



Homology modeling of the cellulose-binding domain of a pollen allergen from rye grass: structural basis for the cellulose recognition and associated allergenic properties[☆]

Annick Barre and Pierre Rougé*

Institut de Pharmacologie et Biologie Structurale, UMR-CNRS 5089, 205 Route de Narbonne, 31077 Toulouse Cedex 4, France

Received 7 August 2002

Abstract

A three-dimensional model of the cellulose-binding domain of the rye-grass pollen allergen Lol pI built by homology modeling is proposed as a structural scaffold for expansins and other expansin-related proteins. A groove and an extended strip of aromatic and polar residues presumably account for the cellulose-binding properties of the protein domain. Two of the four predicted T-cell epitopes readily exposed on the surface of the cellulose-binding domain match with previously reported IgE-binding regions. A close structural relationship occurs between the cellulose-binding and allergenic properties. © 2002 Elsevier Science (USA). All rights reserved.

Keywords: Grass pollen allergen Lol pI; Expansins; Cellulose-binding domain; T-cell epitopes

Grass pollen represents a major cause of immediate hypersensitivity or type I allergy responses in susceptible individuals such as allergic rhinitis, conjunctivitis, hives or seasonal asthma. The impact of type I allergy in industrialized countries has increased tremendously, especially in children, and is now becoming a major health problem. Inhaled pollen, and especially grass pollen, contribute to allergic disorders in up to 25% of adults [1]. Various (glyco)proteins have been identified as major grass pollen allergens, e.g., allergens of group I [2–4]. In this respect, Lol pI has been characterized as the major pollen allergen of group I from the rye-grass (*Lolium perenne*) [3]. However, allergens belonging to different groups share a high degree of amino acid sequence similarity, e.g., Lol pI from group I and Lol pII or Lol pIII from group I/II [5], regardless of their origin, e.g., Lol pI from rye-grass (*L. perenne*) and Phl p2 from

timothy grass (*Phleum pratense*) [5]. Consequently, some cross-reactivity was shown to occur at both the T-cell [6] and IgE-binding [7] levels, even though rather conflicting results were obtained when comparing natural to expressed recombinant allergens.

Like other grass allergens, Lol pI belongs to the β -expansin group [8]. Expansins, which comprises the two closely related families of α - (unglycosylated) and β -expansins (glycosylated) [9], consist of medium-size proteins (~26 kDa), supposed to act as cell wall-loosening proteins involved in the cell enlargement occurring during the acid growth process in plants [10]. Both types of expansins are built of a C-terminal cellulose-binding domain (CBD) associated to an N-terminal domain which exhibits some sequence similarity with the family-45 endoglucanases [11]. The CBD would allow expansins to interact with the cellulose microfibrils of the plant cell wall whereas the endoglucanase domain would be responsible for the catalytic cleavage of the H-bonds linking the other cell wall polysaccharides (e.g., hemicelluloses) to cellulose [12]. To date, no information is available on the three-dimensional structure of Lol pI to account for its unique ambivalent property of being a potent allergen and a cell wall-loosening agent.

[☆] Abbreviations: At-EXP2, α -expansin 2 from *Arabidopsis thaliana*; CBD, cellulose-binding domain; CD, catalytic domain; Cs-EXP1, α -expansin 1 from *Cucumis sativus*; HCA, hydrophobic cluster analysis; Lol pI, pollen allergen from *Lolium perenne*; Os-EXPB, β -expansin from *Oryza sativa*; Phl p2, pollen allergen from *Phleum pratense*.

* Corresponding author. Fax: +33-5-61-17-59-94.

E-mail address: Pierre.Rouge@ipbs.fr (P. Rougé).

Fig. 1. Amino acid sequence alignment of α -expansins (At-EXP2 from *Arabidopsis thaliana*, Cs-EXP1 from *Cucumis sativus*), β -expansins (Os-EXPB from *Oryza sativa*, Lol p1), and Phl p2 (allergen from *Phleum pratense*) performed with ESPrnt [1]. Identical or structurally homologous (according to the Risler's matrix [42]) residues occurring in the cellulose-binding domain are in white letters in black boxes and in letters open boxed, respectively. Identical residues of the catalytic domains are shaded gray and conserved cysteins are indicated by asterisks *. The two basic RR signals occurring between the CBD and the putative endoglucanase domain are black boxed. Secondary structure elements of Phl p2 and of the cellulose-binding domain of Lol p1 are indicated by arrows above and under the sequences, respectively. Strands of β -sheet forming the S1-sheet (β 1, β 2, β 4, β 5, and β 8) and the S2-sheet (β 3, β 4, β 6, and β 7) of both proteins are numbered.

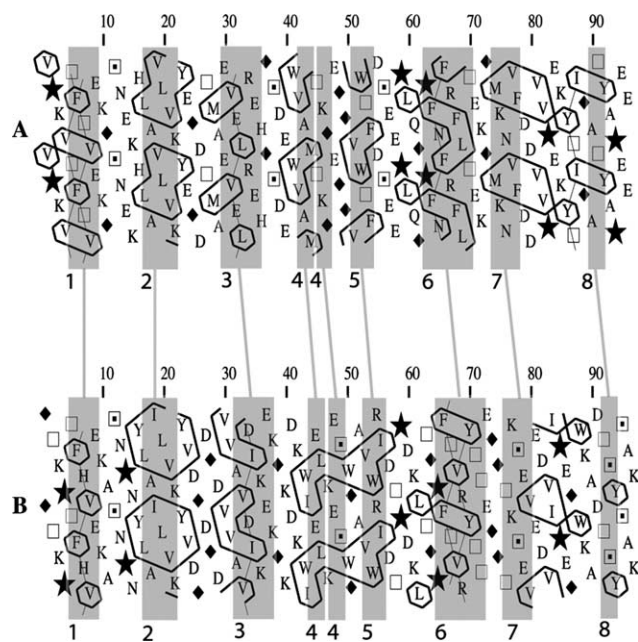


Fig. 2. Comparison of the HCA plots of Phl p2 (A) and the CBD of Lol pI (B). The eight strands of β -sheet (numbered 1–8) occurring along the polypeptide chain of Phl p2 are shaded gray and their structural counterparts are delineated along the polypeptide chain of Lol pI. Glycine (◆), proline (★), threonine (□), and serine (□) are represented by special symbols.

CBD could be built by homology modeling from the X-ray coordinates of Phl p2.

The modeled CBD of Lol pI consists of two β -sheets S1 and S2 arranged into a β -sandwich structure (Fig. 3). β -sheets S1 and S2 comprise 5 (β 1:5–9; β 2:17–22; β 4:47–49; β 5:53–56; and β 8:92–94) and four strands (β 3:32–38;

β 4:44–46; β 6:65–72; and β 7:76–80), respectively. In fact, the fourth strand of β -sheet is kinked and its N-terminal part (residues 44–46) belongs to β -sheet S2 whereas its C-terminal part (residues 47–49) constitutes the β -sheet S1. A similar fold occurs in other proteins belonging to the immune system and possessing an Ig-like domain [26].

Although no direct evidence is yet available on the precise location of the cellulose-binding domain of Lol pI, Lol pI shares in this respect some striking structural similarities with other cellulases, characterized from bacteria [27–32], actinomycetes [33] or fungi [34]. All these proteins possess a typical jelly-roll β -sandwich structure and exhibit an exposed planar surface, presumably responsible for the binding of crystalline cellulose. As proposed for the scaffoldin subunit Cip of the cellulosome of *Clostridium thermocellum* [29], the CBD corresponds to a linear strip of aromatic and polar residues, which extends on a few β -strands forming the flattened face of the β -sandwich. At the opposite side, the β -strands forming the incurved face of the β -sandwich delineate a groove, which mainly consists of aromatic residues, which could interact with soluble cellulose. Lol pI similarly exhibits a linear strip of both polar (Asp25, Lys38, Thr67, Arg69, Glu79, and Glu81) and aromatic residues (Phe66 and Tyr70) on its planar surface and a groove rich in aromatic residues (Phe6, Tyr16, Tyr23, and Trp50) on its opposed face (Fig. 4). Most of the aromatic and hydrophilic residues located in the strip and the groove are well conserved in other expansins and especially in β -expansins. Both regions could be involved in the binding of Lol pI and other expansins to cellulose.

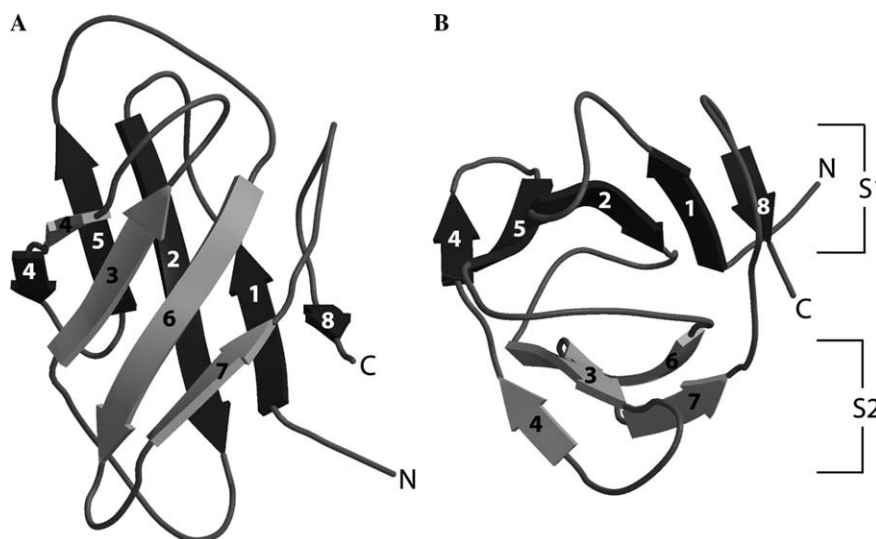


Fig. 3. (A) Ribbon diagram of the modeled CBD of Lol pI showing the two β -sheets S1 (strands 1, 2, 4 [corresponding to β 5 in Fig. 1], 5 and 8, colored black) and S2 (strands 3, 4 [corresponding to β 4 in Fig. 1], 6 and 7, colored light gray) arranged in a β -sandwich structure. Loops, turns, and coil structures are colored gray. (B) Upper view showing the β -sandwich formed by the two β -sheets S1 (black) and S2 (light gray). The N- and C-terminal ends of the polypeptide chain are indicated.

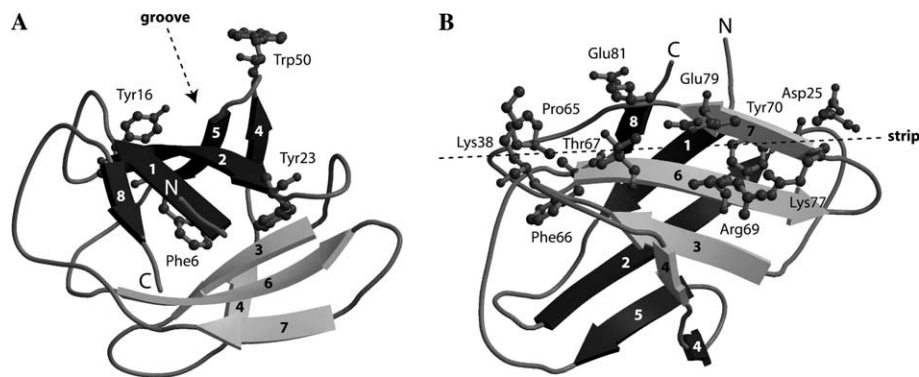


Fig. 4. Ribbon diagrams of the modeled CBD of Lol pI showing the location of the groove (A) and the strip of polar and aromatic residues (B) possibly involved in the binding of cellulose. Strands of β -sheet forming the S1 (colored black) and the S2 (colored light gray) sheets are numbered, and polar and aromatic residues distributed into the groove and along the strip are labeled and represented in gray ball-and-sticks.

Interestingly, an exposed two-basic residue RR cleavage signal almost constantly occurs between the so-called catalytic (CD) and the cellulose-binding domain (CBD) of Lol pI and other α - and β -expansins (Fig. 1) that could allow the CBD to become proteolytically separated from the catalytic domain, once this latter domain has been correctly positioned to attack the cellulose-hemicellulose junction and the separation of both polymers is completed. If occurring *in planta*, such a proteolytic step could induce some modulation in the cell wall-loosening activity of expansins during the plant development.

Four distinct epitopic regions are predicted to occur on the surface of the CBD of Lol pI. They correspond to the amino acid stretches 9–24 (β 2 and loop connecting β 1– β 2), 28–38 (β 3 and parts of loops interconnecting β 2– β 3 and β 3– β 4), 50–63 (β 5 and parts of loops inter-

connecting β 4– β 5 and β 5– β 6), and 66–80 (β 6, β 7 and their connecting loop). They mainly consist of exposed and rather confluent loops and strands of β -sheet (Fig. 5). Interestingly, some of these predicted T-cell epitopes (e.g., β 6 and β 7) are localized in the flattened regions of the β -sandwich, supposed to be involved in the recognition of cellulose by the CBD of Lol pI. In this respect, the allergenic properties would be closely structurally associated to the cellulose-binding properties of Lol pI and expansin-related proteins. The third stretch 50–63 (WGAVWRIDTPDKLT) predicted as a T-cell epitope, coincides with the determinant (ESWGAVWRIDTPDKLTGPFT) of Lol pI recognized by human leukocyte antigen class II-restricted T-cells of individuals allergic to rye-grass pollen [3]. The fourth stretch 66–80 (FTVRYTTEGGTKSEV) predicted as another T-cell

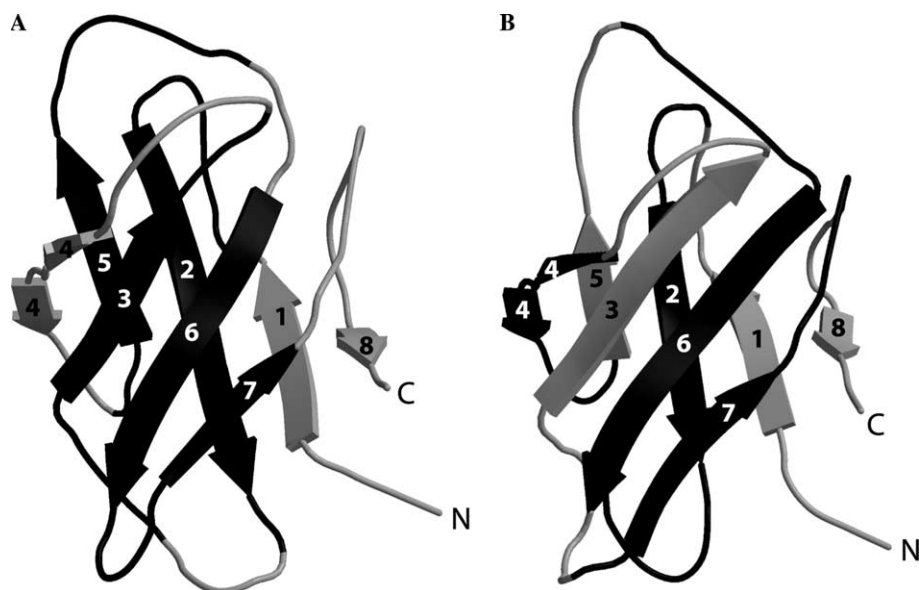


Fig. 5. Location of the predicted T-cell epitope regions (black) on the ribbon diagram of the CBD of Lol pI (A) and Phl p2 (B). Strands of β -sheet are numbered and both proteins are oriented as in Fig. 3. The N- and C-terminal ends of the polypeptide chain are indicated.

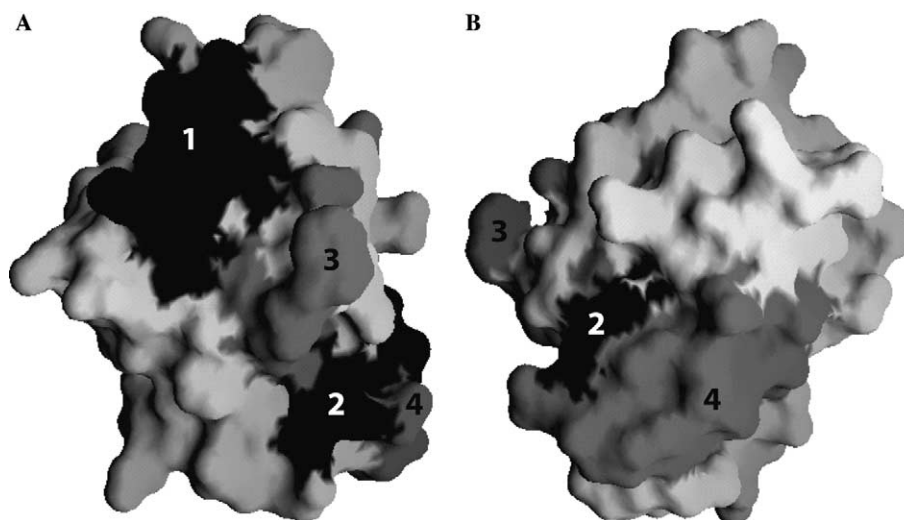


Fig. 6. Mapping of the possible IgE-binding epitopes 1 (black), 2 (black), 3 (gray), and 4 (gray) on the S1 (A) and S2 (B) faces of the molecular surface of Lol pI.

epitope partly coincides with the 29 C-terminal residues of Lol pI (**VRYTTEGGTKSEVEDVIPEGWKADTSYSAK**) recognized as an IgE-binding region [2,35]. Epitope mapping of IgE on recombinant Lol pI has shown that the major IgE-reactive epitope mapped to the C-terminus [36] that roughly corresponds to the CBD of the allergen. The C-terminal region 216–240 (**EGGTKSEVEDVIPEGWKADTSYSAK**) of Lol pI has also been recognized as an IgE- and IgG-binding epitope [37] and binding of rabbit IgG directed against this determinant was shown to reduce ~50% the Lol pI-mediated histamine release from human basophils [38]. In this respect, the predicted epitopes susceptible to interact with IgE antibodies correspond to the less restricted stretches 9–16 (**EKASNPNY**), 28–32 (**GDVVA**), 50–55 (**WGAVWR**), and 69–80 (**RYTTEGGTKSEV**), which are sufficiently exposed on the molecular surface of Lol pI. The first three stretches mainly cluster on strands β 1 and β 2 forming the S1 sheet of Lol pI whereas the last extended stretch covers strands β 6 and β 7 of the S2 sheet of Lol pI (Fig. 6). Among these putative IgE-binding epitopes, stretch 2 displays a more pronounced electronegative character (result not shown).

Rather similarly located T-cell epitopes can be predicted on the surface of Phl p2 (Fig. 4) and other modeled CBD of expansins (results not shown). In this respect, the predicted stretch 41–52 (**WVAMTKGEGGVW**), corresponding to a loop formed by the two short β 4 strands (see Fig. 1) which contains two conserved tryptophan residues (Trp41 and Trp52), has been proposed as a complementary surface for IgE recognition on the basis of both sequence conservation and surface exposition [26]. The Ig-like three-dimensional fold of the cellulose-binding domain of the Lol pI allergen strikingly resembles those of and Der f2 [39] and Der p2

[40], the main group 2 mite allergens. Two over three antibody epitopes (stretches 94–101 and 111–116) have been recently identified on the C-terminal region of Der p2 [41], which is somewhat reminiscent of the location of T-cell epitopes on Lol pI.

References

- [1] B. Knox, C. Suphioglu, Environmental and molecular biology of pollen allergens, *Trends Plant Sci.* 1 (1996) 156–164.
- [2] R.E. Esch, D.G. Klapper, Isolation and characterization of a major cross-reactive grass group I allergenic determinant, *Mol. Immunol.* 26 (1989) 557–561.
- [3] M. Perez, G.Y. Ishioka, L.E. Walker, R.W. Chesnut, cDNA cloning and immunological characterization of the rye grass allergen Lol pI, *J. Biol. Chem.* 265 (1990) 16210–16215.
- [4] I.J. Griffith, P.M. Smith, J. Pollock, P. Theerakulpisut, A. Avjioglu, S. Davies, T. Hough, M.B. Singh, R.J. Simpson, L.D. Ward, R.B. Knox, Cloning and sequencing of Lol pI, the major allergenic protein of rye-grass pollen, *FEBS Lett.* 279 (1991) 210–215.
- [5] C. Dolecek, S. Vrtala, S. Laffer, P. Steinberger, D. Kraft, O. Scheiner, R. Valenta, Molecular characterization of Phl pII, a major timothy grass (*Phleum pratense*) pollen allergen, *FEBS Lett.* 335 (1993) 299–304.
- [6] S. Schenk, H. Breiteneder, M. Susani, N. Najafian, S. Laffer, M. Duchene, R. Valenta, G. Fischer, O. Scheiner, D. Kraft, C. Ebner, T cell epitopes of Phl pI major pollen allergen of timothy grass (*Phleum pratense*). Crossreactivity with group I allergens of different grasses, *Adv. Exp. Med. Biol.* 409 (1996) 141–146.
- [7] S. Laffer, R. Valenta, S. Vrtala, M. Susani, R. van Ree, D. Kraft, O. Scheiner, M. Duchene, Complementary DNA cloning of the major allergen Phl pI from timothy grass (*Phleum pratense*); recombinant Phl pI inhibits IgE binding to group I allergens from eight different grass species, *Allergy Clin. Immunol.* 94 (1994) 689–698.
- [8] D.J. Cosgrove, P. Bedinger, D.M. Durachko, Group I allergens of grass pollen as a cell wall-loosening agents, *Proc. Natl. Acad. Sci. USA* 94 (1997) 6559–6564.

- [9] Y. Lee, D. Choi, H. Kende, Expansins: ever-expanding numbers and functions, *Curr. Opin. Plant Biol.* 4 (2001) 527–532.
- [10] D.J. Cosgrove, Loosening of plant cell walls by expansins, *Nature* 407 (2000) 321–326.
- [11] B. Henrissat, T.T. Teeri, R.A.J. Warren, A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants, *FEBS Lett.* 425 (1998) 352–354.
- [12] S. McQueen-Mason, D.J. Cosgrove, Disruption of hydrogen bonding between wall polymers by proteins that induce plant wall extension, *Proc. Natl. Acad. Sci. USA* 91 (1994) 6574–6578.
- [13] P. Gouet, E. Courcelle, D. Stuart, F. Metoz, ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics* 15 (1999) 305–308.
- [14] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool, *Nucleic Acids Res.* 15 (1997) 4876–4882.
- [15] C. Gaboriaud, V. Bissery, T. Benchetrit, J.P. Mornon, Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences, *FEBS Lett.* 224 (1987) 149–155.
- [16] L. Lemesle-Varloot, B. Henrissat, C. Gaboriaud, V. Bissery, A. Morgat, J.P. Mornon, Hydrophobic cluster analysis: procedure to derive structural and functional information from 2-D-representation of protein sequences, *Biochimie* 72 (1990) 555–574.
- [17] A.A. Fedorov, T. Ball, R. Valenta, S.C. Almo, X-ray crystal structures of birch pollen profilin and Phl p2, *Int. Arch. Allergy Immunol.* 113 (1997) 109–113.
- [18] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic Acids Res.* 28 (2000) 235–242.
- [19] J.W. Ponder, F.M. Richards, Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes, *J. Mol. Biol.* 193 (1987) 775–791.
- [20] M.T. Mas, K.C. Smith, D.L. Yarmush, K. Aisaka, R.M. Fine, Modeling the anti-CEA antibody combining site by homology and conformational search, *Proteins Struct. Func. Genet.* 14 (1992) 483–498.
- [21] P.J. Kraulis, Molscript: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Cryst.* 24 (1991) 946–950.
- [22] R.M. Esnouf, An extensively modified version of Molscript that includes greatly enhanced coloring capabilities, *J. Mol. Graphics* 15 (1997) 132–134.
- [23] E.A. Merritt, D.J. Bacon, Raster3D photorealistic molecular graphics, *Methods Enzymol.* 277 (1997) 505–524.
- [24] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins Struct. Func. Genet.* 11 (1991) 281–296.
- [25] M.K. Gilson, B.H. Honig, Calculation of electrostatic potential in an enzyme active site, *Nature* 330 (1987) 84–86.
- [26] S. De Marino, M.A. Castiglione Morelli, F. Fraternali, E. Tamborini, G. Musco, S. Vrtala, C. Dolecek, P. Arosio, R. Valenta, A. Pastore, An immunoglobulin-like fold in a major plant allergen: the solution structure of Phl p2 from timothy grass pollen, *Structure* 7 (1999) 943–952.
- [27] G.Y. Xun, E. Ong, N.R. Gilkes, D.G. Kilburn, D.R. Muhandiram, M. Harris-Brandts, J.P. Carver, L.E. Kay, T.S. Harvey, Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy, *Biochemistry* 34 (1995) 6993–7009.
- [28] P.E. Johnson, M.D. Joshi, P. Tomme, D.G. Kilburn, L.P. McIntosh, Structure of the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy, *Biochemistry* 35 (1996) 14381–14394.
- [29] J. Tormo, R. Lamed, A.J. Chirino, E. Morag, E.A. Bayer, Y. Shoham, T.A. Steitz, Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose, *EMBO J.* 15 (1996) 5739–5751.
- [30] E. Brun, F. Moriaud, P. Gans, M.J. Blackledge, F. Barras, D. Marion, Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by *Erwinia chrysanthemi*, *Biochemistry* 36 (1997) 16074–16086.
- [31] P.E. Johnson, E. Brun, L.F. MacKenzie, S.G. Withers, L.P. McIntosh, The cellulose-binding domains from *Cellulomonas fimi* β -1,4-glucanase CenC bind nitroxide spin-labeled cellobiosaccharides in multiple orientations, *J. Mol. Biol.* 287 (1999) 609–625.
- [32] E. Brun, P.E. Johnson, A.L. Creagh, P. Tomme, P. Webster, C.A. Haynes, L.P. McIntosh, Structure and binding specificity of the second N-terminal cellulose-binding domain from *Cellulomonas fimi* endoglucanase C, *Biochemistry* 39 (2000) 2445–2458.
- [33] J. Sakon, D. Irwin, D.B. Wilson, P.A. Karplus, Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*, *Nat. Struct. Biol.* 4 (1997) 810–818.
- [34] J. Kraulis, G.M. Clore, M. Nilges, T.A. Jones, G. Pettersson, J. Knowles, A.M. Gronenborn, Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing, *Biochemistry* 28 (1989) 7241–7257.
- [35] R.E. Esch, D.G. Klapper, Identification and localization of allergenic determinants on grass group-1 antigens using monoclonal antibodies, *J. Immunol.* 142 (1989) 179–184.
- [36] C. Sophioglou, P.M. Smith, E.K. Ing, R.B. Knox, M.B. Singh, Recombinant and epitope mapping of grass pollen allergens, *Adv. Exp. Med. Biol.* 409 (1996) 147–155.
- [37] E. Tamborini, S. Faccini, J. Lidholm, M. Svensson, A. Brandazza, R. Longhi, H. Groenlund, A. Sidoli, P. Arosio, Biochemical and immunological characterization of recombinant allergen Lol pI, *Eur. J. Biochem.* 249 (1997) 886–894.
- [38] A.A. Ansari, P. Shenbagamurthi, D.G. Marsh, Complete primary structure of a *Lolium perenne* (perennial rye grass) pollen allergen, Lol p2: comparison with known Lol p1 and2 sequences, *Biochemistry* 28 (1989) 8665–8670.
- [39] S. Ichikawa, H. Hatanaka, T. Yuuki, N. Iwamoto, S. Kojima, C. Nishiyama, K. Ogura, Y. Okumura, F. Inagaki, Solution structure of Der f2, the major mite allergen for atopic diseases, *J. Biol. Chem.* 273 (1998) 356–360.
- [40] U. Derewenda, J. Li, Z. Derewenda, Z. Dauter, G.A. Mueller, G.S. Rule, D.C. Benjamin, The crystal structure of a major dust mite allergen Der p2 and its biological implications, *J. Mol. Biol.* 318 (2002) 189–197.
- [41] G.A. Mueller, A.M. Smith, M.D. Chapman, G.S. Rule, D.C. Benjamin, Hydrogen exchange nuclear magnetic resonance spectroscopy mapping of antibody epitopes on the house dust mite allergen Der p2, *J. Biol. Chem.* 276 (2001) 9359–9365.
- [42] J.L. Risler, M.O. Delorme, H. Delacroix, A. Henaut, Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix, *J. Mol. Biol.* 204 (1988) 1019–1029.