. This structure-sequence analysis should be able to predict the active-site residues of family members with unsolved structures.

This bicovalent flavinylation greatly enhances the enzyme oxidation efficiency *via* modulation of the redox potential of the cofactor (Leferink *et al.*, 2008).

3.4. Putative conserved substrate-binding regions

The VAO members have a conserved FAD-binding domain, perhaps because of a similar FAD-recognition mechanism, and a diverse substrate-binding domain, but have distinct substrate specificity ranging from the small lactate to bulky glycopeptides and long-chain oligosaccharides (Huang *et al.*, 2005; Liu *et al.*, 2011). For example, Cyn d 4 shares 48 and 34% sequence identity to *Ec*BBE in the two domains, respectively. All of the identified bicovalent flavoproteins have a large substrate-binding site. The bicovalent flavoenzymes share a superimposable substrate-binding domain, particularly β11 and β14–16, which make up the majority of the substrate-binding groove (Fig. 3*a*). While the sequences of the substrate-interacting residues are divergent to allow substrate specificity, their spatial positions nevertheless converge on four strands (β11, β14, β15 and β16), in addition to some being near the two flavinylation sites (Fig. 3*b*). Such structure–sequence analysis should help to provide predicative information with respect to the substrate-interacting residues of members with unsolved structures. For example, even though the substrate-binding domains of *At*GOOX and *Fg*ChitO share only 34% sequence identity, they seem to utilize virtually identical residues for substrate interaction (Fig. 3*b*), with the exception of Arg270 in *At*GOOX, while Gln268 in *Fg*ChitO is involved in *N*-acetyl recognition (Heuts *et al.*, 2008). Finally, enzyme catalysis is initiated by proton abstraction from the substrate by a conserved tyrosine residue in *At*GOOX, *Sg*AknOX, *Sg*GilR and Dbv29 (Fig. 3*b*). The size and shape of the substrate-binding site in Cyn d 4 is similar to those in *Ec*BBE and *Sg*GilR, but differs from those in *At*GOOX, *Sg*AknOX and Dbv29. Like *Ec*BBE and *Sg*GilR, Cyn d 4 contains a deep large cavity in close proximity to the flavin ring in which most of the residues are hydrophobic; only a few polar residues are present (Fig. 3*b*). A large number of van der Waals interactions, but only a few hydrogen bonds, are observed between the protein and the bicyclic and tetracyclic complex substrates in *Ec*BBE and *Sg*GilR, respectively. Therefore, we suspect that the substrate of Cyn d 4 is an apolar bulky compound, perhaps one with aromatic rings.

3.5. Putative antibody-binding regions

The IgE-binding activity of Cyn d 4 treated with 1 m*M* sodium periodate was reduced to 80% of that of the unoxidized allergen (Su *et al.*, 1996). Similar results were observed for Phl p 4: treatment with 10 m*M* periodate led to a 20–86% reduction of the IgE reactivity (Stumvoll *et al.*, 2002). These observations indicated that not only the carbohydrates but also the protein are involved in the immune response. Most IgE-binding protein epitopes are identified using overlapping peptides corresponding to the allergen sequences and hence are regarded as sequential (linear) epitopes. Since sequential epitopes are probably less conformationally dependent, we first attempted to identify putative sequential epitopes on the surface-exposed loops of Cyn d 4. Nine decapeptides corresponding to residues 46–55, 106–115, 221–230, 326–335, 346–355, 381–390, 426–435, 431–440 and 446–455 were synthesized and their IgE reactivities were analyzed by a dot-blot immunoassay (Huang *et al.*, 2008) using sera from four Cyn d 4-reactive patients (Su *et al.*, 1996). However, these peptides showed no detectable IgE reactivity. Next, 29 peptides overlapping by five residues covering other solvent-accessible regions with α-helix or β-strand structures were assayed; these also displayed no detectable IgE binding, except for residues 291–300 and 301–310, which showed weak reactivity (data not shown). These results suggest that the IgE-binding protein epitopes of Cyn d 4 are predominantly conformational in nature.

Next, we analyzed the Cyn d 4 structure for surface-exposed residues in order to explore potential conformational IgE-binding epitopes. In our solved Cyn d 4 structure, there are 121 residues with a relative surface accessibility of >50% (24 Lys, 16 Pro, 12 Ala, 10 Thr, nine Ser, eight Glu, seven Asp, seven Gly, six Asn, six Arg, four Gln, four Leu, three Val, one Cys, one Phe, one His, one Ile and one Tyr). This is consistent with a statistical survey of identified surface-exposed epitopes, in which glycine, alanine, serine, asparagine and particularly lysine have a high propensity to be involved in IgE binding (Oezguen *et al.*, 2008). A large number of solvent-exposed positively charged residues produced the high pI of this protein ($\sim$10) and a dominant positive electrostatic potential on the protein surface, except for a few regions in which there are small strongly negative areas (Fig. 4*a*). To identify the putative cross-reactive IgE epitopes, the molecular surfaces of Cyn d 4 and the modelled group 4 allergen structures were analyzed for conserved and solvent-accessible residues. A large coherent conserved stretch extends from the top of the molecule to the active site from Asp12, Lys19, Arg76, Arg24, Lys29, Lys54, Lys110 and Lys112 to Lys335 and Phe405. Another similar conserved patch can be found on the other side of the molecule, stretching from Asp12, Lys19, Asp20, Arg191 and Lys480 to Lys119, Glu200 and Glu472 (Fig. 4*a*). Five conserved clusters could be identified and may comprise putative cross-reactive epitopes, namely Asp12/Lys19/Arg76/His77, Arg24/Lys110/Lys112, Arg195/Lys196/Glu200, Lys321/Lys335/Phe405 and Lys460/Glu464/Glu472/Arg473.

In addition, there are some surface regions with significant structural differences that may account for species-specific epitopes. When only surface amino acids with 50% solvent exposure are compared, Cyn d 4 shares 53% sequence identity with Phl p 4 (Fig. 4*b*). In particular, residues 320–334 in the loop between the αG helix and the β14 strand show very high sequence variability. Interestingly, Phl p 4 contains an extra aspartate residue within this loop and hence possesses a unique region with three consecutive negatively charged residues (Glu323, Asp324 and Asp325; Fig. 4*a*). Cyn d 4 also contains some unique sites, for example Asp50/Lys51/Glu98/Lys99 and Lys144/Lys230/Lys284/Asp285/Glu311.

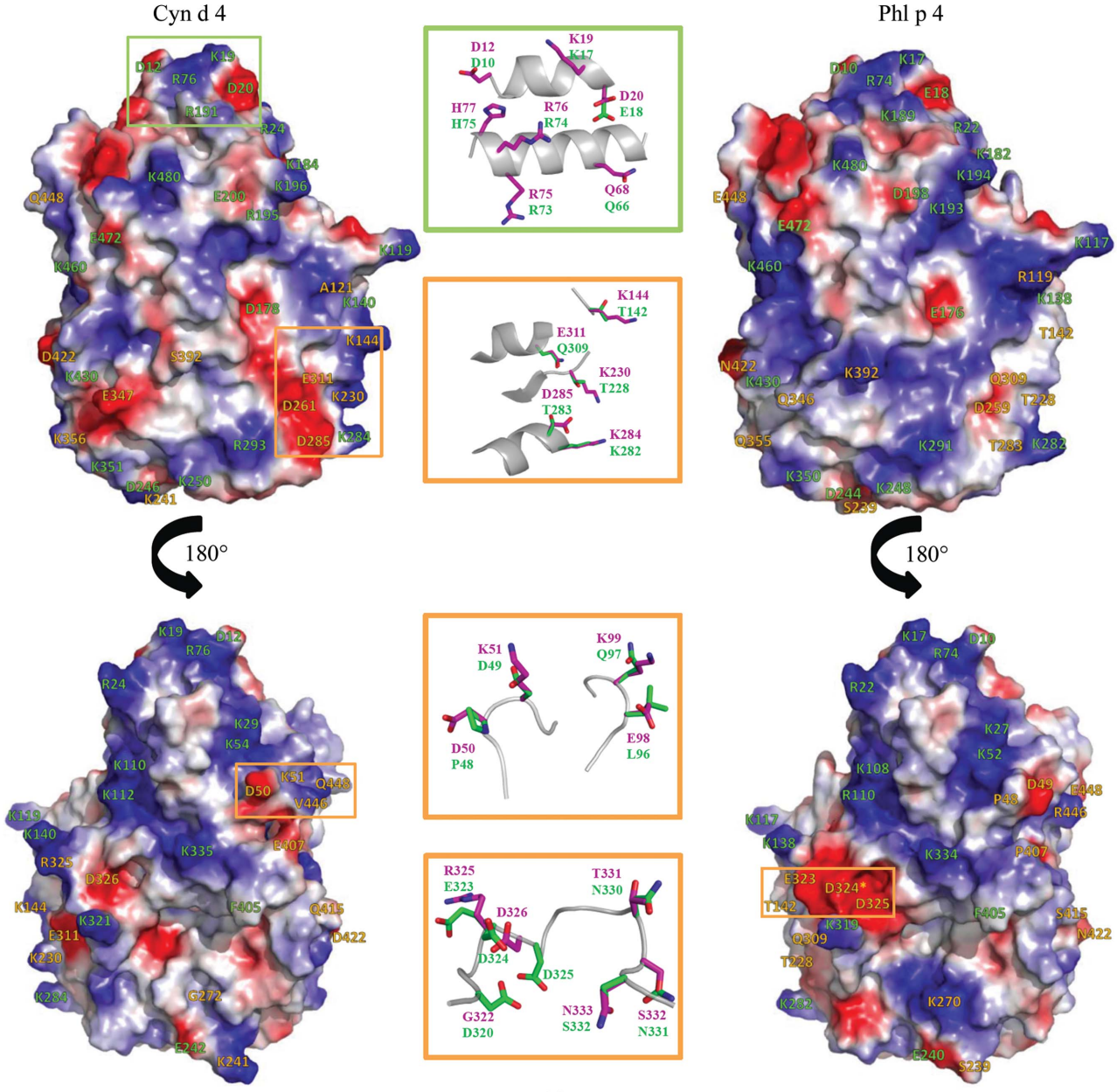

Figure 4
(*a*) Surface comparison of Cyn d 4 and Phl p 4. The surface is coloured according to electrostatic potential from $-60k_BT$ (red) to $60k_BT$ (blue). An interesting feature is the clustering of surface-exposed conserved *versus* variant residues, which are labelled in green and orange, respectively. The higher magnification insets illustrate one common and three variant patches in Cyn d 4 in magenta, while those in Phl p 4 are shown in green. (*b*) Sequence alignment of Cyn d 4 and Phl p 4. The solvent-accessible area for each residue was calculated using *PROCHECK* (Laskowski *et al.*, 1993) and coloured according to the exposure degree from burial (blue) to exposure (white). Amino acids that have a surface accessibility of >50% are highlighted; identical residues are shaded in green, while differences are shaded in orange.

4. Discussion

4.1. The function of Cyn d 4

Structural characterization can provide knowledge of allergen function and this may contribute to understanding the mechanism of allergenicity as, for example, in studies of the proteolytic activities of many mite allergens (Thomas *et al.*, 2010). Recently, Dharajiya and coworkers found that pollen NAD(P)H oxidases induce allergic inflammation through the generation of reactive oxygen species (Dharajiya *et al.*, 2008). All bicovalent flavoproteins identified to date are oxidases. Many other proteins also possess the conserved flavinylated histidine and cysteine residues as well as a proton-shuttle tyrosine residue and these may therefore also be regarded as bicovalent flavoproteins (Fig. 3*b*). To date, the identified substrates of these flavoenzymes are either sugars or species-specific secondary metabolites. In addition to *At*GOOX, *Cc*HEOX and *Fg*ChitO, the identified sugar oxidases include a tobacco nectar protein with glucose oxidase activity (nectarin V; NINEC5; Carter & Thornburg, 2005) and a sunflower defence protein with carbohydrate oxidase activity (*Ha*CHOX; Custers *et al.*, 2004). These sugar oxidases seem to share some conserved residues that are involved in sugar binding (Fig. 3*b*). In contrast, the other identified bicovalent flavoproteins are involved in the biosynthesis of species-specific secondary metabolites that are of potential pharmaceutical interest. For example, in addition to *Ec*BBE, Δ-tetrahydrocannabinolic acid synthase and cannabidiolic acid synthase from *Cannabis sativa* and the tobacco BBE-like proteins are involved in the biosynthesis of plant bioactive alkaloids (Taura *et al.*, 2007). In addition to *Sg*AknOX, *Sg*GilR and some other *Streptomyces* gene products are involved in the biosynthesis of antibiotics. A comparison with the available bicovalent flavoprotein structures reveals various sizes and shapes of the substrate-binding sites owing to the diverse nature of these substrates. The molecular surfaces of these group 4 structures display a conserved hydrophobic deep cavity, like *Ec*BBE and *Sg*GilR. Therefore, based on the above, we suggest that Cyn d 4 may be an oxidase that is involved in the biosynthesis of a pollen-specific apolar secondary metabolite.

4.2. The various cross-antigenicities of the grass pollen group 4 allergens

Grass pollen group 4 allergens are the only allergenic members of the plant BBE-like family to be identified to date. The large number of positively charged surface patches on the group 4 allergens is a unique feature that differs from the situation in other innocuous BBE-like members. For example, while BBE shares high structural homology with Cyn d 4 at 42% sequence identity, they share only 25% sequence identity when the surface-exposed residues are considered. This will lead to distinct surface properties, particularly in terms of electrostatic potential distribution. In contrast, in BBE acid residues (12 Asp and 21 Glu) prevail over basic residues (14 Lys and four Arg), which will lead to the protein surface being markedly negative in this area.

IgE-binding sites on allergen molecules are often classified as either sequential (linear) or conformational epitopes. 38 decapeptides covering the solvent-exposed sequences did not show any strong IgE reactivity, which suggests that the primary protein epitopes of Cyn d 4 are conformational in nature. Two peptides corresponding to residues 291–300 and 301–310 did display weak IgE reactivity. This may imply that residues 291–310 are potential IgE-recognition regions on Cyn d 4. A specific helical or strand conformation in this segment as well as glycosylation at Asn300 may enhance the IgE activity.

Although the grass pollen group 4 allergens display high sequence similarity, structural analysis may reveal fine differences that account for their potential cross-reactive and species-specific conformational epitopes. Together with the model structures of the Pooideae group 4 allergens, the Cyn d 4 crystal structure provides a structural basis for the various cross-allergenicities of these closely related allergens. An analysis of the three-dimensional topography of the grass pollen group 4 allergen structures demonstrated that there is ample space for IgE binding. They also share similar highly hydrophilic positively charged surfaces. This is consistent with a statistical survey in which IgE-binding sites were found to be considerably more hydrophilic and to have prominent electrostatic potential features, particularly lysine residues, compared with other protein–protein interfaces (Oezguen *et al.*, 2008). Interestingly, Phl p 4 possesses a unique loop that contains three consecutive negatively charged residues (Glu323-Asp324-Asp325), which is a likely location of Phl p 4-specific epitopes (Fig. 4*a*). Even in Hor v 4, Sec c 4 and Tri a 4, which share 90–93% sequence identity, there are still fine structural differences at the molecular surface. Thus, in these proteins surface conservation rather than primary-sequence homology may reflect the degree of cross-antigenicity.

Knowledge of the IgE cross-reactivity patterns towards various grass pollens is crucial for improving the diagnosis and treatment of allergies. Hopefully, our Cyn d 4 crystal structure will enhance the understanding of the molecular determinants of grass pollen group 4 allergenicity.

This study was supported by National Science Council (NSC98-2311-B010-006-MY3) and a grant from the Ministry of Education, Aim for the Top University Plan. Portions of this research were carried out at the National Synchrotron Radiation Research Center, a national user facility supported by the National Science Council of Taiwan, Republic of China. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Core Facility Program for Biotechnology.

References

Alexeev, I., Sultana, A., Mäntsälä, P., Niemi, J. & Schneider, G. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 6170–6175.
Andersson, K. & Lidholm, J. (2003). *Int. Arch. Allergy Immunol.* **130**, 87–107.
Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). *Bioinformatics*, **22**, 195–201.

Baer, H. *et al.* (1986). *J. Allergy Clin. Immunol.* **78**, 624–631.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Carter, C. J. & Thornburg, R. W. (2005). *Plant Physiol.* **134**, 460–469.
Custers, J. H., Harrison, S. J., Sela-Buurlage, M. B., van Deventer, E., Lageweg, W., Howe, P. W., van der Meijs, P. J., Ponstein, A. S., Simons, B. H., Melchers, L. S. & Stuiver, M. H. (2004). *Plant J.* **39**, 147–160.
DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.
Dharajiya, N., Boldogh, I., Cardenas, V. & Sur, S. (2008). *Curr. Opin. Allergy Clin. Immunol.* **8**, 57–62.
Ekramoddoullah, A. K., Kisil, F. T. & Sehon, A. H. (1983). *Mol. Immunol.* **20**, 465–473.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Fischer, S., Grote, M., Fahlbusch, B., Müller, W. D., Kraft, D. & Valenta, R. (1996). *J. Allergy Clin. Immunol.* **98**, 189–198.
Fraczkiewicz, R. & Braun, W. (1998). *J. Comput. Chem*, **19**, 319–333.
Gavrović-Jankulović, M., Cirković, T., Bukilica, M., Fahlbusch, B., Petrović, S. & Jankov, R. M. (2000). *J. Investig. Allergol. Clin. Immunol.* **10**, 361–367.
Haavik, S., Paulsen, B. S. & Wold, J. K. (1985). *Int. Arch. Allergy Immunol.* **78**, 260–268.
Heuts, D. P., Winter, R. T., Damsma, G. E., Janssen, D. B. & Fraaije, M. W. (2008). *Biochem. J.* **413**, 175–183.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids. Res.* **38**, W545–W549.
Huang, C.-F., Wu, T.-C., Chu, Y.-H., Hwang, K.-S., Wang, C.-C. & Peng, H.-J. (2008). *Scand. J. Immunol.* **68**, 502–510.
Huang, C.-H., Lai, W.-L., Lee, M.-H., Chen, C.-J., Vasella, A., Tsai, Y.-C. & Liaw, S.-H. (2005). *J. Biol. Chem.* **280**, 38831–38838.
Johansen, N., Weber, R. W., Ipsen, H., Barber, D., Broge, L. & Hejl, C. (2009). *Int. Arch. Allergy Immunol.* **150**, 325–334.
Kelley, L. A. & Sternberg, M. J. (2009). *Nature Protoc.* **4**, 363–371.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leduc-Brodard, V., Inacio, F., Jaquinod, M., Forest, E., David, B. & Peltre, G. (1996). *J. Allergy Clin. Immunol.* **98**, 1065–1072.
Leferink, N. G., Heuts, D. P., Fraaije, M. W. & van Berkel, W. J. (2008). *Arch. Biochem. Biophys.* **474**, 292–301.
Liaw, S.-H., Lee, D. Y., Chow, L.-P., Lau, G.-X. & Su, S.-N. (2001). *Biochem. Biophys. Res. Commun.* **280**, 738–743.
Liaw, S.-H., Lee, D. Y., Yang, S.-Y. & Su, S.-N. (1999). *J. Struct. Biol.* **127**, 83–87.
Liu, Y.-C., Li, Y.-S., Lyu, S.-Y., Hsu, L.-J., Chen, Y.-H., Huang, Y.-T., Chan, H.-C., Huang, C.-J., Chen, G.-H., Chou, C.-C., Tsai, M.-D. & Li, T.-L. (2011). *Nature Chem. Biol.* **7**, 304–309.
Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
Martin, B. G., Mansfield, L. E. & Nelson, H. S. (1985). *Ann. Allergy*, **54**, 99–104.
Nandy, A., Petersen, A., Wald, M., Suck, R., Kahlert, H., Weber, B., Becker, W. M., Cromwell, O. & Fiebig, H. (2005). *Biochem. Biophys. Res. Commun.* **337**, 563–570.
Noinaj, N., Bosserman, M. A., Schickli, M. A., Piszczek, G., Kharel, M. K., Pahari, P., Buchanan, S. K. & Rohr, J. (2011). *J. Biol. Chem.* **286**, 23533–23543.
Oezguen, N., Zhou, B., Negi, S. S., Ivanciuc, O., Schein, C. H., Labesse, G. & Braun, W. (2008). *Mol. Immunol.* **45**, 3740–3747.
Ohsuga, H., Su, S.-N., Takahashi, N., Yang, S.-Y., Nakagawa, H., Shimada, I., Arata, Y. & Lee, Y. C. (1996). *J. Biol. Chem.* **271**, 26653–26658.
Rand, T., Qvist, K. B., Walter, C. P. & Poulsen, C. H. (2006). *FEBS J.* **273**, 2693–2703.
Rossi, R. E., Monasterolo, G. & Monasterolo, S. (2001). *Allergy*, **56**, 1180–1185.
Rossi, R. E., Monasterolo, G., Prina, P., Coco, G., Operti, D. & Rossi, L. (2008). *Allergol. Int.* **57**, 157–164.
Schumacher, M. J., Grabowski, J. & Wagner, C. M. (1985). *Ann. Allergy*, **55**, 584–587.
Stumvoll, S., Lidholm, J., Thunberg, R., DeWitt, A. M., Eibensteiner, P., Swoboda, I., Bugajska-Schretter, A., Spitzauer, S., Vangelista, L., Kazemi-Shirazi, L., Sperr, W. R., Valent, P., Kraft, D. & Valenta, R. (2002). *Biol. Chem.* **383**, 1383–1396.
Su, S.-N., Lau, G.-X., Tsai, J.-J., Yang, S.-Y., Shen, H.-D. & Han, S.-H. (1991). *Clin. Exp. Allergy*, **21**, 449–455.
Su, S.-N., Shu, P., Lau, G.-X., Yang, S.-Y., Huang, S.-W. & Lee, Y.-C. (1996). *J. Allergy Clin. Immunol.* **98**, 486–494.
Taura, F., Sirikantaramas, S., Shoyama, Y., Shoyama, Y. & Morimoto, S. (2007). *Chem. Biodivers.* **4**, 1649–1663.
Thomas, W. R., Hales, B. J. & Smith, W. A. (2010). *Trends Mol. Med.* **16**, 321–328.
Winkler, A., Lyskowski, A., Riedl, S., Puhl, M., Kutchan, T. M., Macheroux, P. & Gruber, K. (2008). *Nature Chem. Biol.* **4**, 739–741.