Detailed insights from microarray and crystallographic studies into carbohydrate recognition by microneme protein 1 (MIC1) of *Toxoplasma gondii*

James A. Garnett,[1] Yan Liu,[2] Ester Leon,[1] Sarah A. Allman,[3] Nikolas Friedrich,[4] Savvas Saouros,[1] Stephen Curry,[5] Dominique Soldati-Favre,[4] Benjamin G. Davis,[3]* Ten Feizi,[2]* and Stephen Matthews[1]*

[1]Division of Molecular Biosciences, Centre for Structural Biology, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom

[2]Glycosciences Laboratory, Faculty of Medicine, Imperial College London, Harrow, Middlesex HA1 3UJ, United Kingdom

[3]Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Oxford OX1 3TA, United Kingdom

[4]Department of Microbiology and Molecular Medicine, Faculty of Medicine, University of Geneva CMU, 1211 Geneva 4, Switzerland

[5]Division of Molecular and Cell Biology, Centre for Structural Biology, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom

Received 18 March 2009; Revised 20 June 2009; Accepted 30 June 2009
DOI: 10.1002/pro.204
Published online 10 July 2009 proteinscience.org

Abstract: The intracellular protozoan *Toxoplasma gondii* is among the most widespread parasites. The broad host cell range of the parasite can be explained by carbohydrate microarray screening analyses that have demonstrated the ability of the *T. gondii* adhesive protein, TgMIC1, to bind to a wide spectrum of sialyl oligosaccharide ligands. Here, we investigate by further microarray analyses in a dose-response format the differential binding of TgMIC1 to 2-3- and 2-6-linked sialyl carbohydrates. Interestingly, two novel synthetic fluorinated analogs of 3′SiaLacNAc$_{1-4}$ and 3′SiaLacNAc$_{1-3}$ were identified as highly potent ligands. To understand the structural basis of the carbohydrate binding specificity of TgMIC1, we have determined the crystal structures of TgMIC1 micronemal adhesive repeat (MAR)-region (TgMIC1-MARR) in complex with five sialyl-*N*-acetyllactosamine analogs. These crystal structures have revealed a specific, water-mediated hydrogen bond network that accounts for the preferential binding of TgMIC1-MARR to arrayed 2-3-linked sialyl oligosaccharides and the high potency of the fluorinated analogs. Furthermore, we provide strong evidence for the first observation of a C—F···H—O hydrogen bond within a lectin-carbohydrate complex. Finally, detailed comparison with other oligosaccharide-protein complexes in the Protein Data Bank (PDB) reveals a new family of sialic-acid binding sites from lectins in parasites, bacteria, and viruses.

Keywords: *Toxoplasma gondii*; sialic acid; crystal structure; MIC1; microneme proteins; carbohydrate microarray

---

James A. Garnett and Yan Liu contributed equally to this work.

Grant sponsor: UK Research Councils' Basic Technology to T.F. and B.G.D; Grant number: GR/S79268; Grant sponsor: BBSRC to S.M., S.C., and T.F.; Grant number: G004608; Grant sponsor: MRC Award to S.M.; Grant number: G0800038; Grant sponsor: Wellcome Trust Equipment Grant to S.M.; Grant number: 085464; Grant sponsor: Swiss National Foundation to D.S-F.

**Correspondence to:* Stephen Matthews, Division of Molecular Biosciences, Centre for Structural Biology, Imperial College London, Biochemistry Building, South Kensington, London SW7 2AZ, United Kingdom. E-mail: s.j.matthews@imperial.ac.uk or Benjamin G. Davis, Department of Chemistry, Chemistry Research Laboratory, Oxford OX1 3TA, United Kingdom. E-mail: Ben.Davis@chem.ox.ac.uk or Ten Feizi, Division of Medicine, Imperial College London, Harrow, Middlesex HA1 3UJ, United Kingdom. E-mail: t.feizi@imperial.ac.uk

Introduction

Toxoplasma gondii is an intracellular protozoan parasite belonging to the phylum *Apicomplexa*, which includes *Plasmodium* and *Cryptosporidium* species. *T. gondii* infection is prevalent worldwide with up to half of the human population having been chronically infected. Although usually asymptomatic, infection of immunocompromised individuals such as patients with HIV/AIDS, results in toxoplasmosis, a condition that encompasses a variety of disease states including focal central nervous system infection.[1,2] In pregnant women, infection can also result in fetal and infant mortality or birth defects.[3] Toxoplasma infection has also been linked to subtle forms of schizophrenia[4] and other psychiatric manifestations.[5]

T. gondii has two distinct parts in its lifecycle: a sexual phase occurring in the primary hosts (domestic and wild cats) and an asexual phase where the parasite can propagate in almost any warm blooded animal. During asexual reproduction, the rapidly dividing form of the parasite, the tachyzoite, quickly establishes infection in the host through recognition and forced entry into a diverse range of cell types. Unlike the typical endocytic host-uptake pathways for viral and bacterial invasion, *T. gondii* and other apicomplexan parasites can actively penetrate host cells via a highly orchestrated process initiated by the binding to receptors on the host cell surface.

Microneme proteins (MICs) are released on the parasite surface at the time of invasion and act as major cellular adhesins,[6] participating in parasite reorientation and entry of the parasite into the host cell.[7–9] *T. gondii* microneme protein 1 (TgMIC1) is an important host cell binding protein that associates with two other MICs, TgMIC4 and TgMIC6.[10–12] The N-terminal region from TgMIC1 possesses two tandemly arranged repeats of a novel cell-binding domain named the micronemal adhesive repeat (MAR).[13,14] Preliminary carbohydrate microarray screening analyses revealed that a wide spectrum of sialylated oligosaccharide structures are recognized by the MAR-region from TgMIC1 (TgMIC1-MARR).[13] These include *N*- and *O*-glycans, gangliosides, and polysialic acid sequences in which sialic acid is 2-3- or 2-6-linked to galactose or 2-8-linked. Crystal structures of TgMIC1 complexed with sialyl carbohydrates revealed major contacts between a conserved threonine residue and the carboxyl group of the sialic acid moiety.[13] The importance of sialic acid recognition for efficient host cell invasion was corroborated by cell invasion assays, showing markedly reduced levels of parasite internalization in the presence of soluble sialic acid or using cells pretreated with neuraminidase.[13]

In this study, we characterize in further detail the binding of TgMIC1-MARR to five sialyl trisaccharides by microarray analyses. Interestingly, differing binding potencies to sialyl ligands are observed when they are presented as multivalent probes that simulate presentation at the host cell surface.[15] We demonstrate that TgMIC1-MARR is not only able to bind a diverse array of sialyl oligosaccharides but can discriminate between different oligosaccharide linkages, which may explain that while *T. gondii* has a predilection for a variety of cell types, the parasite nevertheless exhibits markedly varied virulence *in vivo* in different hosts. Using new crystal structures, we also provide the atomic resolution basis for the differential recognition of 3′ and 6′SiaLacNAc$_{1-4}$ (with the Galβ1–4 linkage known as Type 2 backbone[16]; Scheme 1) by TgMIC1-MARR, and make comparisons with the Type 1 chain isomer of 3′SiaLacNAc$_{1-4}$ (designated 3′SiaLacNAc$_{1-3}$). Furthermore, having observed in exploratory microarray analyses that analogs of 3′SiaLacNAc$_{1-4}$ and 3′SiaLacNAc$_{1-3}$ with a fluorine substitution at C-2 of galactose (Gal) elicit significantly higher binding signals than their nonfluorinated analogs,[17] these compounds are also examined and we observe a C—F···H—O hydrogen bond. We compare results of TgMIC1-MARR with those of the well-characterized sialic acid recognizing plant lectin wheat germ agglutinin (WGA). Furthermore, comparison with structures of other oligosaccharide-protein complexes reveals an analogous mode of interaction in three unrelated lectins, suggesting a new family of binding motifs with diverse scaffolds.

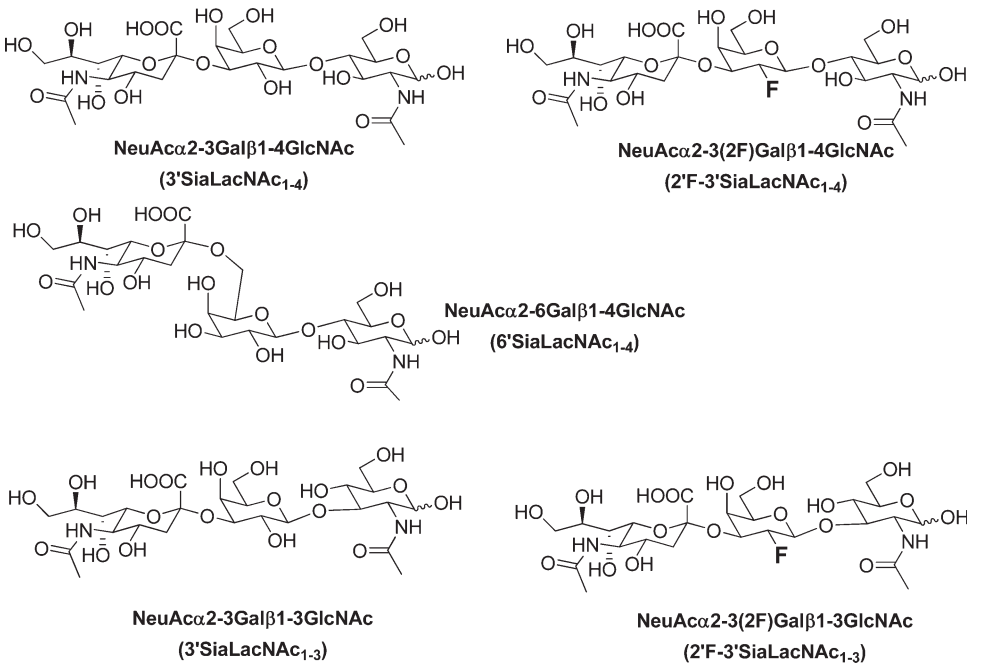
Results

Synthesis of fluorinated sialyl trisaccharides to probe the ligand tolerance of TgMIC1-MARR

Although fluorinated enzyme inhibitors have been elegantly used to investigate sugar processing mechanisms,[18] effective fluorosugar ligands for cell adhesion proteins have not been. We reasoned that, as well as potentially revealing novel features in the mode of action, such unnatural oligosaccharides are also putative adhesion inhibitors. Inhibitors with abiotic substitutions (i.e., 2′-F) are potentially resistant to endogenous enzyme processing. The unnatural synthetic oligosaccharides 2′F-3′SiaLacNAc$_{1-4}$ and 2′F-3′SiaLacNAc$_{1-3}$ were prepared through a chemoenzymatic strategy:[17] suitably protected *N*-acetylglucosamine (GlcNAc) derivatives were chemically glycosylated using trichloroacetimidate methodology to introduce the 2F-Gal or Gal with appropriate connectivity (β1–3 or β1–4). Sialic acid was introduced using tolerant trans-sialidase from *T. cruzi* to create the 3′-siayl trisaccharide variants.

Carbohydrate microarray analyses corroborate the TgMIC1-MARR preference of 2-3-linked sialyl oligosaccharides and reveal superior binding to fluorinated analogs

To compare closely the binding responses of TgMIC1-MARR to the immobilized 3′ and 6′SiaLacNAc$_{1-4}$



Scheme 1. Structures of the five sialyl-*N*-acetyllactosamine analogs investigated.

probes and to the 3′SiaLacNAc$_{1-3}$ isomer, as well as to the fluorinated probes, 2′F-3′SiaLacNAc$_{1-4}$ and 2′F-3′SiaLacNAc$_{1-3}$ (Scheme 1 and Table I), we generated microarrays in a dose-response format [Fig. 1(a,c)]. The two nonsialylated analogs LacNAc$_{1-4}$ and LacNAc$_{1-3}$ were included as negative controls. These oligosaccharides were arrayed as oxime-linked neoglycolipids with ring-closed monosaccharide cores.[19] Also included in the arrays, as controls, were six naturally occurring gangliosides (Table I). In accord with earlier microarray screening data,[13] TgMIC1-MARR gave significant binding signals with all of the sialyl probes tested except GD2 and GD3, which terminate in a 2-8-linked disialyl moiety and the two neutral disaccharides LacNAc$_{1-4}$ and LacNAc$_{1-3}$ [Fig. 1(a,b) and Table I]. As the binding curves were almost linear up to 2 fmol per spot, intensities could be compared at this point, relative to that of 3′SiaLacNAc$_{1-4}$ taken as 1.0 (Table I) to indicate binding potencies.

The relative binding score of 3′SiaLacNAc$_{1-3}$, 1.2, was comparable to that of 3′SiaLacNAc$_{1-4}$, whereas that of 6′SiaLacNAc$_{1-4}$ was considerably lower, 0.3. Strikingly, the strongest binding signals, 4.1 and 3.3, were with the fluorinated compounds 2′F-3′SiaLacNAc$_{1-4}$ and 2′F-3′SiaLacNAc$_{1-3}$, respectively; they even surpassed the signals elicited by the naturally occurring gangliosides GD1a and GT1b.

For comparison, we performed in parallel, microarray analysis of WGA, which is a well-characterized sialic acid-binding lectin.[20,21] As with TgMIC1-MARR, WGA gave very similar microarray responses for the nonfluorinated sialylsugars; that is no discrimination between 3′SiaLacNAc$_{1-4}$ and 3′SiaLacNAc$_{1-3}$ [Fig. 1(b,d) and Table I], and a clear discrimination between 3′SiaLacNAc$_{1-4}$ and 6′SiaLacNAc$_{1-4}$ (1.0 and 0.1, respectively), in accord with previous observations.[19,20] However, unlike TgMIC1-MARR, WGA gave much lower binding signals with the two fluorinated probes, 2F′-3′SiaLacNAc$_{1-4}$ and 2′F-3′SiaLacNAc$_{1-3}$ (0.5 and 0.4, respectively), relative to their unmodified analogs.

To investigate whether the stronger binding of TgMIC1-MARR to the fluorinated sialyl trisaccharides was specific and to compare the binding strengths of 2-3-linked and 2-6-linked sialyl oligosaccharides in the monovalent state, we performed "on-array" inhibition assays using free oligosaccharides 3′ and 6′SiaLacNAc$_{1-4}$. An unrelated neutral trisaccharide, maltotriose (Glcα1-4Glcα1-4Glc) was included as a negative control. We observed that 3′SiaLacNAc$_{1-4}$ inhibited the high avidity binding to the immobilized fluorinated trisaccharide 2′F-3′SiaLacNAc$_{1-4}$, indicating that binding was specific; there was 40% inhibition at 1 mg/mL and almost complete inhibition at the highest concentration tested, 3 mg/mL. Also in accord with the binding data, the 6′SiaLacNAc$_{1-4}$ was less active: there was no inhibition detected at 1 mg/mL; and $\sim$ 70% inhibition at 3 mg/mL [Fig. 2(a)]. The binding of WGA to the immobilized 3′SiaLacNAc$_{1-4}$ probe was also inhibited more strongly by 3′SiaLacNAc$_{1-4}$ than by 6′SiaLacNAc$_{1-4}$ [Fig. 2(b)]. To test whether the fluorosugars were specific inhibitors of the sialic acid-dependant cell binding,[13] a competition cell binding assay was performed with 2′F-3′SiaLacNAc$_{1-4}$. Indeed, effective inhibition of cell binding by TgMIC1 was achieved in 10–500 μM range of 2′F-3′SiaLacNAc$_{1-4}$ [Fig. 2(c)].

Table I. *Comparison of Binding Signals with TgMIC1-MARR (40 μg/mL) and WGA (1 μg/mL) Elicited by Lipid-Linked Sialylated Probes in the Carbohydrate Microarray Binding Analysis*

		Binding signals[a] at 2 fmol/spot			
		TgMIC1-MARR		WGA	
Name	Sequence	Score	Ratio	Score	Ratio
LacNAc (a)[b]	Galß1-4GlcNAc	0	0	514	<0.1
$LacNAc_{1-3}$ (b)	Galß1-3GlcNAc	0	0	0	0
$3'SiaLacNAc_{1-4}$ (c)	NeuAcα2-3Galß1-4GlcNAc	3,488	1.0	15,838	1.0
$6'SiaLacNAc_{1-4}$ (d)	NeuAcα2-6Galß1-4GlcNAc	1,160	0.3	1,696	0.1
$3'SiaLacNAc_{1-3}$ (e)	NeuAcα2-3Galß1-3GlcNAc	4,203	1.2	17,613	1.1
$2'F\text{-}3'SiaLacNAc_{1-4}$ (f)	NeuAcα2-3(2F)Galß1-4GlcNAc	14,287	4.1	7,211	0.5
$2'F\text{-}3'SiaLacNAc_{1-3}$ (g)	NeuAcα2-3(2F)Galß1-3GlcNAc	11,486	3.3	5,669	0.4
Hematoside (h)	NeuAcα2-3Galß1-4Glcß-Cer	2,796	0.8	2,489	0.2
Sial pg	NeuAcα2-3Galß1-4GlcNAcß1-3Galß1-4Glcß-Cer	1,373	0.4	2,723	0.2
GD1a	NeuAcα2-3Galß1-3GalNAcß1-4Galß1-4Glcß-Cer | NeuAcα2-3	3,387	1.0	778	<0.1
GD2	GalNAcß1-4Galß1-4Glcß-Cer | NeuAcα2-8NeuAcα2-3	0	0	504	<0.1
GD3	NeuAcα2-8NeuAcα2-3Galß1-4Glcß-Cer	26	<0.1	816	0.1
GT1b	NeuAcα2-3Galß1-3GalNAcß1-4Galß1-4Glcß-Cer | NeuAcα2-8NeuAcα2-3	4,742	1.4	1,379	0.1

[a] Given as numerical score and ratio relative to 3′SiaLacNAc taken as 1.
[b] a, b, etc., position in Figure 1.

Overall structures

The preference of $3'SiaLacNAc_{1-4}$ over the 2-6-linked analog, and the superior binding of TgMIC1-MARR to the fluorinated sialyl trisaccharides in the microarray analyses prompted us to look into the details of the interactions at the atomic level. The crystal structures of TgMIC1-MARR in complex with either $2'F\text{-}3'SiaLacNAc_{1-4}$, $3'SiaLacNAc_{1-3}$, or $2'F\text{-}3'SiaLacNAc_{1-3}$ were solved at 2.0 Å (Table II). Residues 1–12 and 244–246 were not observed in the crystal structures, probably due to flexibility at the termini. Overall the electron density is well resolved, although within the poorest region of the maps (the disordered loop between Q80 and N82) the atomic positions could not be determined and were either omitted from the final model ($3'SiaLacNAc_{1-3}$ and $2'F\text{-}3'SiaLacNAc_{1-4}$ soaks) or modeled as polyalanine ($2'F\text{-}3'SiaLacNAc_{1-4}$ soak). Although it was possible to model all of the carbohydrate chain in the $3'SiaLacNAc_{1-3}$ and $2'F\text{-}3'SiaLacNAc_{1-3}$ soaks, electron density for the GlcNAc residue of $2'F\text{-}3'SiaLacNAc_{1-4}$ was of insufficient quality to permit its inclusion in the final model [Fig. 3(a–c)].

Each of the two MAR domains is formed from a distorted five strand β-barrel packed against two helices and contains two highly conserved disulphide bonds. The N-terminal MAR domain (MAR1) has an extended helix 1 and subsequent loop region that is anchored via an additional disulfide bond while the C-terminal MAR domain (MAR2) has an extra β-wing at the C-terminus, which is anchored to the main body through two additional disulfide bridges. The crystal structures of TgMIC1-MARR-$3'SiaLacNAc_{1-4}$ (pdb:2Jhd), TgMIC1-MARR-$6'SiaLacNAc_{1-4}$ (pdb:2Jh7)[13] and the structures described here were superimposed onto unliganded TgMIC1-MARR (pdb:2Jh1)[13] giving an average r.m.s.d between C^{α} residues of 0.18 Å. As anticipated the overall secondary structures of TgMIC1-MARR are indistinguishable from those published [Fig. 3(d)], although minor differences occur in the ligand binding site (see below) and deviations are observed within a dynamic loop region corresponding to residues L76-N82.

Carbohydrate-TgMIC1-MARR interactions

Because of the lack of global and local changes to the structure of TgMIC1-MARR upon binding of sialylated oligosaccharides, a detailed analysis of hydrogen bonding between the protein and carbohydrate was undertaken (Table III and Fig. 4) to provide explanations for the differing binding avidities and the intriguing

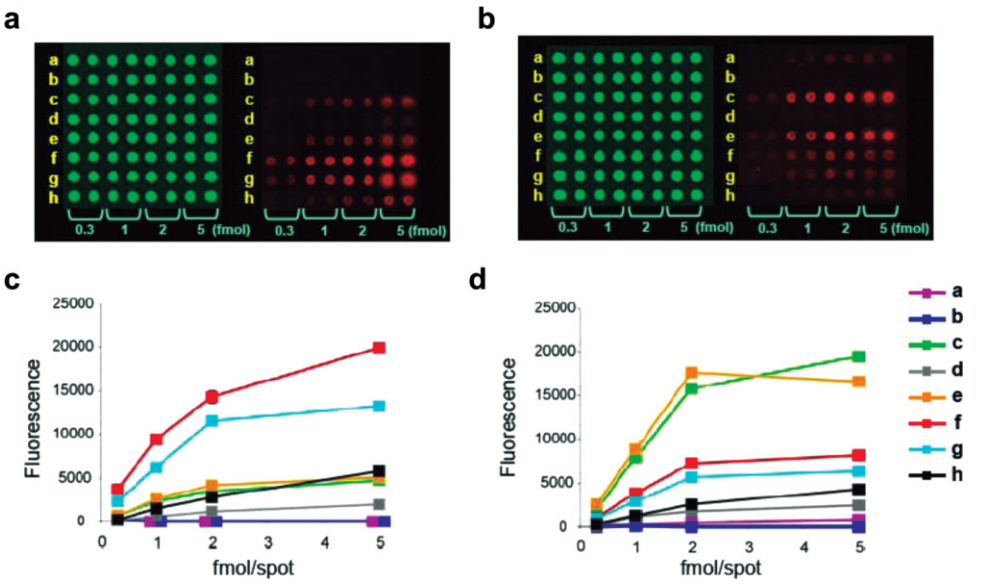


Figure 1. Microarray analyses of TgMIC1-MARR and WGA in a dose response format. Eight lipid-linked oligosaccharide probes, a–h, (sequences in Table I) were printed in duplicate at the indicated levels on nitrocellulose-coated glass slides with Cy3 dye included as a marker (green emission). Binding was detected with Alexa Fluor647-labeled streptavidin (red emission). Images for TgMIC1-MARR (40 μg/mL) and WGA (1 μg/mL) are in panels (a) and (b), respectively. Corresponding binding curves are in (c) and (d), respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

enhancements with fluorinated ligands observed in microarray experiments.

All direct hydrogen bond interactions between the trisaccharides and TgMIC1-MARR are between the sialic acid (NeuAc) residue and three residues in the MAR2 domain: K200 (two hydrogen bonds), H202 (two hydrogen bonds), and T204 (three hydrogen bonds). In all structures except 3′SiaLacNAc$_{1-4}$ a structured water molecule (WAT170) mediates intra-carbohydrate interactions between the NeuAc ring oxygen, NeuAc glycerol chain and Gal, stabilizing the conformation of the glycerol moiety such that its two terminal hydroxyls can hydrogen bond to T204 and E205 via another water molecule (WAT145) [Fig. 4(a–d,g–j)]. Although the glycerol moiety in 3′SiaLacNAc$_{1-4}$ is rotated so that H-bonds to WAT145 cannot be made, the relative orientation of the NeuAc and Gal rings of the carbohydrate is very similar to that observed for 3′SiaLacNAc$_{1-3}$. In all the complexes of all the 2-3-linked sialyl trisaccharides [Fig. 4(a–h)], an addition water molecule (WAT129) mediates interactions between TgMIC1-MARR E206 and the Gal O6 position.

Although the GlcNAc residue at the reducing end is well resolved in the electron density maps of the 3′SiaLacNAc$_{1-3}$ and fluorine substituted structures [Fig. 3(b,c)], it makes no contacts with the MAR2 binding domain. Its position appears to be stabilized by contacts from symmetry-related proteins in the crystal; in solution this group is likely to be disordered. In the 3′SiaLacNAc$_{1-4}$, 2′F-3′SiaLacNAc$_{1-4}$ and 6′SiaLacNAc$_{1-4}$ bound structures,[13] these additional

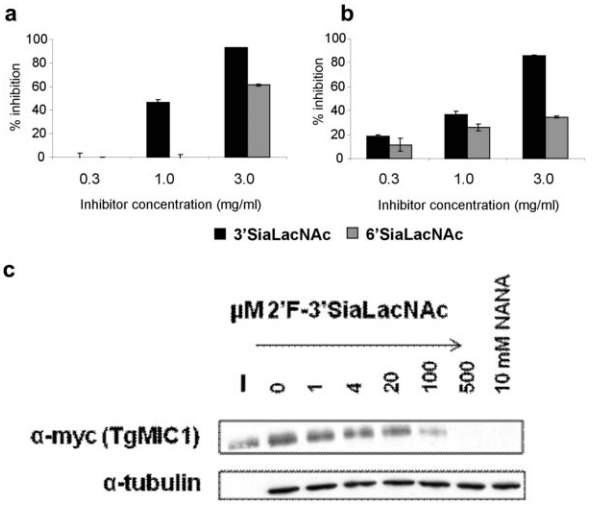


Figure 2. Comparison of the inhibitory activities of free oligosaccharides toward the binding of TgMIC1-MARR. Oligosaccharides 3′SiaLacNAc$_{1-4}$, 6′SiaLacNAc$_{1-4}$, and maltotriose (a negative control) were tested at the indicated concentrations for inhibition of binding of TgMIC1-MARR at 4 μg/mL, to immobilized 2′F-3′SiaLacNAc$_{1-4}$ probe (a), and of WGA at 0.5 μg/mL to immobilized 3′SiaLacNAc$_{1-4}$ probe (b). The results are means of two separate inhibition experiments with error bars. Cell binding competition experiments with soluble 2′F-3′SiaLacNAc$_{1-4}$ were performed using *P. pastoris* culture supernatants expressing TgMIC1myc. Anti-myc antibodies were used as to probe for bound TgMIC1 using Western blots. The input (I) and cell bound fraction at various concentrations of soluble 2′F-3′SiaLacNAc$_{1-4}$ are shown. Tubulin was used as a control for use of equivalent amounts of cell-material (c).

Table II. *Data Collection and Refinement Statistics*

	2′F-3′SiaLacNAc$_{1-4}$ soak	3′SiaLacNAc$_{1-3}$ soak	2′F-3′SiaLacNAc$_{1-3}$ soak
Crystal parameters			
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions (Å, Å)	66.36, 172.37	66.23, 172.19	66.33, 172.09
Data collection			
Beamline	ESRF BM14	ESRF BM14	ESRF BM14
Wavelength (Å)	0.934	0.934	0.934
Resolution (Å)	29.68–2.00 (2.05–2.00)	27.75–2.00 (2.05–2.00)	29.66–2.00 (2.05–2.00)
Unique observations	23,882	23,609	23,986
R_{sym}	0.075 (0.246)	0.074 (0.2228)	0.076 (0.222)
$\langle I/\sigma I\rangle$	12.7 (4.6)	11.2 (4.7)	10.7 (5.6)
Completeness (%)	98.9 (98.8)	99.3 (100)	99.4 (100)
Redundancy	3.7 (3.7)	3.5 (3.6)	3.9 (3.9)
Refinement			
R_{work}/R_{free} (%)	16.8/19.1	16.5/18.9	16.5/20.3
Number of protein residues	231	228	228
Number of carbohydrate residues	1 NeuAc, 1Gal, 5 Glycerol	1 NeuAc, 1Gal, 1 GlcNAc, 5 Glycerol	1 NeuAc, 1Gal, 1 GlcNAc, 5 Glycerol
Number of noncarbohydrate ligands	8 Acetates	8 Acetates	8 Acetates
Number of ions	1 Cl ion	1 Cl ion	1 Cl ion
Number of water molecules	174	189	191
rmsd stereochemistry			
Bond lengths (Å)	1.4	1.4	1.4
Bond angles (°)	0.014	0.013	0.014
Ramachandran analysis (%)			
Residues in most favored regions	89	91.3	91.3
Residues in additionally allowed regions	11	8.2	8.2
Residues in generously allowed regions	0.5	0.5	0.5
Residues in disallowed regions	0	0	0

Numbers in parentheses refer to the outermost resolution shell.
$R_{sym} = \sum |I - \langle I\rangle|/\sum I$ where I is the integrated intensity of a given reflection and $\langle I\rangle$ is the mean intensity of multiple corresponding symmetry-related reflections.
$R_{work} = \sum ||F_o| - |F_c||/\sum F_o$ where F_o and F_c are the observed and calculated structure factors, respectively.
$R_{free} = R_{work}$ calculated using ~ 10% random data excluded from the refinement.
rmsd stereochemistry is the deviation from ideal values.
Ramachandran analysis was carried out using *PROCHECK*.[22]

interactions are not accessible because of steric clashes between the GlcNAc and symmetry-related TgMIC1-MARR; this residue remains disordered and is not seen in the maps [Fig. 3(a)]. These observations are consistent with microarray data, in which TgMIC1-MARR binds equally well to Type 1 and Type 2 backbones.

In contrast to the lack of discrimination of Type 1 and Type 2 backbones, a consistent and reproducible preference for 3′SiaLacNAc$_{1-4}$ relative to 6′SiaLacNAc$_{1-4}$ is observed in the microarray analyses. The structure of 6′SiaLacNAc$_{1-4}$ bound to Tg-MIC1-MARR,[13] reveals a similar network of interactions for the NeuAc moiety to those observed in the 2-3-linked structures [Fig. 4(i)].

However, the 2-6 sialyl linkage results in Gal being flipped by about 180° which puts the ring oxygen (O5) in a position close (~2 Å) to that occupied by the Gal O2/F2 in the other oligosaccharide structures [Fig. 4]. Stabilization of the oligosaccharide conformation is still mediated through WAT170, but this water has moved closer toward Gal O6 and the hydrogen bonding to galactose is via the Gal ring oxygen (O5) instead of Gal O2/F2 as in the other crystal structures. Most striking, however, is the finding that this altered galactose orientation abrogates the WAT129-mediated interactions between TgMIC1-MARR and Gal that is observed in the structures of all the 2-3 linked SiaLacNAc$_{1-4}$ analogs tested [Fig. 4(a–d)].

Fluorination of 3′SiaLacNAc$_{1-3}$ at Gal C2 has no effect on the bound conformation of the carbohydrate moieties [Figs. 4(a–d) and 5(a)] water molecules. However, fluorination of 3′SiaLacNAc$_{1-4}$ at Gal C2 has a significant effect on both carbohydrate structure [Figs. 4 and 5(b)] and hydrogen bond network. In this case, the glycerol side chain is flipped in the nonfluorinated version and WAT170 is absent. The introduction of the fluorine atom appears to stabilize the hydrogen bond formation with the bound water, WAT170 [Fig. 4(g)], with a concomitant change in conformation of

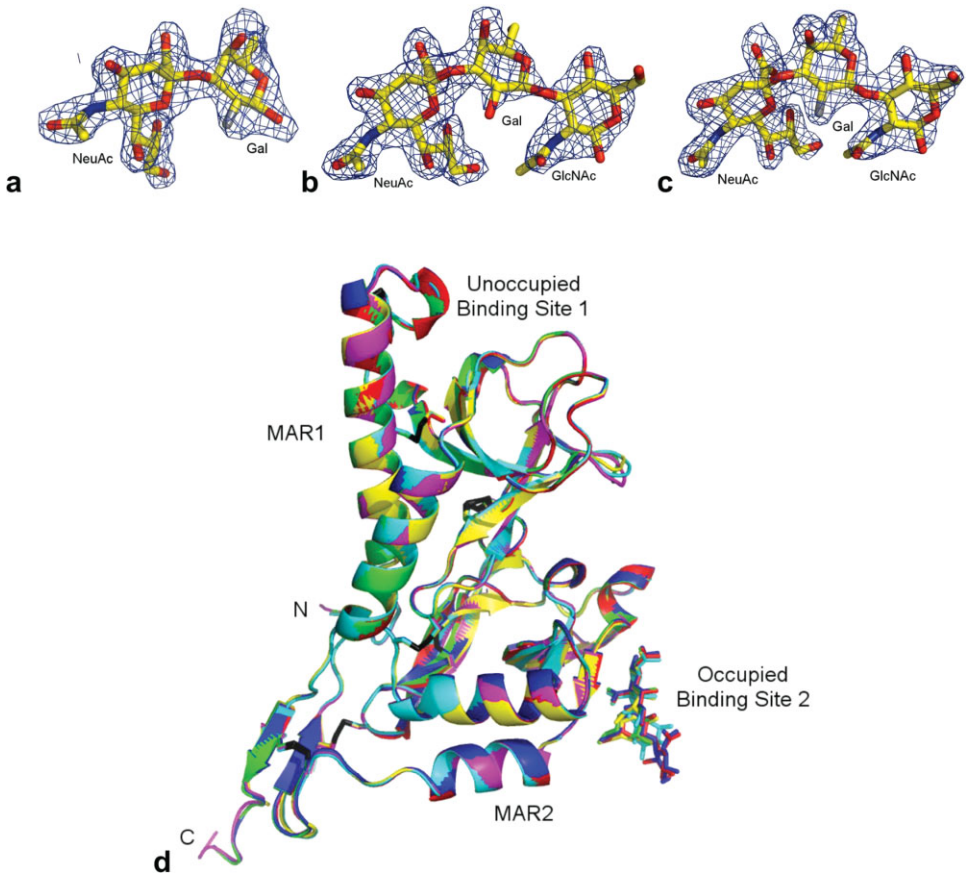


Figure 3. Sialyl trisaccharides bound at the TgMIC1-MAR2 host cell adhesion site. Weighted $2|F_o| - |F_c|$ electron density maps of (a) 2′F-3′SiaLacNAc$_{1-4}$, (b) 3′SiaLacNAc$_{1-3}$ and (c) 2′F-3′SiaLacNAc$_{1-3}$, contoured at 1.0 r.m.s. (d) Overlay of all six TgMIC1-MARR crystal structures: no ligand (magenta; pdb:2Jh1,[23] 3′SiaLacNAc$_{1-4}$ soak (yellow; pdb:2Jhd), 2′F-3′SiaLacNAc$_{1-4}$ soak (green; pdb:3f53), 3′SiaLacNAc$_{1-3}$ soak (red; pdb:3f5a), 2′F-3′SiaLacNAc$_{1-3}$ soak (blue; pdb:3f5e) and.6′SiaLacNAc$_{1-4}$ soak (cyan; pdb:2Jh7). Disulphide bridges are shown as black sticks and the N and C-termini are labeled. The two MAR domains and their binding sites are highlighted, with binding site 2 occupied with sialyated trisaccharide shown as sticks. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table III. *Interatomic Distances Between TgMIC1-MARR and Carbohydrate Ligand*

		Distance (Å)				
		3′SiaLacNAc$_{1-4}$ soak	2′F-3′SiaLacNAc$_{1-4}$ soak	3′SiaLacNAc$_{1-3}$ soak	2′F-3′SiaLacNAc$_{1-3}$ soak	6′SiaLacNAc$_{1-4}$ soak
K200 O	NeuAc O4	2.8	2.6	2.6	2.5	2.8
K200 O	NeuAc N5	3.3	3.3	3.2	3.3	3.3
K200 O	NeuAc N5	2.8	2.7	2.7	2.8	2.8
H202 Nε2	NeuAc O1B	3.4	3.1	3.2	3.2	3.3
T204 Oγ1	NeuAc O1B	2.7	2.7	2.7	2.6	2.7
T204 Oγ1	NeuAc O1A	3.4	3.4	3.5	3.4	3.5
T204 N	NeuAc O1A	2.9	2.8	2.8	2.7	2.8
E205 Oε	NeuAc 09	3.3	—	—	—	—
WAT145	T204 O	—	3.0	3.1	3.1	2.9
WAT145	E205 Oε2	—	3.2	3.1	2.9	2.5
WAT145	NeuAc O8	2.9	2.9	2.6	2.9	2.6
WAT145	NeuAc O9	3.4	3.4	3.3	3.3	3.2
WAT170	Gal O2/F2/O5	—	3.0	2.7	3.0	3.1
WAT170	SIA O6	—	3.4	3.3	3.4	3.2
WAT170	SIA O7	—	3.0	2.9	3.2	3.3
WAT129	Gal O6	3.5	3.3	3.4	3.4	—
WAT129	E206 N	3.0	3.2	3.2	3.1	—

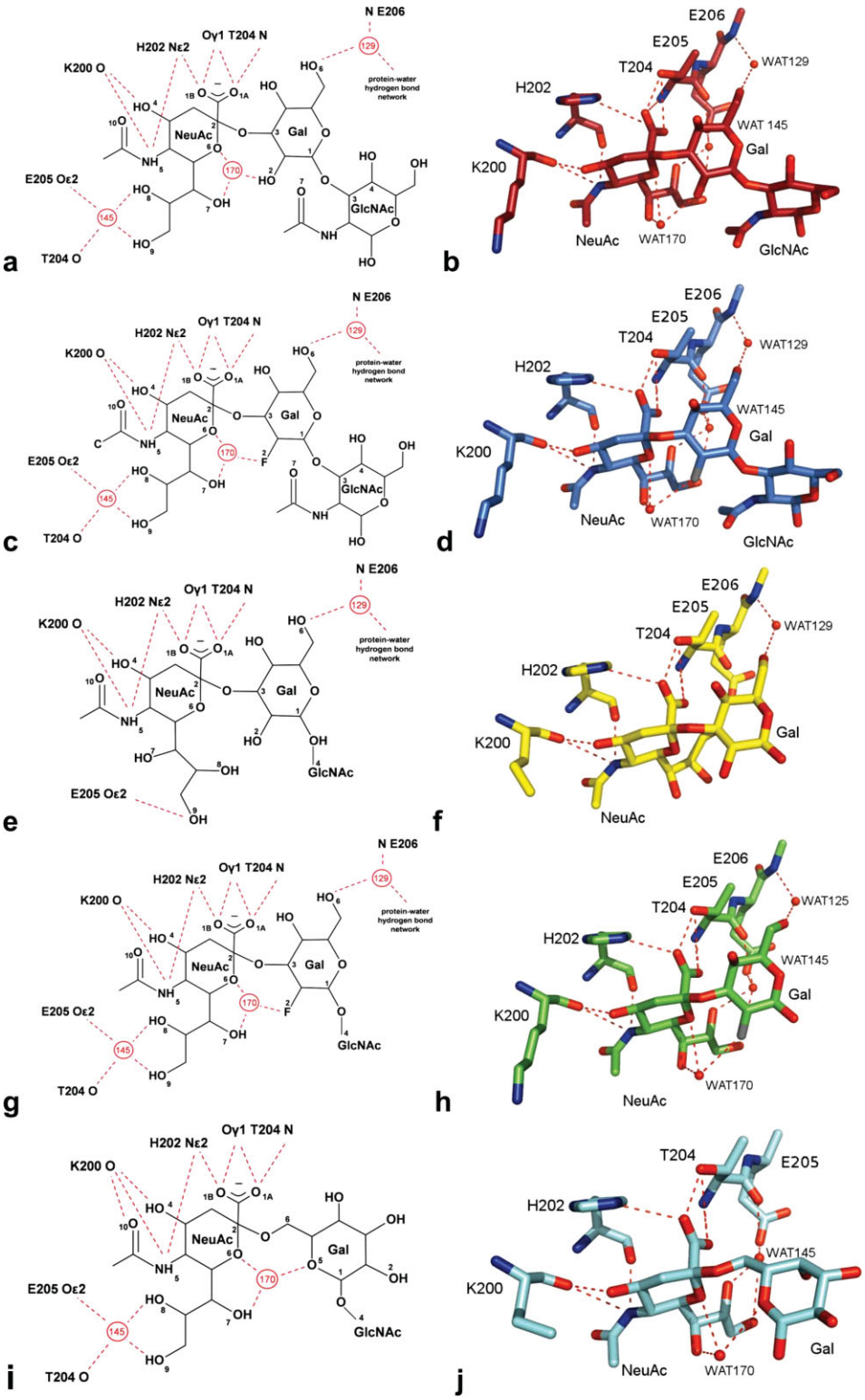


Figure 4. The occupied binding site of TgMIC1-MAR2. (a, b) 3′SiaLacNAc$_{1-3}$ soak, (c, d) 2′F-3′SiaLacNAc$_{1-3}$ soak, (e, f) 3′SiaLacNAc$_{1-4}$ soak, (g, h) 2′F-3′SiaLacNAc$_{1-4}$ soak, and (i, j) 6′SiaLacNAc$_{1-4}$ soak. (a, c, e, g, i) Schematic representation of the occupied TgMIC1-MAR2 binding site with hydrogen bonds shown as red dashed lines, waters as numbered red circles and key atoms within the carbohydrate ligand are numbered (water molecules 1129, 1145, and 1170 in the 2′F-3′SiaLacNAc$_{1-4}$, 3′SiaLacNAc$_{1-3}$ and 2′F-3′SiaLacNAc$_{1-3}$ soak structures are renamed as 129, 145, and 170, respectively, for consistency with the 3′SiaLacNAc$_{1-4}$ and 6′SiaLacNAc$_{1-4}$ soaks). (b, d, f, h, j) Crystal structure of the occupied TgMIC1-MAR2 binding site, with carbon atoms coloured as in Figure 3(d) and hydrogen bonds shown as red dashed lines. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

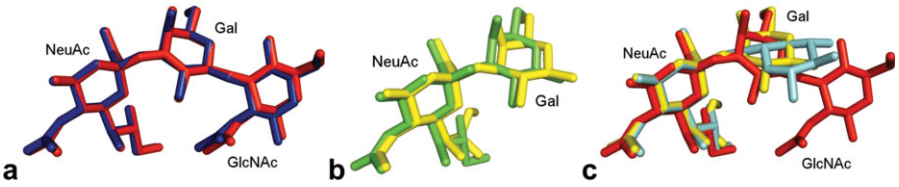


Figure 5. Overlay of sialyl oligosaccharides. (a) Comparison of fluorinated (blue; pdb:3f5e) and nonfluorinated (red; pdb:3f5a) 3′SiaLacNAc$_{1-3}$. (b) Superimposition of fluorinated (green; pdb:3f53) and nonfluorinated (yellow; pdb:2Jhd) 3′SiaLacNAc$_{1-4}$. The glycerol side chain of sialic acid is in two separate conformations. (c) Overlay of 3′SiaLacNAc$_{1-3}$ (red), 3′SiaLacNAc$_{1-4}$ (yellow)[15] and 6′SiaLacNAc$_{1-4}$ (cyan; pdb:2Jh7). Deviations are seen within the alternate conformation of the glycerol side chain of sialic acid in 3′SiaLacNAc$_{1-4}$ and within galactose, where due to the β1–4 linkage in 6′SiaLacNAc$_{1-4}$, it is the ring oxygen (O5) of galactose, rather than O2/F2, which now hydrogen bonds to WAT170. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

the NeuAc glycerol side chain to the same as that observed in the other structures.

Discussion

T. gondii is uniquely adapted to infect a wide range of hosts and to invade virtually all cell types. The remarkable range of sialyl oligosaccharide sequences recognized by TgMIC1-MARR in carbohydrate microarrays accounts for this wide cellular tropism. Sialic acid containing carbohydrates are found at the termini of animal glycoconjugates. Recognition of the sialic acid moiety is often affected by specific structural variations of the monosaccharide and its linkage along the sugar chain.[24] Although TgMIC1-MARR does not discriminate between Type 1 and Type 2 backbones, a consistently observed preference is for 2-3 over 2-6-linked sialyl oligosaccharides.

The major oligosaccharide binding activity lies within the second MAR domain (MAR2), with a major contact occurring between a threonine residue (T204) and the carboxyl group of sialic acid. An extensive database analysis of other oligosaccharide-protein complexes reveals an analogous mode of interaction in three lectins unrelated to TgMIC1, namely the *Bordetella pertussis* toxin, Staphylococcal superantigen-like (SSL) proteins, and the Rotavirus spike-associated carbohydrate binding domain[25–28] (see Fig. 6). These lectins do not possess MAR domains but recognize sialyl oligosaccharides using a similar motif (see Fig. 6) that comprises a short, linear stretch of amino acids containing an essential threonine/serine residue. The structures overlay with great similarity, although the arginine is absent in TgMIC1 this role has been replaced by WAT145, it appears that within a hydrophilic environment a universal binding site exists for sialic acid. Interestingly, although the structures of these binding sequences are very similar, the motif is embedded in unrelated protein scaffolds suggesting convergence to a common binding function. Thus, our structural studies provide detailed insight into carbohydrate recognition by this new family of sialic-acid binding sites. Extending the search of the protein database (PDB) for 3′SiaLac containing motifs bound to unrelated lectins; examples include *Maackia amurensis* leukoagglutinin[29] (green) (pdb:1DBN), mouse sialoadhesin[30] (orange) (pdb:1QFO), Staphylococcal enterotoxin B[31] (blue) (pdb:1SE3), and human Siglec-5[32] (yellow) (pdb:2ZG3). In these cases, the overall conformation of 3′SiaLac in these complexes vary significantly, which is likely due to the inherent flexibility in the linkage between the two sugars.

Structural comparisons between the 2-3- and 2-6-linked sialyl oligosaccharides bound to TgMIC1-MAR2

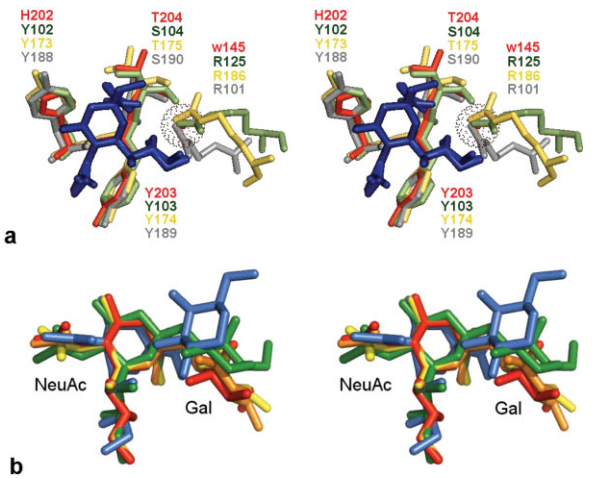


Figure 6. (a) Overlay of a common binding motif for recognition of sialic acid. Stereo image showing superimposition of sialic acid (Neu5Acα2Me in pdb:2i2s) within the sialic acid binding site of TgMIC1-MARR (red; pdb:3f5a), *Bordetella pertussis* toxin (green; pdb:1pto). *Staphylococcus aureus* SSL5 toxin (sand; pdb:3f5a) and Rotavirus VP8* (gray; pdb:2i2s). Amino acid residues, sialic acid (blue) and Neu5Acα2Me (blue) are shown as sticks while water 145 of TgMIC1 (red) is represented as a sphere of dots. (b) Comparison of glycosidic bond in 3′SiaLactose containing oligosaccharides. A search of the protein database (PDB) for 3′SiaLactose containing motifs bound to proteins revealed numerous lectins including *Maackia amurensis* leukoagglutinin (green) (pdb:1DBN), mouse sialoadhesin (orange) (pdb:1QFO), Staphylococcal enterotoxin B (blue) (pdb:1SE3), and human Siglec-5 (yellow) (pdb:2ZG3). The oligosaccharides from these complexes were superimposed onto either the NeuAc or Gal residue of 3′SiaLacNAc$_{1-3}$ (red) bound to TgMIC1-MARR. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

reveal that although recognition is primarily through the terminal NeuAc residue several contributions from coordinated water molecules are important for the Gal ring orientation (see Fig. 4). WAT129 mediates the only interaction between protein (E206) and Gal O6 position in the 2-3-linked sequences examined. In contrast, in the 2-6-linked analog, the Gal O6 position is occupied by the sialic acid linkage and the Gal residue is shifted away from the sialic acid moiety by one bond, leading to the loss of the WAT129 water-coordinated carbohydrate-protein contact. The absence of this structured water molecule explains the observed binding preference for 2-3-linked sialyl oligosaccharides in our microarray analyses. Furthermore, our findings also indicate that E206 is important for determining this binding preference (see Fig. 4). Another contributing factor is the increased flexibility of the 2-6 linkage over their 2-3 counterparts, which would contribute an additional entropic penalty upon ordering in the complex.

Relative binding responses were greatly enhanced to unnatural sialyl trisaccharides that were fluorinated at the C2 position of the Gal residue. Moreover, these binding signals were greater than those with the natural glycosyl ceramides GD1a and GT1b (Table I), which are among the best ligands in our previous screening microarray studies.[13] The increased binding of TgMIC1-MARR to the fluorinated compounds could be fully inhibited by the natural glycan sequence 3′SiaLacNAc$_{1-4}$. Furthermore, host cell binding could be competitively inhibited with soluble fluorinated sugars indicating that the binding is specific. Such enhanced binding of fluorosugar ligand analogs to a lectin is unprecedented. Interestingly, despite recording the highest binding potencies in microarray experiments, many structural aspects of the fluorinated analogs are very similar to their hydroxyl counterparts.

In the structures of 3′SiaLacNAc$_{1-3}$ and 6′SiaLacNAc$_{1-4}$ complexes the sialic acid and galactose rings coordinate a structured water molecule, namely WAT170. WAT170 resides in a highly electronegative environment and while it has no direct contacts with the protein, it is coordinated by intracarbohydrate hydrogen bonds with NeuAc O6 (ring oxygen), NeuAc 7-OH, and Gal O2 positions. The introduction of the fluorine atom at the Gal-2 position preserves the geometry of this hydrogen bond network and the close distance to the water molecule suggests that a hydrogen bond is maintained between WAT170 and the fluorine atom (Table III). The occurrence of C—F···H—O hydrogen bonds in protein ligand complexes is uncommon and their authenticity has been the subject of much debate.[33–35] While a fluorine atom retains the ability to accept a hydrogen bond, the effects of its increased electronegativity are not straightforward. In four examples out of 18 from the PDB the fluorine acts as a hydrogen bond donor.[23,36,37] More striking evidence for C—F...H—O hydrogen bonding can be found in the 3′SiaLacNAc$_{1-4}$ structures. WAT170 is only present in the fluorinated version, which presumably reflects a stabilization resulting from improved geometry or strength of the fluorine-mediated hydrogen bond. It is interesting to note that, despite similar processing and refinement statistics, in all cases the electron density for fluorinated ligands is better defined, indicating perhaps a more ordered state.

Although WGA displays a similar preference for 2-3-linked over 2,6-linked sialyl oligosaccharides, the effect of fluorination was opposite, namely much diminished binding. The structure of WGA in complex with 3′sialyllactose (3′SiaLac)[20,21] reveals a more hydrophobic binding pocket than that of TgMIC1-MARR, with several aromatic side-chains participating in CH-π interactions with the carbohydrate (see Fig. 7). In this case, the galactose ring is oriented by hydrogen bonds between the Gal 2-OH, Tyr66-OH, and a single water molecule. The disruption of these contacts by the 2-6-linkage accounts for the 2-3 preference in microarray experiments, as with TgMIC1-MARR. The negative effect of fluorination observed here may be explained by the loss of a direct hydrogen bond between the 2-position of Gal and the ring oxygen of the NeuAc in the fluorinated structure, as fluorine cannot act as a hydrogen bond donor like Gal 2-OH. In addition to hydrogen bond strength and geometry, altered binding potencies of fluorinated ligands have also been attributed to the increased hydrophobic nature of the fluoride group.[22] A favorable contribution to the entropy is provided by sequestration of fluoride from the aqueous environment. It is plausible that this effect makes a substantial contribution to the increased binding strength observed for TgMIC1-MARR in the microarray analyses while maintaining the geometry of the hydrogen bond network. Furthermore, molecular orbital calculations and empirical observation both suggest that C—F groups are more likely to accept hydrogen bonds when in an electron rich environment.[33,34] The binding site of the bound complex of TgMIC1-MARR is indeed highly electronegative, with numerous oxygen atoms participating in intraoligosaccharide and intermolecular hydrogen bonds, and this most probably contributes to the productive participation of a fluorine atom.

The affinities of individual carbohydrate-protein interactions are low, which can be problematic when probing these interactions in the monovalent state in solution. Polyvalency is a critical feature of cell surface carbohydrate recognition.[24,39] In the carbohydrate microarray system used here, the oligosaccharides are presented as clustered lipid-linked probes, with potential for some lateral mobility, in two dimensional arrays on a chip surface.[15] This might offer a suitable mimic of the arrangement of clustered oligosaccharide structures at the cell surface, and allow for polyvalent binding, resulting in a much more realistic binding response. TgMIC1 is anchored on the parasite surface

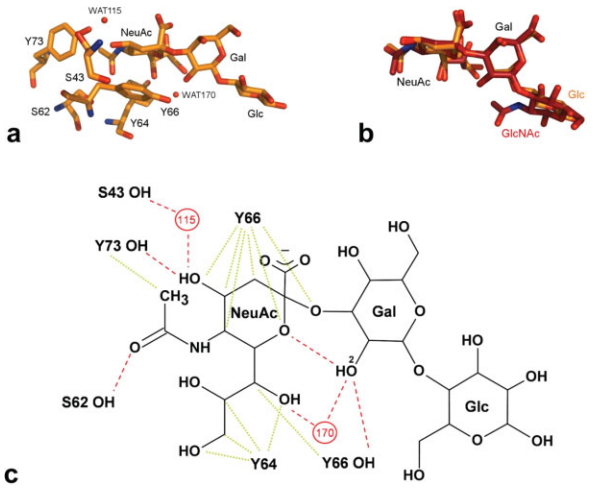


Figure 7. Evaluation of the binding modes of sialyl trisaccharides to WGA and TgMIC1-MAR2. (a) WGA bound to 3′SiaLac[20] (orange; pdb:1wgc) and (b) overlay of 3′SiaLacNAc$_{1-3}$ (red: pdb 3f5a) and 3′SiaLac. Amino acid residues and ligands are shown as sticks while water molecules are represented as spheres (WAT178 in pdb:1wgc has been renamed WAT170 for consistency with the TgMIC1-MARR drug soaks). (c) Schematic representation of WGA bound to 3′SiaLac adapted from Wright.[20] Hydrogen bonds are shown as red dashed lines, van der Waal interactions are green dotted lines, water molecules are numbered red circles and Gal O2 is labeled. TgMIC1-MAR2 interacts with 3′SiaLacNAc$_{1-3}$ on one side of the sialic acid moiety via specific hydrogen bonds [Fig. 4(a,b)], while WGA makes contacts on the opposite face through hydrogen bonds and van der Waals interactions. The only distinction between 3′SiaLac and 3′SiaLacNAc$_{1-3}$ is that galactose is β1–4 linked to glucose (Glc) rather than β1–3 linked to GlcNAc. NeuAc and Gal motifs overlay with little deviation while the terminal Glc or GlcNAc (which do not interact with the protein in either case) show variation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

in abundant multimeric assemblies in which at least two MIC1 molecules are present per complex.[12] The subtle affinity and specificity differences revealed by our structural data are most probably amplified in natural host cell interactions and effectively recapitulated in the microarray experiments.

The high avidity of binding of TgMIC1 to clustered natural ligands such as gangliosides that are abundant on neuronal cells is likely a factor for the tropism of *T. gondii* in the human brain. Furthermore, potent and rapid binding to a variety of sialylated receptors may play a role in a primary interaction, perhaps to sialic acid on the gut epithelial cell wall thereby preventing excretion of the parasite from the intestine after ingestion. Although TgMIC1 binds well to both clustered 2-3 and 2-6-linked sialyl sugars, the differential binding affinity may also contribute to the diverse spectrum of *T. gondii* virulence observed in mammalian hosts.

Our findings on the preferential binding of TgMIC1 to synthetic carbohydrates such as the fluorinated compounds could potentially be exploited in the design of therapeutic receptor analogs. Such treatments would be particularly important for immunocompromised patients who are at especially high risk. Our observations on a distinctive sialic acid-binding motif containing the essential threonine/serine residue shared between TgMIC1-MARR and three unrelated adhesive proteins provides the basis for the design of novel synthetic lectins able to discriminate various sialyl oligosaccharides.

Methods

Materials

Oligosaccharides 3′ and 6′SiaLacNAc$_{1-4}$ were from Dextra (Reading, UK). LacNAc$_{1-3}$, 3′SiaLacNAc$_{1-3}$ and the fluorinated analogs of 3′SiaLacNAc$_{1-4}$ and 3′SiaLacNAc$_{1-3}$ designated 2′F-3′SiaLacNAc$_{1-4}$ and 2′F-3′SiaLacNAc$_{1-3}$, respectively were chemically synthesized.[17] LacNAc and maltotriose were from Sigma. The natural glycolipids (Table I): hematoside and sialyl paragloboside (Sial pg) were gifts of Professor Peter Hanfland (University of Bonn); GD1a was from Sigma; GD2 and GT1b were from HyTest (Turku, Finland); GD3 was from BioCarb (Lund, Sweden). Mouse monoclonal anti-poly-histidine and biotinylated goat anti-mouse IgG antibodies were from Sigma. Biotinylated WGA was from Vector Laboratories (Peterborough, UK).

Expression of recombinant soluble TgMIC1-MARR

His-tagged TgMIC1-MARR (residues 17–262) was expressed and purified as described.[13]

Carbohydrate microarray binding assays

The sialyltrisaccharides (Scheme 1) and the two LacNAc analogs were converted into oxime-linked neoglycolipids[19] for arraying. These neoglycolipid probes, together with natural glycolipids, were robotically printed onto 16-pad nitrocellulose glass slides,[39] the layout of each pad being in a dose-response format [Fig. 1(a,c)]. Microarray binding analyses with the recombinant His-tagged TgMIC1-MARR were performed essentially as described.[13] In brief, the His-tagged TgMIC1-MARR was precomplexed with mouse monoclonal anti-poly-histidine and biotinylated anti-mouse IgG antibodies in a ratio of 1:1.25:1.25 (by weight). A control experiment was carried out in the absence of TgMIC1-MARR and no binding signal was detected. The binding analysis of WGA was performed as described.[19]

On-array inhibition of TgMIC1-MARR and WGA binding using free oligosaccharides as inhibitors

Binding of TgMIC1-MARR at 4 μg/mL and WGA at 0.5 μg/mL were carried out in the presence of free oligosaccharides 3′SiaLacNAc$_{1-4}$ and 6′SiaLacNAc$_{1-4}$, as well as maltotriose (a negative control). The three

oligosaccharides were preincubated, at different concentrations as indicated (see Fig. 2), to the TgMIC1-MARR (precomplexed) or WGA before overlaying on the microarray slides. The percentage of inhibition of binding in the presence of inhibitors was determined as follows using the fluorescence intensity observed with arrayed neoglycolipid probe of 2′F-3′SiaLacNAc$_{1-4}$ printed at 5 fmol/spot in the microarray: % Inhibition = [(means of duplicate spots with maltotriose − means of duplicate spots with 3′ or 6′SiaLacNAc$_{1-4}$)/means of duplicate spots with maltotriose] × 100%.

Competition cell binding assay

These were performed as described previously[13] using 2′F-3′SiaLacNAc$_{1-4}$ as the soluble ligand.

Crystallization and data collection

TgMIC1-MARR was crystallized as described.[13,40] Crystals were soaked overnight in 3.5 *M* ammonium acetate, 100 m*M* bis-tris propane pH 7.0, and either 3.0 m*M* of 2′F-3′SiaLacNAc$_{1-4}$, 3′SiaLacNAc$_{1-3}$ or 2′F-3′SiaLacNAc$_{1-3}$. Before data collection, crystals were soaked in cryoprotectant[13] and then frozen immediately. Data were collected at 100 K on beamline ID14-1 at the European Synchrotron Radiation Facility (ESRF).

Structure solution and refinement

All data were processed with *MOSFLM* and scaled using *SCALA*.[42] Phases were obtained by molecular replacement using *PHASER*[38] with TgMIC1-MARR (pdb:2jh1[13]) as the search model. Approximately 10% of the data were used to calculate R_{free}. Refinement was carried out with *REFMAC5*[43] and model building was performed with *Coot*.[44] During the final stages of refinement TgMIC1-MARR (chain A) was submitted to the TLS motion determination server.[44,45] The modified PDB and TLS input files (with a single TLS parameter describing the chain) were used with *REFMAC5*[43] for a restrained and TLS *B*-factor refinement. The protein structures were validated with *PROCHECK*[47] and the carbohydrate structures validated using *pdb-care*.[48] Data processing and final refinement statistics are given in Table I. Superimposition of molecules was performed with *Lsqkab*[49] and intermolecular distances were measured using *CONTACTS*. The coordinates and structure factors for the TgMIC1-MARR 2′F-3′SiaLacNAc$_{1-4}$, 3′SiaLacNAc$_{1-3}$ and 2′F-3′SiaLacNAc$_{1-3}$ soaks have been deposited with the PDB under ID code 3f53, 3f5a, and 3f5e, respectively.

Acknowledgment

The authors are grateful to Dr. Robert Childs for generating the microarrays of oligosaccharide probes, and to Drs. Henrik Jensen and Balakumar Vijayakrishnan for synthetic assistance. They also thank the European Synchrotron Radiation Facility (ESRF, Grenoble, France) for provision of BM14 beam time.

References

1. Luft BJ, Remington JS (1992) Toxoplasmic encephalitis in Aids. Clin Infect Dis 15:211–222.
2. Montoya JG, Liesenfeld O (2004) Toxoplasmosis. Lancet 363:1965–1976.
3. Hill D, Dubey JP (2002) *Toxoplasma gondii*: transmission, diagnosis and prevention. Clin Microbiol Infect 8: 634–640.
4. Torrey EF, Yolken RH (2003) *Toxoplasma gondii* and schizophrenia. Emerg Infect Dis 9:1375–1380.
5. Kramer W (1966) Frontiers of neurological diagnosis in acquired toxoplasmosis. Psychiatr Neurol Neurochir 69: 43–64.
6. Carruthers VB, Sibley LD (1999) Mobilization of intracellular calcium stimulates microneme discharge in *Toxoplasma gondii*. Mol Microbiol 31:421–428.
7. Carruthers VB, Boothroyd JC (2007) Pulling together: an integrated model of Toxoplasma cell invasion. Curr Opin Microbiol 10:82–89.
8. Jewett TJ, Sibley LD (2003) Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. Mol Cell 11:885–894.
9. Soldati D, Meissner M (2004) Toxoplasma as a novel system for motility. Curr Opin Cell Biol 16:32–40.
10. Reiss M, Viebig N, Brecht S, Fourmaux MN, Soete M, Di Cristina M, Dubremetz JF, Soldati D (2001) Identification and characterization of an escorter for two secretory adhesins in *Toxoplasma gondii*. J Cell Biol 152:563–578.
11. Saouros S, Edwards-Jones B, Reiss M, Sawmynaden K, Cota E, Simpson P, Dowse TJ, Jäkle U, Ramboarina S, Shivarattan T, Matthews S, Soldati-Favre D (2005) A novel galectin-like domain from *Toxoplasma gondii* micronemal protein 1 assists the folding, assembly and transport of a cell-adhesion complex. J Biol Chem 280: 38583–38591.
12. Sawmynaden K, Saouros S, Friedrich N, Marchant J, Simpson P, Bleijlevens B, Blackman MJ, Soldati-Favre D, Matthews S (2008) Structural insights into microneme protein assembly reveal a new mode of EGF domain recognition. EMBO Rep 9:1149–1155.
13. Blumenschein TMA, Friedrich N, Childs RA, Saouros S, Carpenter EP, Campanero-Rhodes MA, Simpson P, Chai W, Koutroukides T, Blackman MJ, Feizi T, Soldati-Favre D, Matthews S (2007) Atomic resolution insight into host cell recognition by *Toxoplasma gondii*. EMBO J 26: 2808–2820.
14. Labbe M, de Venevelles P, Girard-Misguich F, Bourdieu C, Guillaume A, Pery P (2005) Eimeria tenella microneme protein EtMIC3: identification, localisation and role in host cell infection. Mol Biochem Parasitol 140:43–53.
15. Feizi T, Chai WG (2004) Oligosaccharide microarrays to decipher the glyco code. Nat Rev Mol Cell Biol 5: 582–588.
16. Watkins WM (1980) Biochemistry and genetics of the ABO. Lewis and P blood group systems. Adv Hum Genet 10:1–136.
17. Allman SA, Jensen HH, Vijayakrishnan B, Garnett JA, Leon E, Liu Y, Anthony DC, Sibson NR, Feizi T, Matthews S, Davis BG (2009) A Potent fluoro-oligosaccharide probe of toxoplasmosis adhesion. ChemBioChem (in press).
18. Williams SJ, Withers SG (2000) Glycosyl fluorides in enzymatic reactions. Carbohydr Res 327:27–46.
19. Liu Y, Feizi T, Carnpanero-Rhodes MA, Childs RA, Zhang Y, Mulloy B, Evans PG, Osborn HMI, Otto D, Crocker PR, Chai W (2007) Neoglycolipid probes prepared via oxime ligation for microarray analysis of oligosaccharide-protein interactions. Chem Biol 14:847–859.

20. Wright CS (1990) 2.2 a-resolution structure-analysis of 2 refined *N*-acetylneuraminyl-lactose—wheat-germ-agglutinin isolectin complexes. J Mol Biol 215:635–651.
21. Wright CS, Kahane I (1987) Preliminary-X-ray diffraction results on co-crystals of wheat-germ-agglutinin with a sialoglycopeptide from the red-cell receptor glycophorin-A. J Mol Biol 194:353–355.
22. Biffinger JC, Kim HW, DiMagno SG (2004) The polar hydrophobicity of fluorinated compounds. Chembiochem 5:622–627.
23. Hallinan EA, Kramer SW, Houdek SC, Moore WM, Jerome GM, Spangler DP, Stevens AM, Shieh HS, Manning PT, Pitzele BS (2003) 4-Fluorinated L-lysine analogs as selective i-NOS inhibitors: methodology for introducing fluorine into the lysine side chain. Org Biomol Chem 1:3527–3534.
24. Varki A (2007) Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature 446: 1023–1029.
25. Baker HM, Basu I, Chung MC, Caradoc-Davies T, Fraser JD, Baker EN (2007) Crystal structures of the staphylococcal toxin SSL5 in complex with sialyl lewis X reveal a conserved binding site that shares common features with viral and bacterial sialic acid binding proteins. J Mol Biol 374:1298–1308.
26. Blanchard H, Yu X, Coulson BS, von Itzstein M (2007) Insight into host cell carbohydrate-recognition by human and porcine rotavirus from crystal structures of the virion spike associated carbohydrate-binding domain (VP8*). J Mol Biol 367:1215–1226.
27. Chung MC, Wines BD, Baker H, Langley RJ, Baker EN, Fraser JD (2007) The crystal structure of staphylococcal superantigen-like protein 11 in complex with sialyl Lewis X reveals the mechanism for cell binding and immune inhibition. Mol Microbiol 66:1342–1355.
28. Stein PE, Boodhoo A, Armstrong GD, Heerze LD, Cockle SA, Klein MH, Read RJ (1994) Structure of a pertussis toxin sugar complex as a model for receptor-binding. Nat Struct Biol 1:591–596.
29. Imberty A, Gautier C, Lescar J, Perez S, Wyns L (2000) An unusual carbohydrate binding site revealed by the structures of two *Maackia amurensis* lectins complexed with sialic acid-containing oligosaccharides. J Biol Chem 275:17541–17548.
30. May AP, Robinson RC, Vinson M, Crocker PR, Jones EY (1998) Crystal structure of the N-terminal domain of sialoadhesin in complex with 3′ sialyllactose at 1.85 A resolution. Mol Cell 1:719–728.
31. Swaminathan S, Furey W, Pletcher J, Sax M (1995) Residues defining V-beta specificity in Staphylococcal enterotoxins. Nat Struct Biol 2:680–686.
32. Zhuravleva MA, Trandem K, Sun PD (2008) Structural implications of siglec-5-mediated sialoglycan recognition. J Mol Biol 375:437–447.
33. Bartolome C, Espinet P, Martin-Alvarez JM (2007) Is there any bona fide example of O-H center dot center dot center dot F-C bond in solution? The cases of HOC (CF3)(2)(4-X-2,6-C6H2(CF3)(2)) (x = Si(i-Pr)(3), CF3). Chem Commun4384–4386.
34. Dunitz JD, Taylor R (1997) Organic fluorine hardly ever accepts hydrogen bonds. Chem Eur J 3:89–98.
35. Howard JAK, Hoy VJ, Ohagan D, Smith GT (1996) How good is fluorine as a hydrogen bond acceptor? Tetrahedron 52:12613–12622.
36. Kim C-Y, Christianson DW (in press) Binding of fluorine substituted benzenesulfonamides to carbonic anhydrase II (PDB:1IF5 and PDB:1IF6).
37. Muller K, Faeh C, Diederich F (2007) Fluorine in pharmaceuticals: looking beyond intuition. Science 317: 1881–1886.
38. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. Acta Crystallogr Sect D 61:458–464.
39. Crocker PR, Feizi T (1996) Carbohydrate recognition systems: functional triads in cell-cell interactions. Curr Opin Struct Biol 6:679–691.
40. Palma AS, Feizi T, Zhang Y, Stoll MS, Lawson AM, Diaz-Rodriguez E, Campanero-Rhodes MA, Costa J, Gordon S, Brown GD, Chai W (2006) Ligands for the beta-glucan receptor. Dectin-1, assigned using "designer" microarrays of oligosaccharide probes (neoglycolipids) generated from glucan polysaccharides. J Biol Chem 281:5771–5779.
41. Saouros S, Blumenschein TMA, Sawmynaden K, Marchant J, Koutroukides T, Liu B, Simpson P, Carpenter EP, Matthews S (2007) High-level bacterial expression and purification of apicomplexan micronemal proteins for structural studies. Protein Pept Lett 14:411–415.
42. Bailey S (1994) The CCCP4 Suite—programs for protein crystallography. Acta Cryst D 50:760–763.
43. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr Sect D 55:247–255.
44. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D 60: 2126–2132.
45. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr Sect D 62:439–450.
46. Painter J, Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. J Appl Crystallogr 39:109–111.
47. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck—a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283–291.
48. Lutteke T, von der Lieth CW (2004) pdb-care (PDB CArbohydrate REsidue check): a program to support annotation of complex carbohydrate structures in PDB files. BMC Bioinformatics 5.
49. Kabsch W (1978) Discussion of solution for best rotation to relate 2 sets of vectors. Acta Crystallogr Sect A 34: 827–828.