Glycobiology Advance Access published September 10, 2014

The carbohydrate-binding promiscuity of *Euonymus europeaus* lectin is predicted to involve a single binding site

Mark Agostino[2,3,4,5], Tony Velkov[5], Tamir Dingjan[5], Spencer J. Williams[6], Elizabeth Yuriev[1,5] and Paul A. Ramsland[1,2,4,7,8]

[2]School of Biomedical Sciences, CHIRI Biosciences, Curtin University, Perth, WA 6845, Australia

[3]Joint BSC-IRB Research Program in Computational Biology, Life Science Department, Barcelona Supercomputing Centre, Barcelona 08034, Spain

[4]Centre for Biomedical Research, Burnet Institute, Melbourne, VIC 3004, Australia

[5]Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia

[6]School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, VIC 3010, Australia

[7]Department of Surgery Austin Health, University of Melbourne, Heidelberg, VIC 3084, Australia

[8]Department of Immunology, Monash University, Alfred Medical Research and Education Precinct, Melbourne, VIC 3004, Australia

[1]To whom correspondence should be addressed: Tel: +61 3 9282 2178; fax: +61 9282 2100; e-mail: pramsland@burnet.edu.au; elizabeth.yuriev@monash.edu

Abstract

Euonymus europaeus lectin (EEL) is a carbohydrate-binding protein derived from the fruit of the European spindle tree. EEL was first identified for its erythrocyte agglutinating properties and specificity for B and H blood groups. However, a detailed molecular picture of the structural basis of carbohydrate recognition by EEL remains to be developed. In this study, we performed fluorescence titrations of a range of carbohydrates against EEL. Binding of EEL to a wide range of carbohydrates was observed, including a series of blood group-related carbohydrates, mannosides, chitotriose and sialic acid. Affinity was strongest for carbohydrates with H-related structures and the B trisaccharide. A homology model of EEL was produced from templates identified using the HHPred server, which employs hidden Markov models (HMMs) to identify templates. The HMM approach identified that the best templates for EEL were proteins featuring a ricin B-like (R-type) fold. Separate templates were used to model the core and binding site regions of the lectin. Through the use of constrained docking and spatial comparison to a template ligand, binding modes for the carbohydrate ligands were predicted. A relationship between the experimental binding energies and the computed binding energies of the selected docked poses was determined and optimized. Collectively, our results suggest that EEL utilizes a single site for recognition of carbohydrates terminating in a variety of monosaccharides.

Keywords: *Euonymus europaeus* lectin / carbohydrate recognition / molecular docking / blood group carbohydrates / protein-carbohydrate interactions

Introduction

Lectins are valuable tools for studying carbohydrate recognition, and their specificity has been exploited for the identification of cell-surface carbohydrates in both healthy and diseased tissues. Some examples include the use of lectins for blood group typing (Sharon and Lis 2004), for studying glycosylation changes in cancer cells (Kaltner and Gabius 2012), and for targeting glycan epitopes on pathogens such as the human immunodeficiency virus (HIV) (Koharudin and Gronenborn 2014). Lectins are also valuable in a variety of structural and functional studies of important carbohydrate determinants, often as surrogates for anti-carbohydrate antibodies (Grahn et al. 2009; Tempel et al. 2002; Yuriev et al. 2009; Yuriev et al. 2005).

Most lectins with potential biomedical applications are derived from plants, fungi and marine organisms, although there is a range of human lectins that are of interest owing to their prominent roles in immunity and cell biology (Agostino et al. 2011; Crocker et al. 2007; Hardison and Brown 2012; Matsushita 2013). Many plant and fungal lectins have been well characterized and are commercially available for use as reagents. *Euonymus europaeus* lectin (EEL) or agglutinin (EEA) is a plant-derived lectin known to bind to blood group-related carbohydrates (mainly B and H determinants), with a strong preference for carbohydrates with terminal fucose and, to a lesser extent, mannose residues (Fouquaert et al. 2008). Although EEL was initially purified around 40 years ago (Pacak and Kocourek 1975; Petryniak et al. 1981; Petryniak et al. 1977; Petryniak et al. 1980) and is commercially available, it has not been extensively structurally characterized.

As EEL lacks sequence similarity to other structurally characterized proteins, it is difficult to predict if it possesses a similar fold to lectins with known binding preferences for terminal fucose residues (Audette et al. 2000; Mitchell et al. 2002; Wimmerova et al. 2003). EEL is annotated as a ricin B-like lectin in the InterPro database, which analyzes protein

sequences to identify relationships between sequence and function using input from multiple source repositories (Hunter et al. 2012). EEL has been proposed to form part of a novel family of carbohydrate-binding proteins, the *Euonymus* lectins (EUL), to which numerous proteins have been suggested to belong (Al Atalah et al. 2012; Fouquaert et al. 2009; Van Hove et al. 2011).

In this study, we determined binding specificities of EEL for a variety of carbohydrates, and used the data to guide homology modeling to investigate the structural basis of carbohydrate recognition by EEL. We demonstrate that EEL adopts a β-trefoil fold most similar to ricin B-like (R-type) lectins (Cummings and Etzler 2009), in agreement with its annotation in the InterPro database. A combination of hidden Markov model (HMM) and logo analysis of the sequences of R-type lectins most closely matching EEL, enabled the identification of the potential binding sites (α, β and γ) in the EEL sequence. Core-constrained docking was used to determine binding modes of a series of diverse carbohydrate ligands found to bind to EEL with moderate to high affinity. The γ site, which features a solvent-exposed tryptophan residue, was predicted to be the principal carbohydrate-binding site of EEL, capable of interacting with terminal α-fucose or α-galactose, but also with α-mannose, α-sialic acid and *N*-acetylglucosamine residues. The relationship between the experimental and predicted binding energies was used to guide the final selection of binding modes for the ligands.

Results

EEL has broad carbohydrate specificity with preferential binding of H and B saccharides

A range of diverse carbohydrates were evaluated for binding to EEL to investigate its carbohydrate specificity (Table I). The previously reported high affinity for H and B blood groups and α-mannose-terminated glycans (Fouquaert et al. 2008) was replicated in the

affinities derived by fluorescence titration spectroscopy. Affinities for EEL were the highest for H trisaccharide **(12)**, Fucα1-2Gal **(11)**, B trisaccharide **(9)** and Fucα1-OMe **(10)**, with K_d values ranging from 9 to 24 μM. In addition, EEL was shown to bind NeuAcα2-OMe **(15)**, chitotriose **(16)**, αGal-terminated "B-like" saccharides **(5-8)** and the blood group A trisaccharide **(3)**, with moderately high affinities (K_d values of 60 to 75 μM). Low affinity binding of EEL was also observed for Manα1-6Man1-OOctN_3 **(14)**, Manα1-OMe **(13)**, "A-like" αGalNAc-terminated saccharides **(1-2)** and galactose **(4)**, with K_d values ranging from 90 to 140 μM. The lowest affinity carbohydrates tested for binding to EEL were lactose **(17)**, xylose **(18)** and trehalose **(19)**, with K_d values ranging from 280 to 600 μM.

In the blood group A/B/H-related carbohydrate series (Table I), affinity increased primarily with the incorporation of the α-fucose residue. Binding affinities also generally increased with carbohydrate chain lengths, although we only examined carbohydrate chains up to trisaccharide epitopes, which are known to be sufficient to allow discrimination by serology of the A, B and O (H) blood groups (Morgan and Watkins 2000).

Construction of a homology model of EEL

Since EEL shares little sequence similarity with any structurally characterized proteins, the HHPred method (Söding et al. 2005), which utilizes HMMs to identify relationships between protein sequence and structure, was applied to identify likely templates. HHPred indicated that EEL most likely adopts a β-trefoil fold, most similar to that found in ricin B-like (R-type) lectins (Cummings and Etzler 2009) (Table SI).

The R-type fold occurs in proteins from many distantly related species, and there is little sequence conservation among the majority of structurally characterized examples. It features a β-trefoil structure containing three potential carbohydrate-binding sites (termed α,

β and γ), each of which is contained in a single cardioid-shaped loop of approximately 20-30 amino acids in length (Cummings and Etzler 2009). Logo analysis (Figure 1a) on the alignment generated by HHPred (Figure S1), as well as comparison against known structures, allowed for the identification of these regions within EEL and other R-type lectins, building upon an earlier HMM analysis of R-type lectins (Hazes 1996). In all cases, the potential binding sites typically commence with a leucine residue and conclude with a tryptophan residue. From the logo analysis of the alignment, there are numerous positions that appear to feature a conserved leucine and tryptophan. The residue indicating the commencement of a binding site loop can be identified as it is normally aliphatic (leucine, valine, methionine, or isoleucine), as well as being located adjacent to a position largely conserved as an aspartate or a threonine. The conclusion of a binding site loop is typically demarcated by an NQxW motif. The architecture of the R-type fold and location of the binding loops are illustrated in Figure 1b.

A homology modeling strategy utilizing separate templates to build different components of the structure was employed, owing to the limited sequence similarity between EEL and any of the potential template structures. One template was selected and used to build the potential binding sites and another to build the core of the lectin structure. In selecting the template for the EEL binding sites, two features were considered important: (i) the candidate template structure should have been determined in complex with carbohydrates similar or identical to those for which EEL has high affinity, and (ii) the template should feature high sequence homology within the EEL binding site loops and for these loops to be of approximately the same length.

While EEL has a strong binding specificity for fucose terminated carbohydrates, the majority of suitable lectin crystal structures have been determined as complexes with lactose, for which EEL displays a relatively low affinity (Table I). Several structures are available for

lectins in complexes with carbohydrates other than lactose (Table II). The lectin-carbohydrate structures identified as potential templates broadly cover carbohydrates recognized by EEL (Table I). Only one, *Sambucus nigra* agglutinin II (SNA-II), features a bound fucose (Maveyraud et al. 2009), but it was not considered a suitable template since SNA-II contains considerably longer carbohydrate-binding loops compared to EEL. A structure of *Marasmius oreades* agglutinin (MOA) with the B trisaccharide bound is available (Grahn et al. 2009), but MOA does not strongly interact with the terminal fucose; instead MOA tightly binds the core galactose residue, which is a binding mode unlikely to occur with EEL due to its preferential binding to α-fucose terminated carbohydrates.

It is known that fucose and mannose are bound in similar binding modes in DC-SIGN (Guo et al. 2004) and langerin (Feinberg et al. 2011). More recently, structures of a langerin mutant with chitin-derived carbohydrates showed that these ligands also adopt similar binding modes to fucose and mannose (Feinberg et al. 2013). Furthermore, in both cases, branched structures with non-reducing fucose and galactose can bind, but feature the fucose interacting with the same site as mannose in preference to galactose. This suggests that mannose may be a closer structural mimic for fucose than galactose. The only available structures of R-type lectins with mannosides bound are those of actinohivin in complex with Manα1-2Man (Hoque et al. 2012). Furthermore, the actinohivin carbohydrate-binding loops are among the best matches in overall sequence homologies and lengths with respect to the binding site loops of EEL. Thus, actinohivin was identified as an appropriate candidate and the highest resolution complex of actinohivin with Manα1-2Man (PDB ID: 4G1R, resolution 1.57 Å) was selected as the template to model the binding site loops of EEL.

Since the core region of EEL is not likely to be directly involved in ligand recognition, the template affording the greatest sequence homology with the EEL sequence was selected to model the core region. This is the structure of a mosquitocidal holotoxin from

Bacillus sphaericus (MTX(holo)) (PDB ID: 2VSE, resolution 2.50 Å) (Treiber et al. 2008). A chimeric template including the binding sites of actinohivin in complex with Manα1-2Man and the core of MTX(holo) was prepared. The alignment was adjusted to ensure that modeled insertions were largely restricted to the extremities of loop regions in the core structure and to maintain optimal homology between the EEL sequence and the template (Figure 2). Although the sequence homology is low (35%), it is improved over the use of either template.

Two issues with the use of the actinohivin/Manα1-2Man structure as the template for the EEL binding sites required further investigation. The first is that it features the reducing end mannose in the binding site (Figure 3a), whereas the binding specificity of EEL suggests that it has a preference for fucosylated determinants. For both langerin and DC-SIGN, structures of complexes of mannose-containing saccharides have been characterized with mannose in two conformations (Feinberg et al. 2007; Feinberg et al. 2011), with a non-reducing or a reducing mannose fitted into the same region of the electron density maps. In both cases, an identical arrangement of the ring and hydroxyl groups is observed, although different hydroxyl groups are involved in recognition in each case. Thus, it was thought that although the mannoses derived from the actinohivin template were likely to be in the incorrect orientation in the EEL binding site model, they could be used as a template to generate the more probable orientation with the non-reducing end sugar bound. The second issue is that actinohivin features ligands bound to all three binding sites, all of which were included in the model of EEL. Thus, a decision as to which binding site (or sites) is the appropriate to examine needed to be made.

To address both of these issues, the modeled ligands in each site were redocked using core constraints on the ligand. The ligands in both the β and γ sites could be preferentially docked with the non-reducing end bound (Figure 3b). However, the γ site was selected as the likely carbohydrate-binding site on EEL for three reasons. Firstly, the γ site features a greater

sequence homology with the template structure than the β site, thus it is likely to be more accurately modeled. Secondly, complexes of R-type lectins with carbohydrates indicate that the β site is the least commonly utilized site for carbohydrate binding (Table II). Finally, and most importantly, the γ site appears to share some structural features with the binding sites of lectins that interact with terminal fucose residues, particularly *Burkholderia ambifara* lectin (BambL) (Audfray et al. 2012). In the BambL binding site, the hydrophobic face of fucose stacks against a tryptophan residue and hydrogen bonds to a glutamate approximately opposite this stacking arrangement (Figure 3c). The model of EEL has a similar motif, comprised of Trp143 and Asp130 in an approximately perpendicular arrangement (Figure 3a, b). While the modeled mannoside via hydrogen bonding of its 3- and 4-hydroxyls to the carboxylate of Asp130, Trp143 is not heavily engaged in interactions. This suggests that this protein conformation may not be appropriate to bind terminal fucose, and since EEL appears to bind preferentially to fucose-terminated carbohydrates, it is crucial that the model adopts a fucose-binding conformation. Furthermore, since tryptophan fluorescence could be used to monitor ligand binding, it is likely that there is a tryptophan located in the EEL binding site with a direct role in ligand binding, for which Trp143 is the most likely candidate. To generate the likely fucose-binding conformation of the protein, methyl α-fucoside (Fucα1-OMe) was first aligned to the mannoside structure, such that the hydrophobic face of the fucose was directed towards Trp143 while preserving the hydrogen bonding with Asp130. A constrained minimization was then carried out to both maintain the hydrogen bonding with Asp130, while bringing the hydrophobic face of the fucose and Trp143 closer together. The resulting structure more closely resembled fucose recognition in BambL and was used for subsequent docking studies (Figure 3d).

Initial carbohydrate binding modes in the EEL binding site by constrained docking

In the initial binding modes selected for fucoside, mannoside and glucoside ligands, hydrogen bonding interactions appear to occur largely with four residues: Asp130, Asp145, Tyr147 and Gln149 (Table III). Asp130 and Gln149 are involved in hydrogen bonding in all cases. Hydrogen bonds were observed between Lys85 and all of the trisaccharides, as well as Fucα1-2Gal **(11)**. In the model, Lys85 projects from the β site into the γ site, due to being located on a large extension on the loop of the β site (Figure 4). Recognition predominantly involves the terminal non-reducing end residue; the A and B trisaccharides feature two non-reducing termini and the principal interactions occur with the fucose, of which the 2-, 3- and 4-hydroxyl groups are the most frequently involved in recognition. For H-related carbohydrates **(10-12)**, as well as the A **(3)** and B **(9)** trisaccharides, the 6-hydroxyl of the β-galactose is also involved in recognition by EEL. In addition to the hydrogen bonding network, CH-π interactions between the non-reducing end residue and Trp143 occur in all cases, although the strength of these interactions is likely to vary, depending on the carbohydrate and the particular face directed towards Trp143. It is most likely to be strongest with fucose-terminated ligands, where the hydrophobic face of fucose is typically directed towards Trp143. Outside of these dominant interactions, several other protein residues and carbohydrate positions afford supporting roles in carbohydrate recognition by EEL (Table III).

In the initially selected binding mode for the B trisaccharide **(9)**, the α-galactose residue has only a limited involvement in recognition by EEL. However, the fluorescence experiments indicate that "B-like" α-galactosides **(5-8)**, lacking the fucose residue, also bind strongly to EEL. Furthermore, binding modes of the B trisaccharide were identified that feature the α-galactose residue bound to the putative fucose-binding site and satisfied the constraint requirements set during docking (Figure 5a). These findings suggest the possibility

that EEL could bind the non-reducing α-galactose of "B-like" galactosides utilizing the putative fucose-binding site. Therefore, the binding mode of the B trisaccharide with the α-galactose at this site was used to provide core constraints for docking the galactosides. The A-related mono- and disaccharides **(1-2)** were also docked using these core constraints. Binding modes were obtained for all the galactosides using this constraining core (Figure 5b, c), but an initial selection for the binding mode of GalNAcα1-3Gal **(2)** could not be made, as the root-mean-square deviation (RMSD) between the *N*-acetylgalactosamine ring of any of the poses and the templating galactose exceeded the limit of 1.0 Å set for making this selection. The initially selected binding modes for the remaining galactosides (Table IV) reveal similar interactions as in the fucoside ligands. Interactions are restricted to the non-reducing end galactose in all cases, and the 3- and 4-hydroxyl groups are most frequently involved in recognition. Trp143 is again involved in CH-π interactions, but the hydrophobic face of galactose is not oriented towards this residue.

Aside from containing a pyranose ring, NeuAcα2-OMe **(15)** shares little similarity with either fucose or galactose. Therefore, core constraints similar to those used to dock the fucosides and the galactosides could not be used to dock NeuAcα2-OMe. Instead, unconstrained docking of NeuAcα2-OMe was performed and the structural similarity of the pyranose ring of the docked poses and the fucose and galactose template structures was determined and utilized to make an initial selection of the binding mode for this ligand. One binding mode of NeuAcα2-OMe with sufficient similarity in ring placement to the template fucose could be identified (Figure 5d). This binding mode involves hydrogen bonding with a similar set of residues to the fucosides and galactosides, including Asp130, Tyr147, Gln149 and Lys85. Surprisingly, the interaction with Lys85 is not a charge-assisted hydrogen bond with the carboxylate of NeuAcα2-OMe, but with the carbonyl oxygen of the *N*-acetyl group and the 9-hydroxyl. The carboxylate of NeuAcα2-OMe hydrogen bonds to the phenolic

hydroxyl of Tyr147. Although NeuAc does not feature a prominent hydrophobic face, CH-π interactions are observed between the unsubstituted 3-position of the ring and Trp143. The methyl of the *N*-acetyl group engages in hydrophobic contacts with Ile138 and Gly140.

Relationship between predicted and experimental binding energies of carbohydrates interacting with EEL

A weak positive correlation was observed between the binding energies predicted by Glide (Emodel) for the initially selected binding modes and the experimentally determined binding energies (Figure 6a). However, several ligands were outliers to the initially generated relationship (Table SII). The relationship was optimized by removing data points in turn and recalculating the line of best fit (see Materials and Methods). Using this approach, the initially selected modes of the following saccharides were considered outliers: Galα1-OMe **(5)**, the A trisaccharide **(3)**, the linear B trisaccharide **(8)**, Manα1-6Manα1-OMe **(14)**, chitotriose **(16)** and the H trisaccharide **(12)**. Alternative binding modes were then selected for these carbohydrates based on calculating the expected Emodel range from the experimentally observed range in the binding affinity using the optimized relationship (Figure 6b). Since core constraints were employed during docking, it was anticipated that the new binding modes would still bear strong similarities to those initially selected. The expected Emodel range of GalNAcα1-3Gal **(2)** was also calculated and used to select a binding mode from docking, since an initial binding mode could not be selected based on the RMSD comparisons.

Binding modes within the calculated Emodel ranges for Galα1-OMe **(5)**, the linear B trisaccharide **(8)** and GalNAcα1-3Gal **(2)** could be identified (Figure 7). The new binding mode for Galα1-OMe (Figure 7a) is related to the initially identified binding mode by a reflection in the O5-C3 axis. This allows for considerably more hydrogen bonds to occur

between the ligand and EEL, as well as permitting the utilization of the hydrophobic face of galactose for CH-π interactions with Trp143. The new binding mode for the linear B trisaccharide features some small changes from the initial binding mode, resulting in the loss of some hydrogen bonds, but is an overall similar binding mode. The selected binding mode for GalNAcα1-3Gal (Figure 7b) is reflected from that selected for GalNAcα1-OMe (Figure 5c). Similar hydrogen bonds are formed, but the interaction between the *N*-acetyl group and Ile138 is not observed.

From the docking performed, no binding modes could be identified within the Emodel range calculated for the A trisaccharide **(3)**, Manα1-6Manα1-OMe **(14)**, chitotriose **(16)**, or the H trisaccharide **(12)**. All of the binding modes obtained for these ligands gave Emodel values greater in magnitude than the range expected, with the exception of the H trisaccharide, for which the Emodel values were lower in magnitude than expected (Figure 6c). The new binding mode selected for the A trisaccharide features a reflection in the O5-C3 axis of the fucose, resulting in a significant reorientation of the remainder of the molecule and overall fewer interactions than the initial binding mode, including the loss of the stacking interaction between the hydrophobic face of the fucose and Trp143 (Figure 7c). Since the binding mode closest in magnitude to the Emodel range calculated for Manα1-6Manα1-OMe featured the reducing end residue bound, the next closest binding mode with the non-reducing end bound was selected, which was the second-closest to the calculated Emodel value. The new binding mode for Manα1-6Manα1-OMe is reflected from the initial binding mode, but the majority of interactions are preserved (Figure 7d). The new binding mode for chitotriose is almost identical to the initially selected binding mode (Figure 4c). The new binding mode for the H trisaccharide features the fucose bound in an almost equivalent fashion, but the orientation of the *N*-acetyllactosamine portion is different so that the internal galactose participates in more interactions than the reducing end *N*-acetylglucosamine.

A final relationship between Emodel values and experimental binding energies was determined using the selected binding modes from all of the ligands, including the binding modes selected using the optimized relationship (Figure 6c, Table SIII). The resulting relationship was similar to the optimized relationship, although the inclusion of more points resulted in a lower correlation coefficient (0.886 vs. 0.960). Nonetheless, the correlation coefficient still indicates a strong correlation between the experimental binding affinities and Emodel values.

Overlays of the final binding modes for carbohydrate ligands in the EEL binding site show the general recognition pattern of end-on insertion of a terminal non-reducing end monosaccharide, which is fitted firmly in a binding pocket formed between Trp143 and Glu130 (Figure 8). The different carbohydrate ligands are accommodated through the ability of the second and third carbohydrate residues to adopt a variety of conformations with the potential to interact with several additional amino acids in an extended, but more open binding site in EEL. The binding of terminal carbohydrate determinants allows EEL to bind a variety of longer chain and complex glycans, but based on our modeling studies the interactions would be limited to between one and three sugar residues.

Discussion

EEL has been proposed to form the prototypical member of a distinct family of lectins, referred to as *Euonymus* lectins (EUL) (Fouquaert et al. 2008; Fouquaert et al. 2009). The fold recognition approach employed by HHPred, which is capable of identifying relationships between proteins distantly related in sequence (Söding et al. 2005), suggests that EEL adopts a β-trefoil fold most similar to ricin B-like (R-type) lectins, a family of proteins from a diverse range of plants, bacteria and fungi (Cummings and Etzler 2009). This finding is in agreement with the annotation of EEL in the InterPro database. Thus, it is possible that all of

the proteins of the EUL family may adopt ricin B-like folds, which cannot be easily recognized with purely sequence-based alignments. Indeed, R-type lectins have been used as templates to generate models of three EUL family proteins, including EEL (referred to as EEA) in a separate study (Fouquaert and Van Damme 2012). Our docking investigation and the relationship between the computed and experimentally determined binding energies provides support for the γ site as the primary carbohydrate-binding site of EEL. Importantly, the γ site features a solvent exposed tryptophan residue (Trp143) that appears to be involved in CH-π interactions with the ligands and is the most likely residue contributing to the changes in fluorescence observed during the binding experiments.

The preferential binding of EEL to B and H blood group glycoconjugates was previously documented (Pacak and Kocourek 1975; Petryniak and Goldstein 1987; Petryniak et al. 1977; Schmidt 1954). Subsequently, the dominant binding of the non-reducing terminal residues in the B blood group tetrasaccharide, Galα1-3(Fucα1-2)Galβ1-4GlcNAc and a tolerance for binding other Galα1-3 terminated glycolipids was reported (Teneberg et al. 2003). Glycan array screening has been used to demonstrate binding of EEL to a series of mammalian oligosaccharides including those containing B and H determinants, terminal Galα1-3 epitopes, an A-type I glycan as well as high mannose structures containing the core pentasaccharide [Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc] (Fouquaert et al. 2008). In this study, the solution-based fluorescence binding experiments (Table I) show stronger binding of the H trisaccharide over the B trisaccharide, but the general pattern of EEL interacting with H and B blood group and α-mannose terminated saccharides is consistent with earlier studies. In addition, we show that EEL can interact with moderate affinities with α-galactose terminated saccharides (lacking α-fucose), sialic acid (NeuAcα2-OMe), A trisaccharide and an *N*-acetylglucosamine trisaccharide, chitotriose. Thus, the carbohydrate binding specificity of EEL may be broader than originally described.

The Mahal group has screened EEL against a glycan array at 4 concentrations (0.1, 1, 10 and 100 μg/ml). These data are publically available through the Consortium for Functional Glycomics (CFG; http://www.functionalglycomics.org/, accessed 27 August 2014), but not yet published in detail (Agarwal et al. 2012). Importantly, the CFG datasets show that EEL binds strongly to B and moderately to H blood group determinants, which is similar to our observations (see Table I). However, in the glycan arrays, there was no appreciable binding of EEL to monosaccharides, the A blood group or other epitopes. It is possible that the fluorescence titration spectroscopy method we used is more sensitive than the glycan arrays for detecting weaker interactions. Also, the noted differences in EEL specificity could be due to anchoring carbohydrates to a solid support in the glycan arrays, compared to binding in solution of free saccharides, as we report here.

The present docking studies build on our previous work in modeling carbohydrate-protein recognition (Agostino et al. 2014; Agostino et al. 2012; Agostino et al. 2010; Agostino et al. 2011). In this work, we have used core constrained docking and employed comparison to a likely fucose conformation proposed through homology modeling and structural knowledge of fucose recognition by other proteins (Audfray et al. 2012). Previously, we have specifically considered the role of alternative binding modes in recognition (Agostino et al. 2010). Alternate or multiple carbohydrate binding modes have been observed experimentally in a variety of carbohydrate-binding proteins, most notably langerin (Feinberg et al. 2011), DC-SIGN (Feinberg et al. 2007) and SNA-II (Maveyraud et al. 2009). Thus, in this work we selected alternative binding modes, compared to the initially selected binding modes, which afforded more favourable energetic properties in line with the derived relationship between experimental and computed binding affinities. However, it is only possible to optimize binding mode selection in this manner if such a relationship is

known, thus underscoring the importance of both experimental and computational procedures in developing a full understanding of protein-ligand binding.

Crystallography of the human lectins langerin and DC-SIGN demonstrates that fucose and mannose bind to these proteins in almost identical configurations (Feinberg et al. 2011; Guo et al. 2004). Furthermore, a mutant of langerin can bind *N*-acetylglucosamine terminated sugars with a very similar binding mode to both mannose and fucose (Feinberg et al. 2013). This knowledge was exploited in predicting the binding modes for fucose, mannose and *N*-acetylglucosamine terminated ligands with EEL. A binding mode for the B trisaccharide with the α-galactose bound at the fucose-binding site suggested this site could also bind α-galactose residues. The determination of a relationship between the experimental and calculated binding energies for such binding modes for α-galactosides strengthens the notion that EEL utilizes the same binding site for recognition of fucosides and α-galactosides. However, based on the binding modes for the A and B trisaccharides, galactose/*N*-acetylgalactosamine is likely to only interact with the main binding site when fucose is not present in the carbohydrate structure.

In summary, we predict that EEL adopts a β-trefoil fold similar to ricin B-like (R-type) lectins and utilizes the γ site to interact with a broad range of carbohydrate ligands. Our studies have affirmed the preferential recognition by EEL of terminal α-fucose residues, especially in H and B blood group related carbohydrates. Our findings have demonstrated the potential specificity of EEL for other carbohydrate ligands, which may be of relevance for understanding the role of EEL in plant biology.

Materials and methods

Carbohydrates and lectin

Carbohydrates used for binding studies (Table I) were purchased from several suppliers: **(1)**, Toronto Research Chemicals, Ontario, Canada; **(2, 3, 6-9, 16)**, Dextra laboratories, Reading, UK; **(4, 5, 15, 17-19)**, Sigma-Aldrich, Missouri, USA; **(10-12)**, Carbosynth, Berkshire, UK. Mannosides **(13, 14)** were synthesized as previously reported (Cao et al. 2011). *Euonymus europeaus* lectin (EEL) was purchased from Vector laboratories, California, USA.

Fluorescence titration spectroscopy and K_d determination

Fluorescence experiments were adapted from a reported procedure (Chipman et al. 1967). All measurements were carried out using a Varian Cary Eclipse fluorescence spectrophotometer. The Scan application of the Cary Eclipse software was used to carry out fluorescence measurements and perform calculations based on fluorescence data. The excitation wavelength was set to 280 nm, and the emission spectrum was collected from 280 nm to 500 nm. Protein samples were prepared at 2 μM in 1 mL of buffer (10 mM HEPES, 150 mM NaCl and 0.1 mM $CaCl_2$ at pH 7.5) and loaded into a 1.5 mL quartz cuvette with a 1.0 cm path length. Carbohydrates were initially titrated in increments of 20-fold up to 100-fold excess, then increments of 50-fold up to 300-fold excess, which allowed for an approximate establishment of the ligand affinity. Once this was established, the experiment was repeated three times, using one of the titration schedules detailed below. For high affinity ligands (<30 μM), increments of 5-fold up to 50-fold excess were used. For medium-high affinity ligands (30-90 μM), increments of 10-fold up to 100-fold excess were used. For medium-low affinity ligands (90-180 μM), 20-fold up to 100-fold excess were used, followed by increments of 50-fold up to 300-fold excess. For low affinity ligands (>180 μM), increments of 50-fold up to 500-fold excess were used.

To determine the K_d values, the integral from 288 nm (representing the approximate trough between the excitation maximum and the emission maximum) to 500 nm for each scan was calculated and recorded. For each experiment (i.e., one complete titration), the integrals were standardized according to the following equation:

$$\overline{F} = 100 \times \frac{F - F_0}{F_0}$$

where F-bar is the standardized integral from the fluorescence experiment, F is the absolute value of the integral at a given ligand concentration, and F_0 is the absolute value of the integral with no ligand present. When no ligand is present, $F = F_0$ and thus F-bar = 0. Since the absolute values of the integral can vary between experiments, this standardization allows for the relative increase in fluorescence to be monitored in a given titration experiment, which is typically a more consistent measure than the absolute increase. Non-linear regression according to a one-site binding model was used to determine K_d values (as well as the error associated with these values) from the standardized integrals in GraphPad Prism 5. Gibbs energies (ΔG) were calculated from K_d values using the Van 't Hoff equation, $\Delta G = RT\ln(K_d)$.

Preparation of EEL homology model

All molecular modeling tasks were performed with Schrödinger (Suite 2012), unless stated otherwise. The EEL sequence was obtained from the UniProt database (Accession Code: B3SV73) (Fouquaert et al. 2008) and aligned to proteins in the PDB using the HHPred server (Söding et al. 2005). The following HHPred settings were altered from their defaults: the maximum MSA generation iterations was set to eight, realign with MAC was enabled and the E-value threshold for MSA generation was increased to 0.1. Templates featuring high coverage of the EEL sequence (>80%) as well as a probability score greater than 95% were

selected. Since the majority of potential templates identified were ricin B-like lectins (PFAM accessions PF14200 and PF00652) (Table SI), templates that did not display this fold were discarded. An alignment was generated from the remaining templates (Figure S1) and in conjunction with logo analysis, was used to define the locations of the α, β and γ sites of EEL. Logo analysis was performed using the Seq2Logo server (Thomsen and Nielsen 2012). Settings for logo analysis were kept at defaults, with the exception that the weight on prior was set to zero. The potential binding sites and the core structure of the lectin were built from distinct templates. The binding sites were built against the complex of actinohivin with Manα1-2Man (PDB ID: 4G1R) (Hoque et al. 2012), while the core structure was built against the structure of MTX(holo) (PDB ID: 2VSE) (Treiber et al. 2008). A chimeric structure featuring the binding sites of the actinohivin complex on the core of MTX(holo) was built in Maestro 9.3 and used as the template for building the EEL model in Prime 3.1 (Jacobson et al. 2004). The alignment generated by Prime was manually adjusted to better reflect the alignment generated by HHPred, as well as to ensure that modeled insertions were restricted to loop regions in the core while maintaining optimal sequence homology. Bound ligands from the actinohivin complex were retained during the model building process. Structural discontinuities in the backbone of the structure that were not automatically closed by Prime were manually closed by linking the appropriate atoms and minimizing the bond length using the Structure Sculpting tool in Maestro. The built structure was minimized using Prime Minimization.

Identification and optimization of carbohydrate-binding site on EEL

The docking protocol used was based on our earlier approaches, with some modifications (Agostino et al. 2012; Agostino et al. 2009; Agostino et al. 2010; Agostino et al. 2011). The bound mannoside at each site was redocked using Glide 5.8 (Friesner et al. 2004). Grids at

each site were generated using default settings. For the docking step, Glide SP mode was used. The option to sample ring conformations was disabled, as well as the option to add Epik state penalties to the docking score. Up to 100 poses were specified to be returned. Poses with a coulomb-vdW energy greater than 20 kcal/mol were rejected and poses were clustered to an RMSD of 0.5 Å (the default). Core constraints were applied during docking to ensure that the geometry of the hydroxyls pointing towards the site was preserved during docking; the constraining atoms selected from the initial structure are shown in Figure S2. Docking was restricted to the reference position of the defined core, using a tolerance of 0.5 Å. The γ site was selected as the likely carbohydrate-binding site and the structure of methyl α-fucoside was overlaid on the redocked mannoside using the Atom Pair facility of the Superposition tool in Maestro. Atom pairs from the rings of the structures were selected such that the overlay followed that of mannose and fucose in the langerin complexes with these carbohydrates (Feinberg et al. 2011).

The resulting EEL-fucoside complex was then further refined using Prime. Residues within 6.0 Å of the fucoside, as well as the fucoside itself, were selected for minimization. Distance constraints were set to ensure that the minimization resulted in a BambL-like conformation of the site. Specifically, the distances between C3, C4 and C5 of the fucoside and the Cδ2, Cε3 and Cζ3 atoms of Trp143 respectively were constrained to a length of 4.0 Å with a force constant of 350 kcal/mol, and the distances between the hydrogen atoms of the 2- and 3-hydroxyls and Asp130 were constrained to a length of 2.0 Å with a force constant of 700 kcal/mol. The resulting structure was subjected to side-chain prediction in the same region of the protein selected for minimization, excluding the fucoside. Backbone sampling was enabled for side-chain prediction. Following this procedure, the same region of the protein (excluding the fucoside) was further minimized.

Docking of carbohydrate ligands

The fucoside bound to the γ site of EEL was used to centre the grid. A cubic grid box was defined for docking and sized to be appropriate for docking ligands of length less than or equal to 20 Å. The inner box was expanded to the maximum size of 14 Å in all directions. Hydrogen bonding constraints on Asp130 were defined. Ligand structures were generated in Maestro. All ligands except for low affinity ligands were examined in docking.

The same options for docking as previously set were used, with some adjustments. Ligand amides were sampled in the *trans* conformation only. The tolerance on core pattern comparison was increased to 1.0 Å. Depending on the set of ligands being docked, different core patterns were used (see below). The hydrogen bonding constraints on Asp130 defined during grid generation were used for all docking runs.

Initially, all fucose-terminated ligands, including the A and B trisaccharides, were docked using core constraints based on the modeled fucoside structure (Figure S2). Galactose- and *N*-acetylgalactosamine-terminated ligands, excluding the A and B trisaccharides, were docked using core constraints from the pose of the B trisaccharide where the non-reducing galactose afforded the best fit to the modeled fucoside structure (Figure S2). Since NeuAcα2-OMe featured a different scaffold to the other ligands, core constraints were not applied to dock this ligand, however, the hydrogen bonding restraints on Asp130 were applied and the limit on RMSD-based clustering was raised to 2.0 Å.

To make an initial selection of the most likely binding modes, a series of twelve RMSDs between the relevant core structure and the relevant ligand residue was computed for each binding mode, representing all possible RMSDs between the ring atoms for all rotations and reflections of the rings (Figure S3). This was done to identify alternative arrangements of the carbohydrates similar to those identified in carbohydrate binding to DC-SIGN and langerin, where rings of two conformations may appear to align well, but non-equivalent sets

of atoms are overlapping. Special preference was given to the cases of comparing equivalent ring atoms (i.e., O5-O5, C1-C1, C2-C2, etc.) and the reflection of this in the O5-C3 axis (i.e., O5-O5, C1-C4, C2-C3, etc.). Specifically, if structures with a low RMSD (in this case, defined to be less than or equal to 1.0 Å) in the equivalent comparison could be identified, the structure with the lowest RMSD for this comparison was selected. If not, RMSDs in the reflection comparison were then examined, followed by any other comparison. NeuAcα2-OMe was compared to both the template fucose and galactose residues and the pose with the lowest RMSD to either of these in any comparison was selected.

Interactions were identified using the H-Bonds and Contacts tool in Maestro. The default parameters for identifying hydrogen bonds were used. Contacts were examined between carbon and non-polar hydrogen atoms. The cutoff ratios for bad and ugly contacts were set to 1.00 and 0.66, respectively, corresponding to a carbon-hydrogen distance range of 1.9-2.9 Å.

Prediction of experimental binding energy and selection of final poses

A relationship between the experimental binding energies and the calculated binding energies from docking (the Emodel property in Glide) was sought. For each ligand, the Emodel of the initially selected structure from the RMSD comparison was plotted against the experimentally determined binding energy and an initial relationship between these properties was identified. Residuals were calculated for each point. The relationship was optimized by removing the point with the residual greatest in magnitude and regenerating the line-of-best-fit. If removing a point resulted in a new line-of-best-fit with a lower correlation coefficient than the previously generated line, that point was restored and the point with the next greatest residual removed instead. Points were removed from the plot until the correlation coefficient of the resulting line-of-best-fit was greater than 0.95.

To determine likely binding modes for the ligands removed while optimizing the relationship between experimental and calculated binding energy, the optimized relationship was used to determine the expected Emodel value from the experimental binding affinity. Poses within a range of Emodel values were considered, as indicated by the error in the experimental binding affinities. If multiple poses within the Emodel range were identified, the pose fitting the best with the appropriate core structure was selected as the most likely binding mode. Where no pose featured an Emodel within the range expected, the pose with the closest Emodel to the expected value was selected as the most likely pose. A final line-of-best-fit was then generated using all of the selected binding modes.

Acknowledgements

The authors thank Dr Lara K Mahal and Mr Lawrence Meche for helpful discussions relating to the CFG glycan array data for EEL. MA is a recipient of an NHMRC Early Career Fellowship (GNT1054245). TD is supported by an Australian Postgraduate Award. TV is an Australian NHMRC Industry Career Development Fellow. SJW is an Australian Research Council Future Fellow. This work was supported by computational resources provided by the Victorian Life Sciences Computational Initiative through grant VR0250. The authors gratefully acknowledge the contribution toward this study from the Victorian Operational Infrastructure Support Program received by the Burnet Institute.

Abbreviations

DC-SIGN, dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin; EEL/EEA, *Euonymus europaeus* lectin/agglutinin; EUL, *Euonymus* lectin family; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HMM, hidden Markov model; MOA, *Marasmius oreades* agglutinin; MTX, mosquitocidal toxin; $OOctN_3$, O-octyl azide; OMe, O-methyl; PDB, Protein Data Bank; RMSD, root-mean-square deviation; SNA-II, *Sambucus nigra* agglutinin II.

References

Agarwal P, Smith D, Eng W, Cummings R, Mahal L. 2012. Large scale glycan array analysis of commercial lectins and antibodies: 86 lectins and 15 antibodies. *Glycobiology*, 22:1646.

Agostino M, Gandhi NS, Mancera RL. 2014. Development and application of site mapping methods for the design of glycosaminoglycans. *Glycobiology*: doi:10.1093/glycob/cwu1045.

Agostino M, Ramsland PA, Yuriev E. 2012. Antibody recognition of cancer-related gangliosides and their mimics investigated using *in silico* site mapping. *PLoS One*, 7:e35457.

Agostino M, Sandrin MS, Thompson PE, Yuriev E, Ramsland PA. 2009. In silico analysis of antibody-carbohydrate interactions and its application to xenoreactive antibodies. *Mol Immunol*, 47:233-246.

Agostino M, Sandrin MS, Thompson PE, Yuriev E, Ramsland PA. 2010. Identification of preferred carbohydrate binding modes in xenoreactive antibodies by combining conformational filters and binding site maps. *Glycobiology*, 20:724-735.

Agostino M, Yuriev E, Ramsland PA. 2011. A computational approach for exploring carbohydrate recognition by lectins in innate immunity. *Front Immunol*, 2:23.

Al Atalah B, Rouge P, Smith DF, Proost P, Lasanajak Y, Van Damme EJM. 2012. Expression analysis of a type S2 EUL-related lectin from rice in Pichia pastoris. *Glycoconj J*, 29:467-479.

Audette GF, Vandonselaar M, Delbaere LTJ. 2000. The 2.2 Å resolution structure of the O(H) blood-group-specific lectin I from *Ulex europaeus*. *J Mol Biol*, 304:423-433.

Audfray A, Claudinon J, Abounit S, Ruvoen-Clouet N, Larson G, Smith DF, Wimmerová M, Le Pendu J, Romer W, Varrot A*, et al.* 2012. Fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes. *J Biol Chem*, 287:4335-4347.

Cao B, White JM, Williams SJ. 2011. Synthesis of glyconjugate fragments of mycobacterial phosphatidylinositol mannosides and lipomannan. *Beilstein J Org Chem*, 7:369-377.

Chipman DM, Grisaro V, Sharon N. 1967. The binding of oligosaccharides containing N-acetylglucosamine and N-acetylmuramic acid to lysozyme. The specificity of binding subsites. *J Biol Chem*, 242:4388-4394.

Crocker PR, Paulson JC, Varki A. 2007. Siglecs and their roles in the immune system. *Nat Rev Immunol*, 7:255-266.

Cummings RD, Etzler ME. 2009. R-type Lectins. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY).

Feinberg H, Castelli R, Drickamer K, Seeberger PH, Weis WI. 2007. Multiple modes of binding enhance the affinity of DC-SIGN for high mannose *N*-linked glycans found on viral glycoproteins. *J Biol Chem*, 282:4202-4209.

Feinberg H, Rowntree TJ, Tan SL, Drickamer K, Weis WI, Taylor ME. 2013. Common polymorphisms in human langerin change specificity for glycan ligands. *J Biol Chem*, 288:36762-36771.

Feinberg H, Taylor ME, Razi H, McBride R, Knirel YA, Graham SA, Drickamer K, Weis WI. 2011. Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J Mol Biol*, 405:1027-1039.

Fouquaert E, Peumans WJ, Smith DF, Proost P, Savvides SN, Van Damme EJM. 2008. The "old" *Euonymus europaeus* agglutinin represents a novel family of ubiquitous plant proteins. *Plant Physiol*, 147:1316-1324.

Fouquaert E, Peumans WJ, Vandekerckhove TTM, Ongenaert M, Van Damme EJM. 2009. Proteins with an Euonymus lectin-like domain are ubiquitous in Embryophyta. *BMC Plant Biol*, 9.

Fouquaert E, Van Damme EJ. 2012. Promiscuity of the euonymus carbohydrate-binding domain. *Biomolecules*, 2:415-434.

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shaw DE, Shelley M*, et al.* 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*, 47:1739-1749.

Grahn E, Askarieh G, Holmner Å, Tateno H, Winter HC, Goldstein IJ, Krengel U. 2007. Crystal structure of the *Marasmius oreades* mushroom lectin in complex with a xenotransplantation epitope. *J Mol Biol*, 369:710-721.

Grahn EM, Winter HC, Tateno H, Goldstein IJ, Krengel U. 2009. Structural characterization of a lectin from the mushroom *Marasmius oreades* in complex with the blood group B trisaccharide and calcium. *J Mol Biol*, 390:457-466.

Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K. 2004. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat Struct Mol Biol*, 11:591-598.

Hardison SE, Brown GD. 2012. C-type lectin receptors orchestrate antifungal immunity. *Nat Immunol*, 13:817-822.

Hatakeyama T, Unno H, Kouzuma Y, Uchida T, Eto S, Hidemura H, Kato N, Yonekura M, Kusunoki M. 2007. C-type lectin-like carbohydrate recognition of the hemolytic lectin CEL-III containing ricin-type -trefoil folds. *J Biol Chem*, 282:37826-37835.

Hazes B. 1996. The (QxW)3 domain: a flexible lectin scaffold. *Protein Sci*, 5:1490-1501.

Hoque MM, Suzuki K, Tsunoda M, Jiang J, Zhang F, Takahashi A, Ohbayashi N, Zhang X, Tanaka H, Omura S*, et al.* 2012. Structural insights into the specific anti-HIV property of actinohivin: structure of its complex with the alpha(1-2)mannobiose moiety of gp120. *Acta Crystallogr*, D68:1671-1679.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S*, et al.* 2012. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*, 40:D306-312.

Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. 2004. A hierarchical approach to all-atom protein loop prediction. *Proteins*, 55:351-367.

Kadirvelraj R, Grant OC, Goldstein IJ, Winter HC, Tateno H, Fadda E, Woods RJ. 2011. Structure and binding analysis of *Polyporus squamosus* lectin in complex with Neu5Acα(2-6)Galβ1-4GlcNAc human-type influenza receptor. *Glycobiology*, 21:973-984.

Kaltner H, Gabius HJ. 2012. A toolbox of lectins for translating the sugar code: the galectin network in phylogenesis and tumors. *Histol Histopathol*, 27:397-416.

Koharudin LM, Gronenborn AM. 2014. Antiviral lectins as potential HIV microbicides. *Curr Opin Virol*, 7C:95-100.

Matsushita M. 2013. Ficolins in complement activation. *Mol Immunol*, 55:22-26.

Maveyraud L, Niwa H, Guillet V, Svergun DI, Konarev PV, Palmer RA, Peumans WJ, Rouge P, Van Damme EJ, Reynolds CD*, et al.* 2009. Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from Sambucus nigra. *Proteins*, 75:89-103.

Mitchell E, Houles C, Sudakevitz D, Wimmerova M, Gautier C, Perez S, Wu AM, Gilboa-Garber N, Imberty A. 2002. Structural basis for oligosaccharide-mediated adhesion of Pseudomonas aeruginosa in the lungs of cystic fibrosis patients. *Nat Struct Biol*, 9:918-921.

Morgan WT, Watkins WM. 2000. Unravelling the biochemical basis of blood group ABO and Lewis antigenic specificity. *Glycoconj J*, 17:501-530.

Pacak F, Kocourek J. 1975. Studies on phytohemagglutinins. XXV. Isolation and characterization of hemagglutinins of the spindle tree seeds (Evonymus europaea L.). *Biochim Biophys Acta - Protein Struct*, 400:374-386.

Petryniak J, Goldstein IJ. 1987. Evonymus europaea lectin. *Methods Enzymol*, 138:552-561.

Petryniak J, Janusz M, Markowska E, Lisowska E. 1981. Purification of the *Euonymus europaeus* lectin by affinity chromatography on the desialized Mn blood group glycoprotein, and lectin NH_2-terminal analysis. *Acta Biochim Polon*, 28:267-272.

Petryniak J, Pereira ME, Kabat EA. 1977. The lectin of Euonymus europeus: purification, characterization, and an immunochemical study of its combining site. *Arch Biochem Biophys*, 178:118-134.

Petryniak J, Petryniak B, Wasniowska K, Krotkiewski H. 1980. Isolation and immunochemical characterization of the *Euonymus europaeus* lectin receptor from the major sialoglycoprotein of human O erythrocytes. *Eur J Biochem*, 105:335-341.

Schmidt V. 1954. Die hämagglutination, im besonderen menschlicher B-blutzellen durch extrakte aus samen von Euonymus vulgaris (Pfaffenhütchen). *Z Immunitätsforschung*, 11:432-439.

Sharon N, Lis H. 2004. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*, 14:53R-62R.

Skamnaki VT, Peumans WJ, Kantsadi AL, Cubeta MA, Plas K, Pakala S, Zographos SE, Smagghe G, Nierman WC, Van Damme EJ, *et al.* 2013. Structural analysis of the *Rhizoctonia solani* agglutinin reveals a domain-swapping dimeric assembly. *FEBS J*, 280:1750-1763.

Söding J, Biegert A, Lupas AN. 2005. The HHPred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, 33:W244-W248.

Suite. 2012. Glide, version 5.8, Prime, version 3.1, Maestro, version 9.3. New York, NY, USA: Schrödinger, LLC.

Sulzenbacher G, Roig-Zamboni V, Peumans WJ, Rouge P, Van Damme EJM, Bourne Y. 2010. Crystal structure of the GalNAc/Gal-specific agglutinin from the phytopathogenic ascomycete *Sclerotinia sclerotiorum* reveals novel adaptation of a beta-trefoil domain. *J Mol Biol*, 400:715-723.

Tempel W, Tschampel S, Woods RJ. 2002. The xenograft antigen bound to Griffonia simplicifolia lectin 1-B(4). X-ray crystal structure of the complex and molecular dynamics characterization of the binding site. *J Biol Chem*, 277:6615-6621.

Teneberg S, Alsén B, Ångstrom J, Winter HC, Goldstein IJ. 2003. Studies on Gala3-binding proteins: comparison of the glycosphingolipid binding specificities of *Marasmius oreades* lectin and *Euonymus europaeus* lectin. *Glycobiology*, 13:479-486.

Thomsen MC, Nielsen M. 2012. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res*, 40:W281-287.

Treiber N, Reinert DJ, Carpusca I, Aktories K, Schulz GE. 2008. Structure and mode of action of a mosquitocidal holotoxin. *J Mol Biol*, 381:150-159.

Van Hove J, Fouquaert E, Smith DF, Proost P, Van Damme EJM. 2011. Lectin activity of the nucleocytoplasmic EUL protein from Arabidopsis thaliana. *Biochem Biophys Res Comm*, 414:101-105.

Wimmerova M, Mitchell E, Sanchez JF, Gautier C, Imberty A. 2003. Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for Aleuria aurantia lectin. *J Biol Chem*, 278:27059-27067.

Yabe R, Suzuki R, Kuno A, Fujimoto Z, Jigami Y, Hirabayashi J. 2007. Tailoring a novel sialic acid-binding lectin from a ricin-B chain-like galactose-binding protein by natural evolution-mimicry. *J Biochem*, 141:389-399.

Yuriev E, Agostino M, Farrugia W, Christiansen D, Sandrin MS, Ramsland PA. 2009. Structural biology of carbohydrate xenoantigens. *Expert Opin Biol Ther*, 9:1017-1029.

Yuriev E, Farrugia W, Scott AM, Ramsland PA. 2005. Three-dimensional structures of carbohydrate determinants of Lewis system antigens: implications for effective antibody targeting of cancer. *Immunol Cell Biol*, 83:709-717.

Figure legends

Figure 1. Identification and analysis of the ricin B-like fold of EEL. (a) Logo analysis of alignment of ricin B-like lectins. Leucine and aspartate residues are shown in red, asparagine, glutamine and tryptophan residues are shown in green. The locations of α, β and γ sites are highlighted with yellow, green and pink shading, respectively. Numbering used in the analysis is sequential, but includes gaps inserted into the sequence to generate the alignment. (b) α, β and γ sites identified from logo analysis highlighted on the complex of actinohivin with Manα1-2Man (PDB ID: 4G1R). Colors used to highlight the sites on the structure correspond to those used to highlight the sites on the logo analysis. Bound carbohydrates are shown with blue carbon atoms and red oxygen atoms.

Figure 2. Sequence alignment between EEL and the actinohivin-MTX(holo) chimeric template. Numbering follows the EEL sequence; where the alignment of the EEL sequence includes a gap, a corresponding gap in the numbering is made to maintain sequential numbering for the EEL sequence. Gaps in the sequences are indicated by a tilde (~). Consensus symbols shown below alignment indicate identical residues (*), strongly homologous residues (:) and weakly homologous residues (.). The sequences corresponding to the α, β and γ sites are highlighted with yellow, green and pink shading, respectively.

Figure 3. Homology modeling of EEL with ligand-guided improvements in the γ binding site conformation. (a) Initial model of the EEL γ site generated based on the actinohivin-MTX(holo) chimera with Manα1-2Man bound. (b) EEL γ site with Manα1-2Man docked in reversed orientation to initial model. (c) Chain A of BambL in complex with the H-type 2 trisaccharide Fucα1-2Galβ1-4GlcNAcβ (PDB ID: 3ZZV). For clarity, residues 27-39 of BambL are hidden and only the terminal bound fucose residue of the ligand is shown.

Hydrogens were added and optimized using the Protein Preparation Wizard in Maestro. (d) Final model of the EEL γ site induced to fit methyl α-fucoside. In all panels, the neighboring tryptophan and acidic residues are shown. The protein backbones are colored from N- to C-terminus in a blue to red rainbow gradient.

Figure 4. Binding modes of selected fucose-, mannose- and *N*-acetylglucosamine-terminated ligands in complex with EEL. (a) Fucα1-2Gal. (b) Manα1-OMe. (c) Chitotriose. (d) B trisaccharide. Protein residues involved in interaction are labeled in the first panel only and positioned approximately equivalently in the remaining panels. For clarity, Trp143 is not shown.

Figure 5. Binding modes of selected α-galactose-terminated ligands in complex with EEL, and NeuAcα2-OMe. (a) B trisaccharide in alternative binding mode, with α-galactose bound at the fucose site. (b) Galα1-3Galβ1-OMe. (c) GalNAcα1-OMe. (d) NeuAcα2-OMe. Protein residues involved in interaction are labeled in the first panel only and positioned approximately equivalently in the remaining panels. For clarity, Trp143 is not shown.

Figure 6. Correlation of experimental and computed binding energy. (a) Relationship generated considering all initially selected binding modes of all ligands. Outliers removed in optimizing the relationship are highlighted with open points. (b) Relationship generated considering initially selected binding modes, removing outliers (optimized relationship). (c) Final relationship generated, including alternative binding modes of ligands removed from the optimized relationship (shown with open points). Line-of-best-fit is shown in all cases as a dashed line. Equation and correlation coefficient for each line-of-best-fit shown below each graph. All values shown in kcal/mol.

Figure 7. Final binding modes of selected ligands identified using the optimized relationship between experimental and computed binding energy. (a) Galα1-OMe. (b) GalNAcα1-3Gal. (c) A trisaccharide. (d) Manα1-6Manα1-OMe. For clarity, Trp143 is not shown.

Figure 8. Overlays of final selected binding modes in the EEL binding site. (a) Fuc/Man-terminated ligands. (b) Gal/GalNAc/NeuAc-terminated ligands. The non-reducing terminal residue most closely associated with EEL is shown with yellow carbons; the residue adjacent to this is shown with pink carbons. Reducing end residues and non-reducing end residues on branches are shown with blue carbons. EEL is shown as a semi-transparent surface, with Trp143 (blue) and Glu130 (red) highlighted.

Table I. Affinities of carbohydrates for EEL determined by fluorescence titration spectroscopy

Ligand Class	**Ligand**	**K_d (μM)**
A-like (GalNAc-terminated)	GalNAcα1-OMe **(1)**	130 ± 20
	GalNAcα1-3Gal **(2)**	120 ± 20
	GalNAcα1-3(Fucα1-2)Gal **(3)**	73 ± 9
B-like (Gal-terminated)	Gal **(4)**	140 ± 10
	Galα1-OMe **(5)**	70 ± 20
	Galα1-3Gal **(6)**	75 ± 6
	Galα1-3Galβ1-OMe **(7)**	50 ± 20
	Galα1-3Galβ1-4GlcNAc **(8)**	50 ± 10
	Galα1-3(Fucα1-2)Gal **(9)**	22 ± 5
H-like (Fuc-terminated)	Fucα1-OMe **(10)**	24 ± 9
	Fucα1-2Gal **(11)**	11 ± 5
	Fucα1-2Galβ1-4GlcNAc **(12)**	9 ± 3
Man-terminated	Manα1-OMe **(13)**	103 ± 6
	Manα1-6Manα1-OOctN_3 **(14)**	90 ± 10
Other ligands	NeuAcα2-OMe **(15)**	60 ± 10
	GlcNAcβ1-4GlcNAcβ1-4GlcNAc **(16)**	60 ± 20
	Galβ1-4Glc **(17)**	280 ± 40
	Xyl **(18)**	400 ± 200
	Glcα1-1Glc **(19)**	600 ± 100

Table II. Potential templates for modeling the binding sites of EEL

Protein	PDB ID	Ligand (bound at sites)	Resolution (Å)	Reference
Rhizoctonia solani agglutinin (RSA)	4G9N	GalNAc (α)	2.20	(Skamnaki et al. 2013)
Marasmius oreades agglutinin (MOA)	3EF2	Galα1-3(Fucα 1-2)Gal (α, β, γ)	1.80	(Grahn et al. 2009)
	2IHO	Galα1-3Galβ1-4 GlcNAc (α, β)	2.41	(Grahn et al. 2007)
Polyporus squamosus lectin (PSL)	3PHZ	NeuAcα2-6Galβ 1-4GlcNAc (β)	1.70	(Kadirvelraj et al. 2011)
Sclerotinia sclerotiorum agglutinin (SSA)	2X2T	Galβ1-3GalNAc (α)	1.97	(Sulzenbacher et al. 2010)
Lumbricus terrestris lectin mutant	2DS0	NeuAcα2-6Galβ 1-4GlcNAc (α)	1.80	(Yabe et al. 2007)
Actinohivin	4G1S	Manα1-2Man (α, β, γ)	1.30	To be published
	4DEN	Manα1-2Man (α, β, γ)	1.60	(Hoque et al. 2012)
	4G1R	Manα1-2Man (α, β, γ)	1.57	To be published
	4END	Manα1-2Man (α, β, γ)	1.90	To be published
Cucumaria echinata lectin III (CEL-III)	2Z48	GalNAc (α, β, γ)	1.70	(Hatakeyama et al. 2007)
	2Z49	Galα1-OMe (α, β, γ)	1.95	(Hatakeyama et al. 2007)
Sambucus nigra agglutinin II (SNA-II)	3CA1	Gal (α, γ)	1.55	(Maveyraud et al. 2009)
	3CA3	GalNAc (α, γ)	1.55	(Maveyraud et al. 2009)
	3CA5	Galα1-OMe (α, γ)	1.55	(Maveyraud et al. 2009)
	3CA6	GalNAcα1-O-Ser (α)	1.40	(Maveyraud et al. 2009)
	3CAH	Fuc (α)	1.55	(Maveyraud et al. 2009)

Table III. Interactions between EEL and fucoside, mannoside and glucoside ligands for initially predicted binding modes[a]

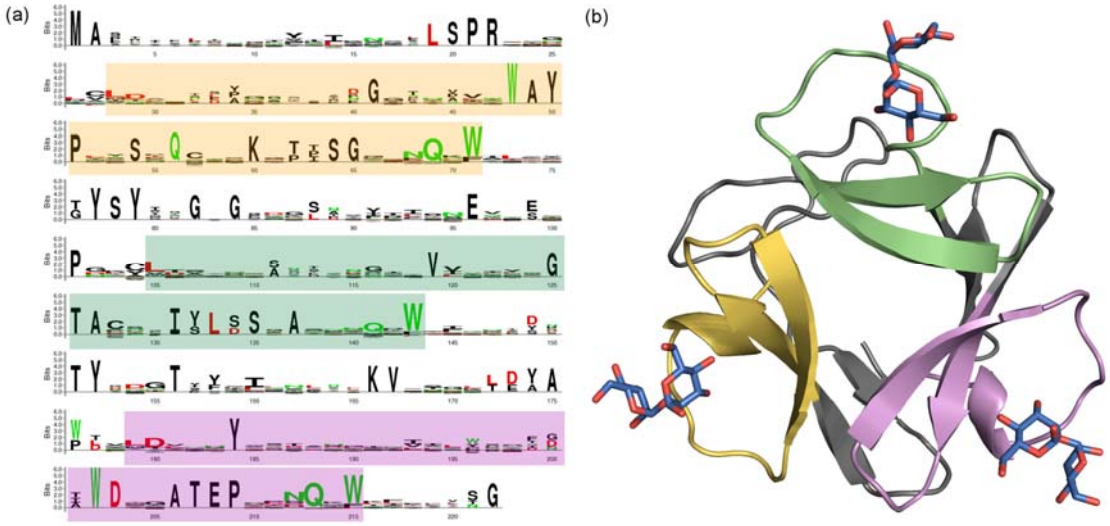
Interaction[b]	**(10)**	**(11)**	**(12)**	**(3)**	**(9)**	**(13)**	**(14)**	**(16)**
Fuc/Man/GlcNAc 1 O2-Asp130 Oδ			X	X	X		X	
GlcNAc 1 C=O-Lys85 NH_3^+								X
Fuc/Man/GlcNAc 1 O3-Asp130 Oδ	X	X		X	X	X		X
Fuc/Man/GlcNAc 1 O3-Gln149 NHε	X	X	X	X	X	X	X	
Fuc/Man/GlcNAc 1 O4-Asp130 Oδ								X
Fuc/Man/GlcNAc 1 O4-Asp145 Oδ	X	X	X			X	X	
Fuc/Man/GlcNAc 1 O4-Tyr147 OH	X	X	X	X	X	X		
Fuc/Man/GlcNAc 1 O4-Gln149 NHε								X
Fuc/Man/GlcNAc 1 O6-Asp145 Oδ								X
Gal 2 O1-Lys85 NH_3^+		X						
Gal 2 O6-Lys85 NH_3^+		X	X	X				
Man 2 O4-Tyr147 OH							X	
Gal 3 O6-Asp130 Oδ					X			
GlcNAc 3 O6-Gly141 NH			X					
GlcNAc 3 O6-Trp143 Hε1			X					
Fuc/Man/GlcNAc 1 C3, C4, C5, C6-Trp143[c]	X	X	X	X	X	X	X	X
Fuc/Man 1 CMe-Ile138 Cδ						X		
GlcNAc 1 CH_3C=O-Ile138 Cδ								X
Gal/GalNAc 3 C6-Ile138 Cδ				X	X			

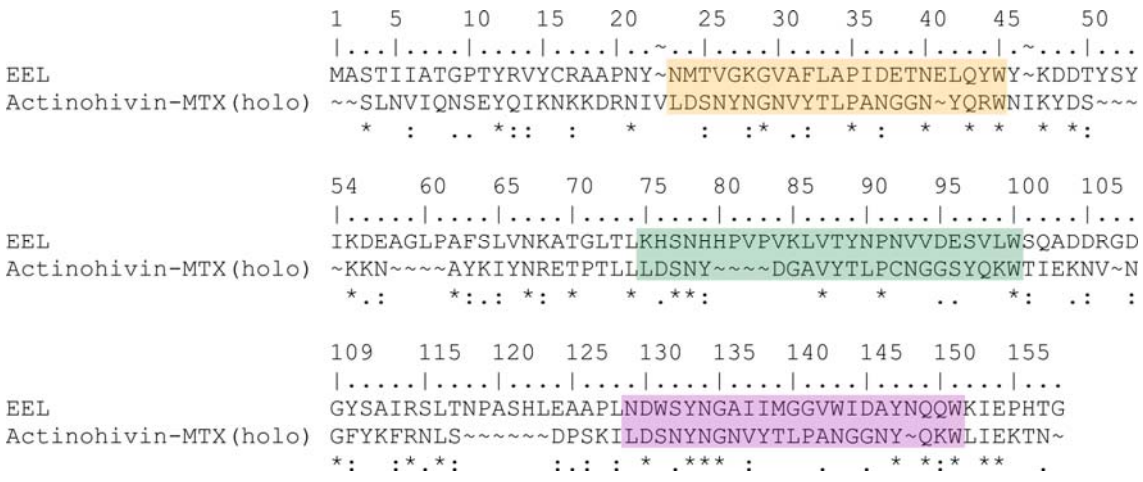
[a]See Table I for full list of compounds with numbers. [b]All interactions shown from ligand to protein. [c]In mannose and *N*-acetylglucosamine, only C4 and C6 are of relevance to this interaction. All atoms in the Trp143 side-chain are considered for this interaction.

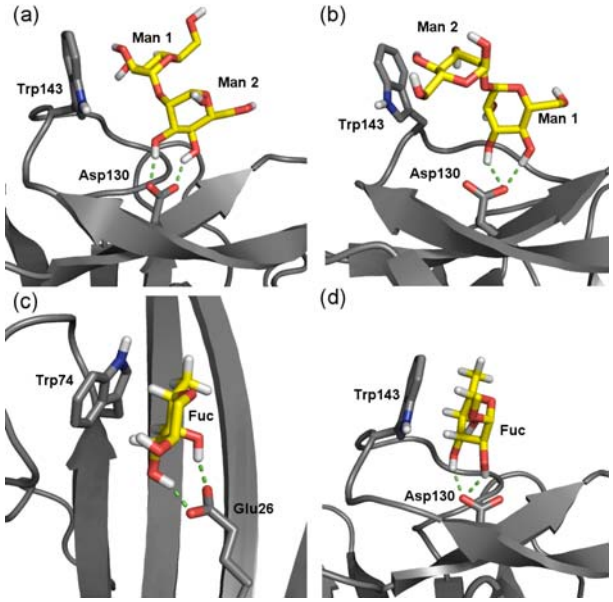
Table IV. Interactions between EEL and galactoside ligands in initial binding modes[a]

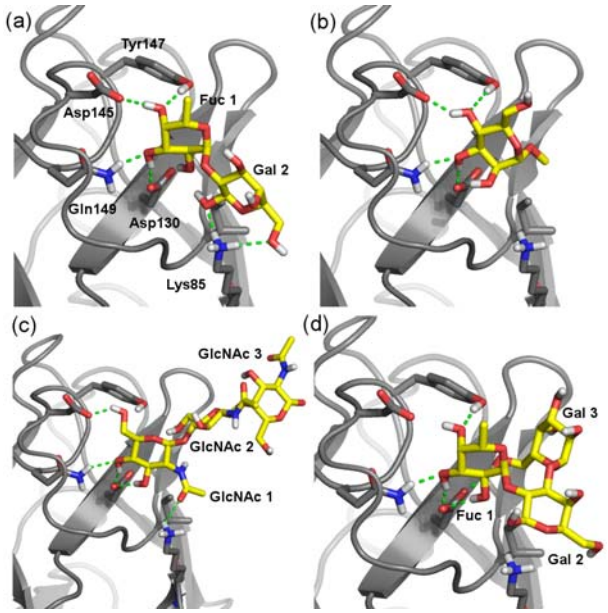
Interaction[b]	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(1)**
Gal/GalNAc 1 O2-Asp130 Oδ				X	X	
Gal/GalNAc 1 O3-Asp130 Oδ	X	X	X	X	X	X
Gal/GalNAc 1 O3-Gln149 NHε	X	X	X	X	X	X
Gal/GalNAc 1 O4-Asp145 Oδ	X	X	X	X	X	X
Gal/GalNAc 1 O6-Asp145 Oδ			X			
Gal/GalNAc 1 O6-Tyr147 Oδ			X		X	
Gal/GalNAc 1 C1, C2-Trp143	X	X	X	X	X	X
GalNAc 1 CH_3C=O-Gly140						X

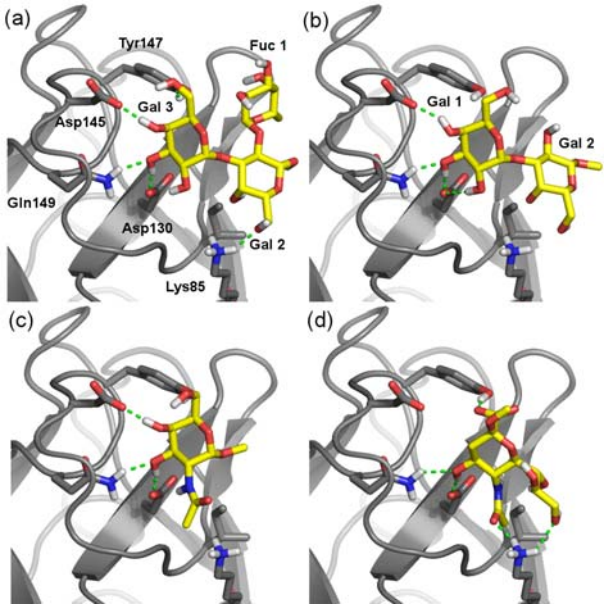
[a]See Table I for full list of compounds with numbers. GalNAcα1-3Gal excluded from this table as no initial binding mode for this ligand could be selected. [b]All interactions shown from ligand to protein. [c]In mannose and *N*-acetylglucosamine, only C4 and C6 are of relevance to this interaction. All atoms in the Trp143 side-chain are considered for this interaction.

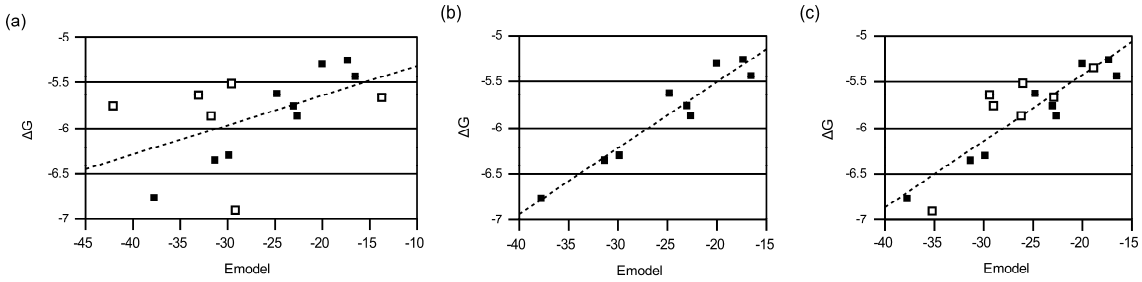












ΔG = 0.032 × Emodel - 4.999
r = 0.522

ΔG = 0.072 × Emodel - 4.072
r = 0.960

ΔG = 0.072 × Emodel - 3.989
r = 0.886