Atomic resolution insight into host cell recognition by *Toxoplasma gondii*

Tharin M. A. Blumenschein[1][§], Nikolas Friedrich[2][§], Robert A. Childs[3], Savvas Saouros[1] , Elisabeth P. Carpenter[1], Maria A. Campanero-Rhodes[3], Peter Simpson[1], Wengang Chai[3], Theodoros Koutroukides[4], Michael J. Blackman[5], Ten Feizi[3], Dominique Soldati-Favre[2] and Stephen Matthews[1][*]

[1]Division of Molecular Biosciences, Imperial College London, South Kensington Campus, London SW7 2AZ UK.

[2]Department of Microbiology and Genetics, Faculty of Medicine, University of Geneva CMU, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland.

[3]Glycosciences Laboratory, Division of Medicine, Imperial College London, Northwick Park & St. Marks Campus, Watford Road, Harrow, Middlesex HA1 3UJ, UK

[4]Medical Science division, Fox Chase Cancer Centre, 333 Cottman Ave. Philadelphia, PA 19111, USA

[5]Division of Parasitology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

[§] These authors contributed equally to the work.

Abstract

The obligate intracellular parasite *Toxoplasma gondii*, a member of the phylum Apicomplexa that includes *Plasmodium* spp., is one of the most widespread parasites and the causative agent of toxoplasmosis. Micronemal proteins (MICs) are released onto the parasite surface just prior to invasion of host cells and play important roles in host-cell recognition, attachment and penetration. Here we report the atomic structure for a key micronemal protein, TgMIC1, and reveal a novel cell-binding motif called the Microneme Adhesive Repeat (MAR). Using glycoarray analyses, we identify a novel interaction with sialylated oligosaccharides that resolves several prevailing misconceptions concerning TgMIC1. Structural studies of various complexes between TgMIC1 and sialylated oligosaccharides provide high-resolution insights into the recognition of sialylated oligosaccharides by a parasite surface protein. We observe that MAR domains exist in tandem repeats which provide a highly specialized structure for glycan discrimination. Our work uncovers new features of parasite-receptor interactions at the early stages of host cell invasion, which will assist the design of new therapeutic strategies.

Introduction

Toxoplasma gondii is a protozoan parasite that is uniquely adapted to infect a wide range of hosts, including virtually all warm-blooded animals and up to 50% of the world's human population. The primary transmission route in humans is via contact with feces from infected domestic cats or ingestion of undercooked contaminated meat, particularly lamb. Toxoplasmosis causes a variety of disease states in humans, including severe disseminated disease in immunosuppressed individuals due to reactivation of encysted parasites and birth defects in infants where mothers are exposed during pregnancy (Hill and Dubey, 2002).

Infection is rapidly established in the host by the fast-replicating form of the parasite, the tachyzoite, which can invade an extremely broad range of cell types. Unlike other pathogens that hijack existing host cell uptake pathways, *Toxoplasma* and other apicomplexan parasites including *Plasmodium* actively force entry into host cells. The process is initiated by contact with the host-cell plasma membrane, followed by reorientation and then the generation of a motive force, which drives penetration into a novel, parasite-induced structure called the parasitophorous vacuole (Carruthers and Boothroyd, 2006). The rapid and smooth transition through these stages requires a highly regulated release of proteins from several parasite organelles, namely micronemes, rhoptries and dense granules (Carruthers and Sibley, 1997). Microneme discharge occurs first and their contents participate in the attachment of parasites to the host cell surface (Carruthers et al., 1999) and the formation of a connection with the parasite actinomyosin system (Jewett and Sibley, 2003), thereby providing the platform from which to drive motility and invasion (Soldati and Meissner, 2004).

One of the first micronemal proteins to be discovered in *T. gondii* was MIC1 (TgMIC1), which functions in cell adhesion (Fourmaux et al., 1996). TgMIC1 is a remarkable, multifunctional protein that in addition to binding host cell receptors, interacts with two other microneme proteins (TgMIC4, TgMIC6) (Brecht et al., 2001; Reiss et al., 2001) and is essential for transport of the whole complex through the early secretory pathways (Huynh et al., 2004; Reiss et al., 2001; Saouros et al., 2005). Deletion of the *mic1* gene in *T. gondii* has also confirmed the specific and critical role played by TgMIC1 in host cell attachment and invasion *in vitro* and has provided evidence for its role in virulence *in vivo* (Cerede et al., 2005). Recent studies have shown that a purified TgMIC1 sub-complex is a potent antigen and acts as an effective vaccine in the mouse model (Lourenco et al., 2006). Unlike the battery of other micronemal proteins, the sequence of TgMIC1 does not exhibit an obvious likeness to vertebrate adhesive motifs. However, a recent nuclear magnetic resonance (NMR) structure revealed a previously unidentified and novel galectin-like domain within the C-terminus of TgMIC1 that promotes proper folding of TgMIC6 and contributes to complex formation (Saouros et al., 2005).

Although studies have clearly highlighted the importance of micronemal proteins in apicomplexan invasion, the finer structural details of host cell recognition remain largely unknown. So far, two main studies have addressed this issue in *Plasmodium falciparum*: erythrocyte binding antigen (EBA-175) binds sialic acid (Tolia et al., 2005) and TRAP binds to heparin (Tossavainen et al., 2006), although high resolution information regarding carbohydrate recognition was not forthcoming. Combining atomic resolution studies with data from carbohydrate microarrays we reveal a novel interaction between *T. gondii* and a variety of sialylated oligosaccharides. The binding mode is attributed to a new family of domains — named the **M**icronemal **A**dhesive **R**epeat (MAR) — that exists in tandem repeats and provides a highly specialized structure for glycan recognition. Our work presented here addresses many long standing issues and uncovers new features regarding parasite-receptor interactions in the early stages of host cell invasion. Furthermore, an understanding of the atomic-resolution details of how *T. gondii* invades host cells opens the way to the design of therapeutic strategies.

Results and Discussion

The overall structure of cell-binding region of TgMIC1

To resolve the atomic structure of the host-cell binding region from TgMIC1 we expressed the *N*-terminal 246 amino acids (residues 17 to 262 hereafter termed TgMIC1-NT) in *Escherichia coli* fused with thioredoxin to aid disulphide bond formation and solubility. Binding assays revealed that our bacterially and *Pichia*-produced TgMIC1-NT bound independently to host-cells (Figure 1A). The high binding efficiency observed for *E. coli*-derived material is likely due to the higher purity of this reagent (Figure 1A). Our earlier experiments on the C-terminal domain from TgMIC1 (TgMIC1-CT) had excluded a role in cell binding for this region, (Saouros et al., 2005), therefore we can conclude that TgMIC1-NT possesses the cell-binding properties of the full-length TgMIC1.

After subsequent removal of the purification tag, we crystallized the recombinant protein and a selenomethionine (SeMet)-substituted form of TgMIC1-NT, and resolved the structure using MAD (multiple-wavelength anomalous dispersion) methods. The atomic

structure for residues 29-259 was solved at a resolution of 1.9 Å (Figure 2A, Supplementary Figure 1 and Table 1). Residues 17-28 and 262 likely exhibit a degree of flexibility and therefore could not be observed in the crystal structure. The new structure reveals a repeated domain consisting of a distorted barrel arrangement of five β-strands, which is flanked on one side by an anti-parallel helical bundle, comprising one helix from each terminus. Two disulphide bonds, C1-C4 and C5-C7 (namely C45-C85 and C103-C113 in repeat 1 and C154-C179 and C193-C203 in repeat 2; Figure 3A), are absolutely conserved between the repeats and stabilize the core structure, one tying helix α1 to the β-barrel and the second between strands β3 and β4. Although the two repeats (amino acid residues 16-144 and 145-237) have a sequence identity of 27% and superimpose with a backbone RMSD of 2.2 Å over 89 residues (Figure 2B), some notable differences are apparent. The first helix and the subsequent loop to strand β1 is significantly longer in the first repeat and is stabilized by an additional disulphide bond, C53-C61 (Figure 3A). Most strikingly, the second repeat is elaborated at its C-terminus by a short "β-finger" (amino acid residues 238-256), which is pinned to the main body of the domain by a new arrangement of two disulphide bonds (C6-C9 and C8-C10; Figures 2C and 3A) replacing the single connection observed in repeat 1 (C6-C8; Figures 2C and 3A). A further intriguing aspect of this region is the presence of a *cis*-proline within the ^{248}N**P**PL^{251} motif (Supplementary Figure 1), which enables unusual positioning of both backbone and side-chain elements that may correspond to an interaction site, possibly for its partner in the complex MIC4 (Saouros et al., 2005).

The MAR domain—a new fold unrelated to thrombospondin type1 repeats

In the initial characterization of TgMIC1 it was postulated that a tandem arrangement of degenerate Thrombospondin type I repeats (TSR1s) was present at the N terminus (Fourmaux et al., 1996). TRS1s adopt an antiparallel, three-stranded fold comprising alternating stacked layers of tryptophan and arginine residues. Our structure of TgMIC1-NT now enables us to reassess this domain classification and reveals a fold that bears no resemblance to the TSR1s. Moreover, it is unrelated to the classical β-sandwich structure of

the prototypic surface antigen, SAG1 from *Toxoplasma* (He et al., 2002) and to the dimeric arrangement of EBA-175 (Tolia et al., 2005). A search of the DALI database reporting no structural hits confirms this and that the structure of TgMIC1-NT presents a previously unknown protein fold (Holm and Sander, 1995). Based on these observations we have named this repeat domain the "MAR domain", after **M**icronemal **A**dhesive **R**epeat.

A search against other apicomplexan genomes identify tandem MAR domains in MIC1 from *Neospora caninum* (Keller et al., 2002) and in three other MIC1-like proteins in *T. gondii* (Figure 3A)*,* which may help to endow this parasite with the ability to invade a wide range of cell types as well as evade host immune responses. MIC3 proteins from *Eimeria tenella*, the cause of Coccidiosis in poultry, contain between 4 and 7 consecutive MAR1 domains constructed from 5 distinct MAR1 sequences (Labbe et al., 2005) (Figure 3A). A model of an uninterrupted stretch of MAR1 domains, as exist in EtMIC3, generated using our structure of TgMIC1-NT, gives rise to a stalk structure comprising a left-handed, helical axis with 70° rotations and 11 Å translations relating adjacent MARs (Figure 3B). This arrangement would project from the parasite surface, presenting an array of MAR domains that could provide increased cell-binding avidity (Supplementary Figure 2). Future studies will be aimed at unveiling other parasite surface proteins that possess members of the MAR family.

The MAR domain—a novel carbohydrate binding domain

A long-standing question is how *T. gondii* can infect and replicate within all nucleated cells. The broad host range suggests that complementary receptors exist on a wide variety of host cell types. Carbohydrate recognition and discrimination provide an excellent means to facilitate such interactions and often play an important role in early recognition events by invasive pathogens. Although carbohydrate-binding properties have been described for microneme proteins from *Toxoplasma* (Cerede et al., 2002; Harper et al., 2004), studies on the nature of these interactions have not been forthcoming. Monteiro et al. showed that sialic acid plays a role in host cell invasion, but the identity of the parasite ligand has been the subject of much speculation (Monteiro et al., 1998). Indirect evidence exists for a lactose-

binding activity for TgMIC1 and the TgMIC1-TgMIC4 sub-complex (Lourenco et al., 2006; Lourenco et al., 2001). However, cell-binding inhibition (Figure 1B) and NMR titration experiments performed in the present study failed to detect lactose recognition by TgMIC1.

To reassess the carbohydrate-binding properties of TgMIC1, microarray analyses (Campanero-Rhodes et al., 2006; Palma et al., 2006) were carried out using the fusion proteins TgMIC1-NT and TgMIC1-CT and lipid-linked oligosaccharide probes (Feizi and Chai, 2004) as described. The microarrays encompassed a panel of >200 oligosaccharide probes representing diverse mammalian glycan sequences and their analogues, as well as some derived from fungal and bacterial polysaccharides (Palma et al., 2006) (Supplementary Table 1). The C-terminal galectin-like domain, TgMIC1-CT, showed no binding to galactose-terminating glycans in the array consistent with our early NMR evidence (Saouros et al., 2005), nor was there binding to any of the other probes in the microarray (results not shown). In contrast, significant binding signals with fluorescence intensities between 150 and ~8000 were elicited by TgMIC1-NT and were observed only among terminally sialylated structures. (Supplementary Table 1 and Fig 4). All the non-sialylated structures tested had numerical scores below 150 (Supplementary Table 1).

More than 40 out of the 69 sialylated probes arrayed gave numerical binding scores greater than 150 with TgMIC1-NT; among them were several *N-* and *O*-glycans and gangliosides in groups D, G and F, respectively, and others representing sialylated capping structures on backbones of mammalian glycoconjugates in groups A, B, C and E (Figure 4). Where close comparisons could be made (Table 2), a mild preference is apparent for the Neu5Acα2-3Gal linkage over the α-2,6 and α-2,8 linkage (Neu5Ac denotes *N*-acetylneuraminic acid). The *N*-acetyl group of the sialic acid was determined to be required for binding (cf for example probes 19 and 27 in Table 2). No effects on binding were observed when downstream residues were either sulfated or fucosylated (Supplementary Table 1).

To locate the receptor binding surfaces on TgMIC1 and investigate the sialic acid binding mode in more detail, crystals were soaked with either sialic acid, α-2,3-sialyl-*N*-acetyllactosamine or α-2,6-sialyl-*N*-acetyllactosamine. The crystallographic structures of the

complexes showed that the protein structure was largely unchanged and in all cases a single glycan was seen bound to the MAR2 domain of TgMIC1-NT (Figure 5 and Supplementary Figure 1). A shallow binding pocket is formed by a contiguous stretch of residues between Lys216 and Glu221, most of which make specific direct contacts to the sialyl moiety that represents a novel binding mode. The interactions between the protein and sugar molecules do not appear to involve any tightly bound water molecules. Most notably, the two oxygen atoms from the carboxyl group are recognized by hydrogen bonds to the amide and side-chain hydroxyl proton of Thr220, rather than the side chain of an arginine or lysine which is often the case in other sialic acid-binding proteins (Alphey et al., 2003; May et al., 1998). Recognition of the acetyl group in sialic acid is enabled by a hydrogen bond between the NH group and the backbone carbonyl of His218. The side chain of Tyr219 contacts the glycerol moiety, while His218 stacks with the underside of the ring.   Electron density was observed for the galactose unit in both α-2,3-sialyl-*N*-acetyllactosamine or α-2,6-sialyl-*N*-acetyllactosamine complexes but no specific protein contacts could account for its preferred orientation (Figure 5A). The glucose ring was not clearly observed in these electron density maps.

To demonstrate whether the cell-binding activity of TgMIC1 (Figure 1) is mediated by sialo-oligosaccharide receptors, assays were repeated in the presence of increasing amounts of sialic acid. Figure 6 shows that cell binding by TgMIC1 displays a striking sensitivity to the presence of soluble sialic acid and was undetectable for either full-length TgMIC1 or TgMIC1-NT in the presence of ~10 mM sialic acid. Furthermore, in support of these findings pre-treatment of cells with neuraminidase abolished cell binding activity of TgMIC1 and TgMIC1-NT (Figure 6). Taken together, these data corroborate the importance of surface exposed sialyl glycans in cell binding and we can also conclude that the tandem MAR domains host all of the cell binding properties of full-length TgMIC1.

TgMIC1 presents a fixed arrangement of two sialic acid binding sites

Results from the carbohydrate microarrays reveal that a variety of sialyl oligosaccharide sequences as found on glycoproteins and glycolipids are recognized with

specific preferences for sialic acid linkage position and some discrimination of oligosaccharide backbone sequences. Interestingly, the most potent binders are branched carbohydrates having two or more terminal sialic acids, raising the possibility that the tandem MAR repeat recognizes specific bidentate ligands. Optimal binding responses are observed when sialic acid termini are separated by 5-8 carbohydrate units, which suggest that this separation could be sufficient to span both MAR domains of one TgMIC1 molecule (Table 2).

Although the key sialic acid-recognizing residues are conserved and located in structurally identical positions in both MAR1 and MAR2 domains (Figure 3A), and despite exhaustive crystal soaking experiments, no oligosaccharides were found bound in MAR1. Instead, an acetate molecule interacts with the equivalent threonine in MAR1 (Thr126) in an identical fashion to the carboxyl group of sialic acid in MAR2, suggesting that MAR1 is able to recognize sialyl oligosaccharides albeit perhaps with a lower affinity (Figure 5B). The most likely reason for the absence of glycan in MAR1 is the extensive crystal contacts in which this region is involved. The binding site in MAR1 is very close to a 2-fold symmetry axis and residues Thr126, Arg127 and Gln129 are involved in crystal contacts. The presence of the glycan in MAR2, and its absence in MAR1, were confirmed with a simulated annealing omit map. To investigate the potential 2:1 glycan:protein stoichiometry and to provide site-specific information we performed a NMR titration experiment using ^{15}N,^{13}C-Ala/Thr labeled TgMIC1-NT in the presence of various sialyl carbohydrates. Using a combination of triple resonance data and information from the site-specific labelling, amide chemical shifts for the two key threonine residues, namely ^{124}HA$\underline{T}$R^{127} and ^{218}HY$\underline{T}$E^{221}, were unambiguously assigned in MAR1 and MAR2, respectively (Figure 7A). Significant amide chemical shift changes are observed for both threonines in the presence of either α-2,3-sialyl-*N*-acetyllactosamine or α-2,6-sialyl-*N*-acetyllactosamine confirming that both MAR domains are active in binding sialic acid (Figure 7A).

To test the importance of both carbohydrate-binding sites on function we performed site-directed mutagenesis and assessed their respective cell-binding capability (Figure 7B). As active site threonine residues make specific side-chain and main-chain contacts to the sialyl carboxylate group, we assumed that their disruption would have a major impact on

9

binding. As predicted the double mutant T126A/T220A-TgMIC1 exhibited no observable

binding to host cells. Additionally, both single mutants, namely T126A-TgMIC1 and T220A-

TgMIC1, were also defective in binding thus confirming the importance of both binding sites.

To assess the role of the "β-finger" extension in MAR2 (Figure 2C) in cell binding, an

experiment was also conducted for a mutant lacking amino acid residues 238-262. No effect

of this truncation on cell binding was observed (Figure 7B). Instead, expression of the

mutant in *T. gondii* followed by immunofluorescence studies suggests that this part of the

molecule is important for binding to its partner protein in the complex TgMIC4 (data not

shown).

A model for glycan recognition by TgMIC1

It is well accepted that the binding of micronemal proteins to host cells provides a

'molecular bridge' to the parasite and initiates invasion. Few detailed structures of

microneme proteins are known and their interactions with the host are especially poorly

characterized. Microarray experiments combined with structural studies reveal the first

insights into the interaction between a key microneme protein complex and the host. Not

only do we identify an interaction between the N-terminal region of TgMIC1 and sialylated

host cell ligands, but we also provide the atomic resolution basis for recognition. The

importance of this interaction is highlighted by observation that the ability of *T. gondii* to

invade lec2, a CHO mutant deficient in sialic acid is reduced to about 30% (Monteiro et al.,

1998). In this study we extend these data by performing cell invasion assays in either the

presence of soluble carbohydrates or using cells pre-treated with neuraminidase. Compared

to controls both conditions severely reduced the levels of parasite internalisation (Figure 7C

and 7D), thus confirming the importance of sialic acid recognition for efficient invasion. The

specific contribution made by the MAR domains has been established by studies reporting

the invasion by *mic1KOs* reduced by ~50% compared to wild-type (Cerede et al., 2005), and

by our earlier work ruling out carbohydrate and cell-binding activities for the C-terminal

portion of TgMIC1, TgMIC4 and TgMIC6 (Saouros et al., 2005). Furthermore, antibodies

directed against the MAR domains from EtMIC3 from *E. tenella* (Figure 3) inhibit invasion

and further development (Labbe et al., 2005), thus emphasising the importance of the MAR domains in penetration.

Sialic acids are found widely distributed in animal tissues especially in glycoproteins and gangliosides and have been shown to play an important role in several protozoa-host cell interactions. A prominent example is the erythrocyte binding antigen (EBA-175) from *P. falciparum*, which recognizes the heavily sialylated receptor glycophorin A during invasion by the malaria parasite (Tolia et al., 2005). Although sialyl lactose binding sites were identified in EBA-175, high resolution insight into glycan recognition was not forthcoming. A model was proposed in which the dimeric receptor-binding domain assembles around a glycophorin A dimer with carbohydrate binding within the channels. Our structural and binding experiments reveal that TgMIC1 possesses two sialic acid binding sites, uniquely arranged in tandemly-repeated MAR domains. These lie on one side of the molecule and, together with their fixed separation and relative geometry, are predicted to bind cognate carbohydrate receptors with high specificity and affinity (Figure 5C). Intriguingly, some of the most potent ligands were multi-valent carbohydrates, possessing two or more sialic acid units, such as those on gangliosides or polysialic acid. Gangliosides belong to a class of glycosphingolipids that are found in abundance within the membranes of neuronal cells (Karlsson, 1998) and the α(2-8) linked polysialic acids are found on the fetal neural cell adhesion molecule (Acheson et al., 1991; Finne et al., 1983). It is tantalizing to speculate that recognition of such a specific glycan might be important during the asexual life cycle of *Toxoplasma* in which cysts are formed within the intermediate host, particularly in the brain. This subunit multivalency may also be required for the formation of cross-links between possible ganglioside receptors and TgMIC1-containing complexes. Our work provides the first atomic resolution insight into the mechanism of cell attachment by *T. gondii* that will provide a foundation for further functional study and the design of novel therapeutics for parasitic infections.

Materials and Methods

Cloning, expression and purification from *E. coli*

TgMIC1-NT, spanning residues 17 to 262 in TgMIC1 (the first 16 residues represent the signal peptide), was expressed using the pET32Xa/LIC plasmid (Novagen) in *E. coli* Origami (DE3) (Stratagene), at 28° C (Saouros et al., 2007). Protein expression was induced with 800 μM isopropyl β-D-thiogalactopyranoside. The hexahistidine-thioredoxin-MIC1-NT fusion protein was purified using nickel-nitrilotriacetic acid HISBind resin (Novagen), and cleaved with Factor Xa (Invitrogen). The cleaved protein was re-applied to the same column, and pure MIC1-N terminal was recovered in the flow-through. The protein was concentrated to ~10 mg/mL, and stored at 4 or -20° C in 1 mM $CaCl_2$, 100 mM NaCl, 50 mM Tris·Cl pH 8.0. Selenomethionine-labelled TgMIC1-NT was expressed in minimal media in *E. coli* Origami (DE3), according to the protocol of Van Duyne *et al.* (Van Duyne et al., 1993), purified and concentrated as described above, and stored in small aliquots at -20° C. ^{15}N,^{13}C-labelled samples of TgMIC1-NT were produced in minimal media, containing 0.07% ^{15}NH$_4$Cl and 0.2% ^{13}C$_6$-glucose. TgMIC1-NT ^{15}N,^{13}C-labelled at Ala and Thr positions was produced according to published protocols (Matthews et al., 1993).

Generation of variants and mutants

Plasmids pPICZα-TgMIC1TSR1 (called pPICZα-TgMIC1NT) and pROP1mycMIC1-TSR1 (called pROP1mycMIC1NT) were obtained as described before (Saouros et al., 2005). C-terminal deletions of these clones were obtained from these plasmids by PCR as follows: The fragment obtained from pPICZα-TgMIC1NT with primers 5'-CGCCTAGGGTTGGGCCAGAAGCATATGGAGAAG-3' (MIC1-1_672) and 5'-CCGGGCGCGGCCGCAGAACATGGGCTGTCGACGGATCC-3' (TgMIC1-17_1717) was digested with NdeI and NotI and cloned back into pPICZα-TgMIC1NT, resulting in pPICZα-TgMIC1NTΔCterm (aa 238-262). Similarly, the fragment obtained from pROP1mycMIC1NT with primers 5'-CCCGCTGCAGGAGCAAAGGCTGCCAATTATTC-3' (ROP1-SP_141) and 5'-GGCGAGCTCTTAATTAAGAACATGGGCTGTCGACGGATCC-3' (TgMIC1-16_1716) was

digested with NsiI and PacI and cloned back into pROP1mycMIC1NT, resulting in pROP1mycMIC1NTΔC2.

Mutants of pPICZα-TgMIC1NT were obtained using the Quickchange Kit (Stratagene). Primers 5'-CAGTAATCACGCAGCGCGCCATGAGATACTGTC-3' (TgMIC1-18_1729) and 5'-GACAGTATCTCATGGCGCGCTGCGTGATTACTG-3' (TgMIC1-19_1730), as well as primers 5'-GACAAGCGGCATTATGCAGAAGAGGAAGGAATTCG-3' (TgMIC1-20_1731) and 5'-CGAATTCCTTCCTCTTCTGCATAATGCCGCTTGTC-3' (TgMIC1-21_1732) were used for generation of the T126A and T220A single mutants and the T126A/T220A double mutant, respectively.

Pichia pastoris expression

P. pastoris transformation and expression was performed using the *Pichia* expression kit (Invitrogen), according to the supplied protocols, and all yeast strains were maintained on yeast extract-peptone-dextrose (YPD)-rich media. Transformation of the supplied host strain GS115 was performed by electroporation following linearization with PmeI for all pPICZα-based vectors. Selection of transformants was then performed on YPD-Zeocin (100 μg/ml). Expression was performed using BMGY and BMMY media according to the manufacturer's instructions.

Cell binding assays

These were performed as described previously (Brecht et al., 2001). Briefly, confluent monolayers of human foreskin fibroblasts (HFF) cells, grown in 12-well plates, were blocked for 1h at 4 °C with 1% BSA in cold PBS,1 mM $CaCl_2$, 0.5 mM $MgCl_2$ (CM-PBS). Excess BSA was removed by two 5-min washes with ice-cold CM-PBS. The proteins to be assayed were then added either in the form of *P. pastoris* culture supernatant (0.25 μg), or bacterially-produced (0.25 μg), together or not with different concentrations of competitors (sialic acid buffered to pH 6.8, lactose, galactose, heparin; all purchased from Sigma-Aldrich) and diluted in cold CM-PBS to a total volume of 500 μl. After incubation at 4 °C for 1 h the supernatant (S) was removed, and the cells were washed four times for 5 min with ice-cold

CM-PBS. The cell-bound fraction (CB) was collected by the direct addition of 50 µl 1x SDS-PAGE loading buffer with 0.1 M dithiothreitol. In some cases, prior to blocking, HFF cells were pretreated with 66mU/ml α-2,3,6,8- *Vibrio cholerae* neuraminidase (Calbiochem) in RPMI1640, 25mM HEPES, L-glutamine (Gibco) for 1h at 37°C in a total reaction volume of 1ml.

Cell invasion assays in the presence of carbohydrate inhibitors

Confluent monolayers of HFF cells on glass coverslips in 12 well plates were washed one time briefly in DMEM (Gibco) and incubated 15 minutes with 250µl DMEM only (control) or with 250µl DMEM containing different double concentrated competitors (sialic acid pH 6.8, lactose, galactose). Then 250µl of *T.gondii* RH parasites were added and invasion was allowed to proceed during 1h at 37°C. Afterwards cells were washed two times with 500µl DMEM, fixed with 4% paraformaldehyde/0.05% glutaraldehyde for 20 min, followed by 2-min incubation with 0.1 M glycine in PBS. Fixed cells were blocked in PBS, 2% bovine serum albumin for 20 min. The cells were then stained with anti-TgSAG1 antibody followed by Alexa 488-conjugated goat anti-mouse IgG (H+L) antibody (Molecular Probes - Invitrogen). Afterwards cells were permeabilized with PBS, 0.2% Triton X-100 for 20 min and blocked again in PBS, 0.2% Triton X-100, 2% bovine serum albumin for 20 min. Staining was performed with anti-TgProfilin antibody followed by Alexa 594-conjugated goat anti-rabbit IgG (H+L) antibody (Molecular Probes - Invitrogen). Invasion assays were carried out as triplicates of two independent experiments. According to their staining, numbers of extracellular (attached) and intracellular (invaded) parasites were counted in three microscopic fields respectively. For the untreated control a total of 200 parasites were counted per microscopic field on average and attachment and invasion were set to 100%. With respect to this untreated control, the percentage of attached and invaded parasites was calculated for each condition.

Cell invasion assays with neuraminidase treament

Confluent monolayers of HFF cells on glass coverslips in 24 well plates were incubated in RPMI1640, 25mM HEPES, L-Glutamine (Gibco) at 37°C with or without 66mU *vibrio cholerae* neuraminidase (Calbiochem) in a total volume of 500µl. After one hour 500µl freshly lysed *T. gondii* RH parasites resuspended in RPMI1640, 25mM HEPES, L-Glutamine (37°C) were added to each well and the plate was centrifuged immediately for 5 minutes at 480g. Then invasion was allowed to proceed for another 5 minutes at 37°C. Afterwards cells were washed two times with 500µl DMEM, fixed with 4% paraformaldehyde, 0.05% glutaraldehyde for 20 min, followed by 2 minutes incubation with 0.1 M glycine in PBS. Fixed cells were blocked in PBS, 2% bovine serum albumin for 20 min. Cells were then stained with anti-TgSAG1 antibody followed by Alexa 488-conjugated goat anti-mouse IgG (H+L) antibody (Molecular Probes - Invitrogen). Afterwards cells were permeabilized with PBS, 0.2% Triton X-100 for 20 min and blocked again in PBS, 0.2% Triton X-100, 2% bovine serum albumin for 20 min. Staining was performed with anti-TgProfilin antibody followed by Alexa 594-conjugated goat anti-rabbit IgG (H+L) antibody (Molecular Probes - Invitrogen).

Four experiments were carried out in parallel. According to their staining, numbers of extracellular (attached) and intracellular (invaded) parasites were counted in five microscopic fields per coverslip respectively. For the untreated control a total of 400 parasites were counted per microscopic field on average and attachment and invasion were set to 100%. With respect to this untreated control, the percentage of attached and invaded parasites was calculated for the enzyme-treated cells. Enzyme activity was monitored in parallel with a binding assay using recombinant TgMIC1-NT.

Crystallisation and data collection

TgMIC1-NT was crystallised in hanging drops by the vapour-diffusion technique. Crystals shaped as tetragonal bipyramids, measuring approximately $50 \times 50 \times 100$ µm, were obtained by adding 1 µl of protein to 1 µl of well solution containing 3.6 M ammonium acetate and 100 mM Bis-tris-propane pH 7.6 at 17° C. Crystals were briefly soaked in a cryoprotectant solution containing all the components of the well solution and 25% glycerol, and cryocooled in liquid nitrogen. X-ray data for both native (1.9 Å) and SeMet-labelled

TgMIC1-NT (2.8 Å) were collected at beamline 10.1 at the Synchrotron Radiation Source, CCLRC Daresbury Laboratory. SeMet-labelled TgMIC1-NT showed a well defined Se K absorption edge by fluorescence scanning, and data were collected from a single crystal at three wavelengths, peak, inflection, and high-energy remote, for MAD phasing. SeMet-TgMIC1-NT crystals were soaked overnight in either 3.6 M ammonium acetate, 100 mM Bis-tris-propane pH 7.8 and 7.4 mM α-2,6-sialyl-N-acetyllactosamine; 3.4 M ammonium acetate, 100 mM Bis-tris-propane pH 7.0, 3.3 mM α-2,3-sialyl-N-acetyllactosamine, 27.5 mM NaCl, and 5.5 mM sodium phosphate, or 3.6M ammonium acetate, 100 mM Bis-tris-propane pH 7.6 and 20 mM sialic acid. The pH and ammonium acetate concentration matched the well solution used in growing the crystals. X-ray data for the crystals soaked in 2,3-sialyl-N-acetyllactosamine (2.3 Å) and 2,6-sialyl-N-acetyllactosamine (2 Å) were collected at beamline 14.1 at the Synchrotron Radiation Source, CCLRC Daresbury Laboratory. The data were indexed, integrated and scaled using the HKL-2000 package (Otwinowski and Minor, 1997).

Structure Solution and Refinement

The unbound structure was phased using the multiple anomalous dispersion method with the program SOLVE (Terwilliger and Berendzen, 1999), producing a mean figure of merit of 0.45 and a Z-score of 8.65. The initial model was built automatically in ARP/wARP (Perrakis et al., 1999). Structures were improved manually using Coot (Emsley and Cowtan, 2004). Refinement and water addition was done with multiple cycles of Refmac5 (Winn et al., 2001) and ARP/wARP. ARP/warp, MOLREP and Refmac5 were used within the CCP4 suite (Bailey, 1994). Fifteen acetate ions, two glycerol molecules and two chloride ions were found in the structure, together with 181 water molecules. Sugar-bound structures were phased by molecular replacement, using MOLREP (Vagin and Teplyakov, 1997) and the protein atoms from the unliganded structure as the model. Clear density was observed for the ligand atoms. The structures were rebuilt and refined as described above. Refinement statistics are listed in Table 1. 99.5% of residues are within the allowed regions of the Ramachandran plot. The quality of the structure was assessed using procheck (Laskowski et al., 1993) and whatcheck

(Hooft et al., 1996). To confirm the presence of the glycan, a simulated annealing FoFc omit map was calculated by omitting the glycan and acetates occupying the binding sites. The models have been deposited in PDB with accession number 2JH1 (free TgMIC1-NT), 2JH7 (TgMIC1-NT in complex with 2,6-sialyl-N-acetyllactosamine) and 2JHD (TgMIC1-NT in complex with 2,3-sialyl-N-acetyllactosamine). Figures were drawn using the program MacPyMol (DeLano Scientific). The model for the sialic acid complex showed density for the sialic acid that matched perfectly the sialic acid moiety in the other two sugars. The sialic acid structure was therefore not refined further as the data was at lower resolution (2.6 Å) than the other two sugars and was deemed not to add any extra information.

Microarray analysis of the binding of TgMIC1-NT fusion protein to diverse oligosaccharide probes.

Two hundred and eighteen, lipid-linked oligosaccharide probes (Supplementary Table 1) were robotically printed on 16-pad nitrocellulose-coated glass slides (FAST slides, Whatman Ltd) as described (Palma et al., 2006). The number of unique sequences printed was 207 as 11 probes were printed more than once. These are grouped (Figure 4 and Supplementary Table 1) according to their backbone sequences, and include numerous mammalian-type carbohydrate sequences: *N*-glycans (neutral and acidic, high mannose- and complex-types), *O*-glycans, blood group-related sequences (A, B, H, Lewis[a], Lewis[b], Lewis[x] and Lewis[y]) on linear or branched backbones and their sialylated and sulphated analogues, gangliosides, glycosaminoglycans, homo-oligomers of sialic acid and fragments of other polysaccharides, ranging in size from two to twenty monosaccharides. The probes were printed at 2 and 7 fmol per spot, in duplicate, using a non-contact arrayer (Piezorray, Perkin Elmer LAS, UK), with Cy3 dye included to enable post-array monitoring of the slides.

For binding analysis with TgMIC1-NT fusion protein, the arrayed pads were overlaid initially for 1 hr with blocking solution containing casein (Pierce) with 1% w/v bovine serum albumin (Sigma) and 10mM calcium (casein/BSA). The slides were rinsed and overlaid for 1h with fusion protein complex in casein/BSA: this contained per 100µl, 10µg mouse anti-

histidine monoclonal antibody, 10µg biotinylated goat anti-mouse IgG and 4 µg of TgMIC1-NT or TgMIC1-CT fusion protein. The overlaid pads were washed with 150 mM sodium chloride, 2mM calcium chloride solution and binding was detected using Alexa Fluor 647-labelled streptavidin (Molecular Probes), 1µg per ml casein/BSA. After washing, the dried slides were scanned using a ProScanArray (Perkin-Elmer LAS) and Alexa Fluor 647-binding signals quantified using ScanArrayExpress software (Perkin-Elmer LAS). Microarray data analysis and presentation were carried out using in-house software (M. S. Stoll, unpublished). The binding signals were glycan dose-related overall. The results shown in Figure 4 represent binding at 7 fmol per spot.

NMR mapping of sialic acid binding sites

For NMR experiments, samples of ^{15}N-labeled TgMIC1-NT, uniformly ^{15}N,^{13}C-labeled TgMIC1-NT or TgMIC1-NT specifically labeled with ^{15}N,^{13}C-labeled Ala and Thr were prepared in 20 mM sodium phosphate buffer pH 7, 100 mM NaCl, 4% Complete^{TM} protease inhibitor cocktail (Roche Diagnostics Ltd, UK, prepared according to instructions), 0.03% sodium azide, and 0.2 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid, in 90% H2O, 10% D2O at approximately 500 µM in 0.25 ml. Backbone assignment for Thr220 was completed using standard double and triple-resonance assignment experiments recorded on uniformly ^{15}N,^{13}C-labelled TgMIC1-NT samples (Sattler et al., 1999). For Thr126, this data was supplemented with HNCO spectra recorded on TgMIC1-NT specifically labeled with ^{15}N,^{13}C-labelled Ala and Thr. Four peaks were identified and could be assigned to AA, TA, TT and ^{124}HA$\underline{T}$R^{127}. Either α-2,3-sialyl-*N*-acetyllactosamine or α-2,6-sialyl-*N*-acetyllactosamine in the same buffer was added in several steps up to a 10 fold molar excess and 2D ^{1}H-^{15}N HSQC spectra were recorded at each stage under identical experimental conditions.

Acknowledgements

References

Acheson, A., Sunshine, J. and Rutishauser, U. (1991) NCAM polysialic acid can regulate both cell-cell and cell-substrate interactions. *J. Cell Biol.*, **114**, 143-153.

Alphey, M.S., Attrill, H., Crocker, P.R. and van Aalten, D.M.F. (2003) High Resolution Crystal Structures of Siglec-7. *Journal of Biological Chemistry*, **278**, 3372-3377.

Bailey, S. (1994) The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography*, **50**, 760-763.

Brecht, S., Carruthers, V.B., Ferguson, D.J.P., Giddings, O.K., Wang, G., Jakle, U., Harper, J.M., Sibley, L.D. and Soldati, D. (2001) The toxoplasma micronemal protein MIC4 is an adhesin composed of six conserved apple domains. *Journal of Biological Chemistry*, **276**, 4119-4127.

Campanero-Rhodes, M.A., Childs, R.A., Kiso, M., Komba, S., Le Narvor, C., Warren, J., Otto, D., Crocker, P.R. and Feizi, T. (2006) Carbohydrate microarrays reveal sulphation as a modulator of siglec binding. *Biochemical and Biophysical Research Communications*, **344**, 1141-1146.

Carruthers, V. and Boothroyd, J.C. (2006) Pulling together: an integrated model of Toxoplasma cell invasion. *Current Opinion in Microbiology*, **In Press, Corrected Proof**.

Carruthers, V.B., Giddings, O.K. and Sibley, L.D. (1999) Secretion of micronemal proteins is associated with toxoplasma invasion of host cells. *Cellular Microbiology*, **1**, 225-235.

Carruthers, V.B. and Sibley, L.D. (1997) Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts. *European Journal of Cell Biology*, **73**, 114-123.

Cerede, O., Dubremetz, J.F., Bout, D. and Lebrun, M. (2002) The Toxoplasma gondii protein MIC3 requires pro-peptide cleavage and dimerization to function as adhesin. *Embo Journal*, **21**, 2526-2536.

Cerede, O., Dubremetz, J.F., Soete, M., Deslee, D., Vial, H., Bout, D. and Lebrun, M. (2005) Synergistic role of micronemal proteins in Toxoplasma gondii virulence. *Journal of Experimental Medicine*, **201**, 453-463.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*, **60**, 2126-2132.

Feizi, T. and Chai, W.G. (2004) Oligosaccharide microarrays to decipher the glyco code. *Nature Reviews Molecular Cell Biology*, **5**, 582-588.

Finne, J., Finne, U., Deagostinibazin, H. and Goridis, C. (1983) Occurrence of Alpha-2-8 Linked Polysialosyl Units in a Neural Cell-Adhesion Molecule. *Biochemical and Biophysical Research Communications*, **112**, 482-487.

Fourmaux, M.N., Achbarou, A., MercereauPuijalon, O., Biderre, C., Briche, I., Loyens, A., OdbergFerragut, C., Camus, D. and Dubremetz, J.F. (1996) The MIC1 microneme protein of Toxoplasma gondii contains a duplicated receptor-like domain and binds to host cell surface. *Molecular and Biochemical Parasitology*, **83**, 201-210.

Harper, J.M., Hoff, E.F. and Carruthers, V.B. (2004) Multimerization of the Toxoplasma gondii MIC2 integrin-like A- domain is required for binding to heparin and human cells. *Molecular and Biochemical Parasitology*, **134**, 201-212.

He, X.-l., Grigg, M.E., Boothroyd, J.C. and Garcia, K.C. (2002) Structure of the immunodominant surface antigen from the Toxoplasma gondii SRS superfamily. *Nature Structural Biology*, **9**, 606-611.

Hill, D. and Dubey, J.P. (2002) Toxoplasma gondii: transmission, diagnosis and prevention. *Clinical Microbiology and Infection*, **8**, 634-640.

Holm, L. and Sander, C. (1995) Dali - a Network Tool for Protein-Structure Comparison. *Trends in Biochemical Sciences*, **20**, 478-480.

Hooft, R.W.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Errors in protein structures. *Nature*, **381**, 272-272.

Huynh, M.H., Opitz, C., Kwok, L.Y., Tomley, F.M., Carruthers, V.B. and Soldati, D. (2004) Trans-genera reconstitution and complementation of an adhesion complex in Toxoplasma gondii. *Cellular Microbiology*, **6**, 771-782.

Jewett, T.J. and Sibley, L.D. (2003) Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. *Molecular Cell*, **11**, 885-894.

Karlsson, K.A. (1998) On the character and functions of sphingolipids. *Acta Biochim. Pol.*, **45**, 429-438.

Keller, N., Naguleswaran, A., Cannas, A., Vonlaufen, N., Bienz, M., Bjorkman, C., Bohne, W. and Hemphill, A. (2002) Identification of a Neospora caninum microneme protein (NcMIC1) which interacts with sulfated host cell surface glycosaminoglycans. *Infection and Immunity*, **70**, 3187-3198.

Labbe, M., de Venevelles, P., Girard-Misguich, F., Bourdieu, C., Guillaume, A. and Pery, P. (2005) Eimeria tenella microneme protein EtMIC3: identification, localisation and role in host cell infection. *Molecular and Biochemical Parasitology*, **140**, 43-53.

Laskowski, R.A., Macarthur, M.W., Moss, D.S. and Thornton, J.M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*, **26**, 283-291.

Lourenco, E.V., Bernardes, E.S., Silva, N.A., Mineo, J.R., Panunto-Castelo, A. and Roque-Barreira, M.C. (2006) Immunization with MIC1 and MIC4 induces protective immunity against Toxoplasma gondii. *Microbes and Infection*, **8**, 1244-1251.

Lourenco, E.V., Pereira, S.R., Faca, V.M., Coelho-Castelo, A.A.M., Mineo, J.R., Roque-Barreira, M.C., Greene, L.J. and Panunto-Castelo, A. (2001) Toxoplasma gondii micronemal protein MIC1 is a lactose-binding lectin. *Glycobiology*, **11**, 541-547.

Matthews, S.J., Jandu, S.K. and Leatherbarrow, R.J. (1993) C-13 Nmr-Study of the Effects of Mutation on the Tryptophan Dynamics in Chymotrypsin Inhibitor-2 - Correlations with Structure and Stability. *Biochemistry*, **32**, 657-662.

May, A.P., Robinson, R.C., Vinson, M., Crocker, P.R. and Jones, E.Y. (1998) Crystal Structure of the N-Terminal Domain of Sialoadhesin in Complex with 3' Sialyllactose at 1.85 A Resolution. *Molecular Cell*, **1**, 719-728.

Monteiro, V.G., Soares, C.P. and de Souza, W. (1998) Host cell surface sialic acid residues are involved on the process of penetration of Toxoplasma gondii into mammalian cells. *Fems Microbiology Letters*, **164**, 323-327.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Macromolecular Crystallography, Pt A*, Vol. 276, pp. 307-326.

Palma, A.S., Feizi, T., Zhang, Y.B., Stoll, M.S., Lawson, A.M., Diaz-Rodriguez, E., Campanero-Rhodes, M.A., Costa, J., Gordon, S., Brown, G.D. and Chai, W.G. (2006) Ligands for the beta-glucan receptor, Dectin-1, assigned using "designer" microarrays

of oligosaccharide probes (neoglycolipids) generated from glucan polysaccharides. *Journal of Biological Chemistry*, **281**, 5771-5779.

Perrakis, A., Morris, R. and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Structural Biology*, **6**, 458-463.

Reiss, M., Viebig, N., Brecht, S., Fourmaux, M.N., Soete, M., Di Cristina, M., Dubremetz, J.F. and Soldati, D. (2001) Identification and characterization of an escorter for two secretory adhesins in Toxoplasma gondii. *Journal of Cell Biology*, **152**, 563-578.

Saouros, S., Blumenschein, T.M.A., Sawmynaden, K., Marchant, J., Koutroukides, T., Liu, B., Simpson, P., Carpenter, E.P. and Matthews, S. (2007) High-level bacterial expression and purification of apicomplexan micronemal proteins for structural studies. *In Press.*

Saouros, S., Edwards-Jones, B., Reiss, M., Sawmynaden, K., Cota, E., Simpson, P., Dowse, T.J., Jäkle, U., Ramboarina, S., Shivarattan, T., Matthews, S. and 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. *Journal of Biological Chemistry*, **280**, 38583-38591.

Sattler, M., Schleucher, J. and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in NMR Spectroscropy.*, **34**, 93–158.

Soldati, D. and Meissner, M. (2004) Toxoplasma as a novel system for motility. *Current Opinion in Cell Biology*, **16**, 32-40.

Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallographica Section D-Biological Crystallography*, **55**, 849-861.

Tolia, N.H., Enemark, E.J., Sim, B.K.L. and Joshua-Tor, L. (2005) Structural Basis for the EBA-175 Erythrocyte Invasion Pathway of the Malaria Parasite Plasmodium falciparum. *Cell*, **122**, 183-193.

Tossavainen, H., Pihlajamaa, T., Huttunen, T., Raulo, E., Rauvala, H., Permi, P. and Kilpelainen, I. (2006) The layered fold of the TSR domain of P. falciparum TRAP contains a heparin binding site. *Protein Sci*, **15**, 1760-1768.

Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography*, **30**, 1022-1025.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J. (1993) Atomic Structures of the Human Immunophilin Fkbp-12 Complexes with Fk506 and Rapamycin. *Journal of Molecular Biology*, **229**, 105-124.

Winn, M.D., Isupov, M.N. and Murshudov, G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallographica Section D-Biological Crystallography*, **57**, 122-133.

Figure Captions

Figure 1:   Host cell binding by TgMIC1-NT

(A) Cell binding assays on human foreskin fibroblasts (HFFs) were performed using supernatant of *P. pastoris* cultures expressing TgMIC1, TgMIC1-NT, or TgMIC1MAR2, bacterially-produced TgMIC1-NT and PfProfillin, the latter as a negative control. Anti-His antibodies are used as the probe for Western blots and the asterisk indicates background from host cells. Samples of input (I), supernatant (S), wash (W) and cell binding fraction (CB) were run on each gel (see Materials and Methods). Molecular weight markers in kDa. The *Pichia* produced material is less efficient as it is a crude cell supernatant while that from *E. coli* is a purified protein. These data show that bacterially-produced TgMIC1-NT retains the cell binding activity of native TgMIC1. Increasing the concentration of the input (I), up to 50 times results in enhanced binding of bacterially produced TgMIC1NT but not of PfProfilin to HFFs. Loading has been adjusted for detection in the linear range. Note: in all cell-binding assays the total volume of the input, supernatant and the wash is 500µl whereas the total volume of the cell bound fraction is 50µl.

(B) Cell binding competition experiments were performed using bacterially-produced hisTgMIC1-NT and supernatants of *P. pastoris* cultures expressing TgMIC1myc. Anti-myc and Anti-his antibodies were used as the probe for Western blots. Samples are named: input (I), supernatant (S), Wash (W) and Cell binding fraction (CB).For the different conditions of competition only the cell bound fraction is shown. These data confirm that <u>no</u> inhibition of TgMIC1 binding to HFFs is observed in the presence of lactose, galactose or heparin. Note: in all cell-binding assays the total volume of the input, supernatant and the wash is 500µl whereas the total volume of the cell bound fraction is 50µl.

Figure 2: Host cell binding and structure of TgMIC1-NT

(A) Ribbon representation of a representative structure for TgMIC1-NT. Repeat 1 (MAR1) is shown in green and Repeat 2 (MAR2) in blue. The orientation shown on the right represents a 180° rotation.

(B) Ribbon representation showing the superimposition of Repeat 1 (MAR1 - green) and Repeat 2 (MAR2 - blue). The orientation shown is the same as (A) left. The calculated RMSD is 2.2 Å over 89 amino acid residues.

(C) Zoomed region illustrating the altered disulphide bond patterns at the C-terminus of Repeat 1 (MAR1 - green) and Repeat 2 (MAR2 - blue). The additional "β-finger" (amino acid residues 238-256) is pinned to the main body of MAR2 by a new arrangement of two disulphide bonds C6-C9 and C8-C10 (namely C197-C242 and C236-C252; See Figure 3A) replacing the single connection observed in repeat 1, C6-C8 (namely C107-C143; Figure 3A).

Figure 3: Micronemal proteins from apicomplexan parasites contain the TgMIC1-repeat

(A) Structure-based sequence alignment of MAR1 and MAR2 domains from other micronemal proteins. Including TgMIC1 (MAR1 aa 16-144, MAR2 aa 145-247), two of the three MIC1-like proteins from *T. gondii* (MAR1 aa 113-222, MAR2 aa 223-336 in TgMIC1a and MAR1 aa 113-222, MAR2 aa 223-335 in TgMIC1b), NcMIC1 (MAR1 aa 30-142, MAR2 aa 143-261) and EtMIC3 (MAR1 aa 42-149, MAR1a aa 150-274, MAR1b aa 291-425, MAR1c aa 36-158 and MAR1d aa 180-280). For clarity the third MIC1-like protein, TgMIC1c, from *T. gondii* has been omitted. Conserved residues are shaded blue. Cysteines and disulphide bond connectivities

are highlighted in orange. *Cis*-proline within the [232]N**PP**L[235] motif is shown in red. Secondary structure elements are indicated above the sequences.

(B) A schematic representation of a model for the seven, sequential MAR1 domains from EtMIC3 is shown in two orientations (left – perpendicular to the helical axis and right – along the helical axis)

Figure 4:   Carbohydrate microarray data on sialyl glycan binding by TgMIC1-NT

Numerical scores are shown of the binding signals, means of duplicate spots at 7 fmol/spot (with error bars) for the 58 sialyl oligosaccharide probes examined.   Sixty nine positions are shown as 11 of the probes were printed more than once (Supplementary Table 1). Selected sialo-oligosaccharide sequences are annotated, with designations of Neu5Acα-2,3-Gal linkage as pink; Neu5Ac α-2,6Gal, blue and Neu5Acα-2,8 linkage yellow. The scores for the non-sialylated probes in the microarray are not shown. These are given in Supplementary Table 1 (positions 70-218).

Figure 5:   Structure of TgMIC-NT in complex with sialyl oligosaccharides

(A) Simulated annealing ($F_o - F_c$) OMIT map contoured at 3σ (left) and 2σ levels (right) for the TgMIC-NT-glycan complex showing the unambiguous orientation of the sialic acid moiety and the position of the galactose unit of α-2,3-sialyl-*N*-acetyllactosamine. The omit map was calculated with the glycan omitted, stick models of key side-chains (green) and sialic acid (orange) are overlaid on the map.

(B) Structure of α-2,3-sialyl-*N*-acetyllactosamine in complex with the MAR2 domain from TgMIC1-NT. Ribbon representation of MAR2 is shown in green. Key interacting side chains and the oligosaccharide are shown as stick representations. Oxygen and nitrogen atoms participating in hydrogen bonds are coloured in red and blue,

24

respectively. Note: The structure of the α-2,6-sialyl-N-acetyllactosamine complex is shown in Supplementary Figure 1.

(C) Structure of the MAR1 domain from TgMIC1-NT show the position of the acetate molecules within the active site. Ribbon representation of MAR1 is shown in blue. Key interacting side chains and the acetate molecule are shown as stick representations.

(D) Ribbon representation of TgMIC1 showing the separation and relative geometries of two sialic acid-binding sites in MAR1 and MAR2.

Figure 6:   Sialic acid competes with TgMIC1 cell binding activity

Cell binding competition experiments with sialic acid were performed on HFFs using supernatants of a *P. pastoris* culture expressing TgMIC1myc (top), bacterially-produced hisTgMIC1-NT (middle) or a *P. pastoris* culture expressing TgMIC1-NTmyc (bottom). Anti-myc and anti-his antibodies are used as the probe for Western blots. Anti-tubulin antibodies were used as a control for equal amount of cell material used in the assay. For the competition experiments with sialic acid only the cell bound fractions are shown. In case of pre-treatment of HFFs with neuraminidase, samples of input (I), supernatant (S), Wash (W) and Cell binding fraction (CB) were run on each gel. These data confirm that TgMIC1 binds to sialic acid exposed receptors on HFF cells. Note: in all cell-binding assays the total volume of the input, supernatant and the wash is 500µl whereas the total volume of the cell bound fraction is 50µl.

Figure 7:   Characterisation of the interaction of sialyl oligosaccharides with MAR domains and their involvement in host cell invasion

(A) ^{1}H-^{15}N 2D TROSY-HNCO (top) and ^{1}H-^{15}N 2D TROSY-HSQC (bottom) NMR spectra for ^{13}C,^{15}N-labelled Ala, Thr TgMIC1-NT in the absence (black) and presence (red) of saturating amounts of a α-2,3-sialyl-*N*-acetyllactosamine. The assignments of the Thr126 and Thr220 were achieved using standard triple resonance spectra.

(B) Cell binding assays using supernatants of *P. pastoris* cultures expressing TgMIC1, TgMIC1-NT single (T126A and T220A) mutants, TgMIC1-NT double (T126A/T220A) mutant and a TgMIC1-NT C-terminal truncation mutant, lacking the "β-finger" (amino acid residues 238-262). Note that all proteins produced in *P.pastoris* have a C-terminal myc-tag. Anti-myc antibodies are used as the probe for Western blots. Cell binding fractions are shown for each protein together with α-tubulin as a control for equal amount of cell material used in the assay. Molecular weight markers in kDa.

(C) Cell invasion assays using *T. gondii* RH parasites in the presence of sialic acid (NANA), lactose and galactose. Invasion assays were carried out in triplicates of two independent experiments. Numbers of intracellular parasites were counted in three microscopic fields.

(D) Cell invasion assays using *T. gondii* RH parasites after pre-treatment of target cells with neuramindase. Four experiments were carried out in parallel. Numbers of intracellular parasites were counted in five microscopic fields.

Table 1 – Data collection, phasing and refinement statistics (Also see Supplementary Figure 1)

Data collection and phasing

	Native	Peak	SeMet Inflection	Remote	2,6-sialyl-N-acetyllactosamine	2,3-sialyl-N-acetyllactosamine
Space group	$P4_32_12$		$P4_32_12$		$P4_32_12$	$P4_32_12$
Cell dimensions a, c (Å)	66.2, 172.3		66.3, 172.6		65.9, 173.0	66.1, 172.7
Beamline	10.1 SRS	10.1 SRS	10.1 SRS	10.1 SRS	14.1 SRS	14.1 SRS
Wavelength	0.980 Å	0.980 Å	0.980 Å	0.972 Å	1.488 Å	1.488 Å
Resolution (Å)*	18.0-1.9 (1.97-1.90)	18.0-2.8 (2.90-2.80)	18.0-2.8 (2.90-2.80)	18.0-2.8 (2.90-2.80)	20.0-2.0 (2.07-2.00)	18.0-2.3 (2.38-2.30)
Total observations	154515	91673	90740	93048	542909	351534
Unique reflections	30153	10069	10030	10070	49275	17796
Redundancy*	5.1 (3.9)	9.1 (9.0)	9.0 (7.5)	9.2 (9.0)	7.6 (7.6)	7.6 (7.8)
R_{sym} (%)†*	7.5 (31.1)	9.1 (23.3)	10.2 (29.1)	9.3 (25.1)	10.5 (53.6)	8.6 (41.3)
$I/\sigma I$ *	19.2 (3.4)	20.5 (9.5)	18.6 (6.4)	20.6 (9.0)	43.54 (6.7)	38.0 (6.7)
Completeness (%)*	96.6 (76.2)	99.9 (100)	99.6 (96.6)	99.9 (100)	99.9 (99.7)	100 (99.9)

Refinement

	Native	2,6-sialyl-N-acetyllactosamine soak	2,3-sialyl-N-acetyllactosamine soak
Resolution (Å)*	17.5-1.9 (1.95-1.90)	18.29-2.07 (2.12-2.07)	17.5-2.3 (2.36-2.30)
No. reflections	28,518	22,744	16,806
R_{factor}§ / R_{free}¶ (%)*	17.8 / 20.4 (29.4 / 31.3)	18.3 / 21.2 (28.4 / 35.6)	17.3 / 21.3 (20.2 / 25.5)
No. atoms			
Protein	1840	1745	1744
Ligand/ion	74	98	83
Water	181	181	153
B-factors			
Protein	26.3	27.1	35.0
Ligand/ion	42.5	52.9	51.8
Water	41.4	39.1	45.8
R. m. s. deviations			
Bond lengths (Å)	0.018	0.021	0.023
Bond angles (º)	1.6	1.8	1.9

[*] Values in parentheses correspond to the highest resolution shell.
† $R_{sym} = \sum_h \sum_i | I_i (h) - <I(h)>| / \sum_h \sum_i (h)$ where $I_i(h)$ is the ith measurement.
§ R factor = $\sum ||F(h)_{obs}| - | F(h)_{calc} || / \sum | F(h)_{obs}$.
¶ R_{free} is calculated in the same way as the R factor, using only 5% of reflections randomly selected to be excluded from the refinement.

Table 2 – Comparisons of TgMIC1 binding signals elicited by selected sialyl probes in the carbohydrate microarray

Name (position in Fig. 4)	Sequence	Fluoresc. Intensity at 7 fmol/spot
SA (3') LacNAc **(3)**	NeuAcα-3Galβ-4GlcNAc	886
SA (6') LacNAc **(6)**	NeuAcα-6Galβ-4GlcNAc	124
Sialyl Le^x **(21)**	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glc | Fucα-3	2520
LSTc **(22)**	NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4Glc	100
6' SU-Sialyl Le^x **(19)**	HSO3-6 | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | Fucα-3	1692
de-*N*-acatyl 6' SU-Sialyl Le^x **(27)**	HSO3-6 | Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | Fucα-3	<1

A2F(A2-3) **(36)**	NeuAcα-3Galβ-4GlcNAcβ-2Manα-6 Fucα-3
	            |
	           Manβ-4GlcNAcβ-4GlcNAc
	NeuAcα-3Galβ-4GlcNAcβ-2Manα-3
		4170

A2 **(37)**	NeuAcα-6Galβ-4GlcNAcβ-2Manα-6 Manβ-4GlcNAcβ-4GlcNAc NeuAcα-6Galβ-4GlcNAcβ-2Manα-3	1197
A3 **(38)**	NeuAcα-6Galβ-4GlcNAcβ-2Manα-6 Manβ-4GlcNAcβ-4GlcNAc NeuAcα-6Galβ-4GlcNAcβ-2Manα-3 | NeuAcα-6Galβ-4GlcNAcβ-2	6777
GT1b **(53)**	NeuAcα-3Galβ-4GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-8 NeuAcα-3	8254
GD1a **(54)**	NeuAcα-3Galβ-4GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-3	4492
GT1a **(55)**	NeuAcα-8NeuAcα-3Galβ-4GalNAcβ-4Galβ-4Glcβ-Cer | NeuAcα-3	204