Supplemental data: Materials and methods

Molecular modeling of ABA orthologs from *Marchantia* and *Tortula*

Hydrophobic cluster analysis, molecular modeling and docking experiments were carried out using standard techniques. Three-dimensional models of the proteins were built using the atomic coordinates of ABA (Carrizo et al., 2005). Multiple amino acid sequence alignments were carried out with CLUSTAL-X (Thompson et al., 1997) and displayed with ESPript (Gouet et al., 1999).

The hydrophobic cluster analysis (Gaboriaud et al., 1987) was performed to delineate the conserved secondary structural features (strands of β-sheet and stretches of α-helix) along the amino acid sequence of MarpoABA2a and TorruABA by comparison with the *Agaricus bisporus* lectin ABA (Carrizo et al., 2005) used as a model. Hydrophobic cluster analysis plots were generated using the program drawhca of L. Canard (http://www.lmcp.jussieu.fr/~soyer/www-hca/hca-form.html).

Molecular modeling of MarpoABA2a and TorruABA was carried out on a Silicon Graphics O2 R10000 workstation, using the programs InsightII, Homology and Discover3 (Accelrys, San Diego CA, USA). The atomic coordinates of ABA (Carrizo et al., 2005) were used to build the three-dimensional model of the proteins. The percentages of both identity (43%/40%) and similarity (76%/74%) of MarpoABA2a and TorruABA, respectively, with ABA allowed building rather accurate three-dimensional models of both proteins using the X-ray coordinates of ABA as a template. Steric conflicts were corrected during the model building procedure using the rotamer library (Ponder and Richards, 1987) and the search algorithm implemented in the Homology program (Mas et al., 1992) to maintain proper side-chain orientation. The geometry of loop regions was corrected using the refine option of TurboFrodo (Roussel and Cambillau, 1989). An energy minimization of the final models was carried out by 100 cycles of steepest descent using Discover3. The program TurboFrodo was run to draw the Ramachandran plot and to perform the superposition of the model with the template protein. PROCHECK (Laskowski et al., 1993) was used to assess the geometric quality of the three-dimensional models. All the residues of both MarpoABA2a and TorruABA were correctly assigned on the best (~ 80%) and more generously allowed regions (~ 20%) of the Ramachandran plot (result not shown). Three-dimensional models very similar to MarpoABA2a were obtained for the other *Marchantia polymorpha* lectin isoforms and, especially, with MarpoABA1a (results not shown).

Docking of the T-antigen disaccharide (Galβ1,3GalNAc) and GlcNAc into the corresponding carbohydrate-binding sites of MarpoABA2a, MarpoABA1a and TorruABA was performed with the program InsightII (Accelrys, San Diego CA, USA). The lowest apparent binding energy (E_{bind} expressed in kcal.mol^{-1}) compatible with the hydrogen bonds (considering Van de Waals interactions and strong [2.5 Å < dist(D-A) < 3.1 Å and 120° < ang(D-H-A)] and weak [2.5 Å < dist(D-A) < 3.5 Å and 105° < ang(D-H-A) < 120°] hydrogen bonds; with D: donor, A: acceptor and H: hydrogen) found in the ABA-T-antigen disaccharide/GlcNAc complex (Carrizo et al., 2005) was calculated with the forcefield of Discover3 and used to anchor the pyranose ring of GalNAc or GlcNAc into the binding sites of the lectins. The positions of the T-antigen disaccharide and GlcNAc observed in the ABA-T-antigen/GlcNAc complex was used as the starting position to anchor the GalNAc residue of the Tn-antigen disaccharide and GlcNAc in the corresponding carbohydrate-binding sites of the modeled lectins.

Cartoons were drawn with PyMOL (W.L. DeLano (http://www.pymol.org).

Literature cited

Carrizo ME, Capaldi S, Perduca M, Irazoqui FJ, Nores GA, Monaco HL (2005) The antineoplastic lectin of the common edible mushroom (*Agaricus bisporus*) has two binding sites, each specific for a different configuration at a single epimeric hydroxyl. J Biol Chem **280**:10614-10623

Gaboriaud C, Bissery V, Benchetrit T, Mornon JP (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett **224**: 149-155

Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics **15**: 305-308

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemistry of protein structures. J Appl Cryst **26**: 283-291

Mas MT, Smith KC, Yarmush DL, Aisaka K, Fine RM (1992) Modeling the anti-CEA antibody combining site by homology and conformational search. Proteins **14**: 483-498

Ponder JW, Richards FM (1987) Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol **193**: 775-791

Roussel A, Cambillau C (1989) TURBO-FRODO. Silicon Graphics Geometry Partners Directory (Committee, S.G., Ed.), Mountain View, California, Silicon Graphics

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. Nucleic Acids Res **25**: 4876-4882