Biophysical characterization and structural determination of the potent cytotoxic Psathyrella asperospora lectin

João P. Ribeiro, Mohamed Ali Abol Hassan, Razina Rouf, Evelin Tiralongo, Tom W. May, Christopher J. Day, Anne Imberty, Joe Tiralongo, Annabelle Varrot

Biophysical characterization and structural determination of the potent cytotoxic *Psathyrella asperospora* lectin

João P. Ribeiro,[1,2] Mohamed Ali Abol Hassan,[3] Razina Rouf,[3] Evelin Tiralongo,[4] Tom W. May,[5] Christopher J. Day,[3] Anne Imberty,[1] Joe Tiralongo,[3,*] and Annabelle Varrot[1,*]

[1] CERMAV, UPR5301, CNRS and Université Grenoble Alpes, 38041 Grenoble, France

[2] DCM, UMR5250, CNRS and Université Grenoble Alpes, 38041 Grenoble, France

[3] Institute for Glycomics, Griffith University, Gold Coast Campus, QLD 4222, Australia

[4] School of Pharmacy and Griffith Health Institute, Griffith University, Gold Coast Campus, QLD 4222, Australia

[5] Royal Botanic Gardens Victoria, Melbourne, Victoria 3004, Australia

ABSTRACT

A lectin with strong cytotoxic effect on human colon cancer HT29 and monkey kidney VERO cells was recently identified from the Australian indigenous mushroom *Psathyrella asperospora* and named PAL. We herein present its biochemical and structural analysis using a multidisciplinary approach. Glycan arrays revealed binding preference towards *N*-acetylglucosamine (GlcNAc) and, to a lesser extent, towards sialic acid (Neu5Ac). Submicromolar and millimolar affinity was measured by surface plasmon resonance for GlcNAc and NeuAc, respectively. The structure of PAL was resolved by X-ray crystallography, elucidating both the protein's amino acid sequence as well as the molecular basis rationalizing its binding specificity.

INTRODUCTION

Lectins are a class of non-immune proteins that recognize specific glycans in a reversible and non-enzymatic fashion. Lectins are known to play key roles in agricultural microecology and in human biology (in both health and disease states), among others.[1] Fungi have attracted considerable interest in the past years since they are a promising source of new lectins with unique specificity and potential applications in the fields of taxonomy, biomedicine and biotechnologies[1,2] Approximately 80 types of mushroom and fungal lectins have been isolated and studied to date.[3] Thus, there is still a considerable limitation on the structural information available for this family of lectins, with only 28 fungal lectins structures being resolved to date (*http://glyco3d.cermav.cnrs.fr*).[1,2]

Psathyrella asperospora is an Australian indigenous species and a member of the family Psathyrellaceae. It contains a functional lectin, PAL, which was shown to have potent anti-proliferative effects against human colon cancer HT29 and monkey kidney VERO cells.[4] This process begins with the PAL binding to terminal *N*-acetyl-D-glucosamine (GlcNAc) present on glycoconjugates expressed at the cell surface. Since terminal GlcNAc is an aberrant epitope characteristic of cancer cells, it is a very interesting antigen to be targeted for diagnosis and therapy.[5,6] Two other GlcNAc-binding lectins from fungi, the *Agrocybe aegerita* lectin II (AAL-II) and *Psathyrella velutina* lectin (PVL), belonging to the same lectin class as PAL, have been characterized by structural and biochemical approaches.[5,7-9] Recently, PVL was demonstrated to be a powerful tool for labeling of O-GlcNAc, a particular protein glycosylation that occurs in the cytoplasm and nucleus,[10] which is of interest in many pathologies.

The amino acid sequence and three-dimensional structure of PAL are still unknown. .Herein we report an in-depth characterization of PAL using a combination of glycan array and surface plasmon resonance (SPR) analysis, with the amino acid sequence and structural basis for its specificity determined from the three-dimensional structure of PAL/GlcNAc complex at 2 Å resolution.

MATERIAL AND METHODS

Mushroom collection and reagents:

The fruiting bodies of *P. asperospora* (Accession no. MEL 2061945) were collected in Melbourne, Australia, identified at Royal Botanic Gardens Victoria and immediately frozen at -20 °C. Unless otherwise stated all the reagents were purchased from Sigma.

Psathyrella asperospora lectin (PAL) purification:

PAL was purified using ammonium sulfate precipitation, chitin-sepharose (New England BioLabs) affinity and size exclusion chromatography (SEC) as previously described (see supplementary information and figure SI-1).[4] All purification steps were carried out at 4 ºC except the final size exclusion chromatography (SEC) step, and PAL purification was monitored using hemagglutination and hemagglutination inhibition assays with rabbit erythrocytes (IMVS Veterinary Services Division Australia) as previously described.[4,11]

Fluorescence labeling of PAL:

Purified PAL (200 µg) was buffer exchanged against PBS pH 7.4 using centrifugal filter device (<10 kDa; Amicon® Ultra Centrifugal Filters) and labeled with 10 µL of Alexa Fluor® 647 succinimidyl ester (Life Technologies). The reaction mixture was wrapped in aluminum foil and incubated for 1 h at room temperature. Subsequently a minimum volume of 10 x TBS, pH 8.5 buffer was added to reconstitute the solution back into TBS buffer and to deactivate any remaining free dye. The integrity of PAL activity following fluorescence labeling was monitored by hemagglutination and hemagglutination inhibition assays.

Glycan Arrays:

Glycan arrays, consisting of 369 diverse glycans with and without the presence of one of the three spacers (sp2, sp3 or sp4), were performed essentially as previously described.[12] Array slides were scanned prior to and following each array experiments using the ProScanArray Microarray 4-laser scanner with the Blue Argon 647 excitation laser (647 nm excitation and 517 nm emission), and subsequent image analysis was carried out using the inbuilt ProScanArray imaging software, ScanArray Express (Perkin Elmer). Raw glycan signals were exported into Microsoft EXCEL, and the mean background was calculated from DMF/DMSO spots on the array plus three standard deviations. This was subtracted from each glycan to generate an adjusted background corrected signal from which a one tailed T-test was performed with significance set at $p = 0.05$, and fold-change over background calculated.

Crystallization, structure solution and refinement:

Solutions of PAL (6.5 mg mL^{-1}) with GlcNAc (2 mM) in PBS where screened in the High Throughput Crystallization laboratory (HTXlab) in Grenoble, France, to find optimal crystallization conditions. The protein/ligand mixture was tested against the JCSG (Qiagen), Wizard I and II (Rigaku reagents), PACT and Classics Suite (Qiagen), Salt and polymer Grids

(Hampton Research) screens. Single plate crystals were observed after two weeks in some of the Salt Grid conditions containing di-Sodium malonate in concentration ranging from 1.0 to 2.9 M, at pH 5, and could be reproduced by hand (supplementary figure SI-2**)**. All the diffraction experiments were performed at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Final data were collected on the ID29 beamline using a Pilatus detector. Diffraction images were integrated using iMosflm[13] and all further processing was done using the CCP4 program suite unless otherwise stated.[14] Five percent of the observations were set aside for cross-validation analysis, and hydrogen atoms were added in their riding positions and used for geometry and structure-factor calculations. The structure was solved by molecular replacement using PHASER[15] and the coordinates of chain A of PVL (PDB ID 2BWR) as search model for one monomer in the asymmetric unit. The model was then submitted to restrained maximum likelihood refinement with REFMAC 5.8[16] iterated with manual rebuilding in Coot[17] Incorporation of GlcNAc molecules was performed after inspection of the 2Fo-DFc weighted electron density map. Water molecules were initially introduced automatically by Coot, and then inspected manually. The quality of the models was assessed using the PDB validation server (*http://wwpdb-validation.wwpdb.org/validservice/*) and coordinates were deposited in the Protein Data Bank with the code 5MB4). The data collection and refinement statistics are listed in Table 1. All figures were drawn with The PyMOL Molecular Graphics System, Schrödinger, LLC.

Surface Plasmon Resonance:
All measurements were performed in HBS-T (Hepes buffer saline pH 7.4 with 0.05% tween 20) at 25˚C using a BIAcore X100 biosensor apparatus (GE Healthcare). Binding was followed by the change resonance units over time after subtraction from the corresponding control. CM5 sensor chips were coated with streptavidin using the amino coupling kit (BIACORE AB, Uppsala, Sweden). Biotinylated GlcNAc-polyacrylamide (PAA), Neu5Ac-PAA, 3'SL-BM (biotinylated monomeric) and 6'SL-BM ligands were immobilized on the sensor chips following the protocol previously described.[18] The reference surface was always in flow cell 1, allowing for the subtraction of bulk effects and non-specific interactions with Fuc-PAA (for the GlcNAc-PAA and Neu5Ac-PAA chips) and streptavidin (for 3'SL-BM and 6'SL-BM). The running buffer was the same HBS-T. Experiments consisted of injecting increasing concentrations of PAL (association 180 s, dissociation 180 s) at a 30 µl min^{-1}. The data was processed and evaluated with Biacore X100 evaluation software, version 2.0.

RESULTS AND DISCUSSION

PAL was purified to homogeneity in three steps: 80% ammonium sulfate precipitation, chitin affinity chromatography and Hiprep Sephacryl S100 SEC. From 200 g of frozen *P. asperospora* fruiting bodies, 28.4 mg of PAL were purified with a purification fold of 9 and a recovery of 13.6% (Supplementary Fig. S1A). The molecular mass of PAL determined by SEC was approximately 36.0 kDa, which correlated well with that observed by SDS-PAGE under reducing conditions (approximately 40 kDa, Supplementary Fig. S1B, lane 6), and that previously determined by DLS (*dynamic light scattering*) analysis (41.8 kDa).[4]
PAL was previously shown to display high binding affinity for mucin (MIC 0.002 mg mL^{-1} in a hemagglutination inhibition assay) and fetuin (MIC 0.0078 mg mL^{-1}), but not for asialofetuin even at 1 mg mL^{-1} concentration.[4] To further explore PAL's carbohydrate specificity, glycan array experiments were performed using Alexa Fluor® 647 labeled PAL. Binding of PAL to glycan structures present on our array was assessed both for statistical significance using T-tests

($p < 0.05$, Fig. 1A, red bars), and by determining fold-change over background. As expected, PAL exhibited significant binding ($p<0.05$, >6 fold-change over background) to glycans containing terminal non-reducing GlcNAc, including terminal GlcNAcβ1-4 such as chitobiose, terminal GlcNAcβ1-3Gal and GlcNAcβ1-3GalNAc, as well as binding (>6 fold-change over background) GlcNAcβ1-2Man (Glycan ID 5A and 5B). The ability of PAL to bind to sialylated glycans with terminal *N*-acetyl-neuraminic acid (Neu5Ac), was only partially confirmed by glycan array analysis with significant binding ($p < 0.05$, > 6-fold change over background) only observed for 3'-sialyl-*N*-acetyllactosamine (3'SLN, Neu5Acα2-3Galβ1-4GlcNAc, glycan ID 298). Nonetheless a number of additional sialoglycoconjugates were also bound by PAL (> 6 fold-change over background) but with $p > 0.05$ (Fig.1A, blue bars). PAL was also found to bind a range of sulfated glycans including 3/4/6-O-Su-Galβ1-3/4/6GlcNAc and 3/4/6-O-Su-GalNAcβ1-4GlcNAc (see Supplementary Table S1 for full list of glycan structures present on the array and access to PVL and AAL-2 data).

PAL specificity towards terminal Neu5Ac and GlcNAc glycans was quantified by surface plasmon resonance (SPR, Fig. 1B-D). Among the tested ligands – GlcNAc, Neu5Ac, 3'SL (NeuAcα2-3Galβ1-4Glc) and 6'SL (NeuAcα2-6Galβ1-4Glc) – the highest affinity was observed for GlcNAc, with a K_d of 0.45 μM. The sialosides show a weaker binding to PAL, with dissociation values of 2.25, 5.16 and 11.93 μM for Neu5Ac, 6'SL and 3'SL, respectively. The SPR experiments with 3'- and 6'SL did not require regeneration with free Neu5Ac, while the experiment with immobilized Neu5Ac did (injection of 100 mM for 60 seconds at a 30 μl min^{-1} flow). These values are comparable to the those recently obtained for PVL (Supplementary table S2).[10] The binding properties of PAL are thus closely related to those of PVL, with PAL having similar specificity but 6 to 7-fold lower binding affinities for surface with multivalent presentation of the ligand. For AAL-2, binding affinities were only measured for monovalent presentation of GlcNAc with immobilization of the protein and not the ligand on the SPR chip and by isothermal calorimetry and are similar to those reported for PVL in the range 34-200 μM (Supplementary table S2).[5,8,9]. It seems that PAL, PVL and AAL-2 present binding affinities in the same order of magnitude and should present similar modes of binding for GlcNAc.

Since the resolution of PAL's three-dimensional structure would allow determination of the amino acid sequence of the protein - unknown to this point - crystallization experiments with PAL in the presence of GlcNAc were performed. The lectin and ligand co-crystallized in rhombohedral space group with hexagonal axis settings R32:h alternatively symbolized H32 with unit cell parameters a = 120.55 Å, b = 120.55 Å and c = 198.95. The final model was refined to R and R_{free} values of 16.9 and 21.6 respectively for all data in the resolution range of 30.91 to 2.00 Å (Table 1). The asymmetric unit encompasses one PAL monomer, six sodium ions, one malonate and 354 water molecules. In the carbohydrate-binding sites, six GlcNAc ligands were assigned to well-defined electronic density. As expected, PAL presents a 7 bladed β-propeller fold where each blade is composed of four β-strands (strands A to D) (Fig. 2). The N-terminus is tucked in the central cavity of the barrel and the barrel is closed via a "Velcro" interaction, a common feature in proteins with this fold as observed for PVL.[19]

The lectin presents 6 sugar binding sites, all located on one face of the β-propeller. This display grants PAL with an avidity effect towards surface presented antigens, a mechanism of molecular recognition often seen in lectins.[20] The binding sites of PAL are located at the interface between

adjacent blades but the pocket between blades 1 and 7 is not occupied as seen for PVL.[5,7] The bound GlcNAc residues adopt the same orientation and relative position in all binding sites. Both α- and β-GlcNAc configurations are seen in sites 2-3 and 3-4. The loop bridging the strands D and A of consecutive β-blades makes the most important interactions with the sugar, in combination with the loop connecting strands B and C. All binding sites present an aromatic amino acid (Trp, Tyr or Phe) involved in stacking interaction with the GlcNAc ring, with the exception of site 2-3, where a Leu residue is found. Two H-bonds are formed between the O3 and O4 hydroxyls of the GlcNAc and the side chain of one asparagine (or aspartic acid). The indole nitrogen of a tryptophan residue located inside the binding site is H bonded to the O3 hydroxyl of the GlcNAc. The nitrogen of the *N*-acetyl group interacts with the main chain O of a non-conserved residue (either a Glu, Gly, Gln or a Ser) while its donor carboxyl oxygen interacts with the main chain nitrogen atom of an Asn or an Ala residue. The neighboring amino acid may also form a water mediated H-bond with the carboxyl group of GlcNAc. In sites 4-5 and 5-6, the O6 hydroxyl of GlcNAc also makes a direct or water-mediated H-bond, respectively, to the nearby Tyr residue. With the exception of these two binding domains, the O6 hydroxyl is always facing towards the solvent. A detailed representation of the interactions in the different binding site can be found in supporting information (supplementary Fig. S3).

The PAL amino acid sequence was unknown but could be deduced from the crystal structure due to the good quality of the electron density for side chains. Assignment of Asp/Asn, Glu/Gln and Val/Thr were based on H-bonding contacts, B factors and sequence homology. Uncertainties remain for about ten residues leading to a degree of confidence of 96%. While the PAL amino acid sequence was initially believed to be considerably different from any other known lectins, the present sequence displays 87% identity with PVL and 60% with AAL-2 (Supplementary Fig. S4). Errors could therefore be corrected in the previously described tryptic peptides obtained from *de novo* sequencing by LC-MS/MS,[4] which revealed an accuracy of only 54%. When searching for other sequences related to PAL, proteins displaying clear tandem repeats and conservation of amino acids involved in glycan binding sites could be identified in the genomes of *Galerina marginata*, *Hebeloma cylindrosporum* and *Laccaria amethystina* (Supplementary Fig. S4).

PAL structure shows higher structural similarity with PVL than AAL-2, with a root-mean-square deviation of 0.33 Å and 0.62 Å respectively, over 401 amino acid residues. AAL-2 has, in fact, several surface loops with different orientation as PAL and PVL. PAL binding sites are almost identical to those of PVL apart for binding sites 2-3 and 3-4 where the residues establishing hydrophobic contacts with the GlcNAc are different, but peculiarly match those of AAL-2: leucine and phenylalanine, respectively (Leu140 and Phe195) are present in place of tyrosines (Fig. 2B-C, and Supplementary Fig. S4). Fewer interactions of PAL with the GlcNAc moiety are therefore made with either the loss of the H bond with the O6 hydroxyl or of the stacking interaction with the pyranose ring. This could explain the lower affinity measured for PAL compared to PVL by SPR. Other variations are observed but involve amino acids interacting via main chain atoms whose substitution has thus no effect on ligand binding.

CONCLUSIONS

Since only 10% of mushroom species have been catalogued and taxonomically classified, fungal lectinology is a field that remains largely unexplored. In this study we present the structural characterization of the *Psathyrella asperospora* lectin, a GlcNAc-binding lectin that displays strong cancer specific cytotoxic properties.[4] Glycan array analysis of PAL showed specificity towards GlcNAc and to a lesser extent Neu5Ac-terminal epitopes. SPR experiments allowed the

quantification of the strength of these interactions with an 11-fold higher binding affinity for GlcNAc over Neu5Ac. The three-dimensional structure of PAL in complex with GlcNAc was determined at 2.0 Å resolution which allowed to determine the amino acid sequence of the protein, unknown until now, with high level confidence. With the increasing need of recombinant lectin, PAL, as well as the related PVL, may be of significant use in basic research, such as in glycoconjugate purification, or in clinical diagnosis, particularly in the early detection, diagnosis and staging of cancer.

ACKNOWLEDGMENTS

AI and AV are CNRS employees. Financial support from Labex ARCANE (ANR-11-LABX-003), Glyco@Alps and from the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement N°283570) is gratefully acknowledged. The authors thank the EMBL-Grenoble staff for assistance in using the High Throughput Crystallisation (HTX) facility and the ESRF, Grenoble, France for the provision of synchrotron beamtime.

REFERENCES

1. Hassan MA, Rouf R, Tiralongo E, May TW, Tiralongo J. Mushroom lectins: specificity, structure and bioactivity relevant to human disease. Int J Mol Sci 2015;16:7802-7838.
2. Varrot A, Basheer SM, Imberty A. Fungal lectins: structure, function and potential applications. Curr Opin Struct Biol 2013;23:678-685.
3. Kobayashi Y, Kawagishi H. Fungal lectins: a growing family. Methods Mol Biol 2014;1200:15-38.
4. Rouf R, Stephens AS, Spaan L, Arndt NX, Day CJ, May TW, Tiralongo E, Tiralongo J. G(2)/M cell cycle arrest by an N-acetyl-D-glucosamine specific lectin from *Psathyrella asperospora*. Glycoconj J 2014;31:61-70.
5. Audfray A, Beldjoudi M, Breiman A, Hurbin A, Boos I, Unverzagt C, Bouras M, Lantuejoul S, Coll JL, Varrot A, Le Pendu J, Busser B, Imberty A. A recombinant fungal lectin for labeling truncated glycans on human cancer cells. PLoS One 2015;10:e0128190.
6. Singh JP, Zhang K, Wu J, Yang X. O-GlcNAc signaling in cancer metabolism and epigenetics. Cancer Lett 2015;356:244-250.
7. Cioci G, Mitchell EP, Chazalet V, Debray H, Oscarson S, Lahmann M, Gautier C, Breton C, Perez S, Imberty A. Beta-propeller crystal structure of *Psathyrella velutina* lectin: an integrin-like fungal protein interacting with monosaccharides and calcium. J Mol Biol 2006;357:1575-1591.
8. Jiang S, Chen Y, Wang M, Yin Y, Pan Y, Gu B, Yu G, Li Y, Wong BH, Liang Y, Sun H. A novel lectin from *Agrocybe aegerita* shows high binding selectivity for terminal N-acetylglucosamine. Biochem J 2012;443:369-378.
9. Ren XM, Li DF, Jiang S, Lan XQ, Hu Y, Sun H, Wang DC. Structural Basis of Specific Recognition of Non-Reducing Terminal N-Acetylglucosamine by an *Agrocybe aegerita* Lectin. PLoS One 2015;10:e0129608.

10. Machon O, Baldini SF, Ribeiro JP, Steenackers A, Varrot A, Lefebvre T, Imberty A. Recombinant fungal lectin as a new tool to investigate O-GlcNAcylation processes. Glycobiology 2017;27:123-128.
11. Rouf R, Tiralongo E, Krahl A, Maes K, Spaan L, Wolf S, May TW, Tiralongo J. Comparative study of hemagglutination and lectin activity in Australian medicinal mushrooms (higher Basidiomycetes). Int J Med Mushrooms 2011;13:493-504.
12. Waespy M, Gbem TT, Elenschneider L, Jeck AP, Day CJ, Hartley-Tassell L, Bovin N, Tiralongo J, Haselhorst T, Kelm S. Carbohydrate Recognition Specificity of Trans-sialidase Lectin Domain from *Trypanosoma congolense*. PLoS Negl Trop Dis 2015;9:e0004120.
13. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr 2011;67:271-281.
14. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 2011;67:235-242.
15. McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr 2007;63:32-41.
16. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 2011;67:355-367.
17. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallographica Section D 2010;66:486-501.
18. Ribeiro JP, Pau W, Pifferi C, Renaudet O, Varrot A, Mahal LK, Imberty A. Characterization of a high-affinity sialic acid-specific CBM40 from *Clostridium perfringens* and engineering of a divalent form. Biochem J 2016;473:2109-2118.
19. Chen CK, Chan NL, Wang AH. The many blades of the beta-propeller proteins: conserved but versatile. Trends Biochem Sci 2011;36:553-561.
20. Arnaud J, Audfray A, Imberty A. Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. Chem Soc Rev 2013;42:4798-4813.

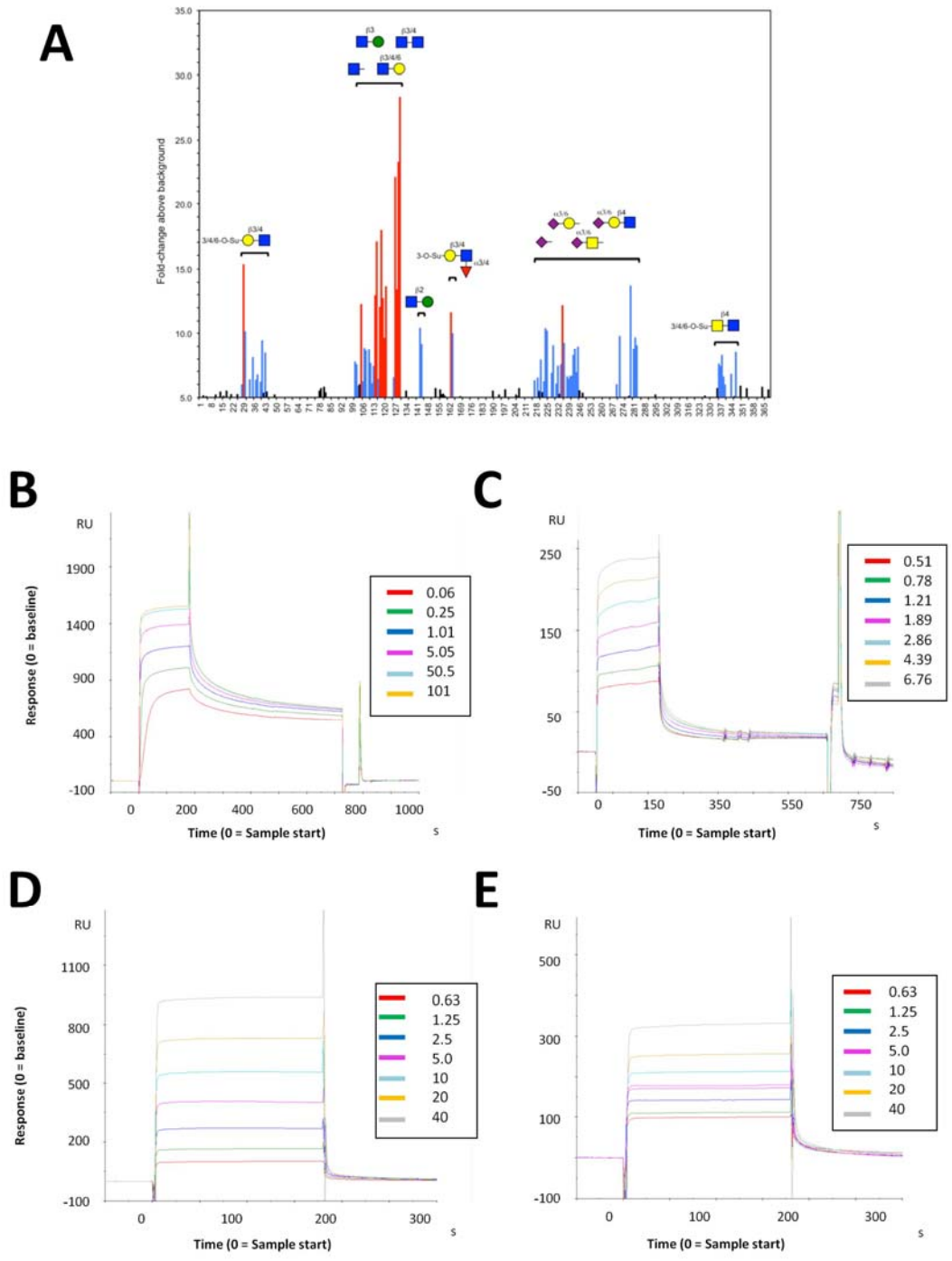

Figure 1 – *On top:* **A)** Glycan array analysis of PAL. Binding of PAL to glycan structures present on the array was assessed both for statistical significance using T-tests ($p<0.05$, red bars), and by determining fold-change over background. Blue bars indicated >6-fold change over background but with $p>0.05$; *Below:* SPR sensorgrams representing binding of PAL to **B)** GlcNAc; **C)** Neu5Ac; **D)** 6'-sialyllactose and **E)** 3'-sialylactose, at 25 °C and with a flow rate of 30 µl min^{-1}. Protein concentration values (inside box) are given in micromolar.

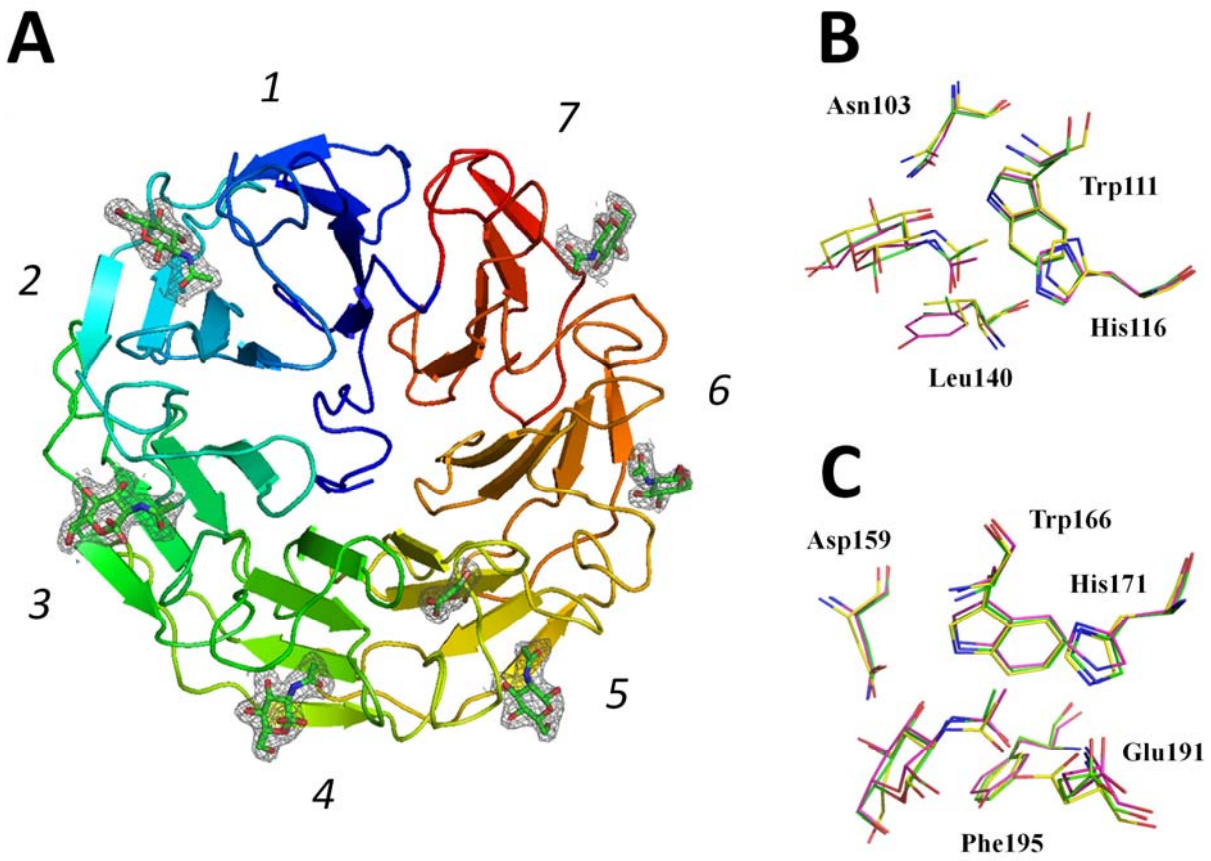

Figure 2 – A) Crystal structure of PAL complexed with GlcNAc (PDB ID 5MB4). The different blades are shown in a different color and identified with a corresponding number. 2Fobs–Fcalc electron density displayed at 1 sigma (0.26 eA^3) around the GlcNAc moieties as well as the malonate ion co-crystalized with the complex. **B, C)** Superposition of binding sites 2-3 and 3-4 of PAL (green), PVL (magenta) and AAL-II (yellow), respectively.