Characterization of a high-affinity sialic acid-specific CBM40 from Clostridium perfringens and engineering of a divalent form

João Ribeiro, William Pau, Carlo Pifferi, Olivier Renaudet, Annabelle Varrot, Lara Mahal, Anne Imberty

Characterization of a high-affinity sialic acid specific CBM40 from *Clostridium perfringens* and engineering of divalent form

João P. Ribeiro[1,2], William Pau[1], Carlo Pifferi[3], Olivier Renaudet[3], Annabelle Varrot[2], Lara K. Mahal[1*] and Anne Imberty[2*]

[1] Biomedical Chemistry Institute, New York University Department of Chemistry, 100 Washington Square East, Room 1001, New York, NY 10003, USA

[2] CERMAV, UPR5301, CNRS and Université Grenoble Alpes, 601 rue de la Chimie, BP 53, 38041, Grenoble, France

[3] Université Grenoble Alpes, DCM, BP 53, 38041 GRENOBLE cedex 9, France

* Corresponding authors : Lara K. Mahal, Biomedical Chemistry Institute, New York University Department of Chemistry, 100 Washington Square East, Room 1001, New York, NY 10003, USA, Tel: +1 212 998 3533; E-mail: lkmahal@nyu.edu, and Anne Imberty, CERMAV, UPR5301, CNRS and Université Grenoble Alpes, 601 rue de la Chimie, BP 53, 38041, Grenoble, France, Tel: +33 (0)47 603 7636; E-mail: imberty@cermav.cnrs.fr

ABSTRACT

Carbohydrate-binding modules (CBMs) are a class of polypeptides usually associated with carbohydrate-active enzymatic sites. We have characterized a new member of the family CBM40, coded from a section of the gene NanI from *Clostridium perfringens*. Glycan arrays revealed its preference towards α(2,3)-linked sialosides, that was confirmed and quantified by calorimetric studies. The CBM40 binds to α(2,3)-sialyllactose with a K_d ~ 30 µM, a leading affinity value for this class of proteins. Inspired by lectins structure and their arrangement as multimeric proteins, we have engineered a dimeric form of the CBM, and using SPR we have observed 6-11 fold binding increases due to the avidity affect. The structures of the CBM, resolved by X-ray crystallography, in complex with α(2,3)- or α(2,6)-sialyllactose explain its binding specificity and unusually strong binding.

Keywords: Carbohydrate binding domain, *Clostridium perfringens*, sialic acid, protein-glycan interactions, x-ray structure.

INTRODUCTION

Sialic acids are the most abundant terminal carbohydrate residues present on glycoconjugates at the surface of eukaryotic cells [1, 2]. The *N*-acetylneuraminic acid (Neu5Ac) monosaccharide is the most common form of sialic acid on mammalian cells. Sialylation is required for the structural stabilization of cells [3] and is especially important in early mammalian development [4, 5]. In cancer, altered surface sialylation is a common hallmark of the malignant phenotype of tumor cells [6]. The analysis of total serum and lipid bound Neu5Ac is valuable in cancer staging, prognosis and progression, and elevated levels have been associated with cancer development [7, 8]. Increase of the negative charges associated with Neu5Ac also contributes to the cell mobility by helping dispersion from the main tumor, influencing the metastatic potential of those cells [9].

Due to the importance of Neu5Ac, specific receptors for this epitope are of interest as biomarkers. Currently, only a handful of Neu5Ac binders are commercially available [10]. This includes proteins purified from plants, as the lectins from wheat germ (WGA), *Maackia amurensis* (MAA) and *Sambucus nigra* (SNA-I) [11, 12], from the mushroom *Polyporus squamosus* (PSL-I) [13, 14] or from invertebrates such as *Limax flavus* (LFA) [15]. While SNA-I and PSL-I are specific for α2,6-linked Neu5Ac (namely terminal 6'-sialyllactosamine, 6'-SLN), the other lectins have a broader range of interactions, for example displaying cross reactivities with N-acetyl-glucosamine – GlcNAc – (WGA, LFA) or with 3'-sulfolactose (MAA). As a consequence, the analysis of complex samples requires the use of several of these lectins. Additionally, these commercial lectins come with restrictions that arise from seasonal depletion of the natural sources, undesired protein glycosylation or batch inconsistency. Recombinant proteins have proven to be a valid and effective alternative, amending for the aforementioned complications [16].

Sialidases, a group of enzymes that cleave sialosides, are secreted by microbial organisms and mammalian cells. These include gut bacteria, either commensals or pathogens [17, 18], viruses such as influenza [19]**,** and human cells [20]. The catalytic domain of these enzymes is often linked to one or more carbohydrate binding modules (CBMs) that reversibly binds the glycan substrate to increase its local concentration, consequently improving the catalytic efficiency of the enzyme [21]. CBMs are small sized, monovalent domains that can generally be easily expressed in recombinant systems [22] and may provide an interesting alternative to lectins. Developed by pathogens for binding and invading human tissues, these receptors could provide specific Neu5Ac binders. In addition, CBMs can be easily produced, and their low affinity binding can be compensated through avidity by fusing multiple CBMs together [23].

CBMs are organized into 73 families based upon their amino acid sequence similarity (in *http://www.cazy.org*) [24]. Most interact with plant cell wall glycans, but five CBM families have been categorized as mammalian glycan binding domains: CBM32, CBM40, CBM47, CBM51 and CBM57. Within these, the members of CBM40 are known to bind to Neu5Ac, what was first verified with the CBM located at the *N*-terminus of GH33 sialidase from *Vibrio cholerae* [25]. Presently, there are only six structurally characterized CBM40s which are associated with the sialidases from *Clostridium perfringens (Cp)* NanJ [26], *V. cholerae* (*Vc*) NanH [23], *Macrobdella decora* (*Md*) NanL [27] and *Streptococcus pneumoniae* (*Sp*) NanA [28], NanB [29] and NanC [30]. They all share a conserved β-sandwich fold consisting of two antiparallel β-sheets, with the exception of *Vc* NanH´s that shows a different architecture, which led researchers to suggest its integration into a new CBM40 subfamily [26].

We herein report the successful characterization of the CBM40 from *C. perfringens* (ATCC 13124) which is associated with NanI [31], for which only the catalytic domain has been structurally characterized [32]. The recombinant CBM40_NanI demonstrates unusual micromolar affinity towards sialylated oligosaccharides. The structural analysis of the complexes with 3'-sialyllactose (3'-SL) and 6'-sialyllactose (6'-SL) provides insights into the molecular details responsible for this affinity. A pseudo-dimeric version of the CBM40 was also engineered as to enhance the affinity from multiple interactions between the CBM and glycans. This work presents a new probe towards sialoconjugates, expanding the repertoire of sialoside binders.

EXPERIMENTAL

Materials

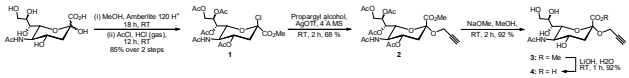
Genomic DNA of *C. perfringens* ATCC® 13124™ was bought from ATCC (#13124D-5™). 3'-Sialyllactose sodium salt (CAS #128596-80-5) and 6'-sialyllactose (#35890-38-1) were purchased from Carbosynth. 3'-sialyl-N-acetyllactosamine (#81693-22-3) and 6'-sialyl-N-acetyllactosamine (#174757-71-2) were bought from Dextra UK and sialic acid (#131-48-6) from AppliChem. Biotinylated 3'-sialyllactose (#0060-BM) and biotinylated 6'-sialyllactose (#0063-BM) were obtained from Lectinity.

Propargyl sialic acid synthesis

All chemical reagents were purchased from Aldrich (Saint Quentin Fallavier, France) or Acros (Noisy-Le-Grand, France). Moisture-sensitive reactions were performed under argon atmosphere by using oven-dried glassware and reactions were monitored by TLC using silica gel 60 F254 pre-coated plates (Merck). Spots were inspected by UV light and visualised by charring with 10% H_2SO_4 in EtOH. Silica gel 60 (0.063-0.2 mm or 70-230 mesh, Merck) was used for column chromatography. ^1H and ^{13}C NMR spectra were recorded on Bruker Avance 400 MHz or Bruker Avance III 500 MHz spectrometers and chemical shifts (δ) were reported in parts per million (ppm). Spectra were referenced to the residual proton solvent peaks relative to the signal of $CDCl_3$ (δ 7.27 and 77.0 ppm for ^1H and ^{13}C) and D_2O (4.79 ppm for ^1H), assignments were done by gCOSY and gHMQC experiments.

A route for the synthesis of the propargyl sialic acid (2-propynyl 5-acetamido-3,5-dideoxy-D-α-D-galacto-2-nonulopyranosidonic acid) starting from *N*-acetylneuraminic acid is shown in scheme 1. All the reaction conditions, mass and NMR spectra and their assignments are in supplementary information.

Scheme 1



Gene cloning

The following oligonucleotide primers were used: 5'-GAC GAC GGA TCC *CAT G*TT AAG TTC ACT AGG AGA ATA TAA GGA TAT-3' (45-mer) and 5'-TCC CAT CTC GAG *TTA* TTT TGT CTC TCC AGT CTT ACT AAG TAA ATA ATC ATC-3' (51-mer) for the construction of pNanI. They were designed as to include BamHI and XhoI restriction sites (underlined sequences) in the CBM40_NanI gene sequence. PCRs were done using the Q5 Hot Start High Fidelity polymerase (New England Biolabs) and genomic DNA from *Clostridium perfringens* strain ATCC® 13124™ as a template. The same cloning protocol was followed for pNanJ, using the following oligonucleotide primers 5'-AC TAC GGA TCC AGG GTA AAT ATA ACA GGT GAT T -3' (33-mer) and 5'-G CAT CTC GAG TTT AGT TTC TCC TGT TTT TCT TA-3' (36-mer) which contain SacI and XhoI restrictions sites respectively. To engineer a CBM40_NanI with a multivalent presentation, we

changed the 5'- and 3'-termini of the gene sequences to integrate 36 additional codons (representing a 12 amino acid long linker that will be present between two CBM copies). The PCRs were done using the oligonucleotides 1-For: 5'-GAC GAC GGA TCC *CAT GT*T AAG TCC ACT AGG ATA TAA GGA TAT-3' (45-mer, BamHI); 1-Rev: 5'-G CAC GAG CTC *GCT GCC GTT CAG CGC* TTT TGT CTC TCC AGT-3' (40-mer, SacI); 2-For: 5'-C GTC GAG CTC *GGC AGC GGC AGC GGC* TTA AGT TCA CTA GGA-3' (40-mer, SacI), 5'-TCC CAT CTC GAG *TTA* TTT TGT CTC TCC AGT CTT ACT AAG TAA ATA ATC ATC-3' (51-mer, XhoI). After the two sequences were made, the digestion by the same endonuclease allowed the ligation of the two CBM40_NanI copies in tandem. All the generated inserts were subjected to endonuclease treatment and introduced in the multiple cloning region of the pET45b(+) vector, yielding to pNanI, pNanJ and p2NanI plasmids that encode for a copy of *Cp* CBM40_NanI, *Cp* CBM40_NanJ or *Cp* di-CBM40_NanI, respectively, fused to a polyhistidine tag at the *N*-terminus. The constructs were verified by DNA sequencing (Genewiz, US or Eurofins, EU). The DNA of the positive clones was amplified in *Escherichia coli* XL1 blue cells and then used to transform *E. coli* BL21(DE3) cells for protein production.

Protein Expression and Purification

E. coli BL21(DE3) cells transformed with plasmids pNanI, pNanJ or p2NanI were cultured in TB medium at 37 °C. When reaching an OD_{600} of 0.5 - 0.7, the cultures were moved to a 16 °C incubator and IPTG was added to a final concentration of 1mM. After overnight incubation, cells were pelleted down, washed and resuspended in cold PBS pH 7.4. The cells were lysed by sonication on ice three times at 20% intensity with 0.5 cycle for 2 minutes or with a Cell Disruption System (Constant Systems Ltd) at 1900 bars. After 30 min centrifugation at 4 °C and 50000 g, the supernatants were submitted to nickel affinity chromatography. After being allowed to bind the matrix (Ni-NTA agarose, Qiagen), the proteins were washed (50 mM $NaH_2PO_4.H_2O$ pH 8.0, 300 mM NaCl and 10 mM imidazole) and eluted (using the same buffer with an increased 250 mM imidazole concentration). The eluted fractions were filtered, checked for absorbance at 280 nm and concentrated by centrifugation using Vivaspin (Startorius) with a 10 kDa cut-off filter, before being further purified by size-exclusion chromatography. A Superdex 75 prep grade column was equilibrated with PBS pH 7.4 and used for the estimation of aggregate states and molecular weights of the purified proteins. A calibration curve for molecular size estimation - done with cytochrome c, myoglobin, ovalbumin and BSA, previously to the experiments – confirmed the expected molecular weight values. Purified proteins were run in denaturing protein gels in a 12% (w/v) acrylamide matrix in alkaline buffer conditions (Tris.HCl, pH 8.8).

Protein estimation

Protein concentrations were determined with a Nanodrop spectrometer ND2000 (Ozyme) with absorbance readings at 280 nm wavelengths, using molar extinction coefficient values of 11.92, 21.89, and 32.78 ($x10^3$ M^{-1} cm^{-1}) for *Cp* CBM40_NanJ, *Cp* CBM40_NanI and *Cp* di-CBM40_NanI, respectively, as well as their corresponding molecular weights of 23.69, 24.62 and 46.55 kDa (values for calculations that include the *N*-terminal tagged amino acids).

Thermostability

To examine their stability, protein solutions of *Cp* CBM40_NanI, *Cp* CBM40_NanJ and *Cp* di-CBM40_NanI were assayed using a Thermal Shift Assay (TSA), adapting a protocol previously described [33]. Samples were diluted to a final protein concentration of 0.1 mg mL^{-1} in the chosen buffer and 5× SYPRO Orange (Molecular Probes, CA) was added to a final volume of 25 µl. The samples were subjected to thermal denaturation

using a Real Time PCR machine (Mini Opticon, Bio-Rad Laboratories) with a temperature gradient starting at 25 °C and ascending to 100 °C at a heating rate of 1 °C/min. Protein unfolding was followed by the increase of the fluorescence values given by the SYPRO Orange probe. Different buffer effects were analyzed in a pH range of 3-10.

Crystallization and data collection

Solutions of CBM40_NanI (5.5 mg mL^{-1}) in PBS where sent to the High Throughput Crystallization laboratory (HTXlab) in Grenoble, France, to screen for crystallization conditions. The protein/ligand mixture was tested against the JCSG (Qiagen), Wizard I and II (Rigaku reagents), PACT (Qiagen), plate 1 (Qiagen), plates 4 and 5 (Hampton) screens. Hits were obtained with the use of the Clear Strategy™ Screens, (Molecular Dimensions Ltd) at pH 4.5. Crystals of CBM40_NanI with 3'-SL appeared in the form of plates after several days in CSSI-12 containing 0.1M sodium acetate pH 4.5, 0.2M calcium acetate and 8% of both polyethylene glycol (PEG) 550 monomethyl ether (MME) and PEG 20000. In the same time period, parallelepiped crystals of CBM40_NanI in complex with 6'-SL grew in 0.1M sodium acetate pH 4.5, 0.3M sodium acetate and 25% PEG 2000 MME. Afterwards, larger crystals were then grown with homemade solutions. All the diffraction experiment data sets were collected at the European Synchrotron Radiation Facility (ESRF). Final data for CBM/3'-SL was collected to 1.9 Å on BM30A beamline using an ADSC Q315r detector whilst the CBM/6'-SL complex was obtained at 2.0 Å on the ID29 beamline using a Pilatus detector. Diffraction images were integrated using program XDS [34, 35] and all further processing was done using the CCP4 program suite [36].

Structure solution and refinement

Coordinates of molecule A of the CBM40 from *C. perfringens* ATCC® 13124™ NanJ (PDB ID 2V73) were used as model for molecular replacement with PHASER [37] in a search for six monomers per asymmetric units for the 6'-SL complex. The coordinates of molecule A of the CBM40/3'-SL were used to search for the three copies of the CBM40/6'-SL complex. 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 initial model was optimized using ARP/wARP [38] prior restrained maximum likelihood refinement with REFMAC 5.8 [39] iterated with manual rebuilding in Coot [40]. Incorporation of the ligand was performed after inspection of the 2Fo-DFc weighted maps. Water molecules, introduced automatically using Coot, were 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 under codes 5FRE and 5FRA.

Hemagglutination and minimal concentration inhibition assays

The hemagglutination activity for the *Cp* CBM40_NanI and *Cp* di-CBM40_NanI was determined using 2-fold serial dilutions of the protein with rabbit erythrocytes (Biomérieux). In those tests, 25 µL of the serial dilutions of the protein were incubated with 25 µL of 2% rabbit erythrocytes suspensions in 100 mM NaCl for 30 min at 37 °C plus 30 min at room temperature. After identifying the highest dilution of the protein that exhibited hemagglutination, a concentration 4 times higher was used in the subsequent minimal inhibition concentration (MIC) experiments. Serially dilutions of 12.5 µL of different saccharide solutions were added to the 12.5 µL solutions of protein of identical concentrations. These 25 µL mixtures were incubated with 25 µL of 2% rabbit erythrocytes suspensions, following the same development steps as above.

Glycan array

Cp CBM40_NanI and *Cp* di-CBM40_NanI were labeled with Alexa Fluor 488 (Invitrogen) using manufacturer's protocols and sent to the Consortium of Functional Glycomics (CFG). Binding specificity was determined using the printed mammalian glycan array v5.2, which contains 609 glycans [41]. The scanner response is linear to a maximum of ~50,000 RFU. Both proteins were incubated with the arrays at 200 μg mL^{-1}. The extracted data was then transformed to give Z scores for each glycan on the array (Z= x-mean/st dev). A threshold of Z=1.645 ($p \leq 0.05$) was used to establish significance.

Isothermal Titration Calorimetry

ITC experiments were performed using a Microcal ITC200 microcalorimeter (Malvern Instruments Ltd). All titrations were carried out in PBS pH 7.4, at 25 °C. Aliquots of 10 μL of each carbohydrate, dissolved in the same buffer were added to the protein solution present in the calorimeter cell, in 2 min intervals. For the different experiments, protein concentrations were prepared in solution with concentrations that varied from 100 to 415 μM, and the carbohydrates samples ranged from 2.1 to 6.0 mM. Data was fitted with MicroCal Origin 7 software, following standard procedures. This yielded the stoichiometry (*N*), the association constant (K_a) and enthalpy of binding (ΔH). Other thermodynamic parameters (changes in free energy, ΔG, and entropy, ΔS, were calculated from the equation $\Delta G = \Delta H - T\Delta S = -RTln\ K_a$, being R = 8.314 J/mol/K.

Surface Plasmon Resonance

SPR experiments were performed using a Biacore X100 biosensor instrument (GE Healthcare) at 25°C. Biotinylated 3'-SL and 6'-SL were immobilized on CM5 chips (GE Healthcare) that were previously coated with streptavidin, following the same protocol as previously described [42]. Each monovalent biotinylated sugar, 3'-SL or 6'-SL, was diluted to 1 μg mL^{-1} in HEPES Buffer with 0.05% Tween 20 (HBS-T) before being injected in one of the flow cells of the chip. Low immobilization levels of 94 and 64 response units were obtained for 3'-SL and for 6'-SL, respectively. A reference surface was always present in flow cell 1, thus allowing for the subtraction of bulk effects and non-specific interactions with streptavidin. The running buffer consisted of the same HBS-T pH 7.4. The purified proteins were injected over the flow cell surface at 30 μL min^{-1} in series of 2-fold dilutions. The dissociation of this analyte was done by passing running buffer during 4 - 6 min. Surfaces were regenerated with one or two consecutive 30 second injections of 100 mM Neu5Ac, also at 30 μL min^{-1}. The information on the affinity was determined by assuming Langmuir 1:1 binding, using the BIAevaluation software.

RESULTS AND DISCUSSION

Sequences identified as belonging to the family 40 CBM were selected from the CAZY (Carbohydrate-Active enZYmes) database [24] and used to build a phylogenetic tree (http://phylogeny.lirmm.fr). The topology of the tree (see supplementary information, *SI 2*), shows the CBM40s cluster into one main group with few outliers. The CBM from *Cp* NanI clusters with the other CBM from the same organism, *Cp* CBM40_NanJ, with a phylogenetic similarity of 0.942. As previously noted by other authors, the CBM of *Vc* NanH is more distantly related to the other CBM40s, among those with validated Neu5Ac specificity and available structural information [26]. The BLAST protein alignment between the *Cp* CBM40_NanI sequence and the other structurally characterized CBM40s – *Cp* NanJ (PDB ID: 2V73), *Sp* NanA (4C1W), NanB (2VW0) and NanC (4YZ5), and *Md* nanL (1SLI) – shows that CBM40_NanI shares 56% identity with the CBM40_NanJ from the same organism (*C. perfringens*), 23%, 30% and 27% homology values with the *S. pneumoniae* proteins, and 33% with the CBM of *Macrobdella decora* leech. *Cp* CBM40 NanI appears therefore sufficiently different from previously characterized CBMs of family 40 to substantiate a full characterization.

After cloning of genomic DNA of *C. perfringens* ATCC® 13124™, recombinant *Cp* CBM40_NanI was purified with a yield of approximately 10 mg L^{-1}. The *Cp* CBM40_NanJ was produced in the same conditions for comparison. The expressed proteins appear as one strong band on SDS-PAGE gels, with apparent molecular weights close to 25 kDa marker, in agreement with the theoretical values of 23.7 and 24.6 kDa for *Cp* CBM40_NanJ and *Cp* CBM40_NanI, respectively. The proteins were separated from aggregates by gel exclusion chromatography and collected as a single peak with a retention time representative of the monomers (see *SI 3*). The thermal stability of the *Cp* CBM40_NanI and *Cp* CBM40_NanJ was assessed by Thermal Shift assays (TSA). The highest stability was obtained in PBS buffer at pH 7, with melting temperatures of 54° C for *Cp* CBM40_NanJ and 65° C for *Cp* CBM40_NanI. One issue with the use of CBMs as glycan probes is their low affinities. One method to overcome this is to engineer constructs containing multiple CBM domains [23]**.** Thus, a divalent version of the protein was engineered. The pseudo-dimeric *Cp* CBM40_NanI has a 12 amino acid-long linker (*ALNGSELGSGSG*) inserted between the two copies to provide flexibility between the binding modules. Under identical expression protocol, a yield of ~ 5 mg L^{-1} purified di-CBM40_NanI was obtained. The expected band for a 46 kDa weight protein was seen in the SDS-PAGE gel (*SI 3*). The engineered *Cp* di-CBM40_NanI was stable in PBS buffer pH 7.0, with an observed T_m of 66 °C (*SI 4*).

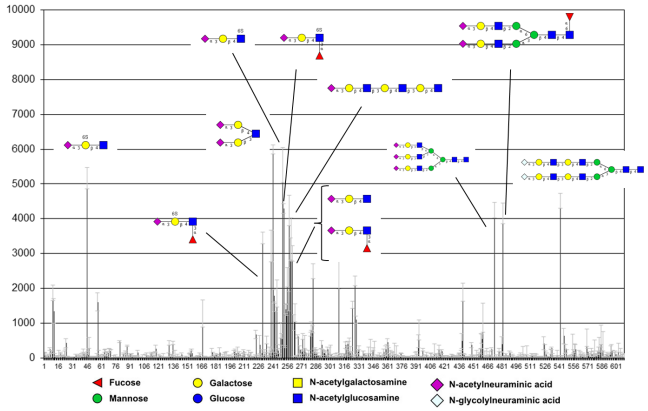


Figure 1 - *Glycan array profile of Cp di-CBM40_NanI. The Alexa Fluor 488-fused protein was tested against a panel of 609 glycans. Some examples of α(2,3)-linked sialosides are shown (more details at CFG request nº16976, supplementary information file CFG_diCBM40_NanI.xls).*

Characterization of binding specificity demonstrates strong preference for 3-linked sialic acids

An initial evaluation of glycan binding was performed by inhibition of hemagglutination using rabbit erythrocytes. As expected, hemagglutinating activity was not observed with monomeric *Cp* CBM40_NanI as agglutination requires a multivalent binding to crosslink the cells. In contrast, *Cp* di-CBM40_NanI showed positive hemagglutination down to a 50 µg mL^{-1} concentration. Among the glycans tested as HA inhibitors, 3'-SL was the strongest inhibitor, with minimal inhibitory concentration (MIC) of 0.78 mM. 6'-SL was also able to inhibit hemagglutination with a MIC value of 1.56 mM. Lactose, galactose, glucose, mannose and fucose (tested up to 50 mM) had no inhibitory effect (see *SI 5*).

To gain further insight into the binding motifs of Cp CBM40_NanI, both monomeric and dimeric CBMs were tested on the Consortium for Functional Glycomics' glycan microarray v5.2. This array contains a panel of 609 mammalian, including α2,3-, α2,6- and α2,8-sialosides [43]. Monomeric *Cp* CBM40_NanI gave only a very weak signal with clear binding to only a few epitopes bearing α2,3-sialic acid on a type II *N*-acetyllactosaminyl core (Neu5Acα2,3Galβ1,4GlcNAc) (*SI 6*). Previous assays involving monomeric *Cp* CBM40_NanJ were also unsuccessful due to low affinity [26]. The signal-to-noise ratio was considerably improved when assaying the engineered dimeric protein. The *Cp* di-CBM40_NanI screen confirms the specificity of this binding domain towards α2,3-sialosides (Figure 1). Of the 27 glycans that had significant binding ($p \leq 0.05$ as determined by Z-score analysis) on the array, 24 contained an α2,3-sialoside. The remaining 3 were an anomalous chitin epitope, sialic acid bound directly to a linker and a single biantennary α2,6-epitope detected. Both N-glycolyl- and 9-O acetyl-containing α2,3-sialosides acids were tolerated. Binding was enhanced by an additional nearby negative charge (sulfation or sialylation).

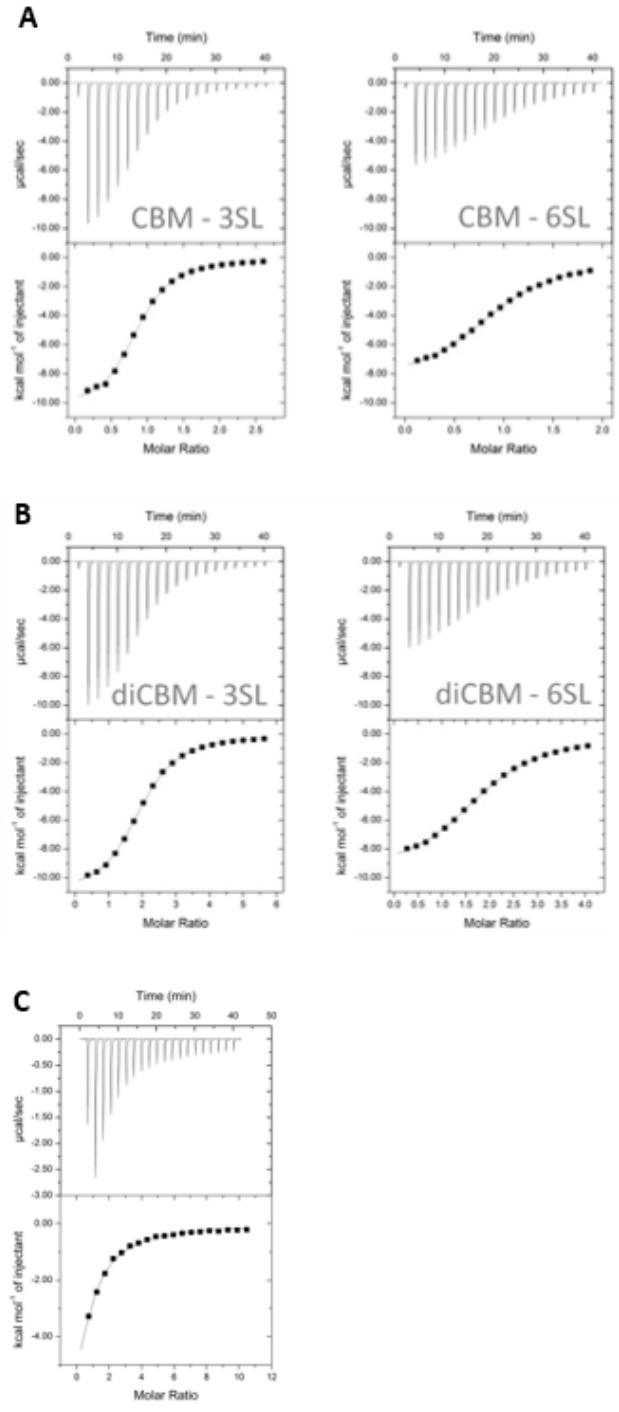


Figure 2 *- ITC isotherms showing the binding of* ***A)*** *CBM40_NanI to 3'-SL and 6'-SL;* ***B)*** *di-CBM40_NanI to 3'-SL and 6'-SL; and* ***C)*** *CBM40_NanI to SiaOPr.*

Table 1 *- Thermodynamics of binding of Cp CBM40_NanI, Cp di-CBM40_NanI Cp CBM40_NanJ to different sialylated ligands, as determined by ITC at 25 °C.*

Protein	[P]	Ligand	[L]	*N*	K_a	K_d	ΔH	TΔS	ΔG
	(μM)		*(mM)*		*(x10⁴ M⁻¹)*	*(μM)*	*(kcal/mol)*	*(kcal/mol)*	*(kcal/mol)*
Cp CBM40_NanI	400	3'-SL	5.0	0.86 ± 0.01	3.16 ± 0.09	32	-10.49 ± 0.14	-4.35	-6.14
	400	6'-SL	3.6	0.95 ± 0.01	1.48 ± 0.05	68	-8.69 ± 0.05	-3.01	-5.68
	50	3'-SLN	0.9	0.98 ± 0.05	2.68 ± 0.20	37	-9.38 ± 0.06	-3.34	-6.04
	50	6'-SLN	0.8	0.92 ± 0.08	2.23 ± 0.25	49	-8.81 ± 0.10	-2.88	-5.93
	100	SiaOPr	5.0	1.00*	0.59± 0.05	170	-19.95± 0.05	-14.81	-5.14
Cp di-CBM40_NanI	185	3'-SL	5.0	1.94 ± 0.01	2.84 ± 0.08	35	-11.20 ± 0.01	-5.16	-6.07
	185	6'-SL	3.6	1.88 ± 0.01	1.71 ± 0.06	59	-9.74 ± 0.01	-3.96	-5.78
	50	SiaOPr	5.0	2.00*	0.51 ± 0.06	196	*n/d*	*n/d*	*n/d*
					(x10³ M⁻¹)	*(mM)*			
Cp CBM40_NanJ	100	3'-SL	10.0	1.00*	0.67 ± 0.04	1.48	-1.51 ± 0.05	2.35	-3.86
	100	6'-SL	10.0	1.00*	0.44 ± 0.06	2.23	-2.72 ± 0.27	0.89	-3.61

Cp CBM40_NanI displays unusually strong affinity for sialylated oligosaccharides

The glycan-binding activities of *Cp* CBM40_NanI and di-CBM40_NanI were confirmed by isothermal titration calorimetry at 25°C. All interactions showed exothermic behaviors with a steep decrease in the exothermic heat of binding until saturation is achieved (see Figure 2). The stoichiometry values (*N*) for sialylated compounds titrated in *Cp* CBM40_NanI and di-CBM40_NanI are close to 1 and 2 respectively, in agreement with the expected presence of one or two binding sites (Table 1). Both proteins have similar behavior with strong affinity for 3'-SL (K_d value ~30 μM) and weaker affinity for 6'-SL (K_d ~ 60 μM). Values in the same data range were obtained for 3'- and 6'-sialyllactosamine (*SI 7*). Free Neu5Ac monosaccharide could not be assayed directly due to pH variations during the titration. To assay the monosaccharide, a derivative of Neu5Ac with a propargyl aglycon (SiaOPr) was synthesized and assayed for binding. The CBMs showed weaker binding (K_d values of 170 and 189 μM for monomeric and dimeric CBM, respectively) to the monosaccharide conjugate than to the sialylated trisaccharides. To directly compare binding affinities, ITC experiments were also performed for *Cp* CBM40_NanJ. As expected, lower binding affinities were obtained with K_d values of 1.48 and 2.67 mM for 3'-SL and 6'-SL, respectively (*SI 7*), confirming the previously published estimations of low affinity [26].

Table 2 *- Binding affinities of Cp CBM40_NanI and Cp di-CBM40_NanI to immobilized 3'-SL and 6'-SL receptors, as determined by SPR at 25 °C.*

Protein	Ligand *(μM)*	K_d *(μM)*
Cp CBM40_NanI	3'-SL	14.4
Cp CBM40_NanI	6'-SL	36.0
Cp di-CBM40_NanI	3'-SL	1.3
Cp di-CBM40_NanI	6'-SL	6.1

The binding interactions were also analyzed by Surface Plasmon Resonance spectroscopy. Commercial available monomeric biotinylated 3'-SL and 6'-SL were immobilized on a streptavidin coated chip to obtain high-density coverage. The sensorgrams show a typical shape for fast association and dissociation events after the injection of increasing concentrations of *Cp* CBM40_NanI or *Cp* di-CBM40_NanI. The binding kinetics of the experiments were too fast to be reliably measured by SPR and steady-state analysis was performed to evaluate the dissociation constants (see *SI 8*). The data confirmed the stronger binding of *Cp* CBM40_NanI and *Cp* di-CBM40_NanI with the 3'-SL- *versus* the 6'-SL-functionalized surface. The monomeric CBM showed a K_d of 14.4 μM for 3'-SL, whereas 6'-SL appears to bind at a 2.5 times lower affinity, with a dissociation value of 36 μM, in agreement with ITC data (Table 2). For the *Cp* di-CBM40, an increase of more than one order of magnitude in affinity (11-fold) is observed for the interaction with 3'-SL and

a 6-fold for 6'-SL, resulting in affinities of 1.3 and 6.1 μM, respectively. Experiments conducted with *Cp* CBM40_NanJ did not yield sensorgrams of sufficient quality to evaluate, most likely due to the low affinity of *Cp* CBM40_NanJ-sialoside interactions.

Structural basis for specificity and affinity

To gain insight into the structural basis for sialoside binding of *Cp* CBM40_NanI, crystal structures were obtained. An original screening of 600 conditions, using a high throughput crystallization robot, led to isolation of thin crystals that diffracted at low resolution. Upon optimization with the Crystal Clear Strategy Screen (Molecular Dimension Ltd) bigger and thicker crystals of *Cp* CBM40_NanI complexed with 3'-SL and 6'-SL were obtained. The protein crystalized in $P2_1$ space group for both complexes, with 3 and 6 independent monomers in the asymmetric unit for crystals of the CBM40_NanI complexed with 3'-SL or 6'-SL, respectively (see table in *SI 9*). The structure of the CBM/6'-SL was solved using PDB ID 2V73 (*Cp* CBM40_NanJ) as a model, and subsequently used to solve the structure of the *Cp* CBM40_NanI/3'-SL. The monomers superimposed with root-mean-square deviation (r.m.s.d.) values between 0.12 and 0.16 Å for *Cp* CBM40/3'SL and between 0.15 and 0.20 Å for *Cp* CBM40_NanI/6'-SL.

The *Cp* CBM40_NanI adopts the characteristic β-sandwich fold of CBM40 members, formed by two β-sheets of five and six antiparallel β-strands and two α-helices, one within the sheets and another at the *C*-terminus, packed against β-strand 7. The carbohydrate recognition site is located on the concave face of the 5-stranded β-sheet, in a shallow depression, in which the ring of the sialic acid lays parallel to the surface of the protein (Figure 3A). Examination of the electron density clearly allowed the identification of the sialic acid and galactose residues in all binding sites in the CBM/3'-SL complex and in half of the molecules in the 6'-SL, whilst only the Neu5Ac moiety could be defined with confidence for the others sites. The glucose moieties could not be modeled as they were too disordered.

The majority of the interactions between the protein and the ligand involve the sialic acid moiety. In both complexes (see Figures 3B and 3C), Ile46, Tyr65 and His93 side chains form a hydrophobic pocket that accommodates the methyl of the N-acetyl group. Arg76 and Arg153 make electrostatic interactions with the carboxylate group. The Asn158 residue establishes hydrogen bonds with the glycerol moiety. The hydroxyl group of Tyr160 makes hydrogen bonds to the carboxylate group and the nitrogen of the *N*-acetyl group. Glu74 is involved in two hydrogen bonds with the nitrogen of the N-acetyl group and the O4 hydroxyl and the NH1 and NH2 atoms of Arg153 make H-bonds with the carboxylate group. The carboxylate makes water mediated interactions with the main chain nitrogen of Asn158 and the NH2 atom of Arg76. Comparison between the binding sites of *Cp* CBM40_NanI and NanJ show that the amino acids making direct contacts with the ligand are conserved in both structures (see *SI 10*).

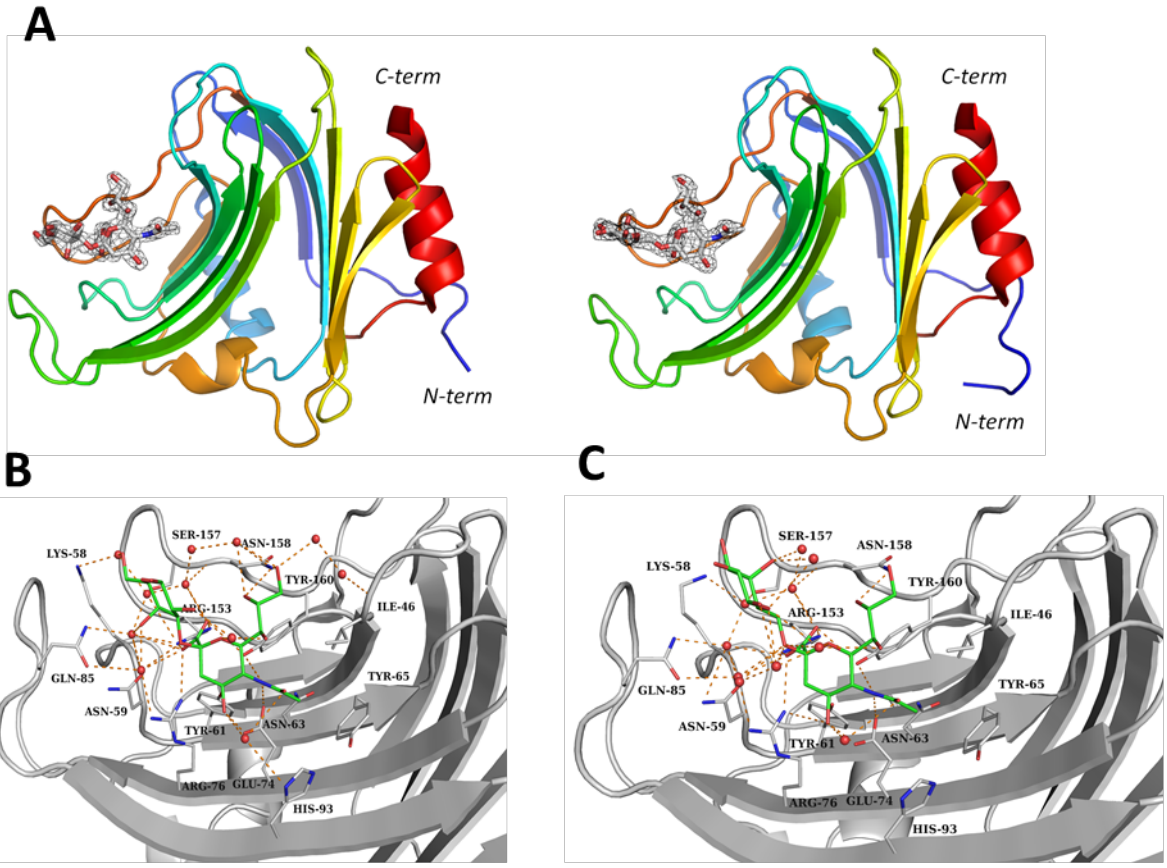

A
C-term
N-term
C-term
N-term
B
LYS-58
SER-157
ASN-158
TYR-160
ARG-153
ILE-46
GLN-85
TYR-65
ASN-59
TYR-61
ASN-63
ARG-76
GLU-74
HIS-93
C
SER-157
ASN-158
LYS-58
TYR-160
ARG-153
ILE-46
GLN-85
TYR-65
ASN-59
TYR-61
ASN-63
ARG-76
GLU-74
HIS-93
D
Cp CBM40_NanI
Cp CBM40_NanJ
Sp CBM40_NanA
Sp CBM40_NanB
Sp CBM40_NanC
Md CBM40_NanL
E
NanI
β1
α1
β2
β3
β4
β5
β6
NanI LSSLGEYKDINLES..SNASNITY..DLEKYKNLDEGTIVVRFNSK..DSKIQSLLGISNS..KTKNGYFNFYVTN.SRVGPFLRNQKNEGNTQNGTENL.VHMY..KDVALN
NanJ VDEIANYGNLKITK.EEERVNITG..DLEKFSSLEEGTIVTRFNMN..DTSIQSLIGLSDG..NKANNYFSLYVSG.GKVGYFLRRQEG......NGDFN.VHHS..ADVTFN
NanA GAMVIEKEDVETNASNGQRVDLSS..ELDKLKKLENATVHMEFKPDAKAPAFYNLFSVSSA..TKKDEYFTMAVY.NNTATLEGRGSD.........GKQFYNNYNDAPLKVK
NanL EGILMEKNNVDIAE..GQGYSLDQEAGAKYVKAMTQGTIILSYKST.SENGIQSLFSVGNSTAGNQDRHFHIYITNSGGIGIELRNTD........GVFN.YTLD..RPASVR
NanC .TPVLEKNNVTLT...GGGENVTKE.LKDK.FTSGDFTVVIKYNQS.SEKGLQALFGISNSKPGQQNSYVDVFLRDNGELGMFARDTS........SNKN.NLVS..RPASVW
NanB ISPIFQGGSYQLN...NKSIDISSL.LLDK.LSGESQTVVMKFKAD.KPNSLQALFGLSNSKAGFKNNYFSIFMRDSGEIGVEIRDAQ........KGIN.YLFS..RPASLW
β7
β8
β9
β10
β11
α2
NanI ......DGDNTVALKIEKN......KGYKLFLNGKMIKEVKDTNTKFLNNIENLDSAFIGKTNRYGQSNEYNFKGNIGFMNIYNEPLGDDYLLSKTGETK..
NanJ ......RGINTLALKIEKG......IGAKIPLNGSLVKTVSDPNIKFLNNAI.NLNSGFIGKTDRANGYNEYLFRGNIDFMNIYDKPVSDNYLLRKTGETK..
NanA P.....GQWNSVTFTVEKPTAELPKGRVRLYVNGVLSRTSLR.SGNFIKDMPDVTHVQIGATKRANN.TVWGSNLQIRNLTVYNRALTPEEVQKRS.....
NanL ALYKGERVFNTVALKADAA.....NKQCRLFANGELLATLDKDAFKFISDITGVDNVTLGGTKRQGK.IAYFFGGTIGDIKVYSNALSDEELIQATGVTTYG
NanC GKYKQEAVTNTVAVVADSV.....KKTYSLYANGTKVVEKKVDNFLNIKDIKGIDYYMLGGVKRAGK.TAFGFNGTLENIKFFNSALDEETVKKMTTNAVT.
NanB GKHKGQAVENTLVFVSDSK.....DKTYTMYVNGIEVFSETVDTFLPISNINGIDKATLGAVNREGK.EHYLAKGSIDEISLFNKAISDQEVSTI.PLSNP.

Figure 3 - A) Structure of the CBM40 from *C. perfringens* ATCC® 13124™ NanI in complex with 3'-SL and 6'-SL (at the left and right, respectively). Ligands are shown with a 2Fo – Fc density map contoured at 1.0 r.m.s.; **B-C)** Zoomed-in view on the binding domain occupied with 3'-SL and 6'-SL; **D)** Comparison of the CBM40 of *C. perfringens* NanI with structurally related structures. *Cp* NanI (in red, boxed), *Cp* NanJ (in cyan), *Sp* NanA (in yellow), *Sp* NanB (in green), *Sp* NanC (in brown) and *Md* NanL (in purple). The loop 5-6 in CBM40_NanI is identified with an arrow; **E)** Multiple amino acid sequence alignment between CBM40_NanI (shown as NanI) and the CBM40s from *C. perfringens* ATCC® 13124™'s NanJ, *S. pneumoniae*'s NanA, NanB and NanC, and *Macrobdella decora* NanL. The secondary structure regions of the CBM40_NanI are identified on top. Shaded residues show the conservation between the proteins. The CBM40_NanI amino acids making key contacts with the sialic acid are identified with blue triangles. The region between β-strands 5 and 6 is boxed within red limits and residue Gln85 is marked with a red star due to their particular importance as part of the binding domain. The sequence of *V. cholerae* NanH is not shown due to its poor alignment (see *SI 8*). The figure was created with Espript 3.0 [53].

In the complex with 3'SL, the galactose does not interact directly with the protein but is involved in a dense network of water-bridged hydrogen bonds. Six water molecules are conserved in the three independent monomers where three of them are directly involved in stabilizing the galactose residue (*SI 11*). These three water molecules are responsible for locking the conformation of the galactosyl moiety and creating a further network with other waters, bridging both inter-amino acid and intra-ligand interactions. Gln85 is a key amino acid in this network, interacting with two water molecules which bridge with the O4 hydroxyl of the galactosyl moiety. Lys58 NH makes another water mediated H-bond with the OH6 group of the galactose. Also the Ser157 interacts with a chain of water molecules connecting to the carboxylate of the NeuAc and the O4 hydroxyl of the galactose.

In the CBM/6'-SL complex, two of the three resolved 6'-SL molecules have the same conformation whereas the third shows different dihedrals angles (*SI 12*). This is an effect of the highest solvent exposure and absence of contacts between the galactose ring of the ligand and the protein. It can be seen in figures 3B-C that the galactose ring is oriented outside the recognition domain of the protein.

Comparison with the other CBM40 proteins

The overall macromolecular architecture of the *Cp* CBM40_NanI is well conserved when comparing to other members of CBM40 family apart NanH which belongs to another branch of the phylogenetic tree as discussed above. Sequence alignment show that the recognition domains of the CBM40s share common key residues that provide the protein its binding specificity (figure 3E and *SI 13*). Within these binding domains, residues Glu74 and Arg153 (*Cp* CBM40_NanI numeration) are particularly important, being spatially conserved in all CBMs (with the exception of *Md* NanL, which has an aspartic instead of glutamic acid), making a total of four H-bonds with the Neu5Ac (two with the N-acetyl and O4 hydroxyl, and two with the carboxylate group, respectively). A highly variable region between all of the CBM40s is the loop between β-strands 5 and 6 (Figure 3D and E, see starred box). In *Cp* CBM40_NanI, this loop is longer by 6-8 amino acids and has a unique spatial arrangement compared to the other CBM40 members. In fact, this is a major point of difference between *C. perfringens* CBM40_NanI and CBM40_NanJ, which share considerable sequence identity and bind sialic acid in virtually the same position on their protein surfaces. Despite this, *Cp* CBM40_NanI displays a much higher affinity to

sialylated oligosaccharides (~ 100-fold) than its counterpart. This difference in affinity appears to be due to an increased water mediated H-bond set-up inside the binding site of *Cp* CBM40_NanI.

As seen in the crystal structures, the loop between β-strands 5 and 6, exclusive to *Cp* CBM40_NanI, introduces an additional surface in the recognition domain. In this loop, residue Gln85 creates water mediated interactions with both the sialyl carboxylate and galactosyl hydroxyl groups. In contrast to NanI, the crystal structure of *Cp* CBM40_NanJ does not show such water-mediated H-bond network (*SI 10*).

Sialylated oligosaccharides are flexible epitopes, due to the possibility of therefore conformations at the α2-3 and α2-6 linkages [44]. It is therefore of interest to compare if different CBMs recognized the same epitope. *Sp* CBM40_NanA and NanC were also successfully co-crystalized with the 3'-SL and 6'-SL ligands (PDB ID: 4C1W and 4C1X, and 4YZ5 and 4YW2, respectively), among others [28, 30]. Probably due to the exposure of the ligands and the few contacts made with the protein, different conformations are seen for the 3'-SL and 6'-SL when bound *Cp* CBM40_NanI, *Sp* CBM40_NanA and *Sp* CBM40_NanC. 3'-SL shares similar conformations in NanI and NanA (in NanC the galactose ring is in a full inverted position relative to the others, whereas 6'-SL structure is alike in NanI and NanC (see *SI 14*).

CONCLUSION

Although the use of CBMs in biotechnology is fairly recent, they have been described in the paper, textile and food industries [45]. Their potential application in health sciences has also been proved in different scenarios as molecular probes [46], in microarrays [47], as tools for the immobilization of proteins and cells [48] or assisting *in vitro* glyco-catalysed reactions leading to enhanced yields [45, 49, 50]. Recent studies have shown the therapeutic potential of multivalent CBMs in the prevention of infection by the influenza virus, through the masking of sialic acid receptors [51, 52].

This study presents a new family 40 CBM that binds specifically to NeuAc with a preference towards α2,3-sialosides. It was easily engineered and expressed in multimeric fashion by fusing consecutives copies of its gene, leading to a substantial increase of its affinity (11-fold) over the monomeric version. Although the previously described *C. perfringens* NanJ is highly similar to NanI, the binding affinity of NanI to sialosides acid was much stronger, making it a better candidate to develop as a glycan binding reagent. Our studies suggest that the variable loop between β-strands 5 and 6 may be responsible for this difference in affinity and may contribute to the unusual ligand specificity observed. The binding *Cp* CBM40_NanI towards sialosides is considerably more restricted than any other published CBM40, with a distinct preference for α2,3-sialosides, and its interactions at micromolar values are comparable only with the CBM of *Vibrio cholerae*'s NanH, which shows a broad specificity to all sialosides [23]. The binding of α2,3-sialosides – but not 3'-O-sulfated ones – makes this reagent more specific for sialic acid residues than the commonly used plant lectin from *Maackia amuerensis* MAL-1 [41].

The present study unravels the molecular basis for the difference in affinity and specificity for two closely related CBM40s from the same organisms. In addition, the high affinity of *Cp* CBM40_NanI for 3'SL and its availability as monomer or as engineered dimer makes it an excellent tool for detection and characterization of sialylation for biotechnology use.

Acknowledgements

The research leading to these results has received funding from the European Community's 7th Framework Programme (FP7/2007-2013) under the Marie Curie International Outgoing Fellowship for Career Development (PIOF-GA-2011-298910), BioStruct-X (grant agreement N°283570), and ERASynbio program SynGlycTis (ANR-14-SYNB-0002-02) grants. The authors are also grateful to the Labex Arcane for the financial support (ANR-11-LABX-0003-01).

We also thank the participation of the Protein-glycan Interaction Resource of the CFG and the supporting grant R24 GM098791.

Crystal data collection was performed at the European Synchrotron Radiation Facility, Grenoble, France and we are thankful for the access to beamlines BM30A-FIP and ID29 as well as for the technical support of David Cobessi on BM30A.

Conflict of Interest: The authors declare that they have no conflicts of interest with the contents of the publication.

Author Contributions:

LKM conceived the project. JPR, LKM and AI coordinated the paper. JPR and WP performed cloning, protein expressions and purifications. JPR ran all the biophysical experiments and provided data analysis. JPR and AV resolved the crystal structures. CP and OR synthesized and characterized the SiaOPr. JPR analyzed, generated the figures and tables and wrote the body of the paper. AV, LKM and AI revised the paper leading to the final version. All authors reviewed the results and approved the manuscript for submission.

References

1 Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. Glycobiology. **3**, 97-130

2 Traving, C. and Schauer, R. (1998) Structure, function and metabolism of sialic acids. Cell Mol Life Sci. **54**, 1330-1349

3 Varki, A. (1999) Essentials of Glycobiology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA

4 Schnaar, R. L., Gerardy-Schahn, R. and Hildebrandt, H. (2014) Sialic Acids in the Brain: Gangliosides and Polysialic Acid in Nervous System Development, Stability, Disease, and Regeneration. Physiological Reviews. **94**, 461-518

5 Wang, B. and Brand-Miller, J. (2003) The role and potential of sialic acid in human nutrition. Eur J Clin Nutr. **57**, 1351-1369

6 Wang, P.-H. (2005) Altered Glycosylation in Cancer: Sialic Acids and Sialyltransferases. Journal of Cancer Molecules. **1**, 73-81

7 Pearce, O. M. and Laubli, H. (2016) Sialic acids in cancer biology and immunity. Glycobiology. **26**, 111-128

8 Dall'Olio, F., Chiricolo, M., Ceccarelli, C., Minni, F., Marrano, D. and Santini, D. (2000) β-galactoside α2,6 sialyltransferase in human colon cancer: contribution of multiple transcripts to regulation of enzyme activity and reactivity with sambucus nigra agglutinin. International Journal of Cancer. **88**, 58-65

9 Schultz, M., Swindall, A. and Bellis, S. (2012) Regulation of the metastatic cell phenotype by sialylated glycans. . Cancer metastasis reviews. **31**, 501-518

10 Ito, S., Hayama, K. and Hirabayashi, J. (2009) Enrichment strategies for glycopeptides. Methods Mol Biol. **534**, 195-203

11 Rogerieux, F., Belaise, M., Terzidis-Trabelsi, H., Greffard, A., Pilatte, Y. and Lambré, C. (1993) Determination of the sialic acid linkage specificity of sialidases using lectins in a solid phase assay. Anal Biochem. **211**, 200-204
12 Nicholls, J., Bourne, A., Chen, H., Guan, Y. and Peiris, J. (2007) Sialic acid receptor detection in the human respiratory tract: evidence for widespread distribution of potential binding sites for human and avian influenza viruses. Respiratory Research. **8**, 73-82
13 Tateno, H., Winter, H. and Goldstein, I. (2004) Cloning, expression in Escherichia coli and characterization of the recombinant Neu5Ac(alpha)2-6Gal(beta)1-4GlcNAc-specific high-affinity lectin and its mutants from the mushroom Polyporus squamosus. Biochemical Journal. **382**, 667-675
14 Kadirvelraj, R., Grant, O., Goldstein, I., Winter, H., Tateno, H., Fadda, E. and Woods, R. (2011) Structure and binding analysis of Polyporus squamosus lectin in complex with the Neu5Ac(alpha)2-6Gal(beta)1-4GlcNAc human-type influenza receptor. Glycobiology. **21**, 973-984
15 Knibbs, R. N., Osborne, S. E., Glick, G. D. and Goldstein, I. J. (1993) Binding determinants of the sialic acid-specific lectin from the slug Limax flavus. Journal of Biological Chemistry. **268**, 18524-18531
16 Hsu, K. L., Gildersleeve, J. C. and Mahal, L. K. (2008) A simple strategy for the creation of a recombinant lectin microarray. Mol Biosyst. **4**, 654-662
17 Lewis, A. L. and Lewis, W. G. (2012) Host sialoglycans and bacterial sialidases: a mucosal perspective. Cell Microbiol. **14**, 1174-1182
18 Juge, N., Tailford, L. and Owen, C. D. (2016) Sialidases from gut bacteria: a mini-review. Biochem Soc Trans. **44**, 166-175
19 Matrosovich, M., Herrler, G. and Klenk, H. D. (2015) Sialic Acid Receptors of Viruses. Top Curr Chem. **367**, 1-28
20 Miyagi, T. and Yamaguchi, K. (2012) Mammalian sialidases: Physiological and pathological roles in cellular functions. Glycobiology. **22**, 880-896
21 Bourne, Y. and Henrissat, B. (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. Curr Opin Struct Biol. **11**, 593-600
22 Ficko-Blean, E. and Boraston, A. B. (2012) Insights into the recognition of the human glycome by microbial carbohydrate-binding modules. Curr Opin Struct Biol. **22**, 570-577
23 Connaris, H., Crocker, P. R. and Taylor, G. L. (2009) Enhancing the receptor affinity of the sialic acid-binding domain of Vibrio cholerae sialidase through multivalency. J Biol Chem. **284**, 7339-7351
24 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. **42**, D490-495
25 Moustafa, I., Connaris, H., Taylor, M., Zaitsev, V., Wilson, J. C., Kiefel, M. J., von Itzstein, M. and Taylor, G. (2004) Sialic acid recognition by Vibrio cholerae neuraminidase. J Biol Chem. **279**, 40819-40826
26 Boraston, A. B., Ficko-Blean, E. and Healey, M. (2007) Carbohydrate recognition by a large sialidase toxin from Clostridium perfringens. Biochemistry. **46**, 11352-11360
27 Luo, Y., Li, S. C., Chou, M. Y., Li, Y. T. and Luo, M. (1998) The crystal structure of an intramolecular trans-sialidase with a NeuAc alpha2-->3Gal specificity. Structure. **6**, 521-530
28 Yang, L., Connaris, H., Potter, J. A. and Taylor, G. L. (2015) Structural characterization of the carbohydrate-binding module of NanA sialidase, a pneumococcal virulence factor. BMC Struct Biol. **15**, 15
29 Xu, G., Potter, J. A., Russell, R. J., Oggioni, M. R., Andrew, P. W. and Taylor, G. L. (2008) Crystal structure of the NanB sialidase from Streptococcus pneumoniae. J Mol Biol. **384**, 436-449
30 Owen, C. D., Lukacik, P., Potter, J. A., Sleator, O., Taylor, G. L. and Walsh, M. A. (2015) Streptococcus pneumoniae NanC: STRUCTURAL INSIGHTS INTO THE SPECIFICITY AND MECHANISM OF A SIALIDASE THAT PRODUCES A SIALIDASE INHIBITOR. J Biol Chem. **290**, 27736-27748
31 Myers, G. S. A., Rasko, D. A., Cheung, J. K., Ravel, J., Seshadri, R., DeBoy, R. T., Ren, Q., Varga, J., Awad, M. M., Brinkac, L. M., Daugherty, S. C., Haft, D. H., Dodson, R. J., Madupu, R., Nelson, W. C., Rosovitz, M. J., Sullivan, S. A., Khouri, H., Dimitrov, G. I., Watkins, K. L., Mulligan, S., Benton, J., Radune, D., Fisher, D. J., Atkins, H. S., Hiscox, T., Jost, B. H., Billington, S. J., Songer, J. G., McClane, B. A., Titball, R. W., Rood, J. I., Melville, S.

B. and Paulsen, I. T. (2006) Skewed genomic variability in strains of the toxigenic bacterial pathogen, Clostridium perfringens. Genome Research. **16**, 1031-1040
32 Newstead, S., Chien, C. H., Taylor, M. and Taylor, G. (2004) Crystallization and atomic resolution X-ray diffraction of the catalytic domain of the large sialidase, nanI, from Clostridium perfringens. Acta Crystallogr D Biol Crystallogr. **60**, 2063-2066
33 Dupeux, F., Rower, M., Seroul, G., Blot, D. and Marquez, J. A. (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. Acta Crystallogr D Biol Crystallogr. **67**, 915-919
34 Leslie, A. G. W. (1992) Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography. **26**
35 Kabsch, W. (2010) Xds. Acta Crystallogr D Biol Crystallogr. **66**, 125-132
36 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. **67**, 235-242
37 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. Journal of Applied Crystallography. **40**, 658-674
38 Langer, G., Cohen, S. X., Lamzin, V. S. and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc. **3**, 1171-1179
39 Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. **67**, 355-367
40 Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) Features and development of Coot. Acta Crystallographica Section D. **66**, 486-501
41 Wang, L., Cummings, R. D., Smith, D. F., Huflejt, M., Campbell, C. T., Gildersleeve, J. C., Gerlach, J. Q., Kilcoyne, M., Joshi, L., Serna, S., Reichardt, N. C., Parera Pera, N., Pieters, R. J., Eng, W. and Mahal, L. K. (2014) Cross-platform comparison of glycan microarray formats. Glycobiology. **24**, 507-517
42 Lameignere, E., Malinovská, L., Sláviková, M., Duchaud, E., Mitchell, E. P., Varrot, A., Sedo, O., Imberty, A. and Wimmerová, M. (2008) Structural basis for mannose recognition by a lectin from opportunistic bacteria Burkholderia cenocepacia. Biochemical Journal. **411**, 307-318
43 Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skehel, J. J., van Die, I., Burton, D. R., Wilson, I. A., Cummings, R., Bovin, N., Wong, C.-H. and Paulson, J. C. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proceedings of the National Academy of Sciences of the United States of America. **101**, 17033-17038
44 Imberty, A. and Pérez, S. (2000) Structure, Conformation, and Dynamics of Bioactive Oligosaccharides: Theoretical Approaches and Experimental Validations. Chemical Reviews. **100**, 4567-4588
45 Oliveira, C., Carvalho, V., Domingues, L. and Gama, F. M. Recombinant CBM-fusion technology - Applications overview. Biotechnol Adv. **33**, 358-369
46 von Schantz, L., Hakansson, M., Logan, D. T., Walse, B., Osterlin, J., Nordberg-Karlsson, E. and Ohlin, M. Structural basis for carbohydrate-binding specificity--a comparative assessment of two engineered carbohydrate-binding modules. Glycobiology. **22**, 948-961
47 Ofir, K., Berdichevsky, Y., Benhar, I., Azriel-Rosenfeld, R., Lamed, R., Barak, Y., Bayer, E. A. and Morag, E. (2005) Versatile protein microarray based on carbohydrate-binding modules. Proteomics. **5**, 1806-1814
48 Shoseyov, O., Shani, Z. and Levy, I. (2006) Carbohydrate binding modules: biochemical properties and novel applications. Microbiol Mol Biol Rev. **70**, 283-295
49 Volkov, I., Lunina, N. A. and Velikodvorskaia, G. A. (2004) [Prospects for practical application of substrate-binding modules of glycosyl hydrolases (A review)]. Prikl Biokhim Mikrobiol. **40**, 499-504
50 Codera, V., Gilbert, Harry J., Faijes, M. and Planas, A. (2015) Carbohydrate-binding module assisting glycosynthase-catalysed polymerizations. Biochemical Journal. **470**, 15-22

51 Govorkova, E. A., Baranovich, T., Marathe, B. M., Yang, L., Taylor, M. A., Webster, R. G., Taylor, G. L. and Connaris, H. (2015) Sialic acid-binding protein Sp2CBMTD protects mice against lethal challenge with emerging influenza A (H7N9) virus. Antimicrob Agents Chemother. **59**, 1495-1504
52 Connaris, H., Govorkova, E. A., Ligertwood, Y., Dutia, B. M., Yang, L., Tauber, S., Taylor, M. A., Alias, N., Hagan, R., Nash, A. A., Webster, R. G. and Taylor, G. L. (2014) Prevention of influenza by targeting host receptors using engineered proteins. Proc Natl Acad Sci U S A. **111**, 6401-6406
53 Robert, X. and Gouet, P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research. **42**, W320-W324