



UNIVERSITÉ
DE GENÈVE

Archive ouverte UNIGE

<https://archive-ouverte.unige.ch>

Thèse

2009

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Molecular players of sialic acid-dependent host cell invasion by toxoplasma gondii

Friedrich, Jan Nikolas

How to cite

FRIEDRICH, Jan Nikolas. Molecular players of sialic acid-dependent host cell invasion by toxoplasma gondii. Doctoral Thesis, 2009. doi: 10.13097/archive-ouverte/unige:5074

This publication URL: <https://archive-ouverte.unige.ch/unige:5074>

Publication DOI: [10.13097/archive-ouverte/unige:5074](https://doi.org/10.13097/archive-ouverte/unige:5074)

UNIVERSITÉ DE GENÈVE

Département de Microbiologie
et Médecine Moléculaire

FACULTÉ DE MÉDECINE
Prof. Dominique Soldati-Favre

Département de Biologie Moléculaire

FACULTÉ DES SCIENCES
Prof. Stéphane Thore

Molecular Players Of Sialic Acid-Dependent Host Cell Invasion By *Toxoplasma Gondii*

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention biologie

par

Jan Nikolas FRIEDRICH
de
Lübeck (Allemagne)

Thèse N° 4140

GENÈVE
Atelier d'impression ReproMail - UniMail
2009



**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES

**Doctorat ès sciences
mention biologie**

Thèse de *Monsieur Jan Nikolas FRIEDRICH*

intitulée :

**" Molecular Players of Sialic Acid-dependent Host Cell Invasion
by *Toxoplasma gondii* "**

La Faculté des sciences, sur le préavis de Madame D. SOLDATI-FAVRE, professeure associée et directrice de thèse (Faculté de médecine, Département de microbiologie et de médecine moléculaire) de Messieurs S. THORE, professeur assistant et codirecteur de thèse (Département de biologie moléculaire), A. HEMPHILL, professeur (Institute of Parasitology, Vetsuisse Faculty, University of Berne, Switzerland), S. MATTHEWS, professeur (Centre for Structural Biology, Division of Molecular Biosciences, Imperial College, London, United Kingdom), O. HARTLEY (Département de biologie structurale et bioinformatique, Faculté de médecine) et de Madame F. TOMLEY, docteure (Institute for Animal Health, Compton, Newbury, United Kingdom), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 3 septembre 2009

Thèse - 4140 -



Le Doyen, Jean-Marc TRISCONE

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Nombre d'exemplaires à livrer par colis séparé à la Faculté : - 5 -

The present thesis resulted in publication of the following peer-reviewed scientific articles:

Blumenschein TM, 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**(11): 2808-2820

Friedrich N, Santos JM, Liu Y, Palma AS, Leon E, Saouros S, Kiso M, Blackman MJ, Matthews S, Feizi T, Soldati-Favre D (2009) Members of a novel protein family containing MAR domains act as sialic acid-binding lectins during host cell invasion by apicomplexan parasites. *J Biol Chem* In Press

Garnett JA, Liu Y, Leon E, Allman SA, Friedrich N, Saouros S, Curry S, Soldati-Favre D, Davis BG, Feizi T, Matthews S (2009) Detailed insights from microarray and crystallographic studies into carbohydrate recognition by microneme protein 1 (MIC1) of *Toxoplasma gondii*. *Protein Sci* **18**(9): 1935-1947

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**(11): 1149-1155

Abstract

Toxoplasma gondii, the causative agent of toxoplasmosis in animals and human, belongs to a diverse phylum of obligate intracellular protozoan parasites, the Apicomplexa which also includes the genus *Plasmodium* responsible for malaria. Invasion of host cells allows these parasites to scavenge host cell metabolites and provides them with partial protection from the host immune system. Apicomplexan parasites carry a machinery for active host cell invasion which relies on the release of proteins from secretory organelles called micronemes onto the parasite surface. Among the microneme proteins (MICs) characterized to date, a subset forms complexes with non-overlapping functions in host cell recognition and invasion. Adhesive domains in these complexes engage with receptors on the host cell surface and are believed to contribute to host cell attachment and specificity. Apicomplexan parasites vary in their capacity to invade different cell types. *P. falciparum* blood stage parasites have specialized for invasion into red blood cells of humans and great apes. In contrast, the *T. gondii* tachyzoite, which is the fast replicating form associated with acute infection in virtually all warm blooded animals, can invade almost any nucleated cell. Interestingly, *T. gondii* shares the capacity to invade a broad range of cells with other tissue-cyst forming enteroparasites like *Neospora caninum* and *Sarcocystis neurona* which are of veterinary importance.

Sialic acid is part of glycoconjugates on the surface of all vertebrate cells and plays a critical role in host cell recognition during bacterial and viral infections. Sialoglycoconjugates are also exploited by the Apicomplexa during host cell invasion. We found that sialic acid is a major determinant for invasion by *T. gondii*. In this study two *T. gondii* MICs, TgMIC1 and TgMIC13, were characterized as the key players of sialic acid-dependent host cell invasion. These two MICs possess domains that adopt a novel fold, termed Microneme Adhesive Repeat (MAR). This domain defines a new family of proteins conserved among the group of enteroparasitic coccidians. We speculate that the binding specificities of these proteins contribute to tissue tropism during infection.

Résumé

Toxoplasma gondii appartient au phylum des Apicomplexes, qui constitue une large collection de parasites eukaryotiques unicellulaires. Parmi ces parasites, plusieurs sont reconnus comme des pathogènes importants chez les animaux ainsi que chez l'humain. Par exemple, les espèces du genre *Plasmodium* provoquent la malaria et *T. gondii* est responsable de la toxoplasmose, une infection normalement sans conséquence chez les individus immunocompétents mais potentiellement dangereuse chez les personnes dont le système immunitaire est affaibli et pour le fœtus. La transmission congénitale peut amener à des complications graves incluant la mort du fœtus, des avortements avant terme et des malformations.

Tous les Apicomplexes se propagent obligatoirement à l'intérieur des cellules hôtes ce qui présente plusieurs avantages, notamment une accessibilité aux métabolites de l'hôte et une protection partielle contre la défense immunitaire de l'hôte. Ces parasites possèdent une machinerie complexe qui leur permet d'envahir activement la cellule hôte. Ce processus d'invasion dépend strictement de la sécrétion de protéines par les organelles nommées micronèmes. Il est connu que ces protéines micronémales (MICs) forment des complexes et ont des fonctions non-redondantes durant l'invasion. Ces complexes protéiques reconnaissent et s'attachent à des récepteurs dans la matrice extracellulaire ou à la surface de la cellule hôte. Ainsi ils peuvent potentiellement influencer la spécificité du pathogène pour un certains types de cellule. Les stades érythrocytaires de *P. falciparum* sont spécialisés pour envahir les érythrocytes humains et ceux des primates. Par contre, la forme tachyzoite du *Toxoplasme* qui se propage rapidement pendant l'infection aigue dans tous les animaux à sang chaud, est capable d'envahir une grande variété de types de cellule. *T. gondii* partage cette capacité d'envahir une variété de types de cellule avec autre parasites entériques comme *Neospora caninum* et *Sarcocystis neurona* qui sont important dans la médecine vétérinaire.

L'acide sialique fait partie des glycoprotéines et des glycolipides et joue un rôle majeur dans la biologie des vertébrés. Comme il est présent à la surface de toutes les cellules des vertébrés, il est exploité par une multitude de bactéries et de virus. Des récepteurs comprenant l'acide sialique sont aussi utilisés par les Apicomplexes pendant l'invasion. Nos résultats montrent que l'acide sialique joue un rôle primordial pour le processus d'invasion chez *T. gondii*. Dans cette étude nous avons caractérisé deux protéines micronémales, TgMIC1 et TgMIC13, qui montrent une reconnaissance spécifique pour l'acide sialique. D'après nos résultats les deux protéines constituent des acteurs clés lors de l'invasion de la

cellule hôte. TgMIC1 et TgMIC13 comprennent un domaine structural nouveau, nommé MAR, qui définit une nouvelle famille de protéines. Les membres de cette famille ont été identifiés uniquement parmi les coccidiens, un group de parasites qui infectent leurs hôtes par la voie entérique. Nous émettons l'hypothèse que la famille de protéines possédant le domaine MAR contribue au tropisme lors de l'infection parasitaire et implique les spécificités d'adhésion individuelles de ces protéines.

Acknowledgements

It's time to summarize four and a half exciting years that have given me the opportunity to work with and to meet so many nice people. First of all I would like to thank my supervisors Dominique Soldati and Mike Blackman. Dominique, it has been simply great to work with you. I admire your knowledge, enthusiasm and inspiration, all of which make science alive and have continuously stimulated my work. Thanks for your guidance, discussions, and not forgetting your criticism, all from which I have learned so much. Mike, thanks for always being supportive during my time in London, the discussions and your thoughtful advice.

A very important role in my Ph.D has been played by our collaborators from Imperial College. I would like to acknowledge Steve Matthews, Tharin Blumenschein, Kovilen Sawmynaden, as well as Ten Feizi, Yan Liu, Angelina Palma and all other people in their labs who contributed to our work. Not only that these collaborations have been very fruitful for the progress of the project, but also the intense exchange of information together with a friendly working atmosphere made it a very interesting and stimulating experience for me.

I also would like to thank Thierry Soldati, Ian Taylor and Steve Martin for helpful discussions and support with experiments. From the lab in Geneva I would like to acknowledge the important contributions that Joana made to this project. I am very grateful for the effort you have made and appreciate you as a colleague and friend. Also the continuous help of Catherine and Natacha has been a big support in the daily work, many thanks to you.

Furthermore I would like to thank Jeremy and Lara for sharing their flat, putting up with my life style and being supportive in difficult moments. Thanks Jeremy for sharing whiskey and cigars on the balcony. Housing is a difficult thing in Geneva and I would like to mention here the hospitality of Julien (Ist), who accommodated me on several occasions and made it possible for me to do urgent experiments. I am also grateful to Deborah, Romain and their friends for a warm welcome during three weeks of thesis writing.

Obviously, my thanks go as well to all past and present lab members in Geneva and London. The friendly and cooperative working atmosphere has been very important to me. Fabienne, Bernardo, Mark, Tim, Mike and Paco, the skiing trips have been wonderful! Special thanks to Paco for beer o'clock. Special thanks as well to Lilach for sharing late working hours and personal discussions. Tim, I am very grateful to you for introducing me to a bunch of nice friends in London. I will never forget. Also I very much appreciated the daily company of Noelle, Karine, Valerie, Wassim, Tobias, Louise, Julien (IInd) and Jean-Baptiste. Noelle and Marta, I must mention that I appreciate your enthusiasm for the German language (Ach ja!). I will always be a member of the Hampstead-gang: Many thanks to Anna, Marta, Nathalie, Robert and Caty. Caty, your humor (and connections) have been of great value to me in the late time of this Ph.D. Andrea, Rob, Chrislaine, Matt, Fiona, Chris, Sharon, Kostas and Malcolm I am very happy to have shared this time with you.

In the time when I was looking for a PhD position I had the support from Robert Ménard, Joel Couprie, Bernard Gilquin, Leif-Alexander Garbe and Lorenz Adrian. Thanks for believing in my capacities and your help.

Finally, Mélanie, my parents, my sister and family, you have been an invaluable support during this time for me. Apart from all other things you have done for me, Mélanie, I am very grateful for your understanding and patience in all matters until the achievement of this Ph.D.

Contents

Abstract	1
Résumé	3
Acknowledgements	5
Contents	7
List of Abbreviations	9
1. Introduction	11
1.1 The Phylum Apicomplexa	11
1.1.1 Life Cycles	13
1.1.1.1 Life cycle of <i>T. gondii</i>	14
1.1.1.2 <i>Plasmodium</i> life cycle	16
1.2 Host cell invasion	17
1.2.1 The “Glideosome”	20
1.2.2 Induced and constitutive secretion	22
1.2.2.1 Microneme secretion	22
1.2.2.2 Ca^{2+} -homeostasis	23
1.2.2.3 Ca^{2+} -release pathways	24
1.2.2.4 Ca^{2+} -signaling	24
1.2.2.5 Rhoptry secretion	25
1.2.2.6 Dense granule secretion	26
1.2.3 Protein trafficking to secretory organelles	27
1.2.4 Microneme protein proteolysis	28
1.2.4.1 Proteolytic maturation in the secretory pathway	28
1.2.4.2 Shedding of MIC-complexes post exocytosis	29
1.2.4.3 Trimming on the parasite surface	31
1.2.5 The role of glycans in host cell invasion by the Apicomplexa	32
1.2.5.1 Glycan recognition during host cell invasion by <i>T. gondii</i>	33
1.2.5.1.1 A lectin family with conserved function throughout Apicomplexa	34
1.2.5.1.2 More parasite-glycan interactions	35
1.2.5.2 Sialic acid-dependent and Sialic acid-independent invasion pathways in <i>Plasmodium</i>	36
1.2.5.2.1 Duffy-binding-like (DBL)-family of proteins	36
1.2.5.2.2 Reticulocyte binding like (RBL)-family of proteins	39
1.2.5.2.3 Switching between invasion pathways	40
1.3 Hypothesis and Aims of Project	42
2. Materials and Methods	43
2.1 Reagents and Suppliers	43
2.1.1 Enzymes	43
2.1.2 Kits	44
2.1.3 Antibodies	44
2.2 Solutions	45
2.2.1 Culture Media	45
2.2.2 General Solutions	46
2.3 Cell lines and microbiological strains	46
2.3.1 Bacteria	46
2.3.2 <i>Pichia pastoris</i>	47
2.3.3 Mammalian cells	47
2.3.4 <i>Toxoplasma gondii</i> strains	47
2.4 Culture conditions	47
2.4.1 Bacterial culture	47

2.4.2	<i>Pichia pastoris</i> culture	47
2.4.3	Mammalian cell culture.....	48
2.4.4	Parasite propagation	48
2.5	Transformations and transfections	48
2.5.1	<i>E. coli</i> transformation.....	48
2.5.2	<i>Pichia pastoris</i> transformation	48
2.5.3	<i>T. gondii</i> transfection.....	49
2.6	Plasmid construction	49
2.6.1	Agarose gel electrophoresis	49
2.6.2	Preparation of nucleic acids	49
2.6.3	Polymerase chain reaction (PCR)	50
2.6.4	Ligations.....	50
2.6.5	Site-directed mutagenesis.....	51
2.7	Sodium Dodecyl Sulphate (SDS) Polyacrylamide Gel (PAGE) Electrophoresis and Western Blotting	51
3.	Results	53
3.1.	Atomic resolution insight into host cell recognition by <i>T. gondii</i>	53
3.2	Structural insights into microneme protein assembly reveal a new mode of EGF domain recognition.....	69
3.3	Detailed insights from microarray and crystallographic studies into carbohydrate recognition by microneme protein 1 (MIC1) of <i>Toxoplasma gondii</i>	79
3.4	A novel protein family containing MAR domains act as sialic acid lectins during Apicomplexa invasion.....	95
4.	Discussion	157
4.1	Function of MAR containing proteins (MCPs) in <i>T. gondii</i>	157
4.1.1	Connection of adhesive complexes to the actin-myosin motor.....	158
4.1.2	Function of TgMCP3 and TgMCP4.....	158
4.1.3	Complex formation by MCPs	159
4.2	Is there a molecular basis for the broad host cell range of cyst-forming coccidians?	161
4.2.1	MAR and apple domains in host cell recognition	162
4.2.2	Chitin-binding-like (CBL) domains in host cell recognition	163
4.2.3	Role of the surface antigen glycoprotein 1 (SAG1)-related superfamily in host cell recognition.....	164
4.3	Benefits and intervention	166
4.4.	Concluding remarks	167
	References	169

List of Abbreviations

ABA	Absciscic acid
AMA	Apical membrane antigen
APS	Ammonium Persulphate
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CAT	Chloramphenicol AcetylTransferase
cADPR	cyclic adenosine di-phosphate ribose
CBL	Chitin-binding-like
CDPK	Calcium-dependent protein kinase
CHO	Chinese Hamster Ovary
CIP	Calf Intestinal alkaline Phosphatase
CPL	Cathepsin L
DAPI	Diamidine-2'-phenylindole dihydrochloride
DBP	Duffy binding protein
DG	Dense granule
DHFR-TS or DHFR	DiHydroFolate Reductase-Thymidylate Synthase gene
EBA	Erythrocyte binding antigen
EBL	Erythrocyte binding ligand
EEF	Exo-erythrocytic forms
EGF	Epidermal Growth Factor
ER	Endoplasmatic reticulum
EST	Expressed Sequence Tag
F-actin	Filamentous actin
FCS	Foetal Calf Serum
GAP	Gliding Associated Protein
GAG	Glycosaminoglycan
GFP	Green Fluorescent Protein
GPI	glycosylphosphatidylinositol
GRA	Dense granule protein
GST	Glutathione S-transferase
HA	Hemagglutinin
HPM	Host cell plasma membrane
HFF	Human Foreskin Fibroblasts
IFA	Immunofluorescence Assay
IFN	Interferon
IMC	Inner membrane complex
IP3	Inositol-1-4-5-triphosphate
LB	Luria-Bertani
MAC	Membrane attack complex
MAR	Microneme Adhesive Repeat
MARR	MAR-region
MCP	MAR containing protein
MIC	Microneme protein
MLC	Myosin Light Chain
MPA	mycophenolic acid
MPP1	Microneme Protein Protease 1

MPP2	Microneme Protein Protease 2
MPP3	Microneme Protein Protease 3
MSP	Merozoite surface protein
MTIP	Myosin Tail Interacting Protein
NJ	Neighbour Joining
ORF	Open Reading Frame
PAF	Paraformaldehyde
PAGE	Polyacrylamide Gel Electrophoresis
PAN	Plasminogen, apple, nematode
PBS	Phosphate Buffered Saline
PBS-T	PBS with 0.05% Tween-20
PCR	Polymerase Chain Reaction
PEG	Polyethelene Glycol
PF	Perforin
PFP	Pore-forming protein
PI-PLC	Phosphoinositol-phospholipase
PM	Plasma membrane
PMCA	Plasma membrane type calcium-ATPase
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
RBC	Red blood cell
RON	Rhoptry neck protein
ROP	Rhoptry protein
RT-PCR	Reverse-Transcriptase PCR
RyR	Ryanodine receptor
RBL	Reticulocyte binding-like
RBP	Reticulocyte binding protein
Rh	Reticulocyte binding protein homologue
SAG	Surface antigen glycoprotein
SDM	Site Directed Mutagenesis
SDS	Sodium Dodecyl Sulphate
SERCA	sarcoplasmic endoplasmic reticulum calcium-ATPase
Sia	Sialic acid
SRS	SAG1-related superfamily
SUB	Subtilisin-like protease
TAE	Tris/Acetate/EDTA
TAP	Tobacco Acid Pyrophosphatase
TdT	Terminal Deoxynucleotide Transferase
TE	Tris EDTA
TEMED	N,N,N',N'-Tetramethyl-1-,2-diaminomethane
TGN	Trans Golgi Network
TM	Transmembrane
TMD	Transmembrane Domain
TPCK	Tosyl Phenylalanine Chloromethyl Ketone
TRAP	Thrombospondin-related anonymous protein
YPD	Yeast extract-Peptide-Dextrose

1. Introduction

1.1 The Phylum Apicomplexa

Toxoplasma gondii belongs to the phylum Apicomplexa which comprises a large and diverse group of obligate intracellular protozoan parasites defined by a characteristic ultrastructural element, the apical complex (Figure 1.1) (Roos, 2005). This apical complex determines the cell polarity of these protozoans and comprises electron-dense cytoskeletal structures termed the conoid and the polar ring, and two sets of secretory organelles, the club-shaped rhoptries and the smaller round micronemes located in the apical part (Dubey et al, 1998). The polar ring is connected to microtubules that laterally run along the parasite. The microtubules support a layer of flattened vesicular structures which underlie almost the entirety of the parasite plasma membrane forming the inner membrane complex (IMC). As another particularity, in addition to a single mitochondrion most Apicomplexa (except *Cryptosporidium spp.*) harbor a plant-like plastid, the apicoplast, which is believed to be the relic of a secondary algal endosymbiont (Vaishnava & Striepen, 2006).

Many apicomplexan parasites are of veterinary or clinical importance such as the *Plasmodium* species that are the etiological agent of malaria and *Cryptosporidium* responsible for opportunistic infections in human. Lifestocks are threatened for example by *Theileria spp.* causing East Coast Fever in cattle and *Eimeria tenella* being responsible for coccidiosis in poultry. *T. gondii* infects all warm-blooded animals and humans. While infection in healthy individuals is usually asymptomatic, the parasite can cause severe disseminated disease in immunocompromised individuals and is a danger for the developing foetus (Montoya & Liesenfeld, 2004). Congenital infection can result in prenatal death, spontaneous abortion and neonatal malformations. Besides being an important pathogen *T. gondii* is an attractive model organism for other members of the phylum as the parasite is comparatively easy to manipulate.

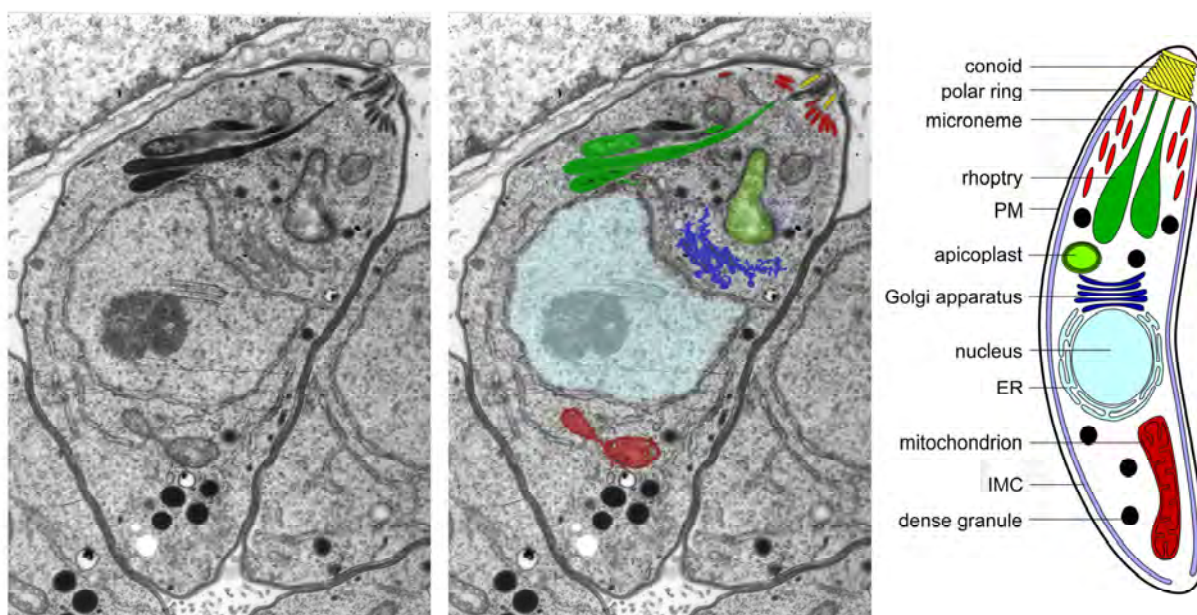


Figure 1.1: Ultrastructure of *T. gondii*.

The left panel shows an electron micrograph of a *T. gondii* tachyzoite growing within a parasitophorous vacuole inside a host cell (courtesy of Dr. Jean-Francois Dubremetz). On the right is a schematic illustrating the structural features of a tachyzoite (courtesy of Dr. Tim Dowse). Colours used in this schematic match with the artificial colouration used in the centre panel to indicate the position of the individual structures. PM: Plasma membrane, ER: Endoplasmatic Reticulum, IMC: Inner membrane complex.

This thesis will focus on *Toxoplasma gondii*, but other coccidians which constitute a group of closely related enteroparasites within the phylum Apicomplexa will also be discussed. Although *Plasmodium* is only more distantly related to *T. gondii*, this parasite will be introduced in some detail because of its importance as a pathogen and because of some unique and remarkable features that are important to keep in mind when considering our findings in *T. gondii*.

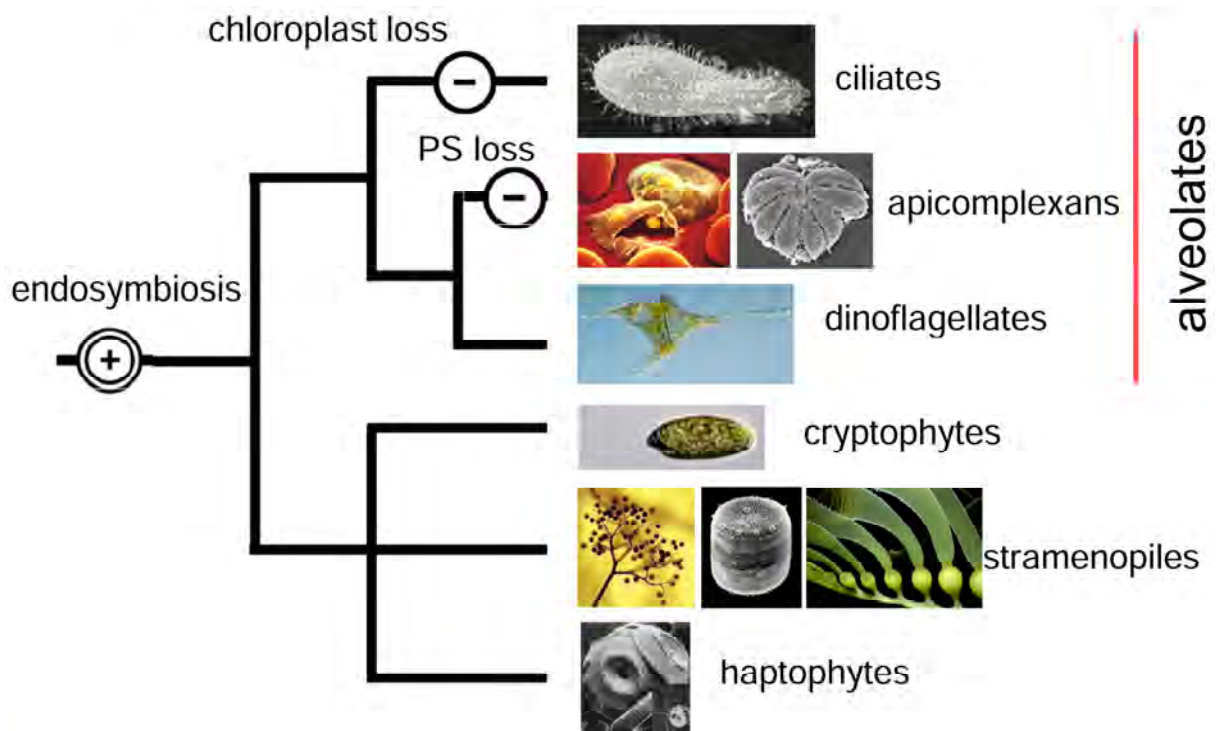


Figure 1.2: Tree of life indicating the position of the phylum Apicomplexa.

Courtesy of Dr. Bernardo Foth. PS: photosynthesis.

1.1.1 Life Cycles

Apicomplexan parasites have complex life cycles comprising alternating phases of asexual and sexual development. For many, but not all Apicomplexa asexual and sexual development takes place in distinct hosts called the intermediate host and the definitive host respectively. These parasites are heteroxenous in contrast to monoxenous parasites which develop and are transmitted between a single host species. In course of the life cycle the Apicomplexa undergo several differentiation steps to be optimally adapted to each niche within a given host. All invasive stages are called zoites.

T. gondii belongs to the coccidia, a large subgroup of apicomplexan enteroparasites (Figure 1.3). In general tissue cyst-forming coccidia are heteroxenous. The group of cyst-forming coccidians includes the genera *Toxoplasma*, *Neospora*, *Hammonida*, *Besnoita*, *Sarcocystis* and *Frenkelia*. Whereas specificity for a certain definitive host appears to be high in these genera, they share the ability to infect a more or less broad range of intermediate hosts (Frenkel & Smith, 2003). Examples for non cyst-forming coccidians are parasites of the

genus *Eimeria*. Each species of this genus infects a very limited number of host species some appearing to be monoxenous (Kvicerova et al, 2007; Levine & Ivens, 1988; Vetterling, 1976). *Eimeria* parasites develop in the host intestine and shedding of oocysts with the feces allows transmission.

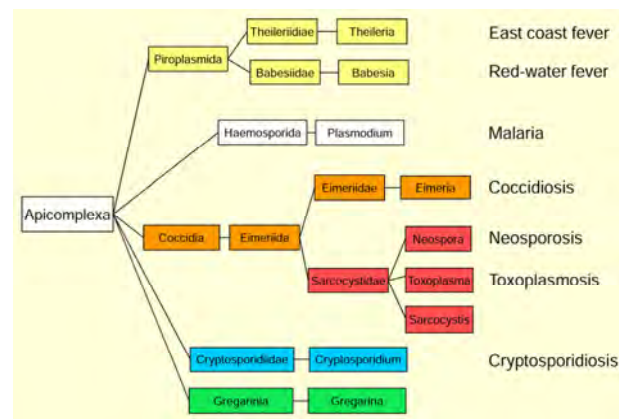


Figure 1.3: Evolutionary relationships between members of the phylum Apicomplexa.

To the right are indicated the pathologies caused by the respective organism.

1.1.1.1 Life cycle of *T. gondii*

For sexual development *T. gondii* relies on felines, such as the domestic cat, as definitive host. In contrast virtually all warm-blooded animals including humans can serve as intermediate hosts. Individuals can get infected through ingestion of sporulated oocysts originating from the feces of cats (Montoya & Liesenfeld, 2004). Sporulated oocysts are shed from the intestines of the cat, which are extremely long-lived (Dubey et al, 1998). Once ingested the sporozoite transforms into the rapidly proliferating tachyzoite form that spreads through all tissues and establishes infection. This involves invasion of a large variety of host cells. Asexual replication of tachyzoites within a cell proceeds through a lytic cycle (Figure 1.4) (Black & Boothroyd, 2000). Following invasion of a host cell, the zoite multiplies within a vacuole by binary fission (endodyogeny (Striepen et al, 2007)) until the progeny are finally released which are able to infect neighboring cells.

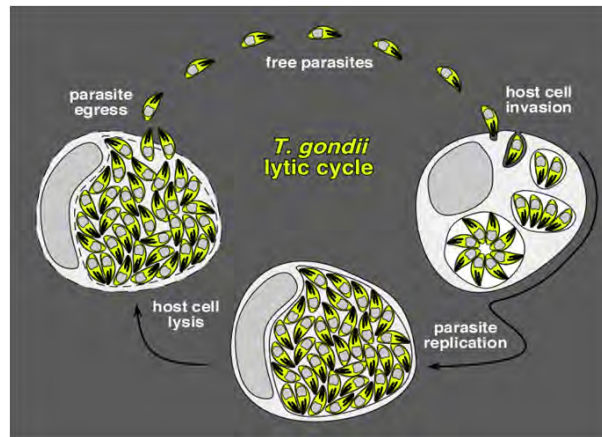


Figure 1.4: Lytic cycle of the *T. gondii* tachyzoite.

The cartoon has been taken from:

www2.bc.edu/~gubbelsj/Toxoplasma.html

The immune response of the host is then believed to bring about the differentiation of the tachyzoite to the bradyzoite form (Weiss & Kim, 2000). Bradyzoites usually persist for the lifespan of the infected individual inside cysts that are mainly found in muscle and brain tissues. In case of a drop in the immune pressure the semi-dormant encysted bradyzoites can be reactivated and redifferentiate into tachyzoites. Stage conversion frequently happens in immunocompromised individuals. Encysted bradyzoites, and to a much lesser extent tachyzoites, are infectious when ingested, for example in undercooked meat (Montoya & Liesenfeld, 2004). In this way the parasite can be transmitted by carnivorous to its definitive host or infect other intermediate hosts (Figure 1.5 summarizes the *T. gondii* life cycle). The latter route of transmission between intermediate hosts without passage through the definitive host is a unique feature of *T. gondii* and *N. caninum* and is believed to have largely contributed to the success of *T. gondii* in being one of the most widespread pathogens (Su et al, 2003).

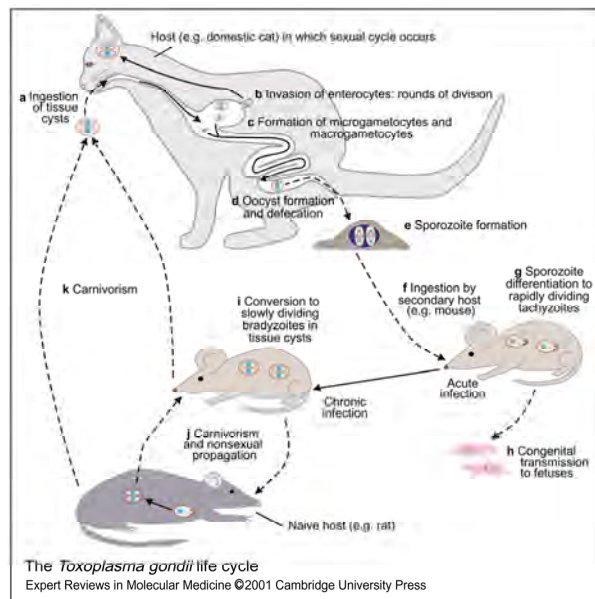


Figure 1.5: Schematic of the life cycle of *T. gondii*.

Taken from (Ajioka et al, 2001).

1.1.1.2 *Plasmodium* life cycle

Infectious *Plasmodium* sporozoites are transmitted by a mosquito and deposited into the skin of an intermediate host during a blood-meal. While different species of mosquitos belonging to the family *Anopheles* are shared by all *Plasmodium* species as definitive hosts, the individual parasite species have adapted to development within different intermediate hosts. Five *Plasmodium* species are known to date to infect humans, which are *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi* (Collins & Barnwell, 2009). Figure 1.6 summarizes the *P. falciparum* lifecycle.

After injection into the skin, the sporozoite traverses the epithelium and tries to get to the liver (Amino et al, 2008). Following invasion of a hepatocyte, the parasite multiplies and develops into exo-erythrocytic forms (EEFs) inside its vacuole (Sturm et al, 2009). Finally the parasites differentiate into merozoites and are released into the blood-stream thereby initiating a lytic cycle of asexual replication in red blood cells (RBCs). Merozoites multiply by schizogony proceeding through several characteristic morphological steps called the ring stage, the trophozoite stage and the schizont stage. The cycle of asexual replication in RBCs also leads to the formation of gametocytes, the precursors of male and female gametes that initiate the sexual stages once taken up by a mosquito into its midgut during a blood meal. Male gamete formation is associated with exflagellation. Male and female gametes fuse to

from the ookinete which invades the mosquito midgut epithelium and develops into an oocyst near the basal lamina. The oocyst stage is the only replicative phase of the parasite that does not take place within a host cell. Instead the oocyst is surrounded by an inner layer of parasite origin and an outer layer derived from the basal lamina. Replication leads to the generation of sporozoites which travel via the hemolymph, the fluid that bathes the mosquito cavity, to the salivary glands where they form a reservoir rendering the mosquito infectious for life (Matuschewski, 2006). From the salivary glands, sporozoites are then injected again into an intermediate host closing the life cycle.

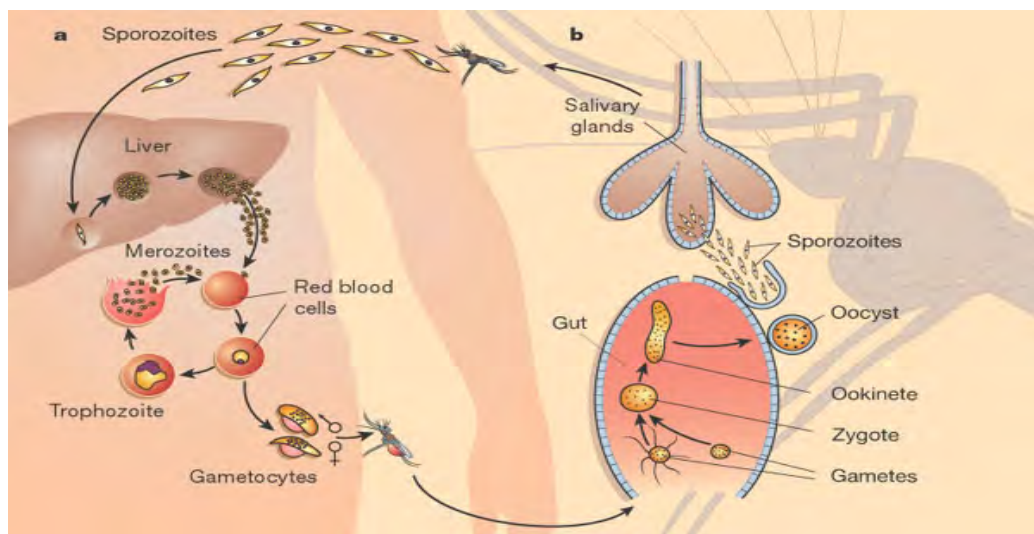


Figure 1.6: Schematic of the life cycle of *P. falciparum*.

(a) The asexual part of the life-cycle in the intermediate host including liver stage and blood stage development. **(b)** Sexual development within the definitive host. Taken from (Menard, 2005).

1.2 Host cell invasion

Invasion of host cells by *T. gondii* and other apicomplexan parasites is a complex process (Figure 1.7) that is fundamentally different from the entry mechanisms of bacteria and viruses which exploit host endocytic uptake pathways. Apicomplexan parasites carry their own machinery for active penetration into the host cell (Carruthers & Boothroyd, 2007). Invasion is initiated by initial attachment of the parasite to the host cell. Attachment may occur in any position of the parasite relative to the host cell and is followed by reorientation such that the parasite's apical tip contacts the surface of the host cell. At this point the attachment becomes

tighter. An electron-dense circular structure termed the moving junction is built which remains at the periphery of the parasite while it actively propels itself into the host cell. During this process the host cell plasma membrane (HPM) progressively invaginates, inducing the formation of the parasitophorous vacuole (PV) which is finally sealed behind the parasite. Penetration of the host cell is driven by a parasite actin-myosin motor system located in the space between the parasite plasma membrane (PM) and the underlying IMC (Soldati-Favre, 2008). This motor complex provides the force for the unique form of substrate-dependant gliding locomotion. The whole machinery that works in gliding motility including this motor complex has been named the “glideosome”. Successful host cell invasion relies on the regulated sequential secretion of proteins from two types of membrane-bound organelles (Carruthers & Sibley, 1997). Discharge of micronemes occurs first, followed by release of proteins from the rhoptries. Proteins stored in these two sets of organelles have been implicated directly in the invasion process. Finally, an additional set of organelles, the dense granules (DGs), secrete their content into the PV in order to establish the parasite in its new environment within the host cell.

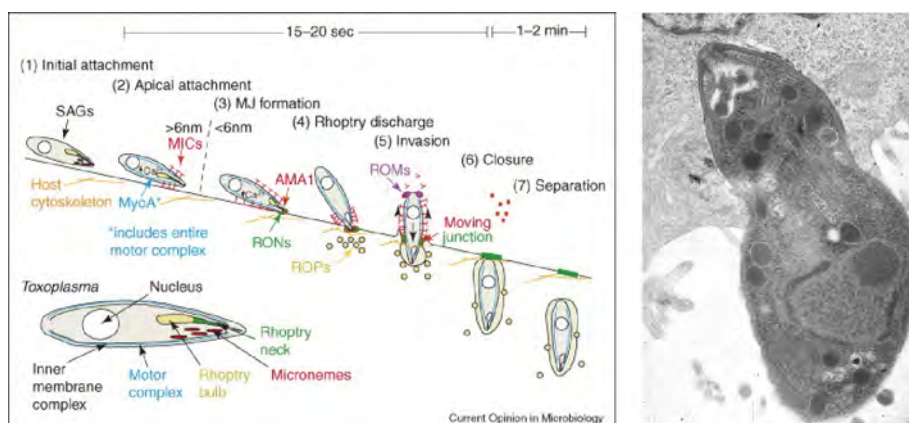


Figure 1.7: Host cell invasion by *T. gondii*.

The schematic on the left represents the various steps of invasion from initial attachment to the host cell until the separation of the parasitophorous vacuole containing the parasite from the host cell plasma membrane. MICs are already present on the parasite surface from the start of the process. Subcellular structures of the parasite are coloured in the cartoon on the bottom left corner. Corresponding colours are used for the name of proteins to indicate the structure they are associated with. Taken from (Carruthers & Boothroyd, 2007). SAG: Surface antigen glycoprotein, MyoA: myosin A, AMA1: Apical membrane antigen 1, ROM: Rhomboid-like protease, MJ: Moving junction. On the right: Electron micrograph of an invading *T. gondii* tachyzoite. Courtesy of Dr. J.-F. Dubremetz.

Microneme proteins (MICs) play essential and non-overlapping roles in the invasion process (Carruthers & Tomley, 2008; Soldati-Favre, 2008). A large repertoire of proteins with adhesive properties has been identified that is delivered onto the parasite surface (Figure 1.8). MICs contain a limited number of adhesive domain types which are arranged in different numbers of copies and in various combinations. For a subset of MICs the formation of complexes has been demonstrated which are already assembled in the endoplasmatic reticulum (ER), travel as an entity through the secretory pathway and are finally released prior to invasion. In *T. gondii*, four complexes composed of transmembrane and soluble MICs and in one case rhoptry neck proteins (RONs) with non-overlapping function in invasion have been described. More complexes are, however, likely to contribute to invasion since additional un-characterized transmembrane microneme proteins (TM-MICs) are encoded in the genome. The specific contribution of each of the four complexes to the invasion process has been uncovered by generating conventional or conditional knockouts of the genes coding for the components of the complexes. TgMIC2 is a TM-MIC that belongs to the thrombospondin-related anonymous protein (TRAP)-family conserved across the phylum Apicomplexa. Members of the TRAP family have attracted much attention due to their central role in gliding motility and host cell invasion (Morahan et al, 2009). TgMIC2 and a soluble partner MIC2-associated-protein, TgM2AP form a multimeric complex (Jewett & Sibley, 2004; Rabenau et al, 2001). This complex grips on receptors present on the extracellular matrix or on the surface of host cells while at the same time being connected to the glideosome via their cytoplasmic tails thus acting as force transducers (Buscaglia et al, 2003; Huynh & Carruthers, 2006; Jewett & Sibley, 2003; Zheng et al, 2009). Parasites depleted for TgMIC2 are markedly deficient in host-cell attachment and gliding motility and are unable to invade host cells (Huynh & Carruthers, 2006). A second complex consists of the TM-MIC TgMIC6, and two soluble MICs, TgMIC1 and TgMIC4. Genetic disruption of any of the three corresponding genes suggests a non-essential role (Reiss et al, 2001), even if the complex has been demonstrated to play an important role for invasion *in vitro* and virulence *in vivo* (Blumenschein et al, 2007; Cerede et al, 2005; Sawmynaden et al, 2008). A third complex is formed by association of the soluble protein TgMIC3 with TgMIC8, a TM-MIC whose genetic disruption interferes with secretion of rhoptry proteins, hence preventing junction formation and completion of invasion (Kessler et al, 2008). Finally, the TM-MIC apical membrane antigen 1 (TgAMA1) anchors a fourth complex to the parasite PM. This complex is unique in that it localizes specifically to the moving junction (Alexander et al, 2005) and

contains the rhoptry proteins TgRON2, TgRON4, TgRON5 and TgRON8 (Alexander et al, 2005; Straub et al, 2009) rather than other MICs, suggesting that TgAMA1 operates downstream of TgMIC8 (Kessler et al, 2008). Parasites depleted in TgAMA1 show normal attachment but are defective in rhoptry secretion, fail to create a junction and are consequently unable to invade a host cell (Mital et al, 2005). Parasite motility is not affected in this mutant. The function, properties, proteolytic processing and trafficking of these four complexes (summarized in table 1.1) will be discussed in detail in the following sections.

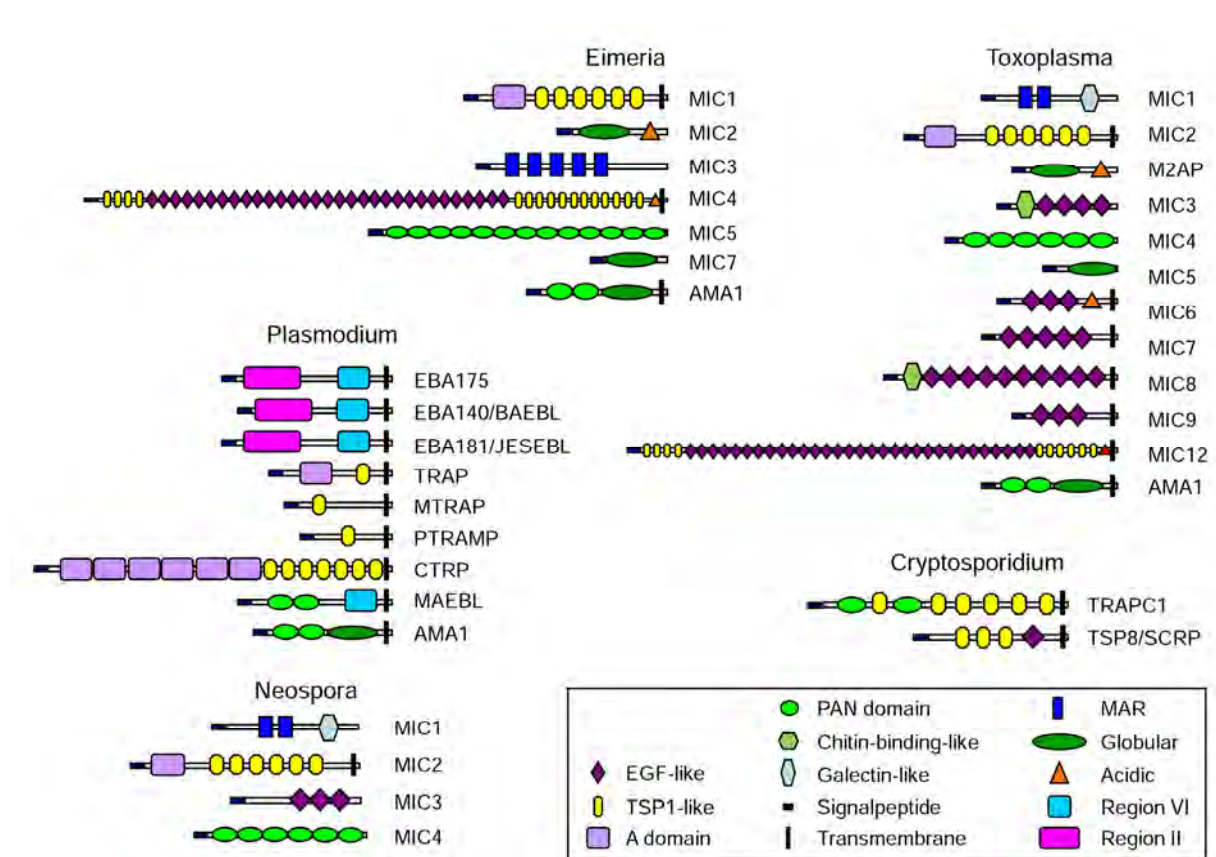


Figure 1.8: Representation of a non-exhaustive repertoire of microneme proteins from apicomplexan parasites. The schematic indicates the domain composition of the individual proteins.

1.2.1 The “Glideosome”

The glideosome is the actin-myosin motor machinery located between the parasite plasma membrane and the underlying IMC which is necessary for substrate dependent gliding motility and active host cell penetration in Apicomplexa (Daher & Soldati-Favre, 2009; Soldati-Favre, 2008). The current model of how this motor works is as follows (Figure 1.9). The association of the IMC with subpellicular microtubules provides a rigid support to anchor

the motor. In *T. gondii* two proteins called gliding associated proteins (GAP) 45 and GAP50 were found to be embedded in the IMC (Gaskins et al, 2004) anchoring Myosin A (TgMyoA) and the myosin light chain 1 (TgMLC1) in a relative fixed position. TgMyoA is essential for gliding motility and invasion (Meissner et al, 2002b) and “walks” on actin-filaments (F-actin) to generate forces and movements (Dobrowolski et al, 1997; Dobrowolski & Sibley, 1996). TgMIC2 and possibly other microneme proteins are translocated by this motor complex over the parasite surface towards the posterior. Rearward translocation of TgMIC2 depends on the interaction of its cytoplasmic tail with the F-actin-binding protein aldolase, which functions as a molecular bridge between TgMIC2 and the motor complex (Buscaglia et al, 2003; Huynh & Carruthers, 2006; Jewett & Sibley, 2003; Starnes et al, 2009; Zheng et al, 2009). Attachment of the TgMIC2-M2AP complex to a receptor on the extracellular matrix or on the host cell surface will then transduce the force generated by the motor complex and bring about a forward movement of the parasite.

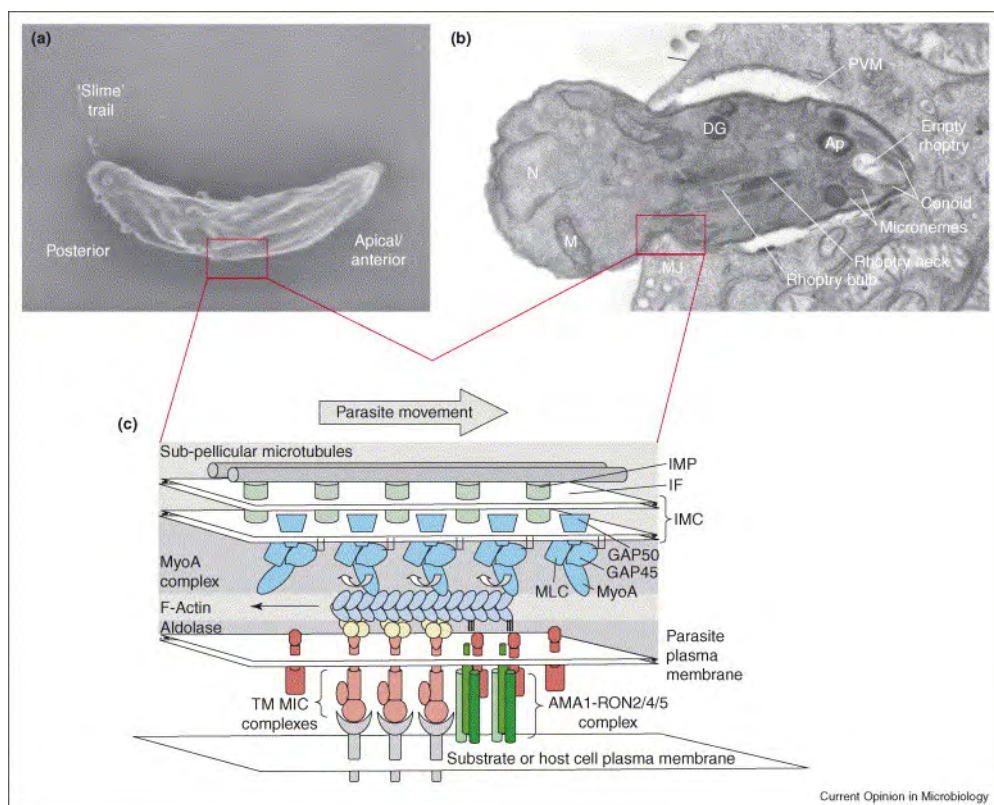


Figure 1.9: The glideosome in *T. gondii*. (a) Scanning electron micrograph of a gliding tachyzoite. (b) Transmission electron micrograph of a tachyzoite in the process of host cell invasion. N: Nucleus, M: Mitochondrion, DG: Dense Granule, Ap: Apicoplast, MJ: Moving junction, PVM: Parasitophorous vacuolar membrane. (c) Current model of the glideosome machinery engaging with receptors present on the extracellular matrix or on the host cell. IF: Intermediate filament-like, IMP: Intramembrane particles. Taken from (Carruthers & Boothroyd, 2007).

1.2.2 Induced and constitutive secretion

1.2.2.1 Microneme secretion

Since a large repertoire of proteins essential for motility and host cell invasion is stored in the micronemes, the nature and timely control of their secretion has been under intense investigation. Microneme secretion in *T. gondii* is thought to resemble Ca^{2+} -mediated exocytosis in other organisms. A major burst of secretion from the micronemes occurs upon contact between the parasite and the host cell (Carruthers & Sibley, 1997). Furthermore, multiple observations on gliding of extracellular parasites show that parasite motility is sustained by oscillating waves of Ca^{2+} (Huynh & Carruthers, 2006; Lovett & Sibley, 2003; Wetzel et al, 2004). More recently direct evidence was obtained that secretion of microneme content occurs already at the point of egress from a host cell because at least one molecular factor critical for egress, the perforin-like protein 1 (TgPLP1), is released from these organelles (Kafsack et al, 2009). This implies that egressing parasites are already fully prepared to enter a neighboring cell, which is important since the parasite does not survive for long in the extracellular environment.

Experiments in *Plasmodium berghei* (Gantt et al, 2000) and *Cryptosporidium parvum* (Chen et al, 2004) provide evidence that the Ca^{2+} -dependent secretion pathway for micronemes is conserved among Apicomplexa. A common mechanism for microneme secretion is further supported by the fact that genes coding for proteins with established or hypothetical function in Ca^{2+} -dependent signaling are conserved among all members of the phylum (Nagamune & Sibley, 2006). Moreover conserved factors of motility and invasion are stored in these organelles indicating functional equivalence. Consequently a common mechanism for microneme release would be intuitively assumed.

Microneme secretion can be artificially induced in a variety of Apicomplexa by Ca^{2+} -ionophore or ethanol treatment, but not in *E. tenella* sporozoites (Wiersma et al, 2004). Instead in this organism microneme secretion can be stimulated by BSA or FCS in a Ca^{2+} -independent manner but the significance of this remains unclear (Bumstead & Tomley, 2000; Wiersma et al, 2004).

Among protozoan parasites, the Ca^{2+} -signaling pathway is best understood in *T. gondii* (for a review see (Nagamune et al, 2008)). Interestingly Ca^{2+} -signaling in apicomplexan parasites shares similarities with the pathways described in plants and these elements are believed to having been acquired from the apicoplast. The four major aspects of research

associated with Ca^{2+} -signalling in *T. gondii* are: (1) Ca^{2+} -homeostasis in the parasite; (2) pathways for intraparasitic Ca^{2+} -release; (3) molecular players of the Ca^{2+} -response leading to microneme secretion; and (4) identification of stimuli leading to Ca^{2+} -signaling and egress. The importance of Ca^{2+} -signaling for parasite biology are underscored by the fact that mutants that do not respond to Ca^{2+} -ionophore treatment are defective in the establishment of infections *in vivo* (Lavine et al, 2007).

1.2.2.2 Ca^{2+} -homeostasis

Microneme secretion in extracellular *T. gondii* parasites relies on release of Ca^{2+} from intraparasitic stores (Lovett et al, 2002). Intraparasitic stores of calcium are the ER, the mitochondrion, the acidocalcisomes and in case of *Plasmodium* also the food vacuole (Nagamune et al, 2008; Rohrbach et al, 2005). However the rapid availability of free Ca^{2+} from the acidocalcisomes has been questioned since most of it is bound to polyphosphates. Ca^{2+} -homeostasis in protozoan parasites is achieved by several Ca^{2+} -ATPases and ion exchangers. A comparative genomic analysis indicated conservation of a sarcoplasmic endoplasmic reticulum Ca^{2+} -ATPase (SERCA) pump, Golgi-type Ca^{2+} -ATPases and a $\text{Ca}^{2+}/\text{H}^{+}$ -exchanger among the phylum Apicomplexa (Nagamune & Sibley, 2006). SERCA pumps are known to be important for Ca^{2+} -homeostasis (Nagamune et al, 2007) and in *Plasmodium* this pump (PfATPase6) is one of the proposed targets of the antimalarial drug artemisinin. In contrast plasma membrane type Ca^{2+} -ATPases (PMCA) and voltage-gated Ca^{2+} -channels were only found in *T. gondii* (Nagamune & Sibley, 2006). The contribution to Ca^{2+} -homeostasis by the PMCA TgA1 expressed on acidocalcisomes and the parasite PM was confirmed. The TgA1 deficient strain shows elevated Ca^{2+} resting levels, altered microneme secretion as well as impaired infectivity *in vitro* and *in vivo* (Luo et al, 2005). The level of complexity has been further enhanced since the discovery of a PM-located $\text{Na}^{+}/\text{H}^{+}$ exchanger (TgNHE1) involved in Ca^{2+} -homeostasis in *T. gondii* (Arrizabalaga et al, 2004), implicating that sodium and proton gradients most likely contribute to the balance. $\text{Na}^{+}/\text{H}^{+}$ -exchanger are present in other Apicomplexa and they were proposed to act either on the plasma membrane or on acidocalcisomes (Arrizabalaga et al, 2004).

1.2.2.3 Ca^{2+} -release pathways

Concerning Ca^{2+} -release, pharmacological studies have demonstrated that *T. gondii* responds to agonists and antagonists of inositol-1-4-5-triphosphate-receptor (IP3R) and ryanodine receptor (RyR) channels (Carruthers et al, 1999b; Chini et al, 2005; Lovett et al, 2002). This suggests that the primary source of free Ca^{2+} for intraparasitic signaling is the ER. Release of Ca^{2+} from intraparasitic IP3-sensitive stores has also been monitored in *Plasmodium* (Passos & Garcia, 1998). Similar to the situation in plants (that also respond to IP3R and RyR agonists), both families of intracellular Ca^{2+} release channels appear to be absent from all apicomplexan genomes studied to date suggesting that the responsive channels are of divergent nature (Nagamune & Sibley, 2006). However, a physiological role of a RyR in *T. gondii* is corroborated by the characterization of cyclase and hydrolase activities involved in the turnover of the secondary messenger cyclic ADP ribose (cADPR) (Chini et al, 2005) which is a natural ligand of RyRs. Furthermore Chini *et al.* confirmed the existence of two separate pathways for Ca^{2+} -signalling involving IP3Rs and RyRs and leading to microneme secretion. Furthermore, microneme secretion in *T. gondii* can be stimulated with ethanol probably through activation of a phosphoinositol-phospholipase C (PI-PLC), which could generate IP3 (Carruthers et al, 1999b). Of note, all Apicomplexans have a delta type PI-PLC. TgPI-PLC has been characterized and localizes to the cytoplasmic face of the parasite plasma membrane (Fang et al, 2006). However a PI-PLC inhibitor previously shown to cause a block in permeabilization-induced egress (Moudy et al, 2001) was ineffective against the recombinant enzyme.

1.2.2.4 Ca^{2+} -signaling

A group of Ca^{2+} -dependent protein kinases (CDPKs) are believed to act downstream of Ca^{2+} -release together with a range of other Ca^{2+} -responsive proteins (calmodulin, calnexin-like and caltractin-like proteins). These CDPKs are otherwise only present in algae and plants and are therefore considered as potential drug targets. Inhibitor studies in *T. gondii* suggest that TgCDPK1 is necessary for microneme secretion and motility (Dobrowolski et al, 1997; Kieschnick et al, 2001). In addition, another class of protein kinases has been implicated in microneme secretion in *T. gondii* and *E. tenella* (Wiersma et al, 2004). Pharmacological

studies suggest that a cyclic GMP-dependant protein kinase (cGMP-PK) is acting downstream of Ca^{2+} -dependant events.

Pharmacological studies in *Plasmodium* support a role for PfCDPK1 which is most closely related to TgCDPK3 in motility. PfCDPK1 is believed to be responsible for the phosphorylation of two components of the motor complex that drives parasite motility: The myosin light chain, called myosin tail interacting protein (PfMTIP) and the membrane-anchoring protein PfGAP45 (Green et al, 2008; Kato et al, 2008). For this purpose, PfCDPK1 is suitably located in being anchored in the plasma membrane. A specific inhibitor (purfalcamin) of this enzyme blocks host cell invasion by *T. gondii* tachyzoites (Kato et al, 2008). Furthermore, a compound related to KT5926 (K252a) shown to be active against PfCDPK1 inhibits invasion of *Plasmodium* merozoites into RBCs at low concentrations (Green et al, 2008). Collectively the data suggests that PfCDPK1 is essential for blood stage parasites.

1.2.2.5 Rhoptry secretion

Rhoptry secretion appears to rely on contact with a host cell since it has been observed only during invasion following discharge of the micronemes and reorientation of the parasite but before formation of the moving junction (Alexander et al, 2005; Carruthers & Sibley, 1997; Mital et al, 2005). The stimulus and exact molecular mechanism by which rhoptry secretion is elicited remains unknown. From analysis of conditional knockout strains for TgAMA1 and TgMIC8 it was concluded that the two complexes defined by TgMIC8 and TgAMA1 are working upstream of rhoptry secretion (Kessler et al, 2008; Mital et al, 2005). Interestingly, recent work in *P. falciparum* has highlighted an essential role for the cytoplasmic domain of PfAMA1 in a step following reorientation (Treeck et al, 2009). It is therefore conceivable that the PfAMA1 cytoplasmic domain is involved in signal transduction connected to the discharge of the rhoptries.

Proteins are stored in two distinct locations within the club-shaped rhoptries. ROPs (for rhoptry protein) are stored in the bulb, whereas those located in the duct of the rhoptries are called rhoptry neck proteins (RONs). Secretion is believed to take place through the duct and an opening at the apical tip of the parasite releasing the proteins into the host cell cytoplasm either within empty-vacuoles (e-vacuoles) or as soluble proteins (Boothroyd & Dubremetz, 2008). The mechanistic details on how injection is achieved remain a mystery,

but patch-clamp experiments have detected a break in the host plasma membrane very early during the invasion process (Suss-Toby et al, 1996). Release of RONS appears to come first and is followed by the release of the ROPs. After secretion, ROPs have been found in various locations either in the lumen of the nascent PV (TgROP1), associated to the PV membrane (TgROP2 family) or inside the host cell nucleus (TgROP16). In contrast, all released RONS have so far been specifically localized to the moving junction being associated with the membrane of the nascent PV (Besteiro et al, 2009). Whereas TgRON4 and TgRON8 have been located on the cytoplasmic face of the host cell PM, TgRON2 and TgRON5 were proposed to be inserted into this membrane through membrane spanning domains however, as for most putative transmembrane rhoptry proteins, their true topology remains to be established.

1.2.2.6 Dense granule secretion

In *T. gondii* proteins secreted through the dense granules are called granule proteins (GRAs) and are anticipated to fulfill important roles in the remodeling of the PV (Coppens et al, 2006; Mercier et al, 2002). Although a major burst of secretion from the dense granules occurs shortly after host cell invasion (Carruthers & Sibley, 1997), release of proteins in a constitutive, Ca^{2+} -independent fashion is widely accepted (Chaturvedi et al, 1999). In numerous examples soluble proteins that do not or are prevented to associate with transmembrane proteins containing specific targeting information (escorters) were delivered to the PV via the dense granules (Karsten et al, 1998; Reiss et al, 2001; Striepen et al, 1998; Striepen et al, 2001). Therefore the dense granules are considered to be the default pathway for secretion of soluble proteins in *T. gondii*. The requirements for transmembrane proteins to be secreted through the DGs are not completely understood but seem to be parasite specific (Gendrin et al, 2008). Since all GRA proteins with predicted TMDs are targeted to specific subcompartments of the PV and do not appear to associate with the parasite PM, it remains unclear if the parasite uses these organelles for delivery of endogenous transmembrane and GPI-anchored proteins onto the PM. However, some foreign and mutant TM-proteins were found to be secreted onto the parasite PM via the dense granules (Hoppe et al, 2000; Karsten et al, 1998; Opitz et al, 2002).

1.2.3 Protein trafficking to secretory organelles

Secretory proteins are dispatched from the trans-Golgi-network (TGN) to the different organelles. Post-Golgi sorting in *T. gondii* is believed to resemble in parts the sorting mechanisms of vesicles to secretory organelles in higher eukaryotes. Both tyrosine-based and dileucine motifs in the cytoplasmic tails of type I transmembrane proteins have been identified to mediate targeting to micronemes and rhoptries (Di Cristina et al, 2000; Hoppe et al, 2000; Ngo et al, 2003). In addition an unusual phenylalanine-based motif exposed to the cytosol was suggested to be a targeting determinant for polytopic transmembrane proteins to the micronemes (Sheiner et al, 2008). The tyrosine-based sorting motif might be conserved in *P. falciparum* as mutation of this motif in the cytoplasmic tail of the micronemal protein PfTRAP resulted in its mistargeting (Bhanot et al, 2003). All soluble rhoptry and microneme proteins are believed to travel to the correct organelle after assembly of complexes including transmembrane proteins that carry the appropriate targeting information. These transmembrane proteins have therefore been termed escorts and need to recruit their cargo at an early stage. Indeed, complex formation of the two soluble proteins TgMIC1 and TgMIC4 with the type I transmembrane protein TgMIC6 was shown to take place already in the ER (Reiss et al, 2001). In this case the appropriate loading of the escorter TgMIC6 with its cargo is controlled through TgMIC1 by assisting in the proper folding of TgMIC6 which otherwise cannot leave the ER (Saouros et al, 2005). Similar kind of quality control checkpoints might exist for other complexes. For example, in the absence of soluble TgM2AP (*m2apko*), its escorter TgMIC2 partially accumulates in the early secretory pathway, which in this case leads to a clear defect in host cell entry by the parasite (Harper et al, 2006).

Several studies on microneme targeting support the idea that some proteins and complexes should associate with yet unidentified escorts (Kessler et al, 2008; Treeck et al, 2006; Treeck et al, 2009). In some cases the protein domains that bring about correct targeting and therefore should mediate the association with the escorter have been identified. For example analysis of deletion mutants of *P. falciparum* EBL proteins showed that a conserved cysteine-rich region (region VI) is required for their microneme targeting (Treeck et al, 2006). In case of the GPI-anchored subtilisin-like protease TgSUB1, the prodomain appears to be essential for correct sorting as its fusion to GFP brings the reporter to the micronemes (Binder et al, 2008).

Accumulating data also suggest a role for propeptide sequences present in several MICs in microneme targeting and were proposed to interact with luminal receptors or lipid rafts supporting correct targeting as has been observed in other organisms (Harper et al, 2006). These propeptides are proteolytically removed post-Golgi but before the proteins reach the micronemes. Deletion of the propeptide in TgM2AP resulted in retention of the TgMIC2-M2AP complex in an endosome-like compartment (Harper et al, 2006). Similarly, a TgMIC5 mutant lacking the propeptide failed to traffic to the micronemes (Brydges et al, 2008). Of note, the TgMIC3 propeptide in combination with any of the EGF domains was shown to be the minimum requirement for targeting to the micronemes (El Hajj et al, 2008).

1.2.4 Microneme protein proteolysis

Proteolytic processing has been observed for several microneme proteins during transport to the micronemes and on the parasite surface. Several of these events are believed to be essential for productive host cell invasion; therefore the molecular identification of the proteases involved is of considerable interest. The proteolytic events that involve the four major invasion complexes in *T. gondii* are recapitulated in figure 1.10.

1.2.4.1 Proteolytic maturation in the secretory pathway

In *T. gondii*, several microneme proteins are known to be proteolytically processed along the secretory pathway at the hands of unknown proteases. TgMIC6 loses its most N-terminal EGF-like domain, probably in the trans-Golgi network (Meissner et al, 2002a), but the significance of this event remains unclear. A mutant in which this domain has been deleted shows no defect in targeting, but the possibility remains that it is involved in “loading” the second EGF domain in TgMIC6 with TgMIC1 and TgMIC4 (Reiss et al, 2001; Sawmynaden et al, 2008).

As mentioned above TgMIC5, TgM2AP and TgMIC3 contain pro-peptides with a potential role in trafficking which are proteolytically removed en route to the micronemes. Processing of the pro-peptides in TgM2AP and TgMIC3 was found to be important for the functionality of the corresponding complexes. Only the processed TgMIC3 is able to function as an adhesin through its chitin-binding-like (CBL) domain (Cerede et al, 2002). Additionally,

it was recently shown that removal of the TgM2AP propeptide is critical for stable assembly and efficient secretion of the TgMIC2-M2AP complex from the micronemes onto the parasite surface and hence is necessary for efficient invasion (Harper et al, 2006).

The cysteine protease cathepsin L (TgCPL) which resides in compartments of the late secretory pathway is a candidate for taking part in proteolytic maturation of secretory proteins (Larson et al, 2009). LHVS (morpholinurea-leucyl-homophenyl-vinyl sulfone phenyl) is a specific inhibitor of this protease that impairs release of proteins from micronemes and consequently parasite attachment and gliding motility (Larson et al, 2009; Teo et al, 2007). This inhibitor should allow the identification of TgCPL substrates in the near future.

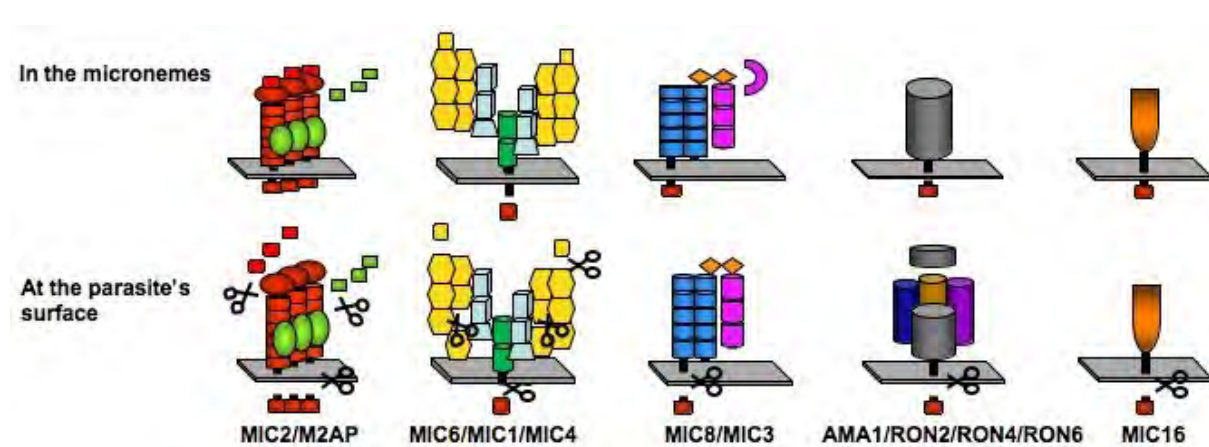


Figure 1.10: Proteolytic processing of microneme protein complexes in *T. gondii*.

Top: Proteolytic maturation taking place within the secretory pathway. Bottom: Proteolytic events taking place after secretion onto the parasite surface involving MPP1 (intramembrane cleavage), MPP2, MPP3 and potentially other protease activities. The cartoon includes a novel transmembrane MIC TgMIC16 currently under investigation in our laboratory (Sheiner *et al.*, in revision). Red: TgMIC2; Light green: TgM2AP; Yellow: TgMIC4; Light blue: TgMIC1; Dark green: TgMIC6; Blue: TgMIC8; Pink: TgMIC3; Grey: TgAMA1; Dark blue, Orange, Violet: TgRONs.

1.2.4.2 Shedding of MIC-complexes post exocytosis

Most single transmembrane proteins on the parasite surface have been shown to be excluded from the forming PV at the level of the moving junction. For this reason MIC complexes that are in excess or are not part of the moving junction are believed to be redistributed towards the posterior end of the parasite during invasion, a phenomenon known as capping, which has been experimentally confirmed at least for TgMIC2 and TgMIC3 (Carruthers et al, 1999a;

Garcia-Reguet et al, 2000). At some point during the penetration process, the tight interactions formed by the different complexes between the parasite and the host cell have to be disengaged and excess adhesive MICs need to be removed. This is effectively achieved by proteolytic shedding of the MIC complexes from the parasite's surface. Shedding was first observed for TgMIC2 and the responsible so far unidentified protease was termed microneme protein protease 1 (MPP1) (Carruthers et al, 2000b). Furthermore this cleavage event was shown to be critical for successful invasion (Brossier et al, 2003). As for MIC2, many other micronemal transmembrane proteins related to adhesive function including MIC6 are also shed from the surface. Cell-based cleavage assays, as well as studies in the parasite, have demonstrated that shedding takes place within the TMDs of the complex-anchoring TM-MICs TgMIC2, TgMIC6, TgMIC12 and TgAMA1 (Brossier et al, 2005; Howell et al, 2005; Opitz et al, 2002; Urban & Freeman, 2003; Zhou et al, 2004). Mass spectrometric analyses mapped the intramembrane cleavage to a conserved IAGG motif (in case of TgAMA1 an IAGL motif) in TgMIC6 (Opitz et al, 2002), TgMIC2 (Zhou et al, 2004) and TgAMA1 (Howell et al, 2005). MPP1 is anticipated to be a constitutively active rhomboid-like protease located at the parasite plasma membrane and to date, the best candidates are TgROM4 and TgROM5 (Brossier et al, 2005; Dowse et al, 2005).

Sequences similar to the MPP1 cleavage motif have been identified in TMDs of MICs in other Apicomplexa and several observations of surface shedding suggest that intramembrane cleavage during invasion is a conserved feature of the phylum (Opitz et al, 2002). Indeed, intramembrane cleavage was demonstrated for PfEBA175 at a site that is conserved across the DBL-EBP family and it was impossible to obtain a parasite line with a mutated cleavage motif (in both 3D7 and W2-mef strains) suggesting that shedding of this protein is a critical event for parasite survival (O'Donnell et al, 2006). Cell-based assays using two *P. falciparum* rhomboid proteases PfROM1 and PfROM4 and a range of parasite adhesins involved in invasion suggest intramembrane cleavage not only for the DBL-EBP family members but also for the reticulocyte binding-like (RBL) family of proteins and members of the thrombospondin-related anonymous protein (TRAP)-family (Baker et al, 2006). However, in *P. falciparum* blood stage parasites the subtilisin-like protease PfSUB2 is known to mediate shedding by juxtamembrane cleavage of the GPI-anchored merozoite surface protein 1 (PfMSP1), and of two transmembrane proteins, *Plasmodium* thrombospondin-related apical merozoite protein (PTRAMP) and PfAMA1 (Green et al, 2006; Harris et al, 2005). Proteolytic shedding of PfAMA1 can also take place within the TMD and the relative importance of these two cleavage events are under investigation

(Howell et al, 2005). The prime candidate for the rhomboid-like protease responsible for intramembrane shedding in *Plasmodium* is PfROM4, that is among all candidates phylogenetically most closely related to TgROM4 and TgROM5 (Dowse & Soldati, 2005).

1.2.4.3 Trimming on the parasite surface

In *T. gondii* several complexes of microneme proteins undergo proteolytic processing after their release onto the parasite surface. At least two distinct proteolytic activities have been noticed and were termed microneme protein protease 2 and 3 (MPP2 and MPP3) (Carruthers et al, 2000b; Zhou et al, 2004). MPP2 activity trims TgMIC2 in several steps from the N-terminus up to its A-domain (Carruthers et al, 2000b) and clips off a 15kDa fragment from the TgMIC4 carboxy-terminal end encompassing the sixth apple domain (Brecht et al, 2001). In addition TgMIC4 is processed at the amino-terminus by an unknown protease. Furthermore proteomic analysis identified TgM2AP and the GPI-anchored subtilisin-like protease TgSUB1 as substrates of MPP2 (Zhou et al, 2004). The significance of most of these processing events remains obscure as parasites cultivated *in vitro* in the presence of MPP2 inhibitors do not show any obvious defect. However, trimming of TgMIC2 enables its interaction with intercellular adhesion molecule 1 (ICAM-1) and was proposed to be important for transmigration through tissues (Barragan et al, 2005). The primary candidate for MPP2 activity is TgSUB1. A knockout strain has been generated and its characterization should clarify this issue (Binder et al, 2008). Interestingly, MPP2-dependent processing is upregulated in a *mic5ko* strain suggesting that TgMIC5 may either act as an inhibitor on MPP2 or somehow protects MPP2 substrates (Brydges et al, 2006).

MPP3 activity is involved in the trimming of TgM2AP (Zhou et al, 2004). Since no inhibitor profile has been established so far it remains unknown whether other proteolytic events correspond to this activity. More work is required to establish whether this activity is of relevance for parasite invasion.

Table 1.1: Properties of *T. gondii* protein complexes functioning in host cell invasion.

Complex and its function	Complex components	Proteolytic processing	Binding specificity	Connection to glideosome	Trafficking
MIC1-4-6 Host cell invasion	MIC1 (soluble)	no	Sialic acid		Relies on escorter MIC6
	MIC4 (soluble)	Trimming post secretion	Galactose (S.Matthews, pers. Communication)		Relies on escorter MIC6
	MIC6 (TM-MIC)	Maturation in secretory pathway, Shedding during invasion	No binding	yes	Contains trafficking determinant, relies on MIC1
MIC2-M2AP Motility and host cell attachment	MIC2 (TM-MIC)	Trimming post secretion, Shedding during invasion	Heparin, ICAM-1	yes	Contains trafficking determinant, relies on M2AP
	M2AP (soluble)	Propeptide cleavage, Trimming post secretion	unknown		Relies on escorter MIC2
MIC3-MIC8 Rhoptry secretion	MIC3 (soluble)	Propeptide cleavage	unknown		Relies on unknown escorter
	MIC8 (TM-MIC)	Shedding during invasion	unknown	unknown	Escorted?
AMA1- RONs Rhoptry secretion, Moving junction formation	AMA1 (TM-MIC)	Shedding during invasion	RON2?	unknown	Escorted (Sheiner <i>et al.</i> , in revision)
	RON2 (TM-protein?)		AMA1?		
	RON4 (soluble)		RON2/5/8?		
	RON5 (TM-protein?)		RON2/4/8?		
	RON8 (soluble)		RON2/4/5?		

1.2.5 The role of glycans in host cell invasion by the Apicomplexa

Glycans are crucial for the biology of many organisms. In vertebrates all cells are decorated with a dense and complex array of glycanstructures collectively termed the glycocalyx comprising glycoproteins and glycolipids. The glycanstructures consist of a chain of sugar molecules which may be branched and often terminate in a sugar unit belonging to the class of sialic acids (Sia). These glycans fulfill diverse functions in fertilization, development, neural plasticity and immune-related processes (Varki, 2008). The ubiquitous distribution of glycans and in particular sialic acids in the body of vertebrates makes them an attractive target

for pathogens. These pathogens express adhesins and toxins that bind to glycans with more or less specificity.

Proteins that bind to glycans are collectively named lectins. For a subset of lectins sialic acid is the critical component for recognition. Sialic acid binding lectins again can be very specific for one particular sialylated glycoconjugate. Recognition may depend on a specific conformation or modification of the sialic acid (such as methylation, acetylation, sulfation and phosphorylation), the linkage to the underlying sugar chain and the composition of this chain (Varki, 1997). In addition the protein or lipid to which the glycan is attached can be important for recognition.

Apicomplexan parasites express an arsenal of adhesins including lectins essential for productive host cell invasion. In the context of specific receptors these interactions are believed to determine the host cell preference for each parasite and each stage in their life cycle (Cerami et al, 1992a; Galinski et al, 1992; Martin et al, 2005; Orlandi et al, 1992). While some of the adhesins are conserved among several species others appear to be restricted to a few or a single species which probably reflects the differences in host and tissue tropism as part of their life cycles (Anantharaman et al, 2007; Templeton, 2007).

1.2.5.1 Glycan recognition during host cell invasion by *T. gondii*

Several studies have implicated recognition of cell surface glycans in host cell invasion by *T. gondii*. In earlier work, sulfated proteoglycans were recognized to play a role (Carruthers et al, 2000a; Ortega-Barria & Boothroyd, 1999). Moreover an important role for sialic acid was inferred from a study on invasion using a CHO cell mutant (lec2) with strongly reduced surface-expression of sialic acid (Monteiro et al, 1998). However in these studies the identity of the parasite ligands and the details of the corresponding biologically relevant receptors on the host cell surface remained undetermined. Glycans are now known to be the target of several *T. gondii* proteins comprising parasite resident surface proteins and MICs.

1.2.5.1.1 A lectin family with conserved function throughout Apicomplexa

TgMIC2 is a member of the thrombospondin-related anonymous protein (TRAP)-family of adhesive proteins conserved throughout the phylum of Apicomplexa. TRAP proteins fulfill essential functions in invasion and substrate-dependent gliding motility in these parasites (Morahan et al, 2009) and are characterized by having one or more copies of an integrin-like A-domain and one or more copies of a type I thrombospondin repeat-like (TSR-1) domain. TgMIC2 binds to the surface of host cells (Carruthers et al, 1999a). More specifically, the A-domain of TgMIC2 was reported to bind to heparin, a highly negatively charged glycosaminoglycan (GAG) composed of a repeated variably sulfated disaccharide unit that is ubiquitously present in the extracellular matrix (Harper et al, 2004). This interaction was shown to be dependent on the multimerization of the protein.

Binding to heparin as well as to sulfatides (sulfo-galactosyl-cerebrosides) was also demonstrated for PfTRAP a protein expressed specifically in the sporozoite stage (Muller et al, 1993; Robson et al, 1995). Numerous studies have attempted to address the relative contribution of the A-domain and the TSR-1 domain to the function of PfTRAP. Recombinant PfTRAP A-domain recapitulates the properties of the full-length protein in binding to heparin, sulfatides and hepatocytes (Akhouri et al, 2004; McCormick et al, 1999). In addition, chemical shift mapping experiments demonstrated weak binding of the TSR-1 domain to heparin indicating possible multiple interaction sites with GAGs (Tossavainen et al, 2006). This dual binding function of the protein has been analysed in *P. berghei*. In this study parasite strains carrying mutations in the PbTRAP A-domain and/or the TSR-1 domain were generated and demonstrated that both structural elements are important for invasion of the mosquito salivary glands, for invasion of hepatocytes and establishment of a liver infection to a different degree (Matuschewski et al, 2002). In contrast to *in vitro* experiments on heparin binding by TgMIC2 and PfTRAP, this study suggested a function for a well conserved metal-ion-dependent-adhesion-site (MIDAS) within the A-domain. MIDAS dependent binding of the PbTRAP A-domain to the glycoprotein fetuin-A was recently reported (Jethwaney et al, 2005). Therefore it appears that the A-domain might have more than one binding partner on the cell surface as suggested by another study (Akhouri et al, 2004).

1.2.5.1.2 More parasite-glycan interactions

Several microneme proteins are predicted to harbor TSR-1 domains (as the TRAP-family of proteins described above) and apple domains (Anantharaman et al, 2007). Some apicomplexan proteins containing these domains have been demonstrated to carry lectin properties (Keller et al, 2004; Klein et al, 1998; Muller et al, 2001; Tossavainen et al, 2006) and therefore, homologues and other proteins predicted to contain these domains are suspected to be involved in interactions with host cell surface carbohydrates. Apple domains belong to the plasminogen-apple-nematode (PAN) domain superfamily, which have been implicated in both protein-protein and protein-carbohydrate interactions (Collins et al, 2009; Muller et al, 2001). Two divergent copies of this fold are present in the N-terminal part of PfAMA-1 and are involved in complex formation with the RON proteins (Collins et al, 2009). NcMIC4, composed of six apple domains, binds to lactose and to the GAG chondroitin-sulfate A (Keller et al, 2004). Its homologue TgMIC4 was also suspected to bind to carbohydrates (Brecht et al, 2001) and recent results confirm it to be a galactose binding lectin on its own (Prof. Stephen Matthews, personal communication).

TgMIC3 and TgMIC8 form a complex during invasion and both contain an N-terminal chitin-binding-like domain as well as several EGF domains. In addition TgMIC8 comprises a TMD necessary for anchoring the complex into the parasite PM as well as a C-terminal cytoplasmic tail. As the name indicates most chitin-binding-like domains seem to be involved in protein-carbohydrate interactions (Shen & Jacobs-Lorena, 1999; Wright et al, 1991). Indeed, binding activity of TgMIC3 was detected to a range of nucleated cells (Garcia-Reguet et al, 2000), and this activity could be assigned to its chitin-binding domain (Cerede et al, 2002; Cerede et al, 2005). Studies on recombinant TgMIC3 and TgMIC8 showed that the binding activity of the TgMIC3 chitin-binding-like domain depends on dimerization of the protein achieved through its EGF domains (Cerede et al, 2002). Binding could also be detected for TgMIC8 upon artificial dimerization. Carbohydrate binding experiments did not reveal any interaction of TgMIC3 with the classical ligands N-acetyl-glucosamine, chitobiose or chitotriose, suggesting a different specificity (Cerede et al, 2005). This is in agreement with the fact, that the aromatic residues of this domain that are typically involved in the interaction with a carbohydrate molecule (Wright et al, 1991) are found at different positions in TgMIC3 (Cerede et al, 2005). Mutational studies identified two aromatic residues crucial for the binding function of TgMIC3 and the corresponding mutant parasite lines showed impaired virulence in mice underscoring the importance of this interaction (Cerede et al, 2005).

Binding activity of unknown specificity has been previously ascribed to TgMIC1 which forms a complex together with TgMIC4 and TgMIC6. The glycanstructures recognized by TgMIC1 and the importance of this interaction for host cell invasion by the parasite is subject of this thesis and will be discussed in detail.

1.2.5.2 Sialic acid-dependent and Sialic acid-independent invasion pathways in *Plasmodium*

Different strains of *P. falciparum* vary in their dependence on sialic acid bearing host cell receptors during invasion into RBCs (Baum et al, 2003; Dolan et al, 1994; Persson et al, 2008; Thompson et al, 2001; Triglia et al, 2005). For example, RBC invasion by parasite lines 3D7 and D10 is to a large extent independent of sialic acid, whereas parasite lines W2-mef and T994 preferentially invade RBCs using sialylated receptors. Based on this phenomenon, two invasion pathways were described each one defined by multiple ligand-receptor interactions involved. Two families of parasite ligands have been implicated in providing an explanation for these alternative invasion pathways at the molecular level. The Duffy-binding-like or erythrocyte-binding-protein (DBL-EBP) family and the reticulocyte-binding-like (RBL)-family of proteins are responsible for the alternative pathways of invasion (Table 1.2).

1.2.5.2.1 Duffy-binding-like (DBL)-family of proteins

The DBL family of proteins all contain one or more cysteine rich regions forming the DBL-domain originally identified in *P. vivax* and *P. knowlesi* (Fang et al, 1991). This domain appears to be specific to the genus *Plasmodium* (Aravind et al, 2003; Templeton, 2007). Despite poor sequence conservation across the family, the fold of the DBL domains appears to be similar in divergent proteins (Singh et al, 2006; Tolia et al, 2005). Proteins of the DBL family recognize receptors of different nature. Whereas some mediate interactions with sialylated oligosaccharides on glycoproteins, others recognize specific epitopes of proteins. The family can be divided into two subgroups, the DBL-EBP family functioning in RBC invasion by merozoites (Adams et al, 1992) and the PfEMP-1 family of variant surface antigens expressed at the surface of infected RBCs and involved in immune-evasion and cytoadherence of parasitized RBCs to the endothelium (Scherf et al, 2008).

There is only one protein of the DBL-EBP family in *P. vivax*, the Duffy-binding protein (DBP) which binds to the Duffy blood group antigen (Wertheimer & Barnwell, 1989). RBC invasion by *P. vivax* and *P. knowlesi* critically depends on the Duffy blood group antigen (Miller et al, 1976; Miller et al, 1975) which explains why Duffy negative or heterozygous human populations in West-Africa and Papua-New-Guinea are resistant to *P. vivax* infections (Kasehagen et al, 2007). The interaction between PvDBP containing a single copy of the DBL domain and the Duffy receptor does not appear to involve glycans (Choe et al, 2005; Hans et al, 2005; Singh et al, 2006).

In *P. falciparum* the DBL family comprises four proteins: Erythrocyte-binding antigen 175 (EBA 175), erythrocyte-binding antigen 140 (BAEBL/EBA 140), erythrocyte-binding antigen 181 (JESEBL/EBA 181) and erythrocyte-binding ligand-1 (EBL-1). These type I transmembrane proteins share a similar domain structure comprised of a signal peptide, a duplicated DBL domain designated F1 and F2 named together region II, another cysteine-rich domain called region VI, a membrane spanning domain and a cytoplasmic tail. All four proteins are stored in the micronemes (Mayer et al, 2009; Sim et al, 1992; Thompson et al, 2001; Treeck et al, 2006). Binding of PfEBA181 to RBCs is dependent on sialic acid (Maier et al, 2009), but the identity of the receptor remains unknown. The other three proteins have been implicated in interactions with the glycoporphins which are major sialoglycoproteins on the surface of the erythrocyte.

PfEBA175 preferentially binds to a cluster of O-linked sialylated oligosaccharide structures on glycoporphin A (Orlandi et al, 1992). The protein is expected to dimerize during receptor recognition forming six glycan binding sites involving both subunits (Tolia et al, 2005). The crystal structure confirmed that all of these sites are able to accommodate the full O-glycan of glycoporphin A. The interaction between PfEBA175 and glycoporphin A is considered to be of importance for Sia-dependent RBC invasion by *Plasmodium* merozoites. This is based on results from several studies on parasite invasion into glycoporphin A deficient (Ena⁻) and glycoporphin A and B deficient (M^kM^k) RBCs (Hadley et al, 1987; Miller et al, 1977; Pasvol & Jungery, 1983). Furthermore, antibodies directed against PfEBA175 were found to strongly inhibit RBC invasion (Sim et al, 1990). In addition, analysis of the invasion phenotype of a PfEBA175 knockout in the Sia-independent 3D7 strain revealed that the protein plays as well a role in the Sia-independent invasion pathway (Duraisingh et al, 2003a).

PfEBA140 recognizes more than one sialylated receptor on the RBC (Maier et al, 2009). Although the binding characteristics of the protein are promiscuous and both glycoporphin A and B were suggested to serve as receptors (Maier et al, 2003), the preferential

binding site appears to be an N-linked glycan on glycophorin C (Maier et al, 2009; Mayer et al, 2006). PfEBA140 is known to exist in at least five variants that are defined by polymorphism in four non-contiguous amino acids in region II encompassing the DBL domains. These variants were proposed to have different receptor specificity (Mayer et al, 2006; Mayer et al, 2002), however a recent study shows that their binding characteristics are rather similar (Maier et al, 2009). Although some evidence supports the notion that polymorphisms in PfEBA140 and PfEBA181 affect their binding affinity (Maier et al, 2009), this sequence variability is more likely to be related to immune selection. Interestingly, the parasite line D10 (cultured for decades in laboratories) lacks the PfEBA140 gene indicating that this protein is not essential for parasite invasion *in vitro* (Thompson et al, 2001), however it remains unknown if such a gene loss has occurred in field isolates.

Glycophorin B was proposed to be the receptor for PfEBL-1 and this interaction was shown to be dependent on sialic acid moieties on the RBC protein (Mayer et al, 2009). However, in about 50% of *Plasmodium* isolates the PfEBL-1 gene was found to contain a frameshift mutation indicating that it is evolving to be a pseudogene. Glycophorin B is highly polymorphic, an indication of strong selective pressure. The frequency of individuals with glycophorin B-null erythrocytes is particularly high in Africans and African-Americans. Interestingly, this frequency is highest (59%) among individuals of the tribe Efé which belong to the Mbuti ethnic group living in the Ituri forests of the Democratic Republic of the Congo. This has led to speculations that glycophorin B loss might have been a consequence of developing resistance against *P. falciparum* (Mayer et al, 2009). In turn, *P. falciparum* might have counteracted this deficiency with the expansion of the DBL-EBP family.

Table 1.2: *P. falciparum* ligands of Sia-dependent and Sia-independent RBC invasion.

Name of protein	localisation	RBC binding	Preferential receptor on the RBC	Invasion pathway
EBA175	Micronemes (Sim et al, 1992)	Sia-dep. (Orlandi et al, 1992)	Glycophorin A (Orlandi et al, 1992)	Sia-dep. and Sia-indep. (Duraisingh et al, 2003a)
EBA140 (BAEBL)	Micronemes (Thompson et al, 2001; Treeck et al, 2006)	Sia-dep. (Maier et al, 2009)	Promiscuous, Glycophorin C (Maier et al, 2009; Mayer et al, 2006)	Sia-dep. (Maier et al, 2003)
EBA181 (JESEBL)	Micronemes (Treeck et al, 2006)	Sia-dep. (Maier et al, 2009)	Unknown	Sia-dep.
EBL-1	Apical (Mayer et al, 2009)	Sia-dep. (Mayer et al, 2009)	Glycophorin B (Mayer et al, 2009)	Sia-dep.
PfRh1	Apical (Rayner et al, 2001)	Sia-dep. (Triglia et al, 2005)	unknown	Sia-dep. (Triglia et al, 2005)
PfRh2a	Rhoptries (Duraisingh et al, 2003b)	Not detected	unknown	Sia-indep. (Desimone et al, 2009)
PfRh2b	Rhoptries (Duraisingh et al, 2003b)	Not detected	unknown	Sia-indep. (Duraisingh et al, 2003b)
PfRh4	Apical (Stubbs et al, 2005)	Sia-indep. (Gaur et al, 2007)	unknown	Sia-indep. (Stubbs et al, 2005)
PfRh5	Rhoptries, Moving junction (Baum et al, 2009)	Sia-indep. (Baum et al, 2009)	unknown	Sia-indep.

1.2.5.2.2 Reticulocyte binding like (RBL)-family of proteins

A second family of erythrocyte binding proteins exists in *Plasmodium* which was first described in *P. vivax* (Galinski et al, 1992). Because of the selective binding of the *P. vivax* proteins to reticulocytes, they were named reticulocyte binding protein (PvRBP) -1 and -2 (Galinski et al, 1992). The corresponding receptors on the RBC are unknown. Homologues were subsequently identified in other *Plasmodium* species and were designated as members of the reticulocyte binding-like or RBL-family (Rayner et al, 2001; Triglia et al, 2001). A group of at least 14 homologues in *P. yoelii* is named the Py235 family. Most interestingly, rodent malaria *P. yoelii yoelii* merozoites originating from a single schizont were found to express distinct members of the Py235 protein family (Preiser et al, 1999). Although the function of the individual proteins has not been described this might indicate that the parasite generates diversity in invasion specificity to overcome immune pressure and receptor heterogeneity.

Similar to the expansion of DBL proteins in *P. falciparum*, the two PvRBPs have five homologues in *P. falciparum* called normocyte binding proteins (NBPs) or RBL-homologues PfRh1, PfRh2a, PfRh2b, PfRh4 and PfRh5. Another gene coding for PfRh3 exists but appears to be a pseudogene. All *P. falciparum* RBLs are large (220-350kD) type I transmembrane proteins of undefined domain structure except PfRh5 which is much smaller and lacks a transmembrane domain, suggesting that it might be part of a protein complex. In contrast to members of the DBL-EBP family, at least some *P. falciparum* RBL proteins have been detected in the rhoptries. Immunoelectron-microscopy demonstrated localization of PfRh2a, PfRh2b and PfRh5 to this type of organelle (Baum et al, 2009; Duraisingh et al, 2003b). PfRh4 colocalizes with both proteins in schizonts but is located more apically in free merozoites as shown by immunofluorescent assays (Stubbs et al, 2005). Apical localization was also detected for PfRh1 in free merozoites (Rayner et al, 2001). Interestingly, PfRh5 was found to follow the tight junction during erythrocyte invasion (Baum et al, 2009). Whereas PfRh1 has been identified as an important player of sialic acid-dependent invasion, PfRh2a, PfRh2b, PfRh4 and PfRh5 have been implicated in the sialic acid-independent pathway as will be explained in the following section. Whereas PfRh2a and PfRh2b were not formally shown to bind to RBCs, PfRh4 and PfRh5 do qualify as adhesins (Baum et al, 2009; Gaur et al, 2007).

1.2.5.2.3 Switching between invasion pathways

The importance of individual interactions between parasite proteins and corresponding receptors for RBC invasion has been the focus of numerous studies, since it has important implications for vaccine development. However, because of the redundancy of invasion pathways, it has been difficult to assess the individual contributions of parasite ligands to this process. First indications that *Plasmodium* is able to change the use of receptors on the RBC surface came from a study where the Dd2 clone (derived from the W2-mef strain which completely relies on sialic acid for invasion) could be selected for invasion into neuraminidase-treated RBCs (Dolan et al, 1990; Dolan et al, 1994). Further evidence for a switch-mechanism between Sia-dependent and Sia-independent invasion pathways in *Plasmodium* came from the functional disruption of PfEBA175 in the W2-mef strain (Reed et al, 2000). These mutant strains were able to invade neuraminidase treated RBCs whereas the parental strain was not. Microarray analysis of the PfEBA175 knockout strain and of parental

W2-mef parasites selected or not on neuraminidase-treated RBCs revealed that the switch of invasion pathways was associated with upregulation of PfRh4 (Stubbs et al, 2005). In this study PfRh4 was shown to be exclusively expressed in several Sia-independent parasite lines including parasites that had undergone the switch to this pathway. Disruption of PfRh4 was impossible in these lines, but was achieved in W2-mef. This knockout mutant was unable to switch to the Sia-independent invasion pathway demonstrating that the switching mechanism relies on PfRh4 (Stubbs et al, 2005). Further experiments showed that switching between these invasion pathways is reversible and is accompanied by silencing of PfRh4. In contrast both PfRh2a and PfRh2b were shown to be dispensable for switching of invasion pathways in the W2-mef strain (Desimone et al, 2009).

PfRh2a and PfRh2b were found to show variant expression among several *P. falciparum* isolates examined (Duraisingh et al, 2003b). In agreement with this observation, both genes could be individually disrupted (in W2mef and 3D7) and a parasite line (D10) lacking Rh2b was identified showing that they are not essential for *P. falciparum* blood stage parasites *in vitro* (Desimone et al, 2009; Duraisingh et al, 2003b; Triglia et al, 2001). However two studies showed that antibodies directed against the two proteins inhibit RBC invasion, suggesting a role for PfRh2a and PfRh2b in this process (Duraisingh et al, 2003b; Triglia et al, 2001). Detailed characterization of a PfRh2b knockout invasion phenotype led to the conclusion that the protein functions in a Sia-independent pathway and its disruption causes increased use of other receptor-ligand interactions for compensation (Duraisingh et al, 2003b). A function in Sia-independent invasion was also suggested for PfRh2a (Desimone et al, 2009). Co-immunoprecipitation experiments suggested that PfRh2 might interact with PfRh1, a sialic acid-dependent erythrocyte binding protein (Rayner et al, 2001), however experiments have shown that PfRh2b functions independently of PfRh1 in the parasite (Duraisingh et al, 2003b).

PfRh1 has been identified as an important player in sialic-acid dependent invasion (Triglia et al, 2005). Similar to PfRh2a and PfRh2b, expression levels of this protein vary among different strains and this is related to gene duplication events that occurred in several parasite lines. In a given strain the relative amount of PfRh1 expression versus Rh2a and Rh2b expression levels correlates with the use of Sia-dependent or -independent pathways (Triglia et al, 2005). Strains with a high expression of PfRh1 (such as T994) were found to be reliant on the Sia-dependent pathway whereas strains lacking or with reduced levels of PfRh1 (like 3D7) use the Sia-independent pathway for invasion. Therefore disruption of the gene had no effect on invasion of 3D7 parasites but in the T994 strain resulted in a shift to Sia-

independent interactions (Triglia et al, 2005). However, microarray as well as Western blots analysis of the T994 parental and knockout strains failed to detect any upregulation of members of the RBL- and EBL-DBL families at the transcript or protein level.

What is the relevance of this remarkable capacity of the parasite to switch between alternative invasion pathways for a natural infection with the malaria parasite? The occurrence of these pathways in field isolates has been investigated in two studies in India and The Gambia as well as in another study with isolates from different origins (Baum et al, 2003; Okoyeh et al, 1999; Perkins & Holt, 1988). In summary, considerable variation was found to exist among the field isolates in their use of Sia-dependent and -independent invasion pathways. Analysis of invasion inhibitory antibodies acquired by malaria patients in Kenya showed that these were directed against ligands of both invasion pathways (Persson et al, 2008). Proteins of the DBL-EBP and RBL families were found to be among the major targets. Therefore it was suggested that *Plasmodium* uses the variation in invasion phenotype for immune evasion during infection in humans. So far no alternative invasion pathways with a switching mechanism have been described in other Apicomplexa.

1.3 Hypothesis and Aims of Project

T. gondii parasites are able to invade essentially any nucleated cell, but the molecular basis that makes recognition and invasion of such a broad spectrum of cells possible is only beginning to be understood. Two general concepts of receptor-ligand interactions could explain this phenomenon. Either the parasite ligand(s) displays low specificity and therefore is able to bind to a range of host cell receptors, or alternatively the parasite ligand(s) should recognize a receptor that is common to all different cell types. The *T. gondii* MIC1-4-6 complex has been partly characterized before and indicated a non-essential but important role for invasion *in vitro* and virulence *in vivo* (Cerede et al, 2005; Reiss et al, 2001). Adhesive properties were ascribed to TgMIC1 (Fourmaux et al, 1996) and TgMIC4 (Brecht et al, 2001), and TgMIC1 was proposed to bind to lactose (Lourenco et al, 2001). We hypothesized that the TgMIC1-4-6 complex functions early in host cell invasion and is able to mediate recognition of a broad spectrum of host cells.

Carbohydrate microarray analyses and structural studies performed on TgMIC1 by our collaborators Prof. Stephen Matthews and Prof. Ten Feizi (both from Imperial College, London) indicated that TgMIC1 binds to sialic acid. In the light of these results the aims of the PhD project were to:

- Functionally characterize the TgMIC1-4-6 complex and in particular TgMIC1.
- Assess the importance of host cell sialylated glycoconjugates for host cell invasion.
- Identify which is/are the important parasite ligand(s) in sialic acid-dependent invasion.
- Characterize any additional parasite ligands involved.
- Identify the relevant host cell receptors.

2. Materials and Methods

This section is intended to give details on the routinely used techniques during this study. For more specific protocols please refer to the Materials and Methods sections of the scientific articles that are part of this thesis.

2.1 Reagents and Suppliers

All chemicals were obtained from Sigma except where stated. Microbiological culture reagents were obtained from BD Biosciences.

2.1.1 Enzymes

All restriction enzymes were from New England Biolabs (NEB).

Calf Intestinal alkaline Phosphatase (CIP; Q-Biogene)

T4 DNA ligase (Roche)

AmpliTaq DNA polymerase (Applied Biosystems)

TaKaRa LA Taq DNA polymerase (TaKaRa)

T4 DNA polymerase (NEB)

Superscript II Reverse Transcriptase (Invitrogen)

2.1.2 Kits

Nucleobond maxi-prep kit for large scale DNA preparation (BD Biosciences);

WizardMini-prep kit for small scale DNA preparation (Promega);

Easy Pure DNA purification kit (Biozym);

pGEM T-easy kit (Promega);

TOPO TA-cloning kit (Invitrogen).

All kits were used according to the manufacturer's instructions.

2.1.3 Antibodies

Table 2.1: Antibodies used in this study

Antibody	Source	Dilution for western blot or immunofluorescence
mouse anti-myc hybridoma GE10	(Invitrogen™)	1:1000
mouse anti-Ty1 hybridoma BB2	(Bastin et al, 1996)	1:1000
mouse anti-HA	Covance™	1:1000
rabbit anti-TgMIC4	(Brecht et al, 2001)	1:1000
rabbit anti-TgMLC	(Herm-Gotz et al, 2002)	1:1000
mouse anti-TgSAG1	kindly provided by JF Dubremetz	1:1000
rabbit anti-TgCatalase	(Ding et al, 2000)	1:1000
goat anti-mouse, horse radish peroxidase (HRP) conjugated	Molecular Probes™	1:3000
goat anti-rabbit, HRP conjugated	Molecular Probes™	1:3000
goat anti-mouse, Alexa 488 conjugated (green)	Molecular Probes™	1:3000
goat anti-mouse, Alexa 594 conjugated (red)	Molecular Probes™	1:3000
goat anti-rabbit, Alexa 488 conjugated (green)	Molecular Probes™	1:3000
goat anti-rabbit, Alexa 594 conjugated (red)	Molecular Probes™	1:3000

2.2 Solutions

2.2.1 Culture Media

Table 2.2: Culture media used in this study

LB (Luria Bertani) liquid media	1% w/v Bacto-Tryptone, 0.5% w/v yeast extract, 1% w/v NaCl
LB Agar	LB liquid with 1.5 % w/v agar
YPD (Yeast extract-Peptone-Dextrose) liquid media	1% w/v yeast extract, 2% w/v peptone, 2% w/v dextrose (glucose)
YPD agar	YPD liquid with 1.5 % w/v agar
Dulbecco's Modified Eagle Medium + 4500mg/L Glucose + L-Glutamine + Pyruvate was supplied by Gibco™ and supplemented with 5% (v/v) Foetal Calf Serum (FCS), 2 mM L-glutamine, and 25 µg/ml Gentamicin	
Alpha Minimal Essential Medium + Ribonucleosides + Deoxyribonucleosides was supplied by Gibco™ and supplemented with 10% (v/v) Foetal Calf Serum (FCS), 2 mM L-glutamine, and 25 µg/ml Gentamicin	
HAM-F12 Medium + pyridoxine + NaHCO ₃ + L-Glutamine was supplied by Gibco™ and supplemented with 5% (v/v) Foetal Calf Serum (FCS), 2mM L-glutamine, and 25 µg/ml Gentamicin	
RPMI 1640 Medium + L-Glutamine + 25mM HEPES was supplied by Gibco™ and supplemented with 5% (v/v) Foetal Calf Serum (FCS), 2 mM L-glutamine, and 25 µg/ml Gentamicin	

2.2.2 General Solutions

Table 2.3: Solutions used in this study

Mini-prep Solution I	50mM glucose, 25mM Tris-Cl, 10mM EDTA, pH 8.8
Mini-prep Solution II	0.2M NaOH, 1% SDS
Mini-prep Solution III	3M potassium acetate, 5M glacial acetic acid
PBS	137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4
SDS-PAGE gel loading buffer (1x)	50mM Tris-Cl pH 6.8, 100mM Dithiothreitol, 2% SDS, 0.1% Bromophenol blue, 10% glycerol
Agarose gel loading buffer (6x)	0.02% Bromophenol blue, 0.02% Xylencyanol FF, 30%, Glycerol in H ₂ O
Cytomix (for <i>T. gondii</i> transfection)	120mM KCl, 0.15mM CaCl ₂ , 10mM K ₂ HPO ₄ /KH ₂ PO ₄ , 25mM Hepes, 5mM MgCl ₂ , pH 7.6, 5mM ATP, 5mM Glutathione
Tris/Acetate/EDTA (TAE)	0.04M Tris-Acetate, 0.001M EDTA, pH 8.0
SDS-PAGE running buffer	25mM Tris, 250mM Glycine, pH 8.3, 0.1% SDS
Semi-dry transfer buffer	2.5mM Tris, 19.2mM Glycine, 20 % Methanol
Tris EDTA (TE) 10x	100 mM Tris-HCl, pH 7.5; 10 mM EDTA, pH 8.

2.3 Cell lines and microbiological strains

2.3.1 Bacteria

Table 2.4: Genotypes of *Escherichia coli* used

Strain name	genotype
XL1-Blue	F'::Tn10 <i>proA</i> ⁺ <i>B</i> ⁺ <i>lacI</i> ^q Δ(<i>lacZ</i>) <i>M15/recA1 endA1 gyrA96</i> (Nal ^r) <i>thi hsdR17</i> (<i>r</i> _K ⁻ <i>m</i> _K ⁺) <i>glnV44 relA1 lac</i>
BL21	F- <i>ompT gal [dcm] [lon] hsdS_B</i> (<i>r</i> _B ⁻ <i>m</i> _B ⁻) λ(DE3)

2.3.2 *Pichia pastoris*

Table 2.5: Strains of *P. pastoris* used

Strain name	genotype	phenotype
GS115	<i>his4</i>	MUT ⁺

2.3.3 Mammalian cells

Human Foreskin Fibroblasts (HFF), the African Green Monkey Kidney Cell line (Vero), and the Chinese Hamster Ovary (CHO) cell line “lec2” were obtained from the American Type Culture Collection (ATCC). CHO-K1, CHO-A745, CHO-B618 and CHO-D677 were a kind gift from Dr. Fiona Tomley (Institute for Animal Health, Berkshire, UK).

2.3.4 *Toxoplasma gondii* strains

Table 2.6: Strains of *Toxoplasma gondii* used

Strain Name	Description
RHΔ <i>hxgp</i> rt	Virulent strain with disruption of <i>HXGPRT</i> gene (Donald et al, 1996)

2.4 Culture conditions

2.4.1 Bacterial culture

Bacteria cultures were grown in LB Broth at 37 °C or on plates of LB agar. Ampicillin was used at a concentration of 100 µg/ml. Kanamycin was used at a concentration of 30µg/ml. Glycerol stocks of bacteria strains were stored at -80 °C in LB Broth with 15% (v/v) glycerol.

2.4.2 *Pichia pastoris* culture

Culture of *P. pastoris* was performed as indicated in the manufacturers instructions (Invitrogen).

2.4.3 Mammalian cell culture

HFF and Vero cells were grown as monolayers in supplemented DMEM. CHO cell lines were grown in conditions as indicated by the ATCC: CHO-lec2 cells were cultured in supplemented Alpha minimum essential medium, other CHO cells were grown in supplemented HAM-F12K medium. C6 rat glioma cells were propagated in supplemented RPMI. Cells were split using Trypsin (Gibco).

2.4.4 Parasite propagation

T. gondii tachyzoites were grown in Vero or HFF cell cultures. Selection and cloning of stably transfected parasites took place in HFF cell cultures, with medium supplemented with appropriate drug (25 µg/ml mycophenolic acid (MPA), 50 µg/ml Xanthine, 1 µM Pyrimethamine, 20 µM chloramphenicol).

2.5 Transformations and transfections

2.5.1 *E. coli* transformation

Competent *E. coli* were prepared using the protocol of Inoue *et al* (Inoue et al, 1990). Transformations were performed by mixing competent cells with DNA, followed by incubation on ice for 10 minutes, 1 minute of heat-shock at 42°C, and a further 5 minutes on ice, followed by plating on LB-Agar supplemented with an appropriate antibiotic.

2.5.2 *Pichia pastoris* transformation

Transformation of *P. pastoris* was performed by electroporation as indicated in the manufacturers instructions (Invitrogen).

2.5.3 *T. gondii* transfection

For *T. gondii* transfection, 5×10^7 extracellular parasites freshly egressed from host cells were used for each transfection. Parasites were centrifuged at 300 g for 10 mins, then resuspended in 700 μ l cytomix per transfection and mixed with 50 μ g of linearised plasmid DNA for stable transfections or 100 μ g of supercoiled plasmid DNA for transient transfections. For random integration, restriction enzyme-mediated integration (REMI) was employed by adding 50-100 units of restriction enzyme to the transfection mix. Parasite/DNA mix was then placed in a 4 mm electroporation cuvette and electroporated with the following settings: Voltage: 2000 V, Resistance, 50 Ω , and Capacitance 25 μ F (Soldati & Boothroyd, 1993). After electroporation, the transfection mix was added to HFF cells.

2.6 Plasmid construction

This is described in the Materials and Methods of the respective publications together with the primers used.

2.6.1 Agarose gel electrophoresis

0.5 – 2 % agarose gels were made by dissolving agarose (Sigma) in heated TAE, and pouring the solution into a gel casting tray (peqLab). Ethidium Bromide (0.5 μ g/ml) or SYBR safe (at 0.4x concentration, Molecular Probes™) was added to the gel to stain nucleic acids. Electrophoresis was performed in TAE buffer.

2.6.2 Preparation of nucleic acids

Total RNA was isolated from $1-5 \times 10^7$ parasites using Trizol reagent (Invitrogen) following the manufacturers instructions. Preparation of cDNA, either using gene specific or general Poly-A primers was performed using Superscript II Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions.

Mini-preps of plasmid DNA from *E. coli* were performed using the alkaline lysis method (Sambrook *et al.*, 1989). Briefly, cell pellets from a 5 ml culture were resuspended in

100 µl mini-prep Solution I. 200 µl of Solution II was added, and the sample was incubated for 5 minutes at room temperature. 150 µl of Solution III was added, and samples were incubated on ice before centrifugation for 10 minutes at 13000g at 4 °C. Supernatants were then mixed with 2.5 volumes of cold ethanol to precipitate nucleic acids. Following centrifugation for 15 minutes at 13000g at 4 °C and washing with 70 % ethanol, nucleic acid pellets were resuspended in water supplemented with RNase A (0.1 mg/ml). Large preparations of plasmid DNA from *E. coli* were performed using the Nucleobond Maxiprep Kit according to the manufacturer's instructions.

2.6.3 Polymerase chain reaction (PCR)

Reverse transcription reactions were performed using Superscript II Reverse Transcriptase (according to the manufacturer's instructions) when required to prepare the first strand for cDNA synthesis. Standard analytical PCR reactions were performed with AmpliTaq, using the supplied buffer, with 0.05 units of enzyme per µl of reaction, primers at 0.4 µM, and dNTPs at 0.1 mM each. Template DNA was added to each reaction in the form of plasmid DNA (0.1-1 ng per reaction), genomic DNA (1-5 µl of genomic preparation) or a bacterial colony. The standard reaction temperatures used were an initial 94 °C for 5 mins, followed by 25 cycles of 94 °C for 30s (denaturing step); 55 °C for 30s (annealing step), 72 °C for 1 min/kb of product (elongation step), followed by a final 7 mins at 72 °C. Preparative PCR reactions for cloning were prepared under similar conditions, with the following exceptions: TaKaRa LA Taq was used, an elongation temperature of 68 °C was used, and when using genomic DNA or cDNA as a template, 35 cycles were performed.

2.6.4 Ligations

TA cloning of PCR products was performed using either pGEM T-easy (Promega) or TOPO (Invitrogen) cloning kits according to the manufacturer's instructions. The correct DNA sequences of clones were confirmed by sequencing. For other plasmids, inserts were prepared either by restriction digest of purified PCR products, or by restriction digest of the appropriate plasmid followed by gel purification of the appropriate band. Vectors were prepared by restriction digest, followed by treatment with 200 units of calf intestinal alkaline phosphatase (C.I.P.; QBiogene) at 37 °C for 1 hour, followed by gel purification of the appropriate band.

Ligations were performed with an insert-vector ratio of ~4:1 in a volume of 10 µl using 0.5 units of T4 DNA ligase (Roche) in the supplied buffer, and incubated at room temperature for at least 1 hour or overnight at 4 °C. Ligation reactions were then used to transform *E. coli*.

2.6.5 Site-directed mutagenesis

Site-directed mutagenesis (SDM) was performed either using the Quikchange II Kit (Stratagene) or using components of the kit bought separately. Complementary sense and antisense primers were designed containing the desired mutation, according to the manufacturer's instructions. A 50 µl polymerase reaction using 2.5 units of *PfuI* Turbo or *PfuI* Ultra (Stratagene) DNA polymerase was performed in the supplied buffer, with 20 ng plasmid DNA, primers at 0.3 µM each, and dNTPs at 0.1 mM each. The reaction temperatures were as follows: an initial 95 °C for 30 s, followed by 12 cycles of 95 °C for 30 s (denaturing step), 55 °C for 1 minute (annealing step); 68 °C for 1 minute per kb of plasmid (elongation step). Following the polymerase reaction, 20 units of *DpnI* enzyme was added directly to the reaction, followed by incubation at 37 °C for one hour. 1 – 10 µl of the reaction was then used to transform *E. coli*. The presence of the desired mutation was confirmed by DNA sequencing.

2.7 Sodium Dodecyl Sulphate (SDS) Polyacrylamide Gel (PAGE) Electrophoresis and Western Blotting

SDS-PAGE electrophoresis was performed according to standard methods (Sambrook *et al.*, 1989). Resolving gel (8-15 % acrylamide; 0.1 % SDS; 375 mM Tris, pH 8.8) and stacking gel (5 % acrylamide, 0.1 % SDS, 200mM Tris, pH 6.8) solutions were polymerized using 0.04-0.1 % N,N,N',N'-Tetramethyl-1-,2-diaminomethane (TEMED) and 0.1 % ammonium persulphate (APS). Electrophoresis was performed in SDS-PAGE running buffer. Semi-dry transfer of proteins to nitrocellulose was performed as described (Sambrook *et al.*, 1989), using semi-dry transfer buffer. Blots were blocked in 5% skimmed milk in 0.05% Tween-20 PBS (PBST) for one hour at room temperature, or overnight at 4 °C. Blots were incubated with primary antibody, diluted in blocking solution. Blots were then washed with PBST, and incubated with HRP conjugated secondary antibody, either goat anti-mouse or goat anti-

rabbit, diluted in blocking solution 1:3000. Blots were then developed using ECL reagent (Amersham), and exposed to chemiluminescence film.

When required, blots were stripped with stripping buffer (2% SDS, 7µl/ml 2-β-mercapto-ethanol in PBS) for 45 minutes, washed thoroughly with PBS-T and a further antibody labelling could be performed again. Samples of *T. gondii* which were to be used for western blotting were prepared as follows: freshly egressed *T. gondii* cells were resuspended in culture medium, centrifuged at 300 g for 10 mins, washed once in PBS, and centrifuged again. The resulting parasite pellet was then resuspended in 200 µl of PBS and snap frozen in liquid nitrogen. The suspension was then sonicated three times for ten seconds, with one minute incubation on ice between each sonication, to lyse the parasites. 50 µl of 5 x SDS-PAGE loading buffer was then added to the lysed parasites, and the samples were loaded onto the gel without boiling.

3. Results

3.1. Atomic resolution insight into host cell recognition by *T. gondii*

Earlier work on TgMIC1 and the TgMIC1-4-6 complex had shown that TgMIC1 possesses host cell adhesive properties and the protein was proposed to bind to lactose (Fourmaux et al, 1996; Lourenco et al, 2001). In the amino-terminal part of the protein a duplicated domain has been identified (Fourmaux et al, 1996). These domains were assumed to be TSR-1 domains of divergent nature due to a CXXXCG motif typically present in this type of domain. However, structural studies on TgMIC1 in the laboratory of our collaborator Prof. Stephen Matthews (Imperial College, London) showed that the N-terminal part adopts a completely unrelated novel fold which was termed Microneme Adhesive Repeat (MAR). Carbohydrate microarray analyses performed in the laboratory of Prof. Ten Feizi (Imperial College, London) revealed that the TgMIC1 MAR domains bind to sialic acid. This carbohydrate has been previously shown to be of importance for *T. gondii* invasion in experiments using a CHO cell mutant (lec2) exhibiting reduced surface expression of sialic acid (Monteiro et al, 1998).

In the light of these results the importance of host cell sialylated glycoconjugates for host cell invasion by *T. gondii* has been reassessed in our laboratory. Cell binding assays were used to determine if the interaction of TgMIC1 with sialylated host cell receptors is physiologically relevant and how this compares to earlier reports of lactose binding. Furthermore mutational studies confirmed the critical role of two threonine residues in TgMIC1 for binding to sialic acid. These results together with results presented in section 3.4. constitute the major personal achievements of this Ph.D..

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

Tharin MA Blumenschein^{1,6},
Nikolas Friedrich^{2,6}, Robert A Childs³,
Savvas Saouros¹, Elisabeth P Carpenter¹,
Maria A Campanero-Rhodes³,
Peter Simpson¹, Wengang Chai³,
Theodoros Koutroukides⁴, Michael J
Blackman⁵, Ten Feizi³, Dominique
Soldati-Favre² and Stephen Matthews^{1,*}

¹Division of Molecular Biosciences, Imperial College London, London, UK, ²Department of Microbiology and Genetics, Faculty of Medicine, University of Geneva CMU, Geneva, Switzerland, ³Glycosciences Laboratory, Division of Medicine, Imperial College London, Middlesex, UK, ⁴Medical Science Division, Fox Chase Cancer Centre, Philadelphia, PA, USA and ⁵Division of Parasitology, National Institute for Medical Research, London, UK

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 before invasion of host cells and play important roles in host cell recognition, attachment and penetration. Here, we report the atomic structure for a key MIC, TgMIC1, and reveal a novel cell-binding motif called the microneme adhesive repeat (MAR). Using glycoarray analyses, we identified 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.

The EMBO Journal (2007) 26, 2808–2820. doi:10.1038/sj.emboj.7601704; Published online 10 May 2007

Subject Categories: microbiology & pathogens; structural biology

Keywords: carbohydrate microarray; crystal structure; micronemal proteins; sialic acid; *T. gondii*

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 (MICs) 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 entire complex through the early secretory pathways (Reiss *et al*, 2001; Huynh *et al*, 2004; 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 subcomplex is a potent antigen and acts as an effective vaccine in the mouse model (Lourenco *et al*, 2006). Unlike the battery of other MICs, 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

*Corresponding author. Division of Molecular Biosciences, Imperial College of Science, Technology and Medicine, South Kensington Campus, London SW7 2AZ, UK. Tel.: +44 207 594 5315;

Fax: +44 207 594 5207; E-mail: s.j.matthews@imperial.ac.uk
⁶These authors contributed equally to this work

Received: 30 November 2006; accepted: 4 April 2007; published online: 10 May 2007

TgMIC1 that promotes proper folding of TgMIC6 and contributes to complex formation (Saouros *et al*, 2005).

Although studies have clearly highlighted the importance of MICs 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 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 micronemal adhesive repeat (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 17–262, hereafter termed TgMIC1-NT) in *Escherichia coli* fused with thioredoxin to aid disulfide 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 the MAD (multiple-wavelength anomalous dispersion) method. The atomic structure for residues 29–259 was solved at a resolution of 1.9 Å (Figure 2A; Table I and

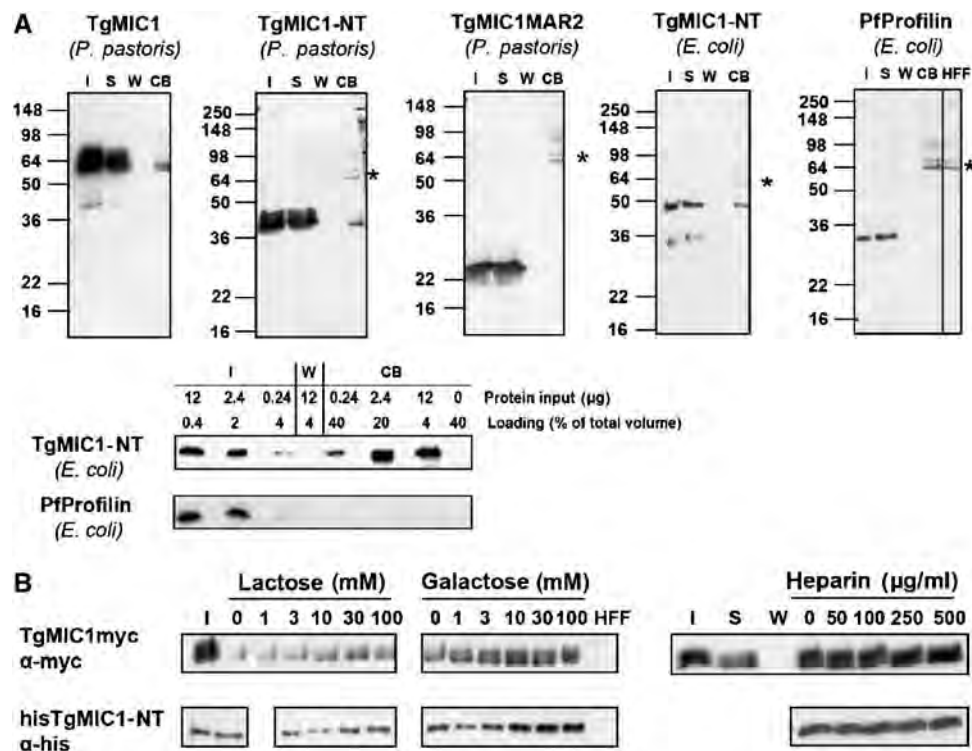


Figure 1 Host cell binding by TgMIC1-NT. (A) Cell binding assays on HFFs were performed using supernatant of *P. pastoris* cultures expressing TgMIC1, TgMIC1-NT, or TgMIC1MAR2, bacterially produced TgMIC1-NT and PfProfilin, the latter being used as a negative control. Anti-His antibodies are used as 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, whereas 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 500 μl, whereas the total volume of the cell bound fraction is 50 μl. (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 probes for Western blots. Samples are named as follows: 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 no 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 500 μl, whereas the total volume of the cell-bound fraction is 50 μl.

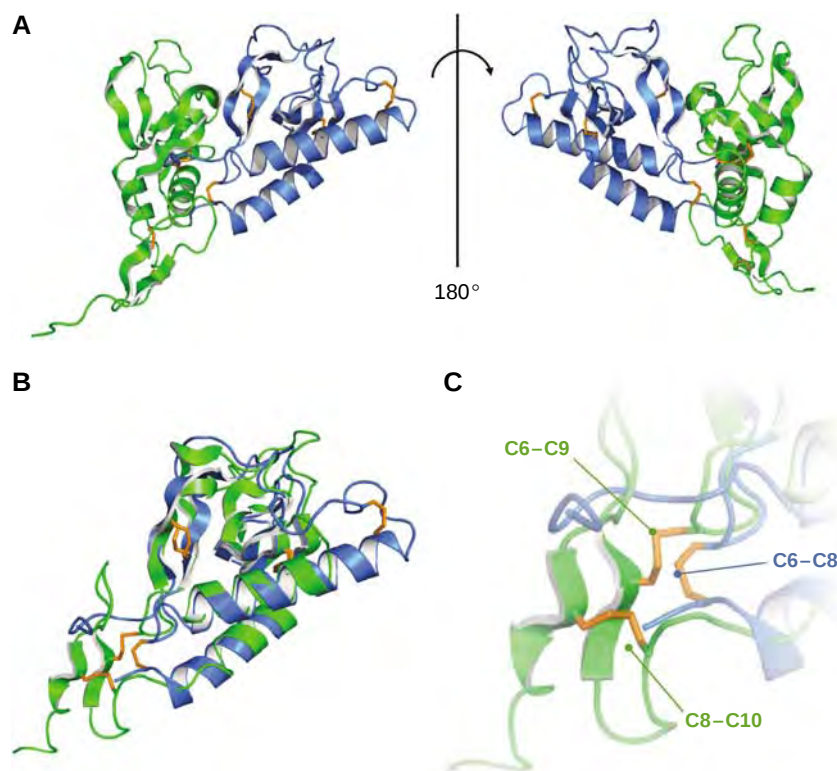


Figure 2 Host cell binding and structure of TgMIC1-NT. (A) Ribbon representation of a representative structure for TgMIC1-NT. Repeat 1 (MAR1) is shown in green and repeat 2 (MAR2) in blue. The orientation shown on the right represents a 180° rotation. (B) Ribbon representation showing the superimposition of Repeat 1 (MAR1, green) and Repeat 2 (MAR2, blue). The orientation shown is the same as the representation on the left in (A). The calculated r.m.s.d. is 2.2 Å over 89 amino-acid residues. (C) Zoomed region illustrating the altered disulfide 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 disulfide 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).

Supplementary Figure 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 antiparallel helical bundle comprising one helix from each terminus. Two disulfide 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 connecting helix α1 to the β-barrel and the other 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 r.m.s.d. of 2.2 Å over 89 residues (Figure 2B), some notable differences are apparent. The first helix and the subsequent loop to strand β1 are significantly longer in the first repeat and are stabilized by an additional disulfide 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 disulfide 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 ²⁴⁸NPPL²⁵¹ 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). TSR1s 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 the fact 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 micronemal adhesive repeat.

A search against other apicomplexan genomes identified 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

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

	Native	SeMet Peak	SeMet Inflection	SeMet Remote	2,6-sialyl- <i>N</i> - acetylactosa- mine	2,3-sialyl- <i>N</i> - acetylactosa- mine
Data collection and phasing						
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Cell dimensions <i>a</i> , <i>c</i> (Å)	66.2, 172.3	66.3, 172.6	66.3, 172.6	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) ^a	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	154 515	91 673	90 740	93 048	542 909	351 534
Unique reflections	301 53	10 069	10 030	10 070	49 275	17 796
Redundancy ^a	5.1 (3.9)	9.1 (9.0)	9.0 (7.5)	9.2 (9.0)	7.6 (7.6)	7.6 (7.8)
<i>R</i> _{sym} (%) ^{a,b}	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>I</i> / <i>σI</i> ^d	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 (%) ^a	96.6 (76.2)	99.9 (100)	99.6 (96.6)	99.9 (100)	99.9 (99.7)	100 (99.9)
<i>Refinement</i>						
	Native	2,6-sialyl- <i>N</i> -acetyl- lactosamine soak	2,3-sialyl- <i>N</i> -acetyl- lactosamine soak			
Resolution (Å) ^a	17.5–1.9 (1.95–1.90)	18.29–2.07 (2.12–2.07)	17.5–2.3 (2.36–2.30)			
Number of reflections	28 518	22 744	16 806			
<i>R</i> _{factor} ^c / <i>R</i> _{free} ^d (%) ^a	17.8/20.4 (29.4/31.3)	18.3/21.2 (28.4/35.6)	17.3/21.3 (20.2/25.5)			
<i>Number of atoms</i>						
Protein	1840	1745	1744			
Ligand/ion	74	98	83			
Water	181	181	153			
<i>B</i> -factors						
Protein	26.3	27.1	35.0			
Ligand/ion	42.5	52.9	51.8			
Water	41.4	39.1	45.8			
<i>R.m.s. deviations</i>						
Bond lengths (Å)	0.018	0.021	0.023			
Bond angles (deg)	1.6	1.8	1.9			

^aValues in parentheses correspond to the highest-resolution shell.^b $R_{\text{sym}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the i th measurement.^c $R_{\text{factor}} = \sum ||F(h)_{\text{obs}}| - |F(h)_{\text{calc}}|| / \sum |F(h)_{\text{obs}}|$.^d*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.

evade host immune responses. MIC3 proteins from *Eimeria tenella*, the cause of coccidiosis in poultry, contain between four and seven consecutive MAR1 domains constructed from five distinct MAR1 sequences (Labbe *et al*, 2005) (Figure 3A). A model of an uninterrupted stretch of MAR1 domains, as 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 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* (1998) 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. Indirect evidence exists for a lactose-binding activity for TgMIC1 and the TgMIC1–TgMIC4 subcomplex (Lourenco *et al*, 2001, 2006). 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).

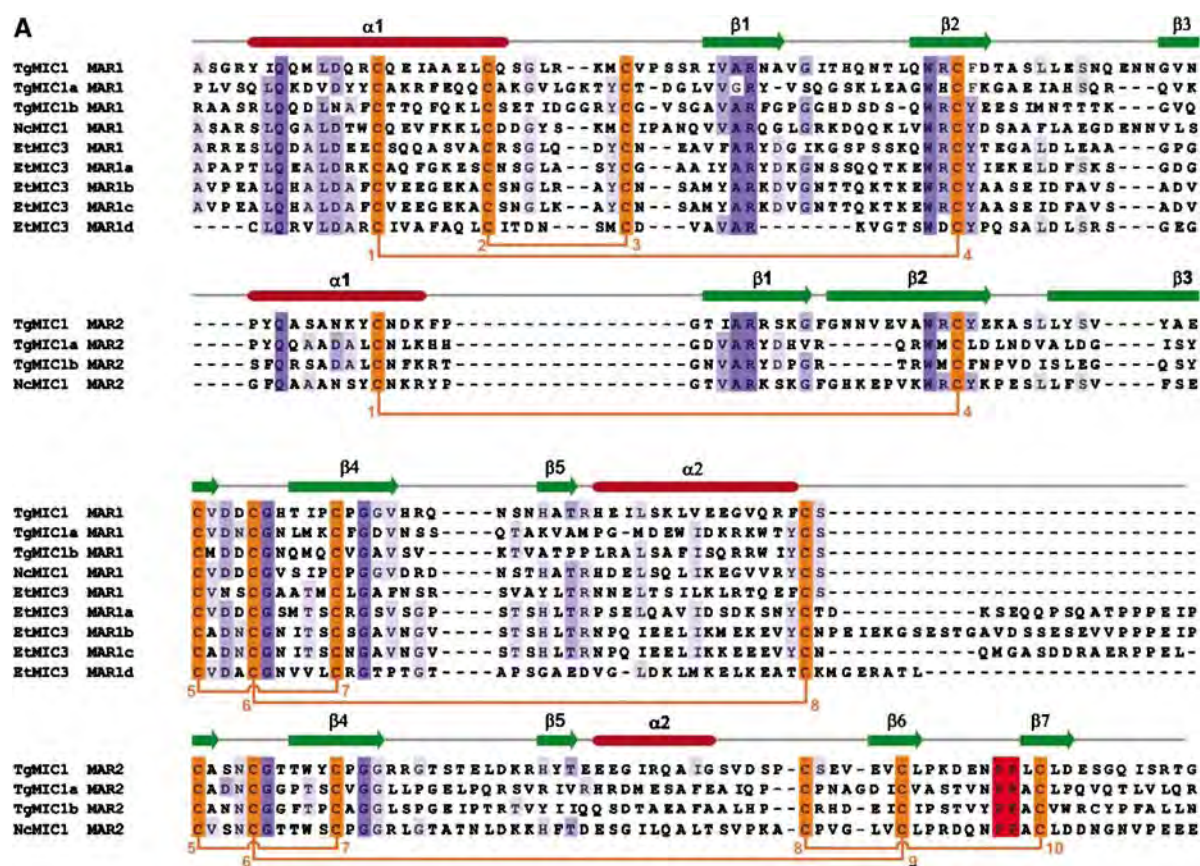


Figure 3 MICs from apicomplexan parasites contain the TgMIC1 repeat. (A) Structure-based sequence alignment of MAR1 and MAR2 domains from other MICs. Including TgMIC1 (MAR1 amino-acid residues 32–144, MAR2 amino-acid residues 145–263), two of the three MIC1-like proteins from *T. gondii* (MAR1 amino-acid residues 114–222, MAR2 amino-acid residues 223–336 in TgMIC1a and MAR1 amino-acid residues 114–222, MAR2 amino-acid residues 223–335 in TgMIC1b), NcMIC1 (MAR1 amino-acid residues 30–142, MAR2 amino-acid residues 143–261) and EtMIC3 (MAR1 amino-acid residues 42–149, MAR1a amino-acid residues 150–274, MAR1b amino-acid residues 291–425, MAR1c amino-acid residues 36–158 and MAR1d amino-acid residues 180–280). For clarity, the third MIC1-like protein, TgMIC1c, from *T. gondii*, has been omitted. Conserved residues are shaded in blue. Cysteines and disulfide bond connectivities are highlighted in orange. *Cis*-proline within the ²³²NPPL²³⁵ 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).

The C-terminal galectin-like domain, TgMIC1-CT, neither showed 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 approximately 8000 were elicited by TgMIC1-NT and were observed only among terminally sialylated structures. (Figure 4 and Supplementary Table 1). 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 II), a mild preference is apparent for the Neu5Ac α 2-3Gal linkage over the α -2,6 and α -2,8 linkages (Neu5Ac denotes *N*-acetylneuraminic acid). The *N*-acetyl group of the sialic acid was found to be required for binding (cf for example probes 19 and 27 in Table II). 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*-acetylglucosamine or α -2,6-sialyl-*N*-acetylglucosamine. The crystallographic structures of the complexes showed that the protein structure was largely unchanged and in all cases a single glycan was shown to be 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 makes specific direct contacts with 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 protons of Thr220, rather than the side chain of an arginine or lysine, which is often the case in other sialic acid-binding proteins (May *et al*, 1998; Alphey *et al*, 2003). 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*-acetylglucosamine and α -2,6-sialyl-*N*-acetylglucosamine 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, pretreatment 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 five to eight carbohydrate units, which suggests that this separation could be sufficient to span both MAR domains of one TgMIC1 molecule (Table II).

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 to be bound in MAR1. Instead, an acetate molecule interacts with the equivalent threonine in MAR1 (Thr126) in a manner identical 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 two-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-labelled 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 $^{124}\text{HATR}^{127}$ and $^{218}\text{HYTE}^{221}$, 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*-acetylglucosamine or α -2,6-sialyl-*N*-acetylglucosamine, 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 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

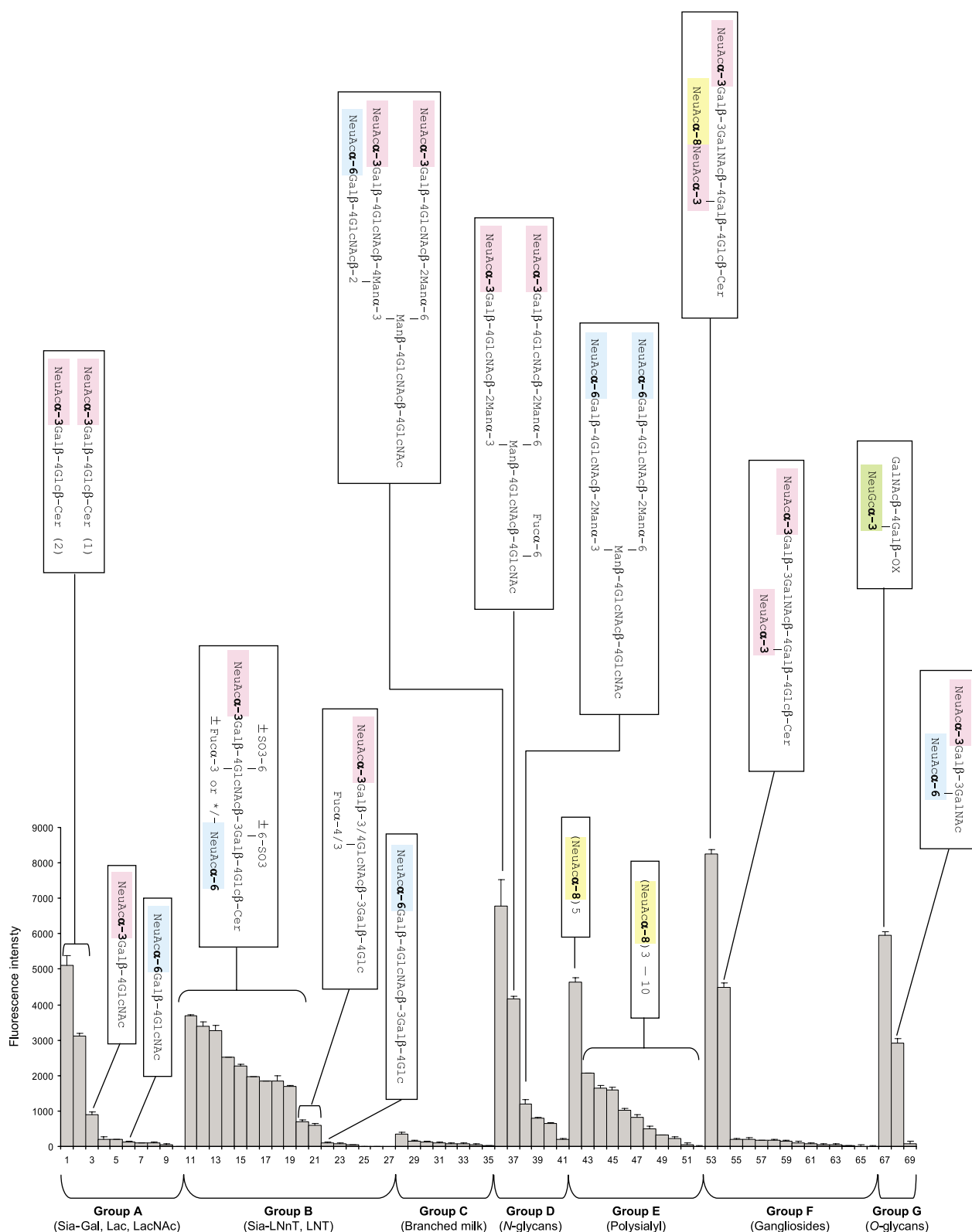


Figure 4 Carbohydrate microarray data on sialyl glycan binding by TgMIC1-NT. Numerical scores are shown of the binding signals, means of duplicate spots at 7 fmol/spot (with error bars) for the 58 sialyl oligosaccharide probes examined. Sixty-nine positions are shown, as 11 of the probes were printed more than once (Supplementary Table 1). Selected sialo-oligosaccharide sequences are annotated, with designations of Neu5Ac α -2,3-gal linkage as pink; Neu5Ac α -2,6Gal, blue; NeuGc α -2-3Gal, green and Neu5Ac α -2,8 linkage yellow. The scores for the non-sialylated probes in the microarray are not shown. These are provided in Supplementary Table 1 (positions 70–218).

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).

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

Name (position in Figure 4)	Sequence	Fluorescence intensity at 7 fmol per spot
SA (3') LacNAc (3)	NeuAc α -3Gal β -4GlcNAc	886
SA (6') LacNAc (6)	NeuAc α -6Gal β -4GlcNAc	124
Sialyl Le ^x (21)	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4Glc $\downarrow$ Fucα-3	2520
LSTc (22)	NeuAc α -6Gal β -4GlcNAc β -3Gal β -4Glc	100
6' SU-Sialyl Le ^x (19)	HSO ₃ -6 $\downarrow$ NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer $\downarrow$ Fucα-3	1692
de-N-acetyl 6' SU-Sialyl Le ^x (27)	HSO ₃ -6 $\downarrow$ Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer $\downarrow$ Fucα-3	< 1
A2F(A2-3) (36)	NeuAc α -3Gal β -4GlcNAc β -2Man α -6 $\downarrow$ Manβ-4GlcNAcβ-4GlcNAc $\downarrow$ Fucα-3	4170
A2 (37)	NeuAc α -6Gal β -4GlcNAc β -2Man α -6 $\downarrow$ Manβ-4GlcNAcβ-4GlcNAc $\downarrow$ NeuAcα-3Galβ-4GlcNAcβ-2Manα-3 $\downarrow$ NeuAcα-6Galβ-4GlcNAcβ-2Manα-6 $\downarrow$ Manβ-4GlcNAcβ-4GlcNAc	1197
A3 (38)	NeuAc α -6Gal β -4GlcNAc β -2Man α -3 $\downarrow$ NeuAcα-6Galβ-4GlcNAcβ-2	6777
GT1b (53)	NeuAc α -3Gal β -4GalNAc β -4Gal β -4Glc β -Cer $\downarrow$ NeuAcα-8 NeuAcα-3	8254
GD1a (54)	NeuAc α -3Gal β -4GalNAc β -4Gal β -4Glc β -Cer $\downarrow$ NeuAcα-3	4492
GT1a (55)	NeuAc α -8NeuAc α -3Gal β -4GalNAc β -4Gal β -4Glc β -Cer $\downarrow$ NeuAcα-3	204

A model for glycan recognition by TgMIC1

It is well accepted that the binding of MICs 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 provide 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 the 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 pretreated with neuraminidase. Compared to controls, both conditions severely reduced the levels of parasite internalization (Figure 7C and D), 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 *mic1KO*s to be 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 emphasizing the importance of the MAR domains in penetration.

Sialic acids are 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 glycoporphin A during invasion by the malarial parasite (Tolia *et al*, 2005). Although

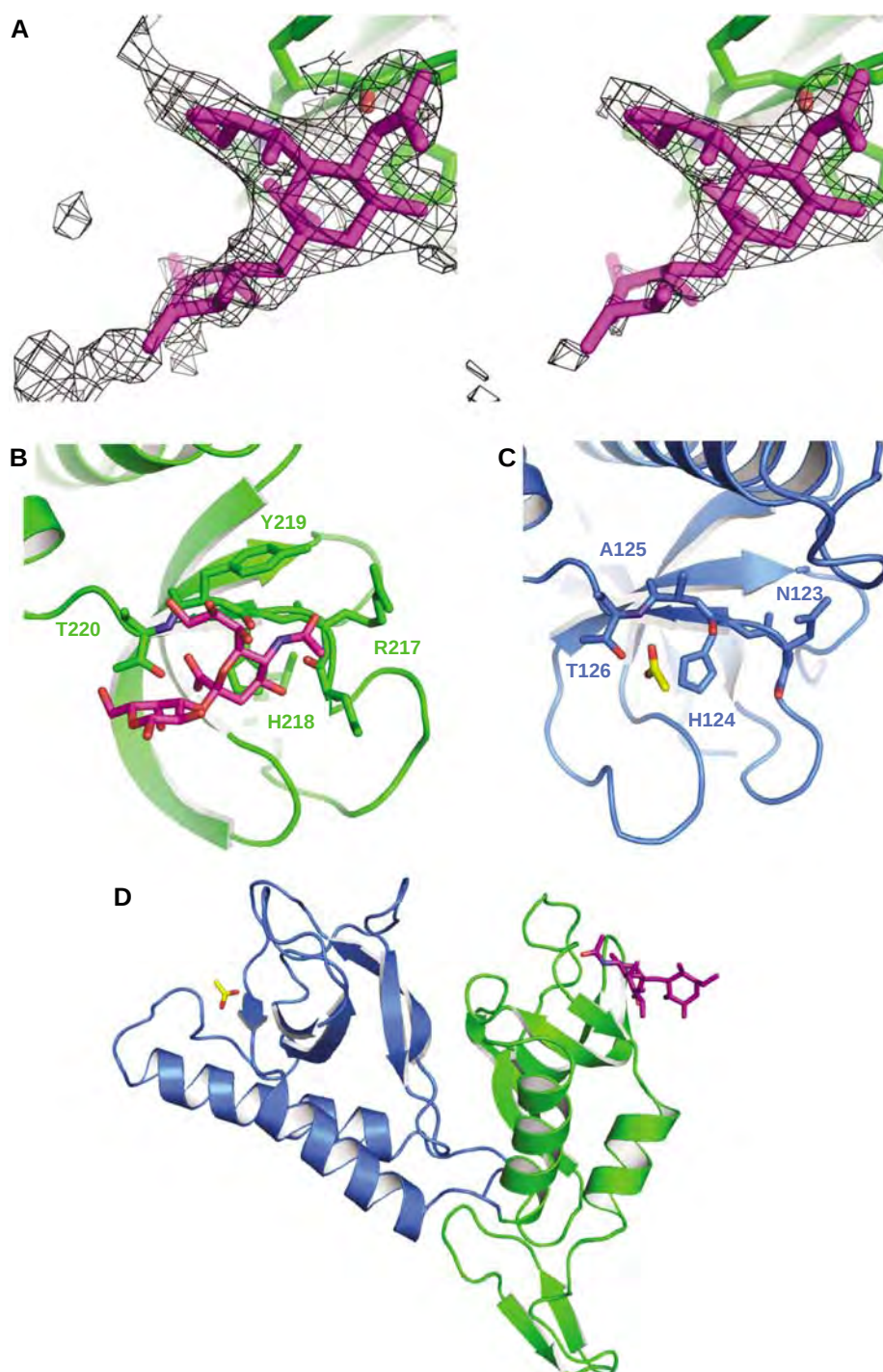


Figure 5 Structure of TgMIC-NT in complex with sialyl oligosaccharides. (A) Simulated annealing ($F_o - F_c$) OMIT map contoured at 3σ (left) and 2σ levels (right) for the TgMIC-NT-glycan complex showing the unambiguous orientation of the sialic acid moiety and the position of the galactose unit of α -2,3-sialyl-N-acetylglucosamine. The omit map was calculated with the glycan omitted; stick models of key side chains (green) and sialic acid (Magenta) are overlaid on the map. (B) Structure of α -2,3-sialyl-N-acetylglucosamine 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 colored in red and blue, respectively. Note: the structure of the α -2,6-sialyl-N-acetylglucosamine complex is shown in Supplementary Figure 1. (C) Structure of the MAR1 domain from TgMIC1-NT shows 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.

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 glycoporphin 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,

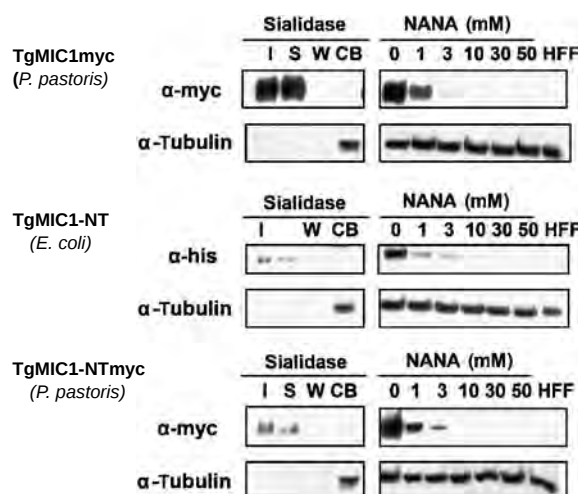


Figure 6 Sialic acid competes with TgMIC1 cell-binding activity. Cell binding competition experiments with sialic acid were performed on HFFs using supernatants of a *P. pastoris* culture expressing TgMIC1myc (top), bacterially produced HisTgMIC1-NT (middle) or a *P. pastoris* culture expressing TgMIC1-NTmyc (bottom). Anti-myc and anti-His antibodies were 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 pretreatment 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 500 μ l, whereas the total volume of the cell-bound fraction is 50 μ l. NANA, N-acetylneuraminic acid.

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 multivalent 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 (Finne et al, 1983; Acheson et al, 1991). 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 crosslinks 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 17–262 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 μ M 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 reappplied to the same column and pure MIC1-N-terminus was recovered in the flow-through. The protein was concentrated to $\sim$ 10 mg/ml and stored at 4 or -20°C in 1 mM CaCl_2 , 100 mM NaCl and 50 mM Tris-Cl pH 8.0. SeMet-labelled TgMIC1-NT was expressed in minimal media in *E. coli* Origami (DE3), according to the protocol of 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% $^{15}\text{NH}_4\text{Cl}$ and 0.2% $^{13}\text{C}_6$ -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 pPICZ α -TgMIC1TSR1 (called pPICZ α -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 pPICZ α -TgMIC1NT with primers 5'-CGCCTA GGGTTGGGCCAGAAGCATATGGAGAAG-3' (MIC1-1_672) and 5'-CCG GCGCGCGCCGAGAACATGGGCTGTCGACGGATCC-3' (TgMIC1-17_1717) was digested with *Nde*I and *Not*I and cloned back into pPICZ α -TgMIC1NT, resulting in pPICZ α -TgMIC1NTACterm (amino-acid residues 238–262). Similarly, the fragment obtained from pROP1mycMIC1NT with primers 5'-CCCGCTGCAGGAGCAAAGGCTG CCAATTATTC-3' (ROP1-SP_141) and 5'-GGCGAGCTCTTAATTAAGAA CATGGGCTGTCGACGGATCC-3' (TgMIC1-16_1716) was digested with *Nsi*I and *Pac*I and cloned back into pROP1mycMIC1NT, resulting in pROP1mycMIC1NTAC2.

Mutants of pPICZ α -TgMIC1NT were obtained using the Quick-change kit (Stratagene). Primers 5'-CAGTAATCACGACGCGCCAT GAGATACTGTC-3' (TgMIC1-18_1729) and 5'-GACAGTATCTCATGGC GCGCTGCGGTGATTACTG-3' (TgMIC1-19_1730), as well as primers 5'-GACAAGCGGCATTATGCAGAAGAGGAAGGAATTCG-3' (TgMIC1-20_1731) and 5'-CGAATTCCTTCTCTTCTGCATAATGCCGCTGTC-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 *Pme*I for all pPICZ α -based vectors. Selection of transformants was then performed on YPD-zeocin (100 μ g/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 1 h at 4°C with 1% bovine serum albumin (BSA) in cold PBS, 1 mM CaCl_2 and 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 μ g), or bacterially produced (0.25 μ g), in combination 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 μ l. 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 direct addition of 50 μ l 1 $\times$ SDS-PAGE loading buffer with 0.1 M dithiothreitol. In some cases, before blocking, HFF cells were pretreated with 66 mU/ml α -2,3,6,8-*Vibrio cholerae* neuraminidase (Calbiochem) in RPMI1640, 25 mM HEPES, L-glutamine (Gibco) for 1 h at 37°C in a total reaction volume of 1 ml.

Cell invasion assays in the presence of carbohydrate inhibitors

Confluent monolayers of HFF cells on glass coverslips in 12-well plates were washed once briefly in DMEM (Gibco) and incubated for 15 min only with 250 μ l DMEM (control) or with 250 μ l DMEM containing different double concentrated competitors (sialic acid

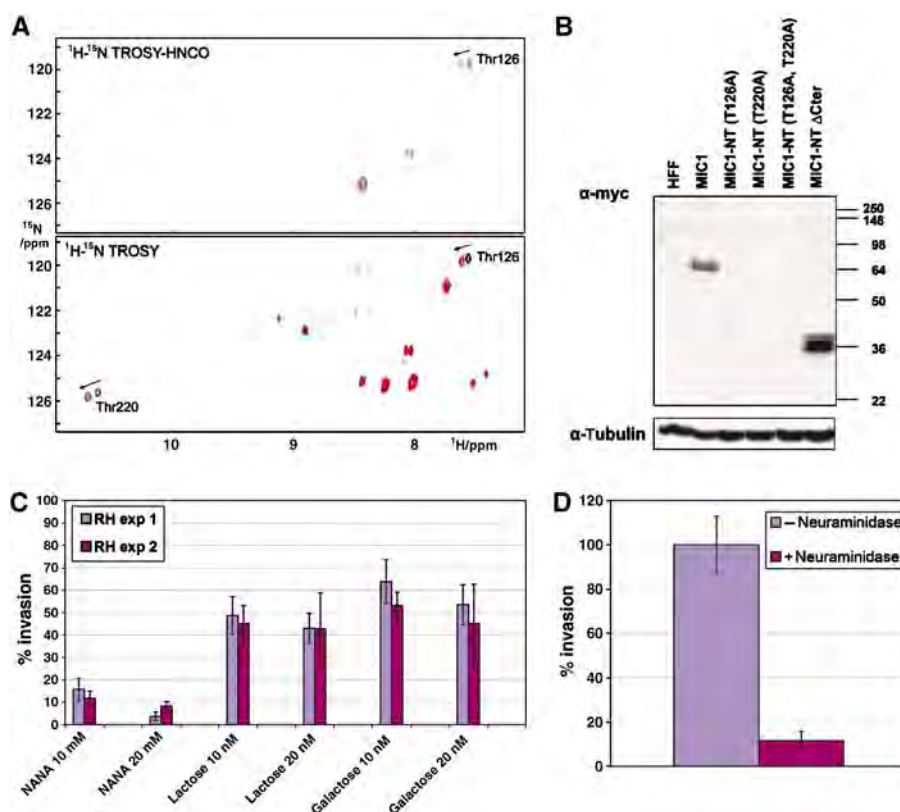


Figure 7 Characterization of the interaction of sialyl oligosaccharides with MAR domains and their involvement in host cell invasion. (A) ^1H - ^{15}N 2D TROSY-HNCO (top) and ^1H - ^{15}N 2D TROSY-HSQC (bottom) NMR spectra for ^{13}C , ^{15}N -labelled Ala, Thr TgMIC1-NT, in the absence (black) and presence (red) of saturating amounts of an α -2,3-sialyl-*N*-acetylglucosamine. 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 pretreatment of target cells with neuraminidase. Four experiments were carried out in parallel. Numbers of intracellular parasites were counted in five microscopic fields.

pH 6.8, lactose, galactose). Then 250 μl of *T. gondii* RH parasites were added and invasion was allowed to proceed for 1 h at 37°C. Afterwards cells were washed twice with 500 μl DMEM, fixed with 4% paraformaldehyde/0.05% glutaraldehyde for 20 min, followed by a 2-min incubation with 0.1 M glycine in PBS. Fixed cells were blocked in PBS and 2% BSA 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 and 0.2% Triton X-100 for 20 min and blocked again in PBS, 0.2% Triton X-100 and 2% BSA 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 procedure, numbers of extracellular (attached) and intracellular (invaded) parasites were counted in three microscopic fields. For the untreated control, a total of 200 parasites was 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 treatment

Confluent monolayers of HFF cells on glass coverslips in 24-well plates were incubated in RPMI1640, 25 mM HEPES and L-glutamine (Gibco) at 37°C, with or without 66 mU *V. cholerae* neuraminidase (Calbiochem) in a total volume of 500 μl . After 1 h, 500 μl freshly lysed *T. gondii* RH parasites resuspended in RPMI1640, 25 mM

HEPES and L-glutamine (37°C) were added to each well and the plate was centrifuged immediately for 5 min at 480 g. Then invasion was allowed to proceed for another 5 min at 37°C. Afterwards cells were washed two times with 500 μl 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% BSA 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 and 0.2% Triton X-100 for 20 min and blocked again in PBS, 0.2% Triton X-100 and 2% BSA 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 procedure, numbers of extracellular (attached) and intracellular (invaded) parasites were counted in five microscopic fields per coverslip. For the untreated control, a total of 400 parasites was 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.

Crystallization and data collection

TgMIC1-NT was crystallized in hanging drops by the vapor-diffusion technique. Crystals shaped as tetragonal bipyramids,

measuring approximately $50 \times 50 \times 100 \mu\text{m}$, 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 in 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-acetylactosamine; 3.4 M ammonium acetate, 100 mM bis-tris-propane pH 7.0, 3.3 mM α -2,3-sialyl-N-acetylactosamine, 27.5 mM NaCl and 5.5 mM sodium phosphate, or 3.6 M 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-acetylactosamine (2.3 Å) and 2,6-sialyl-N-acetylactosamine (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), yielding 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 were carried out 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 provided in Table I. A total of 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 $F_o - F_c$ 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-acetylactosamine) and 2JHD (TgMIC1-NT in complex with 2,3-sialyl-N-acetylactosamine). Figures were drawn using the program MacPyMol (DeLano Scientific). The model for the sialic acid complex showed density for the sialic acid that perfectly matched the sialic acid moiety in the other two sugars. The sialic acid structure was therefore not refined further, as the data were at lower resolution (2.6 Å) than the other two sugar moieties 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 and 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 sulfated analogues, gangliosides, glycosaminoglycans, homo-oligomers of sialic acid and fragments of other polysaccharides, ranging in size from two to 20 monosaccharide units. The probes were printed at 2 and 7 fmol per spot, in duplicate, using a non-contact arrayer (Piezotray, 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 h with blocking solution containing casein (Pierce) with 1% w/v BSA (Sigma) and 10 mM calcium (casein/BSA). The slides were rinsed and overlaid for 1 h with fusion protein complex in casein/BSA: this contained per 100 μl , 10 μg mouse anti-histidine monoclonal antibody, 10 μg 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, 2 mM calcium chloride solution and binding was detected using Alexa Fluor 647-labelled streptavidin (Molecular Probes) and 1 μg 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 ¹⁵N-labelled TgMIC1-NT, uniformly ¹⁵N,¹³C-labelled TgMIC1-NT or TgMIC1-NT specifically labelled with ¹⁵N,¹³C-labelled Ala and Thr were prepared in 20 mM sodium phosphate buffer pH 7, 100 mM NaCl, 4% CompleteTM 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 ¹⁵N,¹³C-labelled TgMIC1-NT samples (Sattler *et al*, 1999). For Thr126, these data were supplemented with HNCQ spectra recorded on TgMIC1-NT specifically labelled with ¹⁵N,¹³C-labelled Ala and Thr. Four peaks were identified and could be assigned to AA, TA, TT and ¹²⁴HATR¹²⁷. Either α -2,3-sialyl-N-acetylactosamine or α -2,6-sialyl-N-acetylactosamine in the same buffer was added in several steps up to a 10-fold molar excess and 2D ¹H-¹⁵N HSQC spectra were recorded at each stage under identical experimental conditions.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

This work was supported by the Medical Research Council (MRC grant numbers G0400423, SJM and G9601454, TF, BBSRC) and Swiss National Foundation and Research Councils UK Basic Technology Grant (GR/S79268 'Glycoarrays', TF). This work is part of the activities of the BioMalPar European Network of Excellence supported by a European grant (LSHP-CT-2004-503578) from the Priority 1 'Life Sciences, Genomics and Biotechnology for Health' in the 6th Framework Programme. We thank Mark Stoll for design of software and collaboration in microarray data analysis, and Kovilen Sawmynaden, Jan Marchant and Bing Lui for useful discussion. The Glycosciences Laboratory acknowledges with gratitude colleagues and collaborators over the years with whom our microarray probes were studied.

References

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

Alphey MS, Attrill H, Crocker PR, van Aalten DMF (2003) High resolution crystal structures of Siglec-7. *J Biol Chem* **278**: 3372–3377

- Bailey S (1994) The Ccp4 suite—programs for protein crystallography. *Acta Crystallogr D* **50**: 760–763
- Brecht S, Carruthers VB, Ferguson DJP, Giddings OK, Wang G, Jakle U, Harper JM, Sibley LD, Soldati D (2001) The *Toxoplasma* micronemal protein MIC4 is an adhesin composed of six conserved apple domains. *J Biol Chem* **276**: 4119–4127
- Campanero-Rhodes MA, Childs RA, Kiso M, Komba S, Le Narvor C, Warren J, Otto D, Crocker PR, Feizi T (2006) Carbohydrate microarrays reveal sulphation as a modulator of siglec binding. *Biochem Biophys Res Comm* **344**: 1141–1146
- Carruthers V, Boothroyd JC (2006) Pulling together: an integrated model of *Toxoplasma* cell invasion. *Curr Opin Microbiol* **10**: 83–89
- Carruthers VB, Sibley LD (1997) Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *Eur J Cell Biol* **73**: 114–123
- Carruthers VB, Giddings OK, Sibley LD (1999) Secretion of micronemal proteins is associated with *Toxoplasma* invasion of host cells. *Cell Microbiol* **1**: 225–235
- Cerede O, Dubremetz JF, Bout D, Lebrun M (2002) The *Toxoplasma gondii* protein MIC3 requires pro-peptide cleavage and dimerization to function as adhesin. *EMBO J* **21**: 2526–2536
- Cerede O, Dubremetz JF, Soete M, Deslee D, Vial H, Bout D, Lebrun M (2005) Synergistic role of micronemal proteins in *Toxoplasma gondii* virulence. *J Exp Med* **201**: 453–463
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**: 2126–2132
- Feizi T, Chai WG (2004) Oligosaccharide microarrays to decipher the glyco code. *Nat Rev Mol Cell Biol* **5**: 582–588
- Finne J, Finne U, Deagostinibazin H, Goridis C (1983) Occurrence of alpha-2–8 linked polysialosyl units in a neural cell-adhesion molecule. *Biochem Biophys Res Comm* **112**: 482–487
- Fourmaux MN, Achbarou A, Mercereau-Puijalon O, Biderre C, Briche I, Loyens A, OdbergFerrat C, Camus D, Dubremetz JF (1996) The MIC1 microneme protein of *Toxoplasma gondii* contains a duplicated receptor-like domain and binds to host cell surface. *Mol Biochem Parasitol* **83**: 201–210
- Harper JM, Hoff EF, Carruthers VB (2004) Multimerization of the *Toxoplasma gondii* MIC2 integrin-like A-domain is required for binding to heparin and human cells. *Mol Biochem Parasitol* **134**: 201–212
- He X-L, Grigg ME, Boothroyd JC, Garcia KC (2002) Structure of the immunodominant surface antigen from the *Toxoplasma gondii* SRS superfamily. *Nat Struct Biol* **9**: 606–611
- Hill D, Dubey JP (2002) *Toxoplasma gondii*: transmission, diagnosis and prevention. *Clin Microbiol Infect* **8**: 634–640
- Holm L, Sander C (1995) Dali—a network tool for protein-structure comparison. *Trends Biochem Sci* **20**: 478–480
- Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* **381**: 272
- Huynh MH, Opitz C, Kwok LY, Tomley FM, Carruthers VB, Soldati D (2004) Trans-genera reconstitution and complementation of an adhesion complex in *Toxoplasma gondii*. *Cell Microbiol* **6**: 771–782
- 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
- Karlsson KA (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, Hemphill A (2002) Identification of a *Neospora caninum* microneme protein (NcMIC1) which interacts with sulfated host cell surface glycosaminoglycans. *Infect Immun* **70**: 3187–3198
- 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
- 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
- Laurenco EV, Bernardes ES, Silva NA, Mineo JR, Panunto-Castelo A, Roque-Barreira MC (2006) Immunization with MIC1 and MIC4 induces protective immunity against *Toxoplasma gondii*. *Microbes Infect* **8**: 1244–1251
- Laurenco EV, Pereira SR, Faca VM, Coelho-Castelo AAM, Mineo JR, Roque-Barreira MC, Greene LJ, Panunto-Castelo A (2001) *Toxoplasma gondii* micronemal protein MIC1 is a lactose-binding lectin. *Glycobiology* **11**: 541–547
- Matthews SJ, Jandu SK, Leatherbarrow RJ (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 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 Å resolution. *Mol Cell* **1**: 719–728
- Monteiro VG, Soares CP, de Souza W (1998) Host cell surface sialic acid residues are involved on the process of penetration of *Toxoplasma gondii* into mammalian cells. *FEMS Microbiol Lett* **164**: 323–327
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In: Carter Jr CW, Sweet RM (eds) *Methods in Enzymology, volume 276A: Macromolecular Crystallography* pp 307–326. New York: Academic Press
- Palma AS, Feizi T, Zhang YB, Stoll MS, Lawson AM, Diaz-Rodriguez E, Campanero-Rhodes MA, Costa J, Gordon S, Brown GD, Chai WG (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
- Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**: 458–463
- 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
- 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** (in press)
- Saouros S, Edwards-Jones B, Reiss M, Sawmynaden K, Cota E, Simpson P, Dowse TJ, Jakle 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
- Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multi-dimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in NMR Spectroscopy* **34**: 93–158. Amsterdam: Elsevier Science
- Soldati D, Meissner M (2004) *Toxoplasma* as a novel system for motility. *Curr Opin Cell Biol* **16**: 32–40
- Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D* **55**: 849–861
- Tolia NH, Enemark EJ, Sim BKL, 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, 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, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30**: 1022–1025
- Van Duyn GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin Fkbp-12 complexes with Fk506 and rapamycin. *J Mol Biol* **229**: 105–124
- Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D* **57**: 122–133

3.2 Structural insights into microneme protein assembly reveal a new mode of EGF domain recognition

The architecture of the TgMIC1-4-6 complex has been partly established previously through conventional knockout of the individual genes showing that TgMIC6 interacts with TgMIC1 which in turn recruits TgMIC4 (Reiss et al, 2001). Structural studies demonstrated the importance of an interaction between TgMIC1-GLD and the membrane proximal TgMIC6-EGF3 encompassing also a downstream stretch of acidic amino acids for assembly and correct trafficking of the TgMIC1-4-6 complex (Saouros et al, 2005). Transfection experiments of TgMIC6 mutants with deletion of individual EGF domains into the *mic6ko* background suggested that TgMIC6-EGF2 might as well be able to interact with TgMIC1-GLD (Reiss et al, 2001). In this study we have investigated on this interaction and the implications for the stoichiometry and functionality of the complex. While all structural studies were performed by the group of Prof. Stephen Matthews (Imperial College, London), the functional dissection of the TgMIC1-4-6 complex in the parasite constitutes the personal contribution from our laboratory. More details on the individual contributions of the authors can be found in the acknowledgements of the article.

Structural insights into microneme protein assembly reveal a new mode of EGF domain recognition

Kovilen Sawmynaden¹, Savvas Saouros¹, Nikolas Friedrich², Jan Marchant¹, Peter Simpson¹, Boris Bleijlevens¹, Michael J. Blackman³, Dominique Soldati-Favre² & Stephen Matthews¹⁺

¹Division of Molecular Biosciences, Imperial College London, London, UK, ²Department of Microbiology and Molecular Medicine, Faculty of Medicine, University of Geneva CMU, Geneva, Switzerland, and ³Division of Parasitology, National Institute for Medical Research, London, UK

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. Adhesive complexes composed of microneme proteins (MICs) are secreted onto the parasite surface from intracellular stores and fulfil crucial roles in host-cell recognition, attachment and penetration. Here, we report the high-resolution solution structure of a complex between two crucial MICs, TgMIC6 and TgMIC1. Furthermore, we identify two analogous interaction sites within separate epidermal growth factor-like (EGF) domains of TgMIC6—EGF2 and EGF3—and confirm that both interactions are functional for the recognition of host cell receptor in the parasite, using immunofluorescence and invasion assays. The nature of this new mode of recognition of the EGF domain and its abundance in apicomplexan surface proteins suggest a more generalized means of constructing functional assemblies by using EGF domains with highly specific receptor-binding properties.

Keywords: apicomplexa; microneme proteins; invasion; solution structure; EGF

EMBO reports advance online publication 26 September 2008;
doi:10.1038/embor.2008.179

INTRODUCTION

The intracellular protozoan parasite *Toxoplasma gondii* is uniquely adapted to infect a wide range of hosts, including virtually all warm-blooded animals and humans. Toxoplasmosis causes a variety of disease states in humans, including severe disseminated disease in immunosuppressed individuals owing to

reactivation of encysted parasites and birth defects in infants where mothers are exposed to the parasite for the first time during pregnancy (Hill & Dubey, 2002). *T. gondii* also causes significant disease and economic loss in the farming industry, principally by inducing abortion and fetal abnormality in sheep (Fusco *et al*, 2007). Raw and undercooked meat from infected animals represents a principal source of human infection with *T. gondii* (Dubey, 2004). Around 500 million of the world's population have been infected with *Toxoplasma*, and toxoplasmosis is considered to be the third leading cause of death attributed to foodborne illness in the United States.

Toxoplasma and other apicomplexan parasites, including *Plasmodium* (the agents of malaria), share a set of apical secretory organelles such as micronemes and rhoptries, which release a large number of soluble and membrane proteins onto the parasite's surface during invasion. Microneme proteins (MICs) are stored in the micronemes before secretion, after which they participate in the attachment of parasites to the host cell surface (Carruthers *et al*, 1999) and in the formation of a connection with the parasite actinomyosin system (Jewett & Sibley, 2003), which drives motility and invasion (Soldati & Meissner, 2004). The integrity of the MIC cargo is controlled by several molecular checkpoints during transport and after secretion, which include conformation-dependent sorting and proteolytic processing events. These processes are particularly well characterized for the TgMIC4–MIC1–MIC6 complex (TgMIC4–1–6) from *T. gondii* (Reiss *et al*, 2001; Saouros *et al*, 2005a).

The TgMIC4–1–6 complex binds tightly to host cell receptors and has a central function in invasion and virulence (Blumenschein *et al*, 2007). TgMIC6 contains a membrane-spanning domain and a carboxy-terminal cytoplasmic tail encompassing the sorting signal that is essential for the accurate targeting of the complex to the micronemes (Fig 1A). TgMIC6 also contains three extracellular epidermal growth factor-like (EGF) domains, the first of which is proteolytically removed during its transport to the microneme organelles and is not present in mature forms of the complex (Meissner *et al*, 2002a). In addition, an interaction between the third EGF-like domain of TgMIC6 (TgMIC6-EGF3) and the C-terminal galectin-like domain of

¹Division of Molecular Biosciences, Imperial College London, Exhibition Road, South Kensington Campus, London SW7 2AZ, UK

²Department of Microbiology and Molecular Medicine, Faculty of Medicine, University of Geneva CMU, 1 Rue Michel-Servet, 1211 Geneva 4, Switzerland

³Division of Parasitology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

⁺Corresponding author. Tel: +44 (0)207 594 5315; Fax: +44 (0)207 594 5207; E-mail: s.j.matthews@imperial.ac.uk

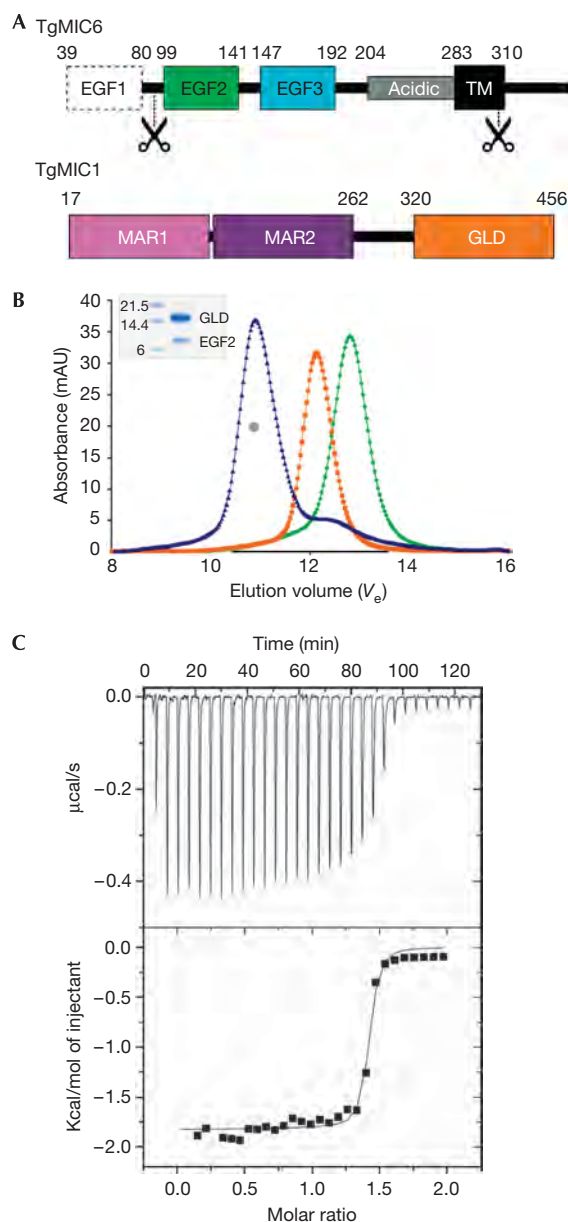


Fig 1 | TgMIC6-EGF2 interacts independently with TgMIC1-GLD. (A) Schematic representation of the domain structure of TgMIC6 (top) and TgMIC1 (bottom). Microneme adhesive repeat (MAR) domain, galectin-like domain (GLD), epidermal growth factor-like (EGF), transmembrane (TM) and acidic domains are shown. The first EGF domain is proteolytically removed during transport to the micronemes. (B) Gel filtration profiles for TgMIC6-EGF2 (green), TgMIC1-GLD (orange) and TgMIC1-GLD–TgMIC6-EGF2 complex (blue). (C) Isothermal titrations of TgMIC1-GLD with TgMIC6-EGF2. Analysis of the titration data shows $K_d = 53 \pm 13$ nM; $\Delta H = -1820 \pm 19$ cal mol⁻¹ and $\Delta S = 27.2$ eu. MIC, microneme proteins.

TgMIC1 (TgMIC1-GLD; Fig 1A) is crucial for the correct folding of TgMIC6-EGF3 and the transport of the whole complex to the micronemes (Reiss *et al*, 2001; Saouros *et al*, 2005a). TgMIC1

comprises a tandem repeat of a new cell-binding motif called the microneme adhesive repeat (MAR), which has been shown to bind to sialylated oligosaccharides on the host cell surface and recruitment of the third component of the complex, TgMIC4 (Blumenschein *et al*, 2007). TgMIC4 is also a soluble cell-adhesion protein composed of six apple domains that interacts with TgMIC1 and is processed extracellularly, releasing a 15-kDa C-terminal fragment (Brecht *et al*, 2001).

Here, we describe a new interaction between TgMIC1 and the second EGF domain from TgMIC6 (TgMIC6-EGF2), and provide a high-resolution basis for assembly of this complex. Experiments based on immunofluorescence studies and host cell invasion assays indicate that both MIC1–MIC6 interactions contribute to the formation of active host cell-binding sites for the invasion of *T. gondii*. Our study provides a new insight into the mode of interaction between TgMIC1–4–6 and host receptors, as both EGF domains contribute to MIC1 recruitment, recognition of host cell and subsequent invasion.

RESULTS AND DISCUSSION

EGF2 from TgMIC6 interacts independently with TgMIC1

Although interactions between MICs are of central importance to their integrity and surface presentation, little is known about the high-resolution determinant of complex assembly. For example, the mature extracellular portion of TgMIC6 contains two canonical EGF domains (EGF2 and EGF3; Fig 1A), and although TgMIC6-EGF3 is known to interact with the C-terminal galectin-like domain of TgMIC1 (TgMIC1-GLD; supplementary Fig 1 online; Reiss *et al*, 2001; Saouros *et al*, 2005a), no role has been reported for the second EGF domain (TgMIC6-EGF2). To address this issue, the second EGF domain from the extracellular portion of mature TgMIC6 (TgMIC6-EGF2) was produced in *Escherichia coli* and assayed for binding to TgMIC1-GLD by gel filtration (Fig 1B). Examination of the gel filtration profiles showed the co-elution of a protein complex of around 25 kDa molecular weight, corresponding to TgMIC1-GLD with TgMIC6-EGF2. To determine the affinity of this interaction, the binary complex was characterized by using isothermal calorimetry. The binding data for TgMIC1-GLD and TgMIC6-EGF2 followed a standard binding curve. Recorded measurements for enthalpy changes during the titration were fit to a model for a single binding event yielding $K_d = 53 \pm 13$ nM with an observed stoichiometry consistent with a 1:1 interaction (Fig 1C).

A new and specific mode of EGF domain recognition

The high-resolution solution structure of the TgMIC1-GLD–TgMIC6-EGF2 complex was solved by using heteronuclear multi-dimensional nuclear magnetic resonance (NMR) spectroscopy (Fig 2A,B; supplementary Table 1 online). In agreement with the structure of TgMIC1-GLD (Saouros *et al*, 2005a), this region shows the classic β -sandwich of a galectin-like domain when in complex with TgMIC6-EGF2. The structure of the EGF domain within the complex shows a canonical EGF domain topology (Fig 2B), in which two double-stranded β -hairpins stack in a staggered arrangement that is fixed by the three disulphide bonds present within the core. The two β -hairpins from TgMIC6-EGF2 clamp onto one side of the β -sandwich from TgMIC1-GLD burying an intimate and large surface area at the interface (~ 1450 Å²), which

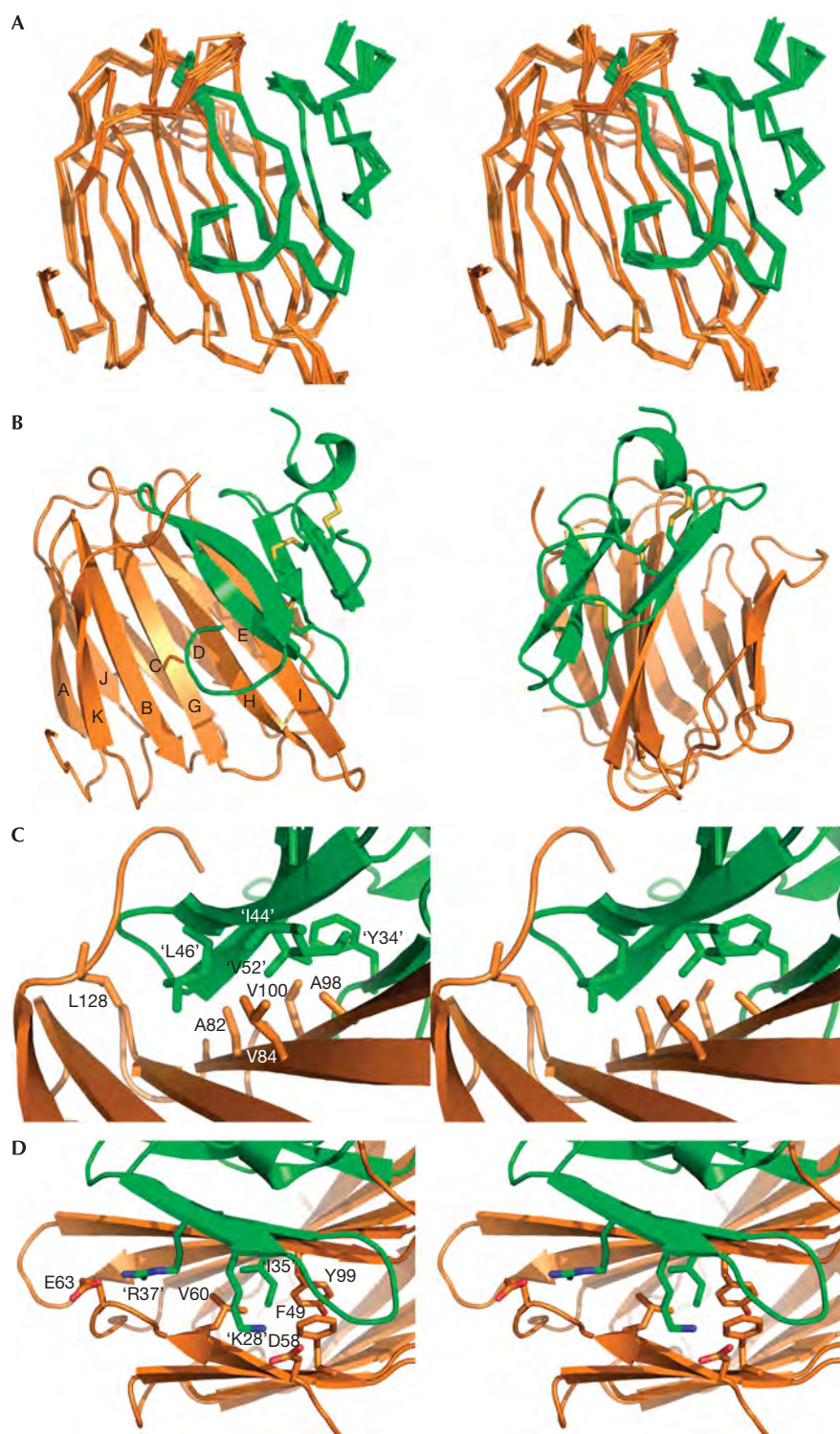


Fig 2 | Solution structure of the complex between TgMIC1-GLD and TgMIC6-EGF2. (A) Stereo diagram showing C_{α} traces representing the ensemble of nuclear magnetic resonance-derived structures (TgMIC1-GLD is shown in orange and TgMIC6-EGF2 is shown in green). (B) Ribbon representation of a representative structure for the TgMIC1-GLD-TgMIC6-EGF2 complex (TgMIC1-GLD is shown in orange and TgMIC6-EGF2 is shown in green). Strand assignments are indicated. The orientation shown on the left is identical to that in (A), whereas the orientation shown on the right represents a 90° rotation about the y axis. (C) Stereo diagram showing an enlarged view of interface I from the TgMIC1-GLD-TgMIC6-EGF2 complex illustrating crucial hydrophobic residues at the interface. (D) Stereo diagram showing an enlarged view of interface II from the TgMIC1-GLD-TgMIC6-EGF2 complex. Residues are numbered for both TgMIC1-GLD and TgMIC6-EGF2 (labelled with quotes). EGF, epidermal growth factor-like; GLD, galectin-like domain; MIC, microneme proteins.

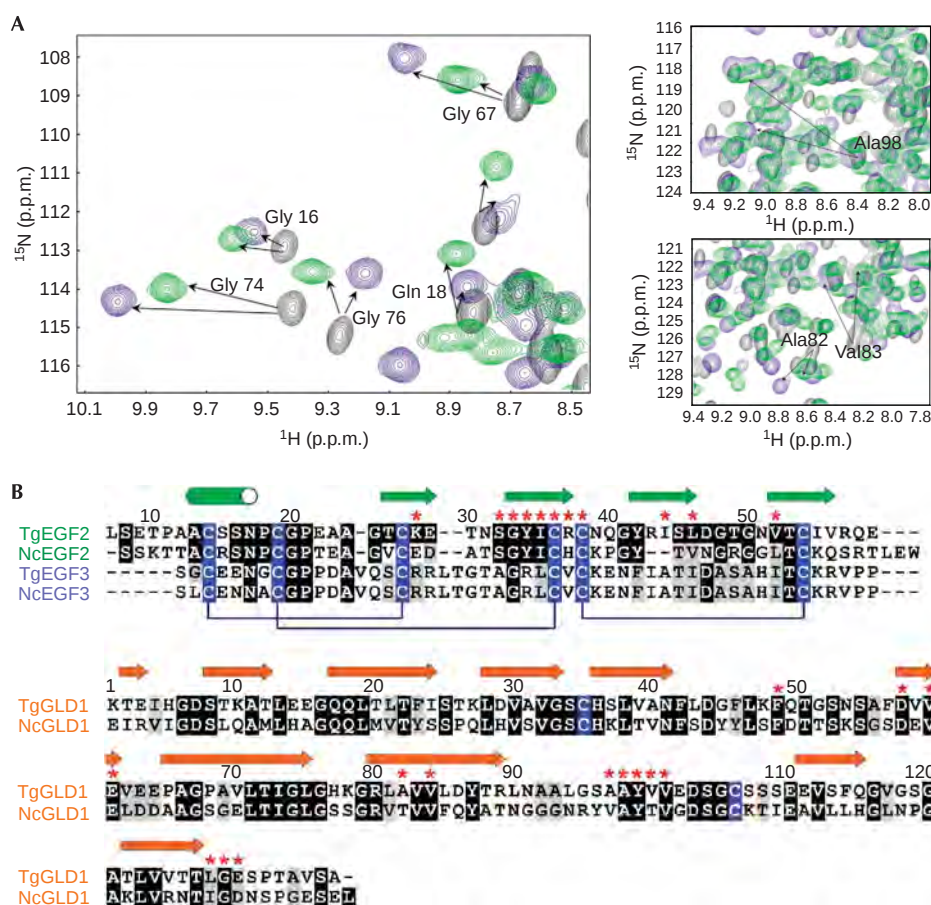


Fig 3 | MIC6 and MIC1 from *Toxoplasma gondii* and *Neospora caninum* assemble in an analogous manner. (A) Regions of two-dimensional ^1H - ^{15}N HSQC spectra for ^{13}C - ^{15}N -labelled TgMIC1-GLD alone (black) and in the presence of unlabelled TgMIC6-EGF2 (green) or TgMIC6-EGF3 (blue) at a molar ratio of 1:1. Assignments for resonances experiencing large chemical shift deviations are labelled. (B) Sequence alignments for the EGF domains from TgMIC6 and from a putative MIC6 homologue from *N. caninum* (top), and for the GLD domains from TgMIC1 and NcMIC1 (bottom), including TgMIC6 (EGF2 aa 93–146, EGF3 aa 146–196, TgMIC1 (GLD aa 320–456) and NcMIC1 (GLD aa 323–460)). Cysteines are highlighted in blue and disulphide bond connectivities are indicated (brackets) for the EGF domains. Secondary structure elements are indicated above the sequence alignments; β -strands are shown as arrows and the α -helix is shown as a cylinder. Prominent interfacial side chains are indicated by red asterisks. EGF, epidermal growth factor-like; GLD, galectin-like domain; HSQC, heteronuclear single quantum coherence; MIC, microneme protein.

explains the high affinity of the complex. Gap volume index for the complex is 1.1 Å, indicating highly complementary interacting surfaces (Laskowski, 1995).

Although TgMIC1-GLD is in contact with many residues along the length of the TgMIC6-EGF2 sequence, two principal interfacial regions can be delineated. The first comprises Ala 82, Val 84, Ala 98 and Val 100 from β -strands H and I in TgMIC1-GLD, which forms an exposed hydrophobic surface on one of the large β -sheets (strands KBGHI; Fig 2B) and interacts with residues Ile 44, Leu 46 and Val 52 from the underside of the C-terminal β -hairpin from EGF2 (Fig 2C; interface I). A second important point of contact is presented by TgMIC1-GLD at one edge of the β -sandwich (Fig 2D; interface II). In the centre of this region are the hydrophobic residues Phe 49, Val 60 and Tyr 99 from TgMIC1-GLD, which form a binding pocket for Ile 35 from TgMIC6-EGF2 (Fig 2D). Classical antiparallel hydrogen bonding patterns are also observed between main chain atoms of the second β -strand

(residues Cys 38, Cys 36 and Tyr 34) from the EGF domain and the edge of the KBGHI β -sheet (Ser 96, Ala 98 and Val 100 from strand I) to form a contiguous seven-stranded sheet.

To provide a complete structural insight into the mode of interaction, the high-resolution solution structure of TgMIC6-EGF2 in its unbound form was solved. Although the absence of principal rearrangements of secondary structure between the free and bound forms (r.m.s.d. of 1.8 and 1.4 Å over 102 and 41 C_α atoms within secondary structure elements of TgMIC1 and TgMIC6, respectively) supports a lock-and-key mode of interaction, several loop regions become more ordered on complex formation. For example, the disordered C terminus of TgMIC1-GLD (residues 128–137) provides a ‘latch’ by folding back over the KBGHI β -sheet and making numerous intermolecular contacts with the EGF domain (supplementary Fig 2 online).

To propose a model for the assembly of mature, full-length forms of TgMIC1 and TgMIC6, we used NMR spectroscopy to

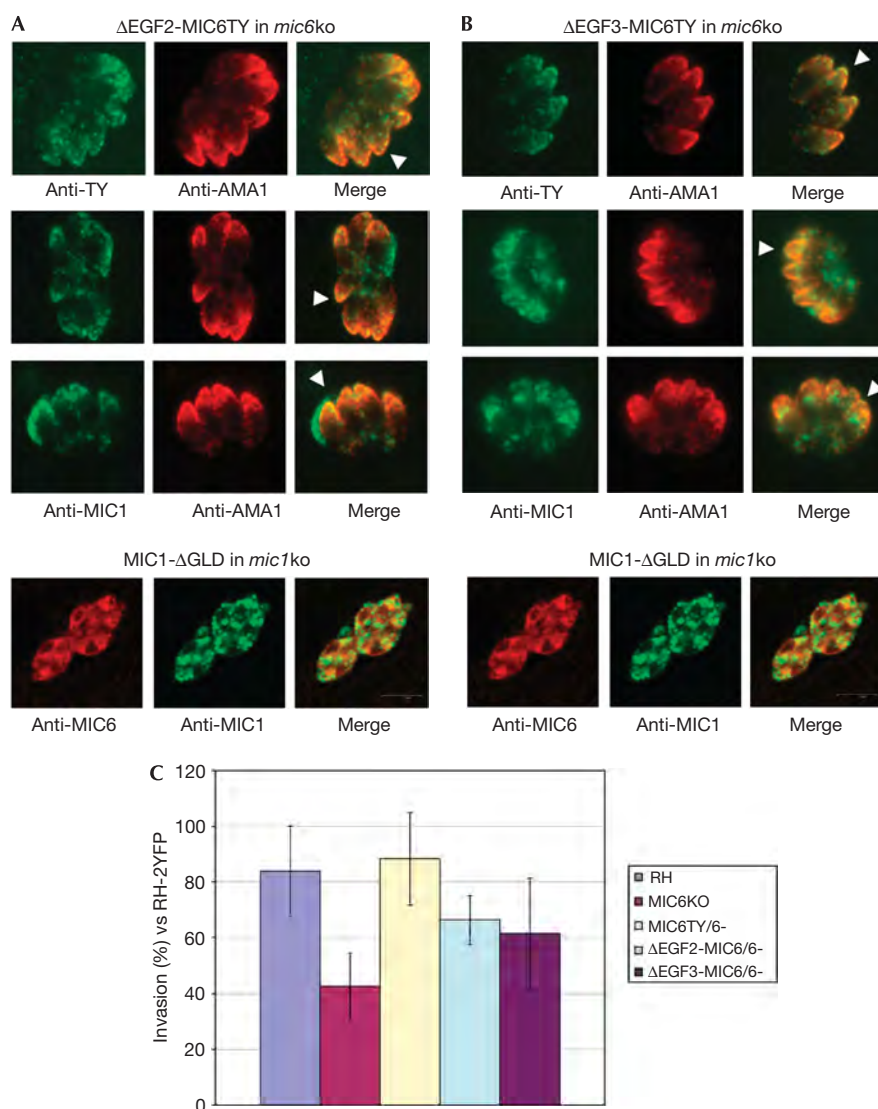


Fig 4 | Functional characterization of the *mic6* knock out strain and *mic6* knock out parasites complemented with either full-length TgMIC6 or Δ EGF2-MIC6TY and Δ EGF3-MIC6TY mutant proteins. Immunofluorescence assays showing that (A) Δ EGF2-MIC6TY and (B) Δ EGF3-MIC6TY are stably expressed and correctly targeted to the micronemes (anti-TY; green). Transport of TgMIC1 to the micronemes is partly restored in these parasites (anti-MIC1; green), as shown by colocalization with the microneme marker AMA1 (anti-AMA1; red). White arrowheads indicate examples of correct targeting to the micronemes. Note that Δ EGF2-MIC6TY is TY-tagged. A control assay is shown (bottom in A,B) in which MIC1- Δ GLD is unable to interact with either EGF domain from MIC6 and rescue correct microneme localization. TgMIC6 and TgMIC1 are retained in the endoplasmic reticulum/Golgi and dense granules, respectively. A green/magenta version of this figure is shown as supplementary Fig 4 online. (C) Functional assay comparing various parasite strains for their invasion efficiency using RH-2YFP parasites as an internal standard for parasite fitness and human foreskin fibroblasts as host cells. Invasion data were compared by one-way ANOVA followed by a Newman-Keuls test. Invasion data for Δ EGF2-MIC6/6 and Δ EGF3-MIC6/6 parasites are statistically different from MIC6KO and RH-2YFP parasites, supporting the partial complementation of invasion activity in these mutants. ANOVA, analysis of variance; EGF, epidermal growth factor-like; GLD, galectin-like domain; ko, knock out; MIC, microneme protein; RH-2YFP, RH strain of *Toxoplasma gondii* transformed with a plasmid expressing yellow fluorescent protein.

establish whether TgMIC1-GLD binds to TgMIC6-EGF3 in the same mode as EGF2. Most of the TgMIC1 amides (>85%) experiencing chemical shift changes in the EGF2 complex are also perturbed when TgMIC1 is bound to EGF3 (Fig 3A). Furthermore, the amides showing the largest chemical shift changes, indicative of principal changes in the chemical environment as a result of binding, are of comparable magnitude in both complexes

(supplementary Table II online). We also recorded an isotope ^{13}C -filtered/edited nuclear Overhauser enhancement spectroscopy (NOESY) spectrum on a hybrid ^{13}C - ^{15}N -labelled sample and confirmed several predicted intermolecular nuclear Overhauser effects (NOEs) based on the TgMIC1-GLD-TgMIC6-EGF2 complex (supplementary Fig 3 online). Therefore, we conclude that the modes of recognition of the two EGF domains by TgMIC1

are analogous, consistent with a sequence alignment that shows a conservation of crucial interfacial residues between the two EGF domains (20% identity between TgEGF2 and TgEGF3; Fig 3B).

EGF domains are often found in cell surface proteins where they are involved in protein–protein interactions that promote intercellular signalling, often in multivalent interactions to increase specificity. EGF domains are also present in other apicomplexan MICs and in several malarial surface proteins (Anantharaman *et al*, 2007). A notable example includes a homologue of TgMIC6 identified in the *Neospora caninum* genome sequence. The alignment of the EGF domains from these two sequences, together with the alignment of galectin-like domains from TgMIC1 and NcMIC1, shows a high degree of conservation at positions identified as crucial interacting residues in the MIC1–MIC6 complex, suggesting that an equivalent complex is assembled in *N. caninum* (Fig 3B). Further examples containing tandem repeats of EGF domains include TgMIC3, TgMIC7, TgMIC8 and TgMIC9 (Meissner *et al*, 2002b), MIC4 from *Eimeria*, *Plasmodium falciparum* Pf25, which is expressed in zygotes and ookinetes, and the merozoite surface protein MSP1. Our structure for the TgMIC1–GLD–TgMIC6–EGF2 complex provides the first structural insight into the recognition of the EGF domain at the surface of apicomplexans and might represent a generalized mode of interaction, in which the first β -hairpin of the EGF domain extends an intermolecular β -sheet and the second β -hairpin clamps the surface. It is conceivable that β -sheet-rich domains other than GLDs interact with EGF domains in a similar manner.

EGF2 and EGF3 promote host cell receptor recognition

Our observation of two independent TgMIC1 interactions within TgMIC6 has profound consequences for the recognition of host cell receptors, as TgMIC1 is able to discriminate oligosaccharides in a highly specific manner (Blumenschein *et al*, 2007). To determine the relative importance of the twin TgMIC1-binding sites within TgMIC6, transport of the components of the complex was analysed by immunofluorescence assay in a *mic6* knock out (ko) strain, as well as in *mic6*ko parasites complemented with constructs expressing full-length TgMIC6, TgMIC6- Δ EGF3 or TgMIC6- Δ EGF2. It has been shown previously that, in the absence of TgMIC6, TgMIC1 is mistargeted to the dense granules but that the expression of full-length TgMIC6 in the *mic6*ko background fully rescues the correct targeting of TgMIC1 to the micronemes (Reiss *et al*, 2001; Saouros *et al*, 2005a). Complementation of the *mic6*ko with either TgMIC6- Δ EGF3 or TgMIC6- Δ EGF2 also restores the transport of TgMIC1 to the micronemes (Fig 4A,B), suggesting that both TgMIC6 EGF domains are able to interact independently with TgMIC1 within the parasite. It has recently been shown for TgMIC3 that the minimal requirement for microneme delivery is the presence of its pro-peptide together with any one of the adjacent EGF domains (El Hajj *et al*, 2008). In other studies, it has been suggested that short α -helical segments are important for secretory organelle sorting (Dikeakos *et al*, 2007). Interestingly, our structural data show the presence of an α -helix within the amino terminus of TgMIC6-EGF2, which is immediately adjacent to the predicted TgMIC6-EGF1 cleavage site (Fig 1A); therefore, the subsequent exposure of the helix by the proteolytic removal of TgMIC6-EGF1 could reflect a biological

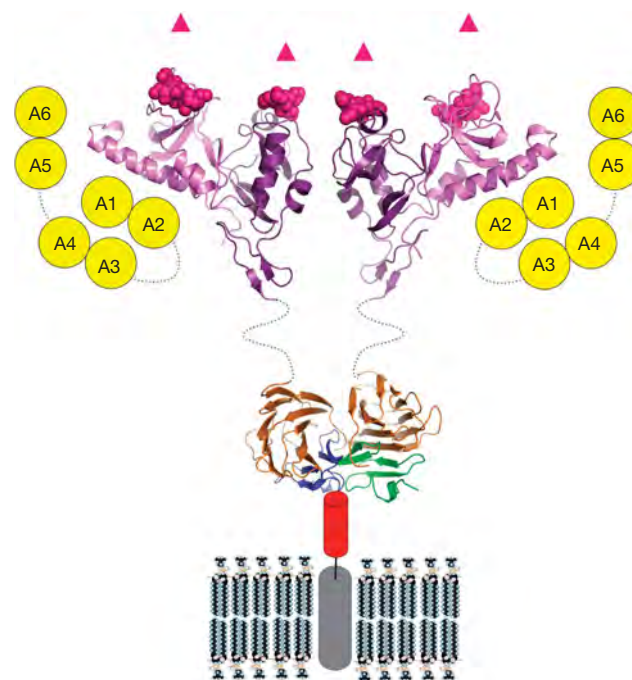


Fig 5 | A new model for the architecture of the TgMIC4–1–6 complex. A schematic representation of a model for the TgMIC4–1–6 complex illustrating the relative location of the domains and host cell-binding sites. The sialic acid-containing receptor(s) are shown as pink space-filling spheres. Colour code: yellow, apple domains; purple, microneme adhesive repeat (MAR) domain; orange, galectin-like domain (GLD); green and blue, second and third epidermal growth factor-like (EGF) domains; grey, transmembrane (TM); and red cylinder, acidic region. The six apple domains of TgMIC4 are represented as yellow circles, as no structural information is available. MIC, microneme protein.

signal that indicates the correct assembly of the complex and, in turn, allows correct targeting.

To determine whether the mutant complexes in the TgMIC6- Δ EGF3 and TgMIC6- Δ EGF2 parasites are functional in invasion, we characterized the relative invasion efficiency of the *mic6*ko and the various complemented strains. The host cell invasion efficiency of the *mic6*ko strain was reduced to around 50% of the level of the parental RH strain (Fig 4C). Invasion efficiency was partly restored when the *mic6*ko strain was complemented with TgMIC6, TgMIC6- Δ EGF3 or TgMIC6- Δ EGF2, providing evidence that both EGF domains allow the presentation of TgMIC1 and recognition of host cell receptors for invasion (Fig 4C). Microneme targeting and subsequent invasion by the mutant parasites are not as efficient as with full-length TgMIC6, which most likely reflects the perturbed stoichiometry of the complex (Fig 4).

Speculation

Fig 5 shows a model of the relative arrangement of the two EGF domains based on a structural alignment with the tandem EGF domains from MSP1. Both TgMIC1-binding sites lie on opposite sides of the complex and are available for interaction, which projects the host cell-binding MAR domains of TgMIC1 and the associated TgMIC4 into the extracellular milieu. Previous results

from carbohydrate microarray experiments have shown that the most potent binders of TgMIC1 were branched carbohydrates having two or more terminal sialic acids, which raised the possibility that tandem MAR regions provide highly specialized ligands for recognizing bidentate receptors (Blumenschein *et al*, 2007). Our new discovery extends this model for TgMIC4–1–6 complex function in that four receptor-binding sites could be presented through two branches of TgMIC1 molecules bound to one TgMIC6. This subunit multivalency is further reminiscent of the erythrocyte-binding antigen (EBA-175) from *P. falciparum* (Blumenschein *et al*, 2007; Hager & Carruthers, 2008), which is dimeric and comprises tandem Duffy-binding ligand domains that recognize several sialyl glycans (Tolia *et al*, 2005).

METHODS

Recombinant TgMIC1-GLD, TgMIC6-EGF2 (residues 87–147) and TgMIC6-EGF3 (residues 141–209) were expressed and purified as described previously (Saouros *et al*, 2005a,b).

T. gondii *mic6ko* parasites, as well as the *mic6ko*-complemented stable parasite strains MIC6TY, MIC6-ΔEGF2 and MIC6-ΔEGF3, were obtained as described previously (Reiss *et al*, 2001). Indirect immunofluorescence assays were used to probe the transport of TgMIC6 and TgMIC1. Comparison of the various *T. gondii* strains for invasion efficiency was carried out using the RH strain transformed with a plasmid expressing yellow fluorescent protein (RH-2YFP) as the internal standard.

The interaction between TgMIC1-GLD and either TgMIC6-EGF2 or TgMIC6-EGF3 was probed using two-dimensional ¹H–¹⁵N heteronuclear single quantum coherence (HSQC) NMR mapping experiments, gel filtration chromatography and isothermal titration calorimetry.

For the structural determination, NMR spectra on the complex were recorded on hybrid labelled samples of ¹⁵N, ¹³C- or ¹⁴N, ¹²C-TgMIC1-GLD (residues 320–456 in the full-length TgMIC1 sequence) and ¹⁴N, ¹²C- or ¹⁵N, ¹³C-TgMIC6-EGF2 (residues 87–147 in the full-length TgMIC6 sequence). Backbone and side-chain assignments were completed using standard methodology (Sattler *et al*, 1999). Intermolecular NOE identification was aided by filtered (¹²C,¹⁴N)H-NOESY–¹³C-HSQC experiments (Zwahlen *et al*, 1997). Coordinates for the NMR structures of unbound TgMIC6-EGF2 and the TgMIC1-GLD–TgMIC6-EGF2 complex have been deposited at the Protein Databank under the accession codes 2K2T and 2K2S. For detailed descriptions, see the supplementary information online.

Supplementary information is available at *EMBO reports* online (<http://www.emboreports.org>).

ACKNOWLEDGEMENTS

This paper is dedicated to our friend and colleague Laurent Bonomo, who died in July 2008. This study was supported by the Medical Research Council (MRC grant numbers G0400423 and G0800038, S.M.), the Biotechnology and Biological Sciences Research Council (grant number E02520X, S.J.M.) and Swiss National Foundation (31-116722, D.S.-F.). This study is part of the activities of the BioMalPar European Network of Excellence supported by a European grant (LSHP-CT-2004-503578). K.S. identified the TgMIC1–TgMIC6-EGF2 interaction and carried out all MIC6-related structural work; S.S. carried out all MIC1-related structural work; and N.F. carried out the localization and invasion assays.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- Anantharaman V, Iyer LM, Balaji S, Aravind L (2007) Adhesion molecules and other secreted host-interaction determinants in apicomplexa: insights from comparative genomics. *Int Rev Cytol* **262**: 1–74
- Blumenschein TM *et al* (2007) Atomic resolution insight into host cell recognition by *Toxoplasma gondii*. *EMBO J* **26**: 2808–2820
- Brecht S, Carruthers VB, Ferguson DJP, Giddings OK, Wang G, Jakle U, Harper JM, Sibley LD, Soldati D (2001) The *Toxoplasma* micronemal protein MIC4 is an adhesin composed of six conserved apple domains. *J Biol Chem* **276**: 4119–4127
- Carruthers VB, Giddings OK, Sibley LD (1999) Secretion of micronemal proteins is associated with *Toxoplasma* invasion of host cells. *Cell Microbiol* **1**: 225–235
- Dikeakos JD, Lacombe M-J, Mercure C, Mireuta M, Reudelhuber TL (2007) A hydrophobic patch in a charged α -helix is sufficient to target proteins to dense core secretory granules. *J Biol Chem* **282**: 1136–1143
- Dubey JP (2004) Toxoplasmosis—a waterborne zoonosis. *Vet Parasitol* **126**: 57–72
- El Hajj H, Papoin J, C  r  de O, Garcia-R  guet N, So  te M, Dubremetz JF, Lebrun M (2008) Molecular signals in the trafficking of the *Toxoplasma gondii* protein MIC3 to the micronemes. *Eukaryot Cell* **7**: 1019–1028
- Fusco G, Rinaldi L, Guarino A, Proroga YTR, Pesce A, Giuseppina DM, Cringoli G (2007) *Toxoplasma gondii* in sheep from the Campania region (Italy). *Vet Parasitol* **149**: 271–274
- Hager KM, Carruthers VB (2008) MARveling at parasite invasion. *Trends Parasitol* **24**: 51–54
- Hill D, Dubey JP (2002) *Toxoplasma gondii*: transmission, diagnosis and prevention. *Clin Microbiol Infect* **8**: 634–640
- 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
- Laskowski RA (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* **13**: 323–330
- Meissner M, Schluter D, Soldati D (2002a) Role of *Toxoplasma gondii* myosin A in powering parasite gliding and host cell invasion. *Science* **298**: 837–840
- Meissner M, Reiss M, Viebig N, Carruthers VB, Torsel C, Tomavo S, Ajioka JW, Soldati D (2002b) A family of transmembrane microneme proteins of *Toxoplasma gondii* contain EGF-like domains and function as escorts. *J Cell Sci* **115**: 563–574
- 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
- Saouros S *et al* (2005a) 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
- Saouros S, Chen HA, Simpson P, Cota E, Edwards-Jones B, Soldati-Favre D, Matthews S (2005b) Complete resonance assignments of the C-terminal domain from MIC1: a micronemal protein from *Toxoplasma gondii*. *J Biomol NMR* **31**: 177–178
- Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog NMR Spectrosc* **34**: 93–158
- Soldati D, Meissner M (2004) *Toxoplasma* as a novel system for motility. *Curr Opin Cell Biol* **16**: 32–40
- Tolia NH, Enemark EJ, Sim BKL, Joshua-Tor L (2005) Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. *Cell* **122**: 183–193
- Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex. *J Am Chem Soc* **119**: 6711–6721

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

Lectins display variable ligand specificities (Varki, 1997). Discrimination between ligands occurs by type and conformation of the sugar and by sugar modifications such as acetylation and sulphation. It may also involve the linkage to the underlying sugar chain and its composition. Additional contacts to the protein moiety of a glycoconjugate may further enhance the specificity of the interaction.

Carbohydrate microarray analyses revealed a preference of TgMIC1-MARR for 2-3 linked sialyl-oligosaccharides (Blumenschein et al, 2007). The molecular and structural details that confer this specificity are analysed in this study. Dose-response assays confirmed the preference of TgMIC1-MARR for 2-3 linked over 2-6 linked sialyl-oligosaccharides. The comparison of crystal structure complexes of TgMIC1-MARR with several oligosaccharides comprising 2-3 or 2-6 linkages identified the amino acid residues involved in contacts with the sugar and demonstrated the importance of a glutamate residue for the preference of 2-3 linked sialyl-oligosaccharides. In addition two fluorinated analogues were found to be highly potent ligands. The personal contribution to this study is restricted to a cell binding competition assay using one of the fluorinated analogues to demonstrate that this compound is able to compete for TgMIC1 binding with endogenous host cell receptors.

The following article is reprinted with permission of John Wiley & Sons, Inc.

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*}

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

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

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

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

⁵Division 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'SiaLacNAc₁₋₄ and 3'SiaLacNAc₁₋₃ 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*-acetylglucosamine 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 schizophrenia⁴ 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 presen-

tation 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 Gal β 1–4 linkage known as Type 2 backbone¹⁶; 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.

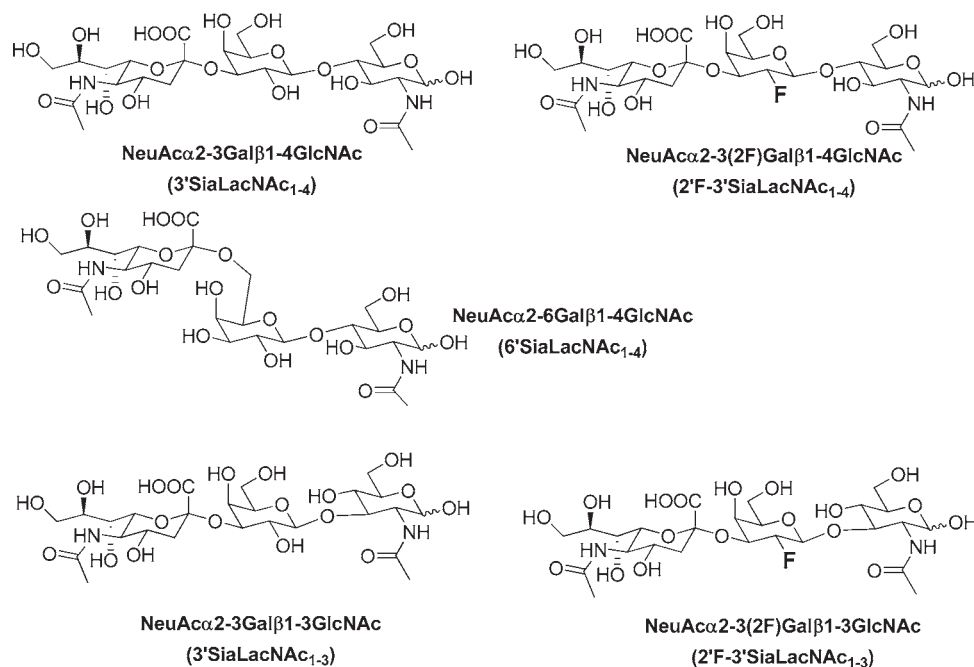
Results

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 (β 1–3 or β 1–4). Sialic acid was introduced using tolerant trans-sialidase from *T. cruzi* to create the 3'-sialyl 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'SiaLacNAc₁₋₃ isomer, as well as to the fluorinated probes, 2'F-3'SiaLacNAc₁₋₄ and 2'F-3'SiaLacNAc₁₋₃ (Scheme 1 and Table I), we generated microarrays in a dose-response format [Fig. 1(a,c)]. The two nonsialylated analogs LacNAc₁₋₄ and LacNAc₁₋₃ 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 LacNAc₁₋₄ and LacNAc₁₋₃ [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'SiaLacNAc₁₋₄ taken as 1.0 (Table I) to indicate binding potencies.

The relative binding score of 3'SiaLacNAc₁₋₃, 1.2, was comparable to that of 3'SiaLacNAc₁₋₄, whereas that of 6'SiaLacNAc₁₋₄ was considerably lower, 0.3. Strikingly, the strongest binding signals, 4.1 and 3.3, were with the fluorinated compounds 2'F-3'SiaLacNAc₁₋₄ and 2'F-3'SiaLacNAc₁₋₃, 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'SiaLacNAc₁₋₄ and 3'SiaLacNAc₁₋₃ [Fig.

1(b,d) and Table I], and a clear discrimination between 3'SiaLacNAc₁₋₄ and 6'SiaLacNAc₁₋₄ (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, 2'F-3'SiaLacNAc₁₋₄ and 2'F-3'SiaLacNAc₁₋₃ (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'SiaLacNAc₁₋₄. An unrelated neutral trisaccharide, maltotriose (Glcα1-4Glcα1-4Glc) was included as a negative control. We observed that 3'SiaLacNAc₁₋₄ inhibited the high avidity binding to the immobilized fluorinated trisaccharide 2'F-3'SiaLacNAc₁₋₄, 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'SiaLacNAc₁₋₄ was less active: there was no inhibition detected at 1 mg/mL; and ~ 70% inhibition at 3 mg/mL [Fig. 2(a)]. The binding of WGA to the immobilized 3'SiaLacNAc₁₋₄ probe was also inhibited more strongly by 3'SiaLacNAc₁₋₄ than by 6'SiaLacNAc₁₋₄ [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'SiaLacNAc₁₋₄. Indeed, effective inhibition of cell binding by TgMIC1 was achieved in 10–500 μM range of 2'F-3'SiaLacNAc₁₋₄ [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

Name	Sequence	Binding signals ^a at 2 fmol/spot			
		TgMIC1-MARR		WGA	
		Score	Ratio	Score	Ratio
LacNAc (a) ^b	Galβ1-4GlcNAc	0	0	514	<0.1
LacNAc ₁₋₃ (b)	Galβ1-3GlcNAc	0	0	0	0
3'SiaLacNAc ₁₋₄ (c)	NeuAcα2-3Galβ1-4GlcNAc	3,488	1.0	15,838	1.0
6'SiaLacNAc ₁₋₄ (d)	NeuAcα2-6Galβ1-4GlcNAc	1,160	0.3	1,696	0.1
3'SiaLacNAc ₁₋₃ (e)	NeuAcα2-3Galβ1-3GlcNAc	4,203	1.2	17,613	1.1
2'F-3'SiaLacNAc ₁₋₄ (f)	NeuAcα2-3(2F)Galβ1-4GlcNAc	14,287	4.1	7,211	0.5
2'F-3'SiaLacNAc ₁₋₃ (g)	NeuAcα2-3(2F)Galβ1-3GlcNAc	11,486	3.3	5,669	0.4
Hematoside (h)	NeuAcα2-3Galβ1-4Glcβ-Cer	2,796	0.8	2,489	0.2
Sial pg	NeuAcα2-3Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ-Cer	1,373	0.4	2,723	0.2
GD1a	NeuAcα2-3Galβ1-3GalNAcβ1-4Galβ1-4Glcβ-Cer NeuAcα2-3	3,387	1.0	778	<0.1
GD2	GalNAcβ1-4Galβ1-4Glcβ-Cer NeuAcα2-8NeuAcα2-3	0	0	504	<0.1
GD3	NeuAcα2-8NeuAcα2-3Galβ1-4Glcβ-Cer	26	<0.1	816	0.1
GT1b	NeuAcα2-3Galβ1-3GalNAcβ1-4Galβ1-4Glcβ-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'SiaLacNAc₁₋₄ 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-3'SiaLacNAc₁₋₄, 3'SiaLacNAc₁₋₃, or 2'F-3'SiaLacNAc₁₋₃ 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'SiaLacNAc₁₋₃ and 2'F-3'SiaLacNAc₁₋₄ soaks) or modeled as polyalanine (2'F-3'SiaLacNAc₁₋₄ soak). Although it was possible to model all of the carbohydrate chain in the 3'SiaLacNAc₁₋₃ and 2'F-3'SiaLacNAc₁₋₃ soaks, electron density for the GlcNAc residue of 2'F-3'SiaLacNAc₁₋₄ 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'SiaLacNAc₁₋₄ (pdb:2Jhd), TgMIC1-MARR-6'SiaLacNAc₁₋₄ (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

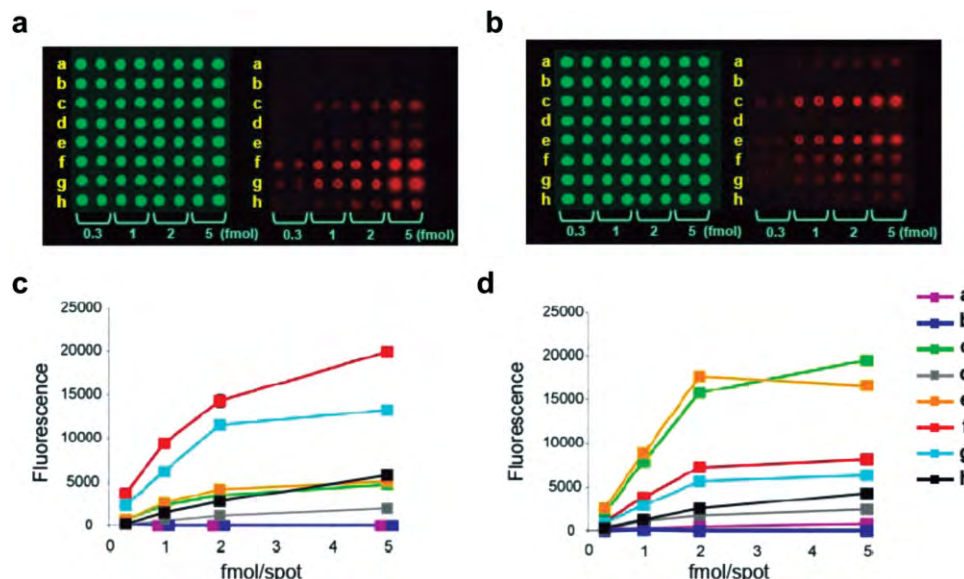


Figure 1. Microarray analyses of TgMIC1-MARR and WGA in a dose response format. Eight lipid-linked oligosaccharide probes, a–h, (sequences in Table I) were printed in duplicate at the indicated levels on nitrocellulose-coated glass slides with Cy3 dye included as a marker (green emission). Binding was detected with Alexa Fluor647-labeled streptavidin (red emission). Images for TgMIC1-MARR (40 $\mu\text{g/mL}$) and WGA (1 $\mu\text{g/mL}$) 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'SiaLacNAc₁₋₄ 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'SiaLacNAc₁₋₄ 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'SiaLacNAc₁₋₃. 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'SiaLacNAc₁₋₃ 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'SiaLacNAc₁₋₄, 2'F-3'SiaLacNAc₁₋₄ and 6'SiaLacNAc₁₋₄ bound structures,¹³ these additional

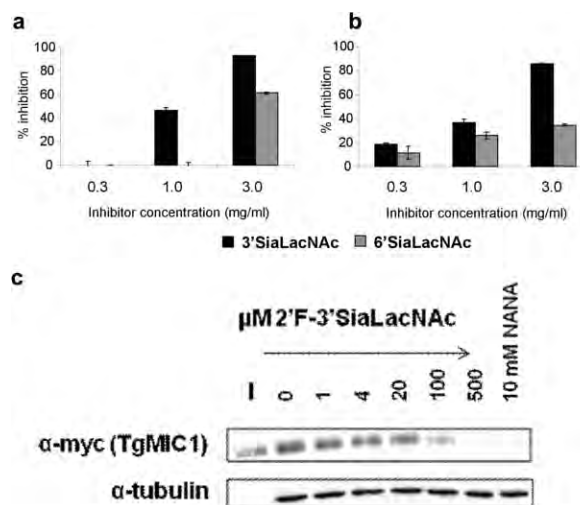


Figure 2. Comparison of the inhibitory activities of free oligosaccharides toward the binding of TgMIC1-MARR. Oligosaccharides 3'SiaLacNAc₁₋₄, 6'SiaLacNAc₁₋₄, and maltotriose (a negative control) were tested at the indicated concentrations for inhibition of binding of TgMIC1-MARR at 4 $\mu\text{g/mL}$, to immobilized 2'F-3'SiaLacNAc₁₋₄ probe (a), and of WGA at 0.5 $\mu\text{g/mL}$ to immobilized 3'SiaLacNAc₁₋₄ probe (b). The results are means of two separate inhibition experiments with error bars. Cell binding competition experiments with soluble 2'F-3'SiaLacNAc₁₋₄ 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'SiaLacNAc₁₋₄ 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	<i>P</i> _{4₃} 2 ₁ 2	<i>P</i> _{4₃} 2 ₁ 2	<i>P</i> _{4₃} 2 ₁ 2
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
<i>R</i> _{sym}	0.075 (0.246)	0.074 (0.2228)	0.076 (0.222)
<i>I</i> / <i>σI</i>	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			
<i>R</i> _{work} / <i>R</i> _{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.

$R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$ where *I* is the integrated intensity of a given reflection and *I* is the mean intensity of multiple corresponding symmetry-related reflections.

$R_{\text{work}} = \sum ||F_o| - |F_c|| / \sum F_o$ 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′SiaLacNAc₁₋₄ relative to 6′SiaLacNAc₁₋₄ is observed in the microarray analyses. The structure of 6′SiaLacNAc₁₋₄ 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 SiaLacNAc₁₋₄ analogs tested [Fig. 4(a–d)].

Fluorination of 3′SiaLacNAc₁₋₃ 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′SiaLacNAc₁₋₄ 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

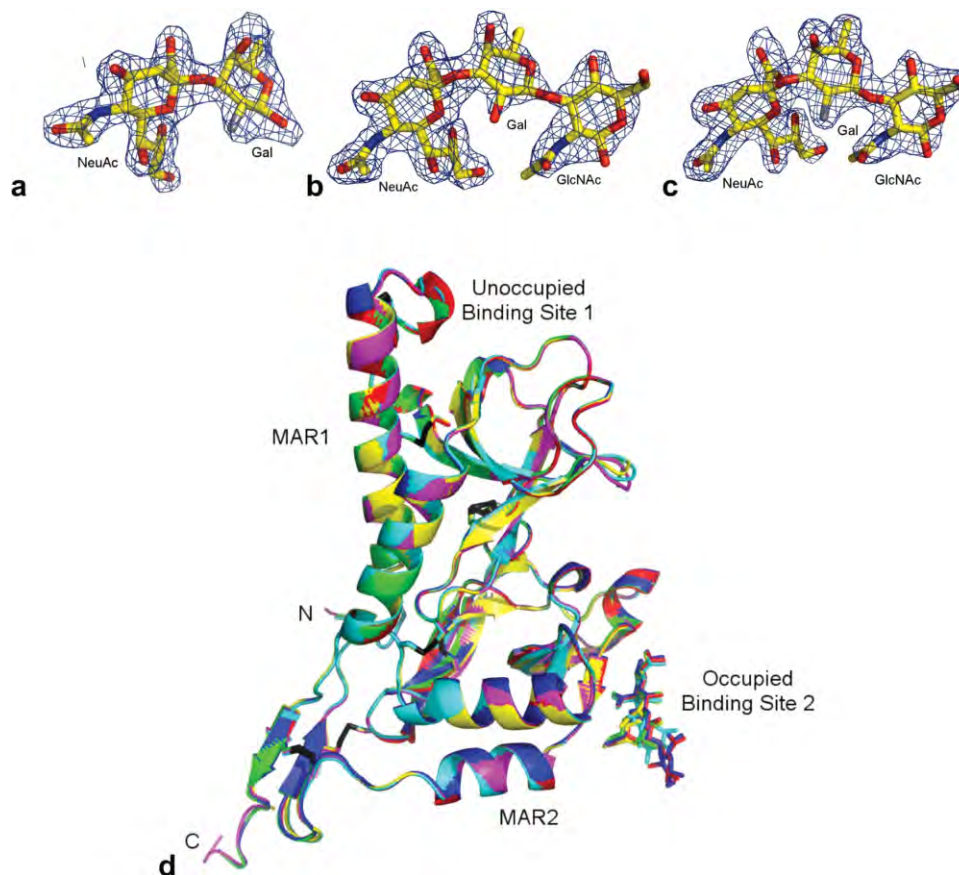


Figure 3. Sialyl trisaccharides bound at the TgMIC1-MAR2 host cell adhesion site. Weighted $2|F_o| - |F_c|$ electron density maps of (a) 2'-F-3'SiaLacNAc₁₋₄, (b) 3'SiaLacNAc₁₋₃ and (c) 2'-F-3'SiaLacNAc₁₋₃, contoured at 1.0 r.m.s. (d) Overlay of all six TgMIC1-MARR crystal structures: no ligand (magenta; pdb:2Jh1,²³ 3'SiaLacNAc₁₋₄ soak (yellow; pdb:2Jhd), 2'-F-3'SiaLacNAc₁₋₄ soak (green; pdb:3f53), 3'SiaLacNAc₁₋₃ soak (red; pdb:3f5a), 2'-F-3'SiaLacNAc₁₋₃ soak (blue; pdb:3f5e) and 6'SiaLacNAc₁₋₄ 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 sialylated 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 O9	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	—

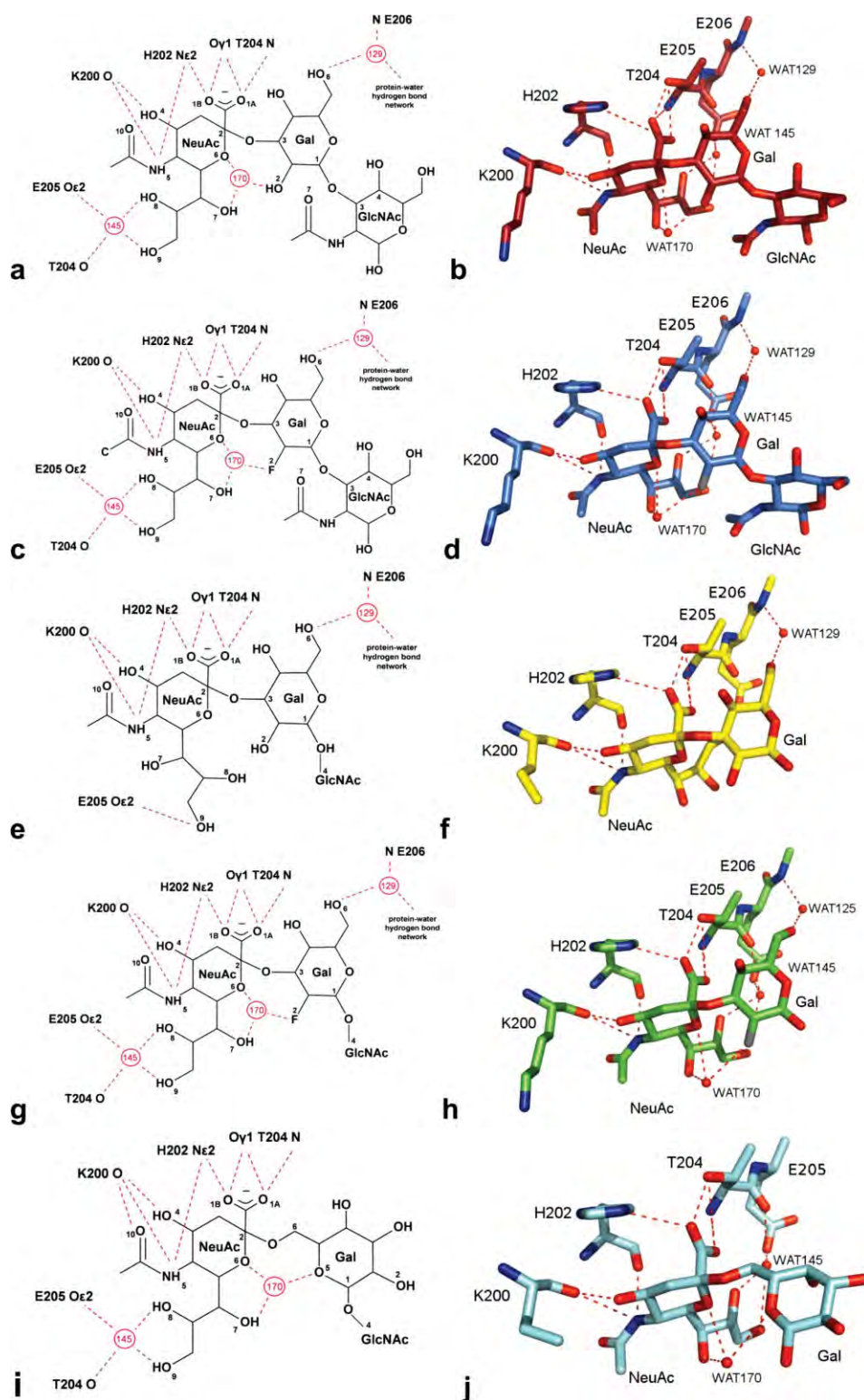


Figure 4. The occupied binding site of TgMIC1-MAR2. (a, b) 3'SiaLacNAC₁₋₃ soak, (c, d) 2'F-3'SiaLacNAC₁₋₃ soak, (e, f) 3'SiaLacNAC₁₋₄ soak, (g, h) 2'F-3'SiaLacNAC₁₋₄ soak, and (i, j) 6'SiaLacNAC₁₋₄ soak. (a, c, e, g, i) Schematic representation of the occupied TgMIC1-MAR2 binding site with hydrogen bonds 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'SiaLacNAC₁₋₃, 3'SiaLacNAC₁₋₃ and 2'F-3'SiaLacNAC₁₋₃ soak structures are renamed as 129, 145, and 170, respectively, for consistency with the 3'SiaLacNAC₁₋₄ and 6'SiaLacNAC₁₋₄ 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.]

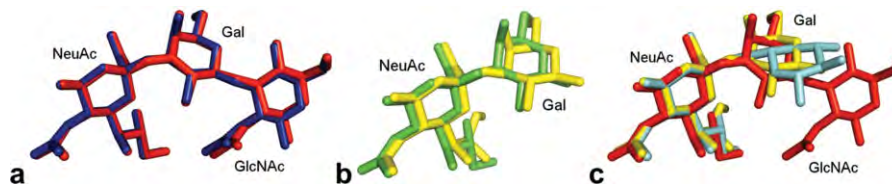


Figure 5. Overlay of sialyl oligosaccharides. (a) Comparison of fluorinated (blue; pdb:3f5e) and nonfluorinated (red; pdb:3f5a) 3'SiaLacNAc₁₋₃. (b) Superimposition of fluorinated (green; pdb:3f53) and nonfluorinated (yellow; pdb:2Jhd) 3'SiaLacNAc₁₋₄. The glycerol side chain of sialic acid is in two separate conformations. (c) Overlay of 3'SiaLacNAc₁₋₃ (red), 3'SiaLacNAc₁₋₄ (yellow)¹⁵ and 6'SiaLacNAc₁₋₄ (cyan; pdb:2Jh7). Deviations are seen within the alternate conformation of the glycerol side chain of sialic acid in 3'SiaLacNAc₁₋₄ and within galactose, where due to the β 1–4 linkage in 6'SiaLacNAc₁₋₄, 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 proteins revealed numerous lectins including *Maackia amurensis* leucoagglutinin (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'SiaLacNAc₁₋₃ (red) bound to TgMIC1-MARR.

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ensis leucoagglutinin²⁹ (green) (pdb:1DBN), mouse sialoadhesin³⁰ (orange) (pdb:1QFO), Staphylococcal enterotoxin B³¹ (blue) (pdb:1SE3), and human Siglec-5³² (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

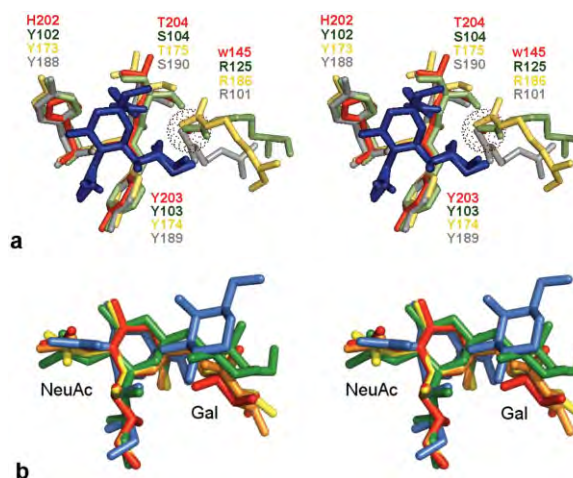


Figure 6. (a) Overlay of a common binding motif for recognition of sialic acid. Stereo image showing superimposition of sialic acid (Neu5Ac α 2Me in pdb:2i2s) within the sialic acid binding site of TgMIC1-MARR (red; pdb:3f5a), *Bordetella pertussis* toxin (green; pdb:1pto), *Staphylococcus aureus* SSL5 toxin (sand; pdb:3f5a) and Rotavirus VP8* (gray; pdb:2i2s). Amino acid residues, sialic acid (blue) and Neu5Ac α 2Me (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* leucoagglutinin (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'SiaLacNAc₁₋₃ (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'SiaLacNAc₁₋₄. 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'SiaLacNAc₁₋₃ and 6'SiaLacNAc₁₋₄ 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'SiaLacNAc₁₋₄ 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

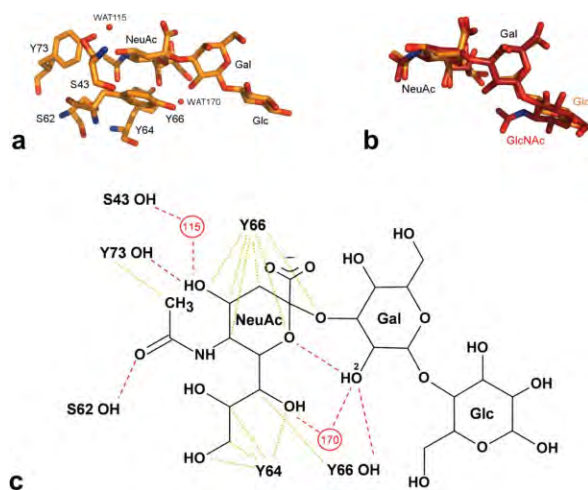


Figure 7. Evaluation of the binding modes of sialyl trisaccharides to WGA and TgMIC1-MAR2. (a) WGA bound to 3'SiaLac²⁰ (orange; pdb:1wgc) and (b) overlay of 3'SiaLacNac₁₋₃ (red; pdb:3f5a) and 3'SiaLac. Amino acid residues and ligands are shown as sticks while water molecules are represented as spheres (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'SiaLacNac₁₋₃ 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'SiaLacNac₁₋₃ is that galactose is β 1-4 linked to glucose (Glc) rather than β 1-3 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 fluori-

nated 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'SiaLacNac₁₋₄ were from Dextra (Reading, UK). LacNac₁₋₃, 3'SiaLacNac₁₋₃ and the fluorinated analogs of 3'SiaLacNac₁₋₄ and 3'SiaLacNac₁₋₃ designated 2'F-3'SiaLacNac₁₋₄ and 2'F-3'SiaLacNac₁₋₃, 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 neoglycolipids¹⁹ 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 μ g/mL and WGA at 0.5 μ g/mL were carried out in the presence of free oligosaccharides 3'SiaLacNac₁₋₄ and 6'SiaLacNac₁₋₄, 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'SiaLacNAc₁₋₄ printed at 5 fmol/spot in the microarray: % Inhibition = [(means of duplicate spots with maltotriose – means of duplicate spots with 3' or 6'SiaLacNAc₁₋₄)/means of duplicate spots with maltotriose] × 100%.

Competition cell binding assay

These were performed as described previously¹³ using 2'F-3'SiaLacNAc₁₋₄ 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 mM bis-tris propane pH 7.0, and either 3.0 mM of 2'F-3'SiaLacNAc₁₋₄, 3'SiaLacNAc₁₋₃ or 2'F-3'SiaLacNAc₁₋₃. Before data collection, crystals were soaked in cryoprotectant¹³ 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:2jhi¹³) 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'SiaLacNAc₁₋₄, 3'SiaLacNAc₁₋₃ and 2'F-3'SiaLacNAc₁₋₃ 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, Campanero-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 Å-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 glycoporphin-A. *J Mol Biol* 194:353–355.
22. Biffinger JC, Kim HW, DiMaggio 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 Å 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 Commun* 4384–4386.
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 CCP4 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.

3.4 A novel protein family containing MAR domains act as sialic acid lectins during Apicomplexa invasion

The results presented in section 3.1. demonstrated that sialic acid is a major determinant of host cell invasion by *T. gondii*, but the parasite ligands important for sialic acid-dependent invasion remained undetermined. Since we showed that TgMIC1 specifically recognizes sialylated oligosaccharides on the host cell surface, this protein is an obvious candidate. A previously generated TgMIC1 knock-out strain (*mic1ko*) showed a 50% reduction in invasion efficiency compared to the wild-type strain (Cerede et al, 2005). In this strain the TgMIC1-4-6 complex is disrupted, ablating transport of TgMIC4 and TgMIC6 to the micronemes and instead resulting in their retention in the early secretory pathway (Reiss et al, 2001). In consequence, the contribution of the individual components to host cell invasion remained unresolved. In this study we have addressed this question and investigated on further candidate parasite proteins involved in sialic acid-dependent host cell invasion. This led us to the identification of a novel protein family defined by the presence of the MAR domain and the characterization of TgMIC13 and NcMIC1, the TgMIC1 homologue in *N. caninum*. Carbohydrate microarray analyses and structural modeling were employed for detailed analysis of binding specificities of TgMIC1, TgMIC13 and NcMIC1. These experiments were performed by the groups of Prof. Stephen Matthews and Prof. Ten Feizi (both from Imperial College, London). More details on the individual contributions of the authors can be found in the acknowledgements of the article.

Note added before final printing of this thesis: The final version of the following article is now available online at <http://www.jbc.org/cgi/doi/10.1074/jbc.M109.060988>

A novel protein family containing MAR domains act as sialic acid lectins during Apicomplexa invasion

Nikolas Friedrich¹, Joana M. Santos^{1*}, Yan Liu^{2*}, Angelina S. Palma^{2§}, Ester Leon³, Savvas Saouros³, Makoto Kiso⁴, Michael J. Blackman⁵, Stephen Matthews³, Ten Feizi^{2**} and Dominique Soldati-Favre^{1**}

*These authors contributed equally.

** **Corresponding authors** t.feizi@imperial.ac.uk and dominique.soldati-favre@unige.ch

¹Department of Microbiology and Molecular Medicine, Faculty of Medicine, University of Geneva CMU, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland.

²Glycosciences Laboratory, Division of Medicine, Imperial College London, Northwick Park Campus, Harrow, HA1 3UJ, UK.

³Division of Molecular Biosciences, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK.

⁴Department of Applied Bioorganic Chemistry, Gifu University, Gifu 501-11, Japan

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

[§]On leave from REQUIMTE, Centro de Química Fina e Biotecnologia, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal.

Running Title: A family of MAR domain proteins in Apicomplexa

Character count 59285

Key words: *Toxoplasma gondii* / micronemal proteins / carbohydrate microarray / sialic acid

Abbreviations: Abs: antibodies, ESA: Excreted secreted antigens fraction, GLD: galectin-like domain, GRAs: dense granule proteins, HFFs: human foreskin fibroblasts, MAR: microneme adhesive repeat, MARR: MAR region, MCP: MAR containing protein, MICs: microneme proteins, NANA: N-acetylneuraminic acid

Abstract

Numerous intracellular pathogens exploit cell surface glycoconjugates for host cell recognition and entry. Unlike bacteria and viruses, *Toxoplasma gondii* and other apicomplexans actively invade host cells. Micronemal adhesins (MICs) play crucial roles in this process. We recently described a new structural module in *T. gondii* MIC1 (TgMIC1) termed Microneme Adhesive Repeat (MAR), which recognizes sialyl-oligosaccharides. Here, we establish TgMIC1 as one important player in sialic acid-dependent invasion and provide evidence that another sialic acid-binding lectin is involved. This led us to identify a family of MAR containing proteins in *T. gondii* and to characterize a novel sialic acid-binding lectin, TgMIC13. Binding characteristics of TgMIC13 were compared to TgMIC1 and its homologue *Neospora caninum* MIC1 by carbohydrate microarray analysis and molecular modelling. While all three share a preference for α 2-3-linked sialyl-oligosaccharides, binding specificities are distinct, and this might have implications for tissue tropism. Conservation of the MAR family among coccidians suggests that all these enteroparasites exploit sialylated glycoconjugates on the host cell surface as major determinants for a broad host cell range and enteric invasion.

Introduction

Sialic acids occur abundantly in glycoproteins and glycolipids on the cell surface, and are exploited by many viruses and bacteria for attachment and host cell entry. Recognition of carbohydrates and in particular sialylated glycoconjugates is important also for host cell invasion by the Apicomplexa (Blumenschein et al, 2007; Persson et al, 2008; Sim et al, 1990; Takabatake et al, 2007), a phylum which comprises several thousand species of obligate intracellular parasites. Among them the *Plasmodium* spp. are responsible for malaria, *Eimeria* spp. for coccidiosis in poultry, *Neospora* spp. for neosporosis in cattle, and *Toxoplasma* is the causative agent of toxoplasmosis in warm blooded animals and humans.

The host range and cell type specificity varies widely across the phylum. In contrast to *Plasmodium falciparum* merozoites that exclusively invade erythrocytes of humans and great apes (Martin et al, 2005), *T. gondii* rapidly establishes infection in the host by the fast-replicating tachyzoite form which disseminates into deep tissues and invades an extremely broad range of cell types in humans and virtually all warm blooded animals (Montoya & Liesenfeld, 2004). The molecular basis for this broad host cell recognition is not well understood.

Many intracellular pathogens have evolved to manipulate the phagocytic pathways of host cells during invasion. This contrasts with invasion by the apicomplexans, which express their own machinery for active host cell entry. Invasion is a multistep process requiring the tightly regulated discharge of parasite organelles called micronemes and rhoptries (Carruthers & Sibley, 1997). Micronemes release adhesins (MICs) onto the parasite surface, which form multi-protein complexes with non-overlapping roles in mediating motility, host cell attachment, secretion of rhoptry organelles and cell penetration (Soldati-Favre, 2008). After attachment and reorientation of the parasite, invasion induces the formation of a non-fusogenic parasitophorous vacuole derived in large part from host cell plasma membrane (Carruthers & Boothroyd, 2007). MICs across the phylum Apicomplexa share a limited

number of adhesive domains arranged in various combinations and numbers (Carruthers & Tomley, 2008). These domains are implicated in host cell recognition and attachment and are believed to contribute to host cell type specificity and hence disease pathology.

T. gondii microneme protein 1 (TgMIC1) forms a complex with TgMIC4 and TgMIC6 (Brecht et al, 2001; Reiss et al, 2001) and binds to sialylated glycoconjugates on the host cell surface (Blumenschein et al, 2007). Previous studies based on gene disruption have established a critical role for the complex in host cell invasion in tissue culture and its contribution to virulence *in vivo* (Cerede et al, 2005). The N-terminal region of TgMIC1 interacts with TgMIC4, a protein comprising six ‘apple’ domains that has been shown to bind to host cells in the presence of TgMIC1 (Brecht et al, 2001). TgMIC6 contains three EGF-like domains and is a type I membrane protein, which serves as an escorter and anchors the TgMIC1-MIC4-MIC6 (TgMIC1-4-6) complex to the parasite surface during invasion (Reiss et al, 2001). The first EGF-like domain (TgMIC6-EGF1) is cleaved off during secretory transport of the complex, probably in a post Golgi compartment (Meissner et al, 2002). Each of the remaining two EGF-like domains is able to recruit one molecule of TgMIC1 via interaction with its C-terminal galectin-like domain (Reiss et al, 2001; Saouros et al, 2005; Sawmynaden et al, 2008) (for a schematic of the wild-type complex see **Fig. 1A**). Correct trafficking of the complex to the micronemes depends not only on a sorting determinant in the C-terminal tail of TgMIC6, but also on the interaction between the galectin-like domain of TgMIC1 with the third membrane-proximal EGF-like domain of TgMIC6 (TgMIC6-EGF3). This interaction is crucial for transport of the entire complex through the early secretory pathway as it assists proper folding of TgMIC6-EGF3, providing a quality control check point (Huynh et al, 2004; Reiss et al, 2001; Saouros et al, 2005).

Several studies have shown that recognition of carbohydrate structures on the host cell surface is critical for efficient invasion by *T. gondii* (Carruthers et al, 2000; Monteiro et al, 1998; Ortega-Barria & Boothroyd, 1999). We have recently demonstrated that the N-terminal

region of TgMIC1 contains two copies of a novel domain Microneme Adhesive Repeat (MAR) and that this region termed TgMIC1-MARR binds specifically to sialylated oligosaccharides as shown by cell binding assays and carbohydrate microarray analyses (Blumenschein et al, 2007). Also, we observed a 90% reduction of invasion efficiency when N-acetylneuraminic acid (NANA) was used as a competitor or when host cells were treated with neuraminidase (Blumenschein et al, 2007). This has suggested that sialic acid is a major determinant of host cell invasion by *T. gondii* and that the effect observed in the invasion assays may be attributed, at least in part, to the inhibition of the interaction between TgMIC1 and its host cell receptor(s).

Here, we compare the invasion efficiency of several *T. gondii* knockout strains and complemented mutants and demonstrate that this interaction is indeed important for efficient host cell invasion. However, we also show that the TgMIC1-4-6 complex is not the only molecular player involved in sialic acid-dependent invasion by *T. gondii* tachyzoites. This has led us to identify a complete family of MAR domain containing proteins (MCPs) in *T. gondii* and related apicomplexan parasites. Among these we characterize TgMIC13, and we compare the binding specificities of recombinant TgMIC13, TgMIC1 and its homologue from the closely related organism *Neospora caninum*, NcMIC1, by carbohydrate microarray and cell binding assays. Our results reveal that the MAR domain is unique to and conserved among coccidian parasites, and acts as an important determinant in host cell recognition by this subclass of Apicomplexans through selective binding to sialylated glycoconjugates.

Results

1. TgMIC1 is an important player in sialic acid-dependent host cell invasion by *T. gondii*

A previously generated TgMIC1 knock-out strain (*mic1ko*) showed a 50% reduction in invasion efficiency compared to the wild-type strain (Cerede et al, 2005). In this strain the TgMIC1-4-6 complex is disrupted, ablating transport of TgMIC4 and TgMIC6 to the micronemes and instead resulting in their retention in the early secretory pathway (Reiss et al, 2001). In consequence, the contribution of the individual components to host cell invasion remained unresolved (for a schematic of the complex see **Fig. 1A**). To address this question we complemented the *mic1ko* strain with three different TgMIC1-mutant constructs, then examined whether expression of these proteins is able to improve or restore invasion efficiency. A first parasite mutant line named *mic1ko/mycMIC1-GLD* expresses the MIC1 carboxy-terminal galectin-like domain (TgMIC1-GLD) on the *mic1ko* background. A second parasite line called *mic1ko/MIC1^{TTAA}myc* was generated using a construct coding for full length TgMIC1 carrying threonine to alanine substitutions at two positions (126 and 220) that were previously shown to be critical for the host cell binding activity of the TgMIC1-MAR region (TgMIC1-MARR) (Blumenschein et al, 2007). A third complemented line called *mic1ko/MIC1^{T220A}myc* expresses MIC1 containing just a single T220A substitution that also abolishes the host cell binding activity of the protein (Blumenschein et al, 2007). As a control the *mic1ko* strain was complemented with full length wild-type TgMIC1 carrying a myc tag epitope at the carboxy-terminus (parasite line *mic1ko/MIC1myc*). Western blots revealed comparable levels of expression for the respective mutant proteins (**Fig. 1B**). In addition, the previously generated *mic4ko* line (Reiss et al, 2001) was included in the analysis. **Figure 1F** shows a comparison of invasion efficiency of the parental wild-type and the *mic1ko* lines and

confirms the 50% reduced invasion phenotype previously reported for the *mic1ko* (Cerede et al, 2005).

Given that TgMIC4 and TgMIC6 are retained in the early secretory pathway in the *mic1ko* strain, assessment of the sub-cellular localisation of the TgMIC1-mutant proteins, as well as of TgMIC4 and TgMIC6, in the complemented *mic1ko* parasite lines was important for interpretation of the invasion assay results. Expression of TgMIC1myc in the *mic1ko* strain rescued the targeting of TgMIC4 and TgMIC6 to the micronemes [(Reiss et al, 2001); and **suppl. Fig.1**]. Invasion efficiency in this strain was restored to a level slightly higher than the parental wild-type line (**Fig. 1F**), confirming that a complex composed of endogenous TgMIC4 and TgMIC6 together with the epitope-tagged TgMIC1myc is functional in invasion. In agreement with previous studies (Saouros et al, 2005) expression of TgMIC1-GLD in the *mic1ko* strain brought TgMIC6 to the micronemes but TgMIC4 remained in the ER (**Fig. 1C**). In the invasion assay, when compared to *mic1ko*, *mic1ko*/mycMIC1-GLD showed no rescue of phenotype (**Fig. 1F**), demonstrating that the TgMIC6-TgMIC1-GLD mutant complex does not contribute to invasion, likely due to the absence of adhesive domains.

In the *mic1ko*/MIC1^{T220A}myc line, TgMIC1^{T220A} targeted to the micronemes and restored proper trafficking of TgMIC4 and TgMIC6 (**Fig. 1D**). This mutant invaded significantly better than *mic1ko*/mycMIC1-GLD (**Fig. 1F**) despite the fact that the mutation in TgMIC1^{T220A} abrogates the adhesive properties of the protein (Blumenschein et al, 2007). This suggested that the invasion enhancement observed upon expression of this protein might be solely a result of its capacity to recruit TgMIC4 to the membrane-bound complex and/or residual weak binding of the first TgMIC1 MAR domain. Invasion efficiency of this line was similar to the wild-type but lower than the control line *mic1ko*/MIC1myc, underlining the importance of the adhesive properties of TgMIC1. In the *mic1ko*/MIC1^{TTAA}myc line, TgMIC1^{TTAA} was found mainly in the early secretory pathway probably due to incorrect

folding of this mutant protein (**Fig. 1E**), precluding any confirmation of the above results with the single T220A mutant. No improvement of invasion efficiency compared to the *mic1ko* was found for this line, as expected (**Fig. 1F**).

In the *mic4ko* strain, the partial TgMIC1-6 complex was delivered to the micronemes (Reiss et al, 2001). This strain invaded more efficiently than *mic1ko* and *mic1ko/mycMIC1-GLD* (t-tests: $p < 0.05$ respectively), confirming the critical role of TgMIC1-MARR in invasion (**Fig. 1F**). In addition the *mic4ko* strain did not invade as well as the wild-type (t-test: $p < 0.05$), suggesting a role for TgMIC4 in invasion in agreement with our observations with the *mic1ko/MIC1^{T220A}myc* line, or alternatively that the presence of MIC4 impacts on the proper function of the MAR domains.

Collectively, the characterization of these mutants suggests a cooperative function for TgMIC1-MARR and TgMIC4 in receptor binding and confirms the absence of any adhesive function for TgMIC6 and TgMIC1-GLD. Furthermore, these experiments establish that the binding activity of TgMIC1-MARR to sialylated structures recapitulates TgMIC1 function and its contribution to efficient invasion.

2. *T. gondii* possesses more than one sialic acid binding factor involved in host cell invasion

To address the question of whether other parasite lectins bind to host sialylated glycoconjugates during invasion, we tested the effect of free NANA on invasion by the *mic1ko* strain. Interestingly, this assay showed that host cell invasion by the *mic1ko* strain was still considerably impaired in the presence of free NANA (**Fig 2A**), indicating that TgMIC1 is not the only sialic acid-binding parasite lectin contributing to invasion by *T. gondii* tachyzoites. This result was confirmed by performing an invasion assay with neuraminidase-

treated host cells which resulted in a substantial reduction of invasion by the *mic1ko* strain (Fig 2B).

3. Identification of a novel family of MAR containing proteins (MCPs) in Coccidia

In the light of the observation that host cell invasion by *mic1ko* parasites is still sensitive to inhibition by free NANA, we hypothesized that *T. gondii* has at least one additional sialic acid-binding parasite lectin. A survey of the *T. gondii* genome sequence revealed that it encodes a family of four MAR-domain containing proteins (MCPs), herein named TgMIC1 (80.m00012), TgMCP2 (55.m04865), TgMCP3 (25.m00212) and TgMCP4 (25.m01822) (see **Fig. 3A and suppl. Fig. 2**). The presence in all cases of a predicted amino-terminal signal peptide suggests that these proteins are delivered to the secretory pathway. In contrast to TgMIC1, which comprises two MAR domains (TgMIC1-MARR) followed by a galectin-like domain (TgMIC1-GLD), the other three sequences contain four consecutive MAR domains but lack a galectin-like domain. In addition, TgMCP4 possesses a novel N-terminal repeat region (16 repeats, 17 to 22 amino acids per repeat), with the number of repeats varying between different strains of *T. gondii*. Examination of the EST-data set indicate that all these MCP genes are transcribed both in tachyzoite and bradyzoite stages and analysis of available proteomics data (Xia et al, 2008) confirmed the expression of TgMCP2 in tachyzoites.

Our earlier studies have established that TgMIC1-MARR has the potential to bind to two molecules of sialic acid through its two binding sites, one in each MAR domain, characterized by the presence of conserved His and Thr residues. In both binding sites the threonine residue (T126 and T220 respectively) makes principle contacts to the sialic acid moiety (Blumenschein et al, 2007). Sequence comparison reveals that these important threonine residues are not present in either TgMCP3 or TgMCP4, but are conserved in three

out of the four MAR domains in TgMCP2 (**suppl. Fig. 2**). Consequently, we hypothesized that TgMCP2 may be another sialic acid binding parasite lectin.

An examination of related apicomplexan parasites allowed identification of a complete set of homologues in *N. caninum* (**Table I and suppl. Fig. 2**). NcMIC1, the homologue of TgMIC1 has been previously described (Keller et al, 2002), whereas TgMCP2 is highly similar to NC_LIV_092400 (73% id. and 86% sim.), and genes coding for proteins highly similar to TgMCP3 (67% id. and 75% sim.) and TgMCP4 (73% id. and 80% sim. in the MAR region) are present on *N. caninum* chromosome 1b, contig 996 (**suppl Fig. 3 and 4** show full amino acid sequence alignments). Like TgMCP4, NcMCP4 displays an N-terminal region with short repeats, but comprising only 8 units. Interestingly, the His and Thr signature residues of the MAR sialic acid binding site are conserved when comparing the *N. caninum* and *T. gondii* homologues, suggesting that these could be functionally equivalent and hence orthologues (**suppl. Fig. 2**). This view is supported by a phylogenetic analysis comparing the first two predicted MAR domains of each member of the family (**Fig. 3B**). In this analysis the putative orthologues are indeed most closely related. TgMCP3, TgMCP4, NcMCP3 and NcMCP4 lacking the critical Thr residues form a separate cluster, consistent with an expected functional divergence.

Unfortunately, there was insufficient genomic data available to establish whether a similar gene family exists in *Sarcocystis* spp. but EST-data (SnEST4a79g11.y1, SnEST4a34d02.y1, SnESTbab01e09.y1, SnEST4a69h07.y1 and SnESTbab30h02.y1) indicates the presence of several MAR-containing proteins, one of which (termed SnMCP5) has a homologue in *N. caninum* (NC_LIV_144780 on chr.12, named NcMCP5 hereafter, see **Fig. 3A and suppl. Fig. 2 and 4**). Despite a high level of synteny between the two parasites, the gene encoding NcMCP5 is absent at the corresponding locus in *T. gondii*. Two additional loci coding for putative MCPs were found in the *N. caninum* genome (NC_LIV_131070 and on contig 1031, named NcMCP6 and NcMCP7 respectively) both coding for proteins

composed of a signal peptide and two MAR domains (**Fig. 3A and suppl. Fig. 2 and 5**). In *Eimeria tenella*, the previously described EtMIC3 contains 7 MAR domains, in which repeats three, four and five are identical (Labbe et al, 2005). EtMCP2, lying downstream of EtMIC3 on contig 29262, was annotated *in silico* by comparison with EST data from sporulated oocyst and sporozoite stages. EtMCP2 is composed of a signal peptide and a single MAR domain (**Fig. 3A and suppl. Fig. 2 and 5**). In addition, this parasite possesses a number of hypothetical MCPs (two gene models respectively on contigs 29652 and 14843). EST data supports expression of at least one of them in first generation merozoites. Our BLAST searches failed to identify any MCPs in *Plasmodium*, *Theileria* or *Cryptosporidium*. Accession numbers and proposed orthologous relationships for all proteins are depicted in **Table I**.

A comparison of MAR domains from *Toxoplasma*, *Neospora* and *Sarcocystis* highlights differences in disulphide bond patterns between two tandemly arranged MAR domains in a given MCP (**suppl. Fig. 2**). As in the TgMIC1-MARR prototype, the beginning of the second domain (referred to as type II) lacks a stretch of amino acids containing two cysteine residues shown to participate in the formation of a disulphide bond in the TgMIC1-MARR crystal structure, but possesses an additional C-terminal extension (with the notable exception of the second MAR domain in TgMCP2 and NcMCP2) that adopts a beta-hairpin or ‘beta-finger’ (Blumenschein et al, 2007). Curiously, in this regard all the MAR domains in EtMIC3 and in all other predicted MCPs in *Eimeria* resemble the first of the two tandemly arranged MAR domains, referred to as type I (**Fig. 3A and suppl. Fig. 2**). This suggests that these proteins have evolved differently in *Eimeria* compared to the MCPs found in *Toxoplasma*, *Neospora* and *Sarcocystis*. Therefore, despite the fact that some of the MCPs in *Eimeria* display the binding site His-Thr or a similar motif, these proteins might possess different properties and functions.

4. TgMIC13, a novel sialic acid binding lectin

A parasite line expressing a C-terminal epitope-tagged copy of TgMCP2 was generated. TgMCP2ty was co-localised with TgMIC4 to the micronemes by immunofluorescence assay (IFA) and was named TgMIC13 accordingly. IFA on wild-type parasites using antibodies raised against recombinant TgMIC13 expressed in *Pichia pastoris* confirmed localization to the micronemes, although only partial co-localization with TgMIC3 was observed (**Fig. 4A**). Analysis of lysates from wild type and wt/TgMIC13ty parasites by Western blot indicated that TgMIC13 migrates on SDS-PAGE as a single protein species with an apparent molecular mass somewhat higher than the predicted (51.5kDa) (**Fig. 4B**). TgMIC13ty migrates slightly faster than endogenous TgMIC13. The solubility profile of TgMIC13 is identical to TgMIC4 and is consistent with it being a soluble protein within the parasite organelles (**Fig. 4C**, upper panel). To assess whether TgMIC13 displays cell binding activity, excreted-secreted antigen fraction (ESA) prepared from wt/TgMIC13ty parasites was tested in a cell binding assay; the protein was found to bind to the cell surface (**Fig. 4C**, lower panel). Furthermore, transient expression of TgMIC13ty in *mic1ko* and *mic6ko* recipient strains indicated that TgMIC13 traffics to the micronemes independently of the TgMIC1-4-6 complex and does not associate with its components (**Fig. 4D**). These results establish that expression of TgMIC13 cannot functionally rescue the *mic1ko*, strongly suggesting that TgMIC13 belongs to a distinct complex.

The cell binding activity of TgMIC13 was reproduced with recombinant protein expressed in *P. pastoris* (**Fig. 5A**). Similar to TgMIC1, TgMIC13 binding was found to be abolished by pre-treatment of cells with neuraminidase (**Fig. 5B**). Binding of TgMIC13 to host cell receptor(s) could be competed out with free NANA, but was not affected by increasing concentrations of a control acidic monosaccharide, glucuronic acid (**Fig. 5C**).

However, the binding properties of TgMIC13 differed from TgMIC1 in the following ways: (1) at equivalent concentrations, more TgMIC13 appeared to bind to HFFs (**Fig. 5A**); (2) compared to TgMIC1, higher concentrations of NANA were required to fully inhibit receptor binding of TgMIC13 (10 mM for TgMIC1 (Blumenschein et al, 2007) versus 50 mM for TgMIC13) (**Fig. 5C**); (3) only extensive treatment of cells with neuraminidase completely abolished the binding of TgMIC13 (data not shown); and (4) TgMIC1 did not bind to CHO lec2 cells that display strongly reduced surface expression of sialic acid (Deutscher et al, 1984), whereas TgMIC13 can weakly bind to them (**Fig. 5D**). No other differences were observed in cell binding assays with CHO mutants deficient in glycosaminoglycan synthesis (CHO-pgsA, CHO-pgsB) or with CHO cells lacking two glycosyltransferase activities (CHO-pgsD) (Esko et al, 1988) (**Fig. 5D**). Taken together these observations suggested that TgMIC1 and TgMIC13 have different sialic acid binding properties and may bind to different sialylated receptors.

In order to assess precisely the role of TgMIC13 in invasion, we repeatedly tried to knock-out the gene in wild-type (RH strain) and in *mic1ko* parasites without success. In addition, repeated knock-out attempts failed in a delta-ku80 strain efficiently amenable to gene disruption by dramatically enhanced frequency of homologous recombination (Fox et al, 2009; Huynh & Carruthers, 2009).

5. Binding properties of other *T. gondii* proteins of the MAR domain family

Given that the MAR domains on TgMIC1 and TgMIC13 are involved in carbohydrate recognition, we postulated that TgMCP3 and TgMCP4 could belong to a distinct functional class of lectins. In cell binding assays using ESA prepared from transgenic parasites expressing TgMCP3ty and TgMCP4ty, neither of the two epitope-tagged proteins showed

detectable binding activity (not shown). Full-length TgMCP3 as well as the MAR region of TgMCP4 (TgMCP4-MARR) encompassing all four MAR domains without the N-terminal repeat region were expressed in both *P. pastoris* and *E. coli*. No binding to HFFs could be observed with the recombinant proteins in cell-binding assays (**Fig. 5A** and data not shown). Furthermore, no binding was detected for recombinant TgMCP3 expressed and purified from *E. coli* in carbohydrate microarray analysis using neutral and sialylated probes (data not shown). While strong binding was observed for TgMIC13 expressed in *P. pastoris*, the same protein expressed in *E. coli* showed only weak binding activity reflecting the limitations of this expression system when large proteins containing numerous disulphide bonds are produced. Considering this limitation, we cannot exclude mis-folding of bacterial recombinant TgMCP3 and TgMCP4, although the lack of binding to sialic acid containing moieties would be consistent with the absence of the critical Thr residues in the MAR domains of TgMCP3 and TgMCP4 (**suppl. Fig. 2**). Homology modelling and sequence alignments suggest that while TgMCP3 and TgMCP4 adopt the MAR fold, they present a largely hydrophobic surface in the equivalent position to the hydrophilic carbohydrate binding site in TgMIC1 and therefore would not be expected to bind sugars (data not shown).

6. Carbohydrate microarray analysis reveals differing binding characteristics for TgMIC13, TgMIC1 and NcMIC1

To examine in detail the binding specificities of TgMIC13, we performed cell-independent carbohydrate microarray studies. Preliminary screening analyses with more than 300 lipid-linked oligosaccharide probes (Blumenschein et al, 2007; Feizi & Chai, 2004) showed that recombinant TgMIC13 expressed in *P. pastoris* bound exclusively to sialylated

oligosaccharide probes (data not shown). This is in accord with results of the cell binding assays, and shows that sialic acid is a requirement for TgMIC13 binding.

The fine binding specificity of TgMIC13 was compared with that of TgMIC1 and NcMIC1 using an array of 88 oligosaccharide probes. These comprised 82 diverse sialylated oligosaccharide probes with different sialyl linkages and backbone sequences (**Fig. 6; suppl. Table I**); 6 neutral oligosaccharide probes served as negative controls. In these experiments we used full-length TgMIC13 expressed in *P. pastoris*, as well as TgMIC1-MARR (aa 17-262) and NcMIC1-MARR (aa 17-259), both expressed in *E. coli*. TgMIC1 lacking its C-terminal galectin-like domain is known to reflect the binding specificities of the full-length protein (Blumenschein et al, 2007; Sawmynaden et al, 2008) and we assumed that this is as well the case for its homologue NcMIC1. The three proteins bound to various sialylated, but not to neutral oligosaccharides in the microarrays. Stronger binding was observed to α 2-3-linked sialyl probes compared with α 2-6 linked sialyl probes. This is in agreement with a recent study revealing the importance of a glutamic acid residue (E222) in TgMIC1 for this preference, which is conserved in TgMIC13 and NcMIC1 (**suppl. Fig. 2**) (Garnett et al, 2009).

Interestingly, for each protein a distinct binding profile was observed. Among the broad range of sialylated probes bound by TgMIC1-MARR, the preferred ligands are α 2-3-linked sialyl-*N*-acetylglucosamine (3'SiaLacNAc) sequences with or without fucosylation/sulphation on the *N*-acetylglucosamine residue (highlighted probes in **Fig. 6A**). This is in overall agreement with previous findings (Blumenschein et al, 2007). Although the oligosaccharide sequences of probes 18, 20, 21 and 22 are short tri- or tetrasaccharides, their recognition demonstrates the effective presentation of these as oxime-linked neoglycolipids on the array surface (Liu et al, 2007).

TgMIC13 bound to a more restricted spectrum of probes compared to TgMIC1-MARR (**Fig. 6B**). Several 3'SiaLacNAc-based probes were recognized, but there was little or

no binding to sialyl sequences having long chain polyLacNAc or *N*-glycan backbones (probes 38-41), and to polysialyl sequences (probes 76-85). Strikingly, the strongest binding of TgMIC13 was to a disialyllactose probe having α 2-9 linkage between the two sialic acid residues (probe 88). The binding signal for this probe was extremely high (saturated) at the protein concentration tested. In addition, TgMIC13 bound to an uncommon 4-*O*-acetylated 3'sialyllactose (probe 16) which was not recognized by TgMIC1. It is interesting that the intensity of binding to probe 16 was stronger than that to the non-*O*-acetylated analogue (probe 12). For NcMIC1-MARR, there was strikingly high binding to two sialyl Le^x-related probes (33 and 37, **Fig. 6C**), which have in common a sulphate group at the 6 position of the galactose residue. In contrast, if the sulphation is on the *N*-acetylglucosamine residue as in probe 35, binding of NcMIC1-MARR was much weaker. NcMIC1-MARR gave stronger binding signals than TgMIC1 and MIC13 with polysialyl sequences.

7. Molecular modelling corroborates the findings of microarray analyses

To better understand the basis of the binding preference revealed by the microarray studies, we used the optimised sequence alignment to build homology models for the TgMIC13 MAR domains, based on the crystal structure of TgMIC1-MARR (Bordoli et al, 2009). While the first three MAR domains of TgMIC13 possess the sialic acid-binding signature, we have assumed that the type II domains (MAR2) make the dominant contribution to sialic acid recognition as this was observed for TgMIC1-MARR (Blumenschein 2007). Strikingly, the hydroxyl group on C4 of the sialic acid ring packs against the two large side chains of K216 and F169 from loops β 1- β 2 and β 4- β 5 of TgMIC1-MAR2 and would hinder any substantial modification of the sugar at this position i.e. 4-*O*-acetylation (**Fig. 7A**). In the equivalent MAR domain from TgMIC13 these bulky side-chains are replaced by smaller Ser and Thr residues, respectively, which creates space to accommodate the additional acetyl group and provide potential hydrogen-bonding partners.

In 2-9-linked disialyl oligosaccharides the two sialyl moieties are separated by three additional bonds compared to the separation between sialic acid and galactose in 3'SiaLacNAc. Assuming a similar mode of recognition for the outermost sialic acid moiety, the longer linkage would shift the second sialic acid to a position directly over the end of helix II (**Fig. 7B**). In TgMIC1 three consecutive Glu residues (E221-223) cap this helix, which might induce electrostatic repulsion of the negatively charged carboxylate group and hinder any preference for 2-9 disialyl sugars. In TgMIC13, Glu 221 and 223 are replaced by positively charged Arg and Lys residues, which would not only reduce any electrostatic repulsion but provide opportunities for the formation of specific charge-charge interaction with the second sialic acid.

We also derived homology models for the MAR domains of NcMIC1 based on the crystal structure of TgMIC1-MARR. The preference of NcMIC1-MARR for the two sialyl Le^x-related probes sulphated in the 6-position of galactose could be explained by the presence of a positive charged Lys (K176) in NcMIC1, which is replaced by an Ala in TgMIC1 (**Fig. 7C**). In our homology model K176 is spatially closely positioned to the 6-position of galactose and could form a specific charge-charge interaction with the negative charged sulphate. K176 in NcMIC1 is not conserved in TgMIC13 but is present in NcMCP2.

8. Gangliosides and 3'SiaLacNAc are potential determinants for *T. gondii* binding to host cells

Earlier data from carbohydrate microarray analyses on the binding preferences of TgMIC1-MARR suggested that gangliosides GD1a and GT1b are among the most potent binding molecules (Blumenschein et al, 2007); these were also bound by TgMIC13 in the carbohydrate microarray analyses (**suppl. Table I**). In order to determine whether gangliosides are able to compete with the authentic host cell receptor(s) for binding of

TgMIC1 and TgMIC13, we performed cell-binding assays using gangliosides GD1a and GT1b incorporated into liposomes as competitors. Binding of recombinant TgMIC1 to the cells was hardly detectable in the presence of 500 μ M GD1a or GT1b (**Fig. 8A**). Binding of TgMIC13 was also abolished in the presence of 500 μ M GD1a, and the presence of GT1a had a clear inhibitory effect (although weak binding could be detected even at 2mM GT1b upon long exposure of the blot) (**Fig. 8A**). Lipid-linked oligosaccharide ligands displayed on liposomes can be much more potent inhibitors than free oligosaccharides due to the multivalent presentation. Therefore the nominal concentrations of the liposome inhibition assays can not be compared with concentrations of free oligosaccharides used in our cell binding assays.

Binding assays using C6 rat glioma cells deficient in complex gangliosides (such as GD1a and GT1b) showed that both TgMIC1 and TgMIC13 bind to these cells (**Fig. 8B**) implying – in agreement with our carbohydrate microarray data - that glycoconjugates other than complex gangliosides can be bound by these molecules. As mentioned above, several of the probes in this array strongly bound by TgMIC1 and TgMIC13 are based on the 3'SiaLacNAc sequence. In order to determine if such high affinity is physiologically relevant, we examined if this interaction could be reproduced in a cell-binding assay. Interestingly, binding of TgMIC1 and TgMIC13 was abolished by concentrations of free 3'SiaLacNAc as low as 1mM and 5mM respectively, whereas increasing concentrations of LacNAc had no effect (**Fig. 8C**). In a parallel assay, the competition of NANA for TgMIC13 binding was reproducible, suggesting that the inhibitory concentrations of free 3'SiaLacNAc for TgMIC1 and TgMIC13 are lower than those of free NANA required for equivalent binding inhibition (Blumenschein et al, 2007). This increased affinity can be explained by an additional contact observed in the crystal structure of TgMIC1-MARR in complex with 3'SiaLacNAc between the galactose and the protein which is involved in the preference of the protein for 2-3 linked sialylated glycoconjugates (Garnett et al, 2009).

Discussion

Several studies and our previous results have highlighted the critical contribution of carbohydrate recognition on the host cell surface during invasion by *T. gondii*. TgMIC2, a member of the thrombospondin-related anonymous protein (TRAP) family, conserved throughout the Apicomplexa, is an essential component for parasite motility and invasion and has been shown to interact with heparin via its integrin-A-like domain upon multimerization (Harper et al, 2004). TgMIC3 and TgMIC8 are part of another major complex involved in invasion. Dimerization of a chitin-binding-like domain present in TgMIC3 and TgMIC8 allows host cell binding and has been associated with parasite virulence (Cerede et al, 2002; Cerede et al, 2005). The TgMIC1-4-6 complex, critical for invasion and virulence in mice, binds sialylated glycoconjugates, and a synergy-effect has been demonstrated between this complex and TgMIC3 (Blumenschein et al, 2007; Cerede et al, 2005).

In this study we have characterized key parasite molecular players involved in sialic acid-dependent host cell invasion by *T. gondii*. The salient findings in this investigation are first, the characterization of TgMIC1-4-6 complex as one important player in sialic acid-dependent host cell invasion, second, the reliance of the parasite on an additional sialic acid-binding lectin for invasion, identified as TgMIC13 and third, the determination of distinct sialic acid-dependent binding specificities for three members of a novel family of MAR domain containing proteins (MCPs) conserved in coccidians. In *T. gondii* all the four MCPs are expressed in tachyzoites. Among them, only TgMIC1 and TgMIC13 possess the sequence characteristics of a sialic acid-binding MAR domain, and only these specifically recognize sialyl oligosaccharides. The results of our study now add a hypothetical TgMIC13-containing complex to the list of carbohydrate binding parasite surface complexes.

Distinct binding specificities might determine tissue tropism of parasitic infections

With carbohydrate microarray technology, we examined details of carbohydrate-binding specificities of TgMIC13, in comparison with MAR regions of TgMIC1 and NcMIC1. The preponderant binding activity of all three proteins was to α 2-3-linked sialyl probes. However, the relative binding intensities to specific sialyl probes were quite different for the three proteins. TgMIC13 bound to two sialyl oligosaccharide probes, namely the 4-*O*-acetylated sialyllactose and the α 2-9-linked disialyl sequences, which were not recognized by TgMIC1 or NcMIC1. The 4-*O*-acetyl-substituted sialic acid has been described in various tissues in mice, especially in the gut; its presence has been documented in a number of other animal species and in trace amounts in humans (Iwersen et al, 2003; Rinninger et al, 2006). There is so far limited information on the occurrence of 2-9-linked sialic acid in animals. It has been reported as a component on a murine neuroblastoma cell line (Inoue et al, 2003) and a human embryonal carcinoma-cell line (Fukuda et al, 1985). A special feature of NcMIC1 as revealed by carbohydrate microarray analyses is the strong binding to the two sulphated sialyl Le^x-related probes. Sulphation was previously reported to play a role in cell surface binding by NcMIC1 (Keller et al, 2002); in that study, glycosaminoglycans were proposed to be targeted by NcMIC1; however the influence of sialic acid was not investigated. The distinct binding specificities of the three MIC proteins might have implications for tissue tropism, and the cellular targets for these proteins warrant further investigation.

Our results suggest that TgMIC1 and TgMIC13 recognize sialic acid on several types of glycoconjugates with differing affinities. In general 3'SiaLacNAc sequences are bound with high affinity by TgMIC1 and TgMIC13. This is in marked contrast to the *P. falciparum* erythrocyte binding antigen 175 (PfEBA175), which binds selectively to a unique sialylated glycoprotein, glycophorin A (Orlandi et al, 1992; Tolia et al, 2005). As our results suggest cooperativity between TgMIC1 and TgMIC4 in receptor binding by the TgMIC1-4-6 complex, we cannot exclude the possibility that TgMIC4 discriminates further in receptor

recognition and that some of the oligosaccharide sequences proposed are actually not recognized by this complex. Identification of the host cells receptor(s) recognized by TgMIC1 and TgMIC13 will be necessary to resolve this issue. Given the ubiquitous distribution of sialylated glycoconjugates in tissues, the binding preferences of TgMIC1 and TgMIC13 are likely to be important for parasite dissemination. This view is supported by the fact that TgMIC1 contributes to parasite virulence in mice (Cerede et al, 2005). The finding that gangliosides GD1a and GT1b are able to compete with host cell receptors for TgMIC1 and TgMIC13 binding is of particular interest for parasite biology regarding the colonization of the ganglioside-rich brain tissue prior to cyst formation. In addition, considering the fact that the MAR family is restricted to enteroparasites, these adhesins might be of importance in the context of the natural route of infection via the intestine. To our knowledge no study has so far addressed the importance of receptor-ligand interactions for invasion of the midgut epithelium.

The role of TgMIC13 in host cell invasion by *T. gondii*

To contribute to host cell invasion, TgMIC13 needs to be targeted accurately to the micronemes and then delivered in a timely fashion, firmly anchored on the parasite surface. Since TgMIC13 does not contain a predicted membrane spanning domain or GPI-anchor motif, it likely belongs to a yet undefined complex. Expression of TgMIC13 in the *mic1ko* and *mic6ko* strains ruled out any interaction with the TgMIC1-4-6 complex. A recent study on TgMIC3 has shown that a signal peptide and an EGF domain can be sufficient for microneme targeting of soluble proteins, presumably by the means of an unidentified escorter (El Hajj et al, 2008; Kessler et al, 2008). However, TgMIC13 does not possess EGF modules, nor does it contain a galectin-like domain like TgMIC1, which precludes a similar architecture for the putative TgMIC13-containing complex. Unravelling the mechanism of TgMIC13 sorting to the micronemes must await the identification of its escorter protein, as described for other

soluble microneme proteins (Meissner et al, 2002; Reiss et al, 2001). Despite several attempts, it has not been possible to disrupt the *TgMIC13* gene, even in the delta-ku80 parasite line (Fox et al, 2009; Huynh & Carruthers, 2009), which strongly suggests that this gene is crucial for parasite propagation.

Concluding remarks

The presence of the MAR domain containing proteins in *Toxoplasma*, *Neospora*, *Sarcocystis* and *Eimeria* parasites and its absence in *Plasmodium*, *Theileria* and *Cryptosporidium* suggests that this domain is exclusively conserved among the Coccidia. We hypothesize that sialylated receptors may be exploited as a common target by a subset of MCPs carrying the appropriate binding motif in this group of enteroparasites. As suggested in this study, not all MCPs could bind to sialic acid or even to host cells which raise questions regarding their alternative functions. We have identified a complete set of homologues of *T. gondii* MCPs in *N. caninum*, a parasite causing significant economic loss by infecting cattle. Similar to *T. gondii*, *N. caninum* is also a promiscuous pathogen, infecting very different cell types. It is tempting to speculate that this novel family of lectins is instrumental to infection of such a broad range of cells.

Materials and Methods

Generation of *T. gondii* mutant parasite strains

All *T. gondii* strains were grown in human foreskin fibroblasts (HFF) or Vero cells in DMEM (GIBCO) supplemented with 10% fetal calf serum (FCS), 2 mM glutamine and 25 mg/mL gentamicin. The RH-strain is referred to as “wild-type”.

The *mic1ko* parasite strain, generated previously (Reiss et al, 2001), was stably complemented with linearized plasmids pM2MIC1myc, pM2MIC1^{T220A}myc, pM2MIC1^{T126A,T220A}myc and pROP1mycMIC1-GLD coding for the expression of TgMIC1myc, TgMIC1^{T220A}myc, TgMIC1^{TTAA}myc and mycTgMIC1-GLD (encompassing amino acids 299-456 of TgMIC1) respectively using a standard electroporation transfection protocol with restriction enzyme mediated insertion (REMI). For selection plasmids were cotransfected with p2854_DHFR or pTUB5-CAT (carrying the pyrimethamine and the chloramphenicol resistance marker respectively) at a 5:1 ratio. pM2MIC1^{T220A}myc was generated from pM2MIC1myc using the Quickchange kit (Stratagene) and primers TgMIC1-20_1731 and TgMIC1-21_1732. pM2MIC1^{T126A,T220A}myc was obtained by replacing the fragment between restriction sites NdeI and EcoNI in pM2MIC1myc with the equivalent region in plasmid pPICZα-TgMIC1NT^{T126A,T220A} (see (Blumenschein et al, 2007)). To obtain a clonal line, parasites were cloned at least twice by limiting dilution from the drug-resistant pool of parasites obtained from transfection and selection.

Cloning of *T. gondii* MCPs

Restriction enzymes were purchased from New England Biolabs. All primers are listed in **suppl. Table 2**. Accession numbers are listed in **Table 1**. TgMIC13, TgMCP3 and TgMCP4 were amplified by PCR from a *T. gondii* tachyzoite cDNA pool. TgMIC13 with or without

the signal peptide was amplified using primer pairs MIC1/2-5/1047 and MIC1/2-6/1048 or MIC1/2-1/1147 and MIC1/2-2/1148 respectively. TgMCP3 was cloned using primers MIC1/3-7_1981 and MIC1/3-7_1982. To obtain the coding sequence for TgMCP4 with or without the signal peptide, primer pairs TgMIC1/4-7_1983 and TgMIC1/4-7_1984 as well as TgMIC1/4-3_1847 and TgMIC1/4-4_1848 were used for amplification respectively. The fragments were cloned into pGEM and sequenced.

The sequence encoding full-length TgMIC13 was digested with EcoRI and SbfI and sub-cloned into pTUB8TgMLCTy_HX between EcoRI and NsiI restriction sites.

pTUB8TgMLCTy_HX was cut with EcoRI, blunt ended with endonuclease and cut again with NsiI prior to insertion of full-length TgMCP3 coding sequence cut from pGEM-TgMCP3 with EcoRV and PstI. TgMCP3 without signal peptide (SP) was cloned into pROP1mycMIC1-GLD in between NsiI and PacI (blunt end cloning) resulting in pROP1mycTgMCP3(3430).

Full-length TgMCP4 was cloned into pTUB8TgMLCTy_HX using EcoRI and NsiI restriction sites. The sequence corresponding to TgMCP4 without SP was sub-cloned into pROP1mycMIC1-GLD using NsiI and PacI restriction sites.

Constructs were used in transient transfection experiments in either RHΔHx, *mic1ko* or *mic6ko* parasites. Stable parasite lines were derived from transfection of the pTUB8TgMLCTy_HX based vectors carrying the mycophenolic acid/xanthine resistance marker into the RHΔHx strain.

Expression and purification of MCPs from *E. coli*

TgMIC1 (aa 17-262), TgMIC13 (aa 23-468), TgMCP3 (aa 54-565), TgMCP4 (aa 526-1016) and NcMIC1 (aa 17-259) were 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

fusion proteins were purified using nickel-nitrilotriacetic acid HISBind resin (Novagen). For glycoarray analysis proteins were concentrated into PBS, pH 7.2.

Expression and purification of MCPs from *P. pastoris*

P. pastoris transformation and expression was performed using pPICZ α -based plasmids according to the protocols supplied with the *Pichia* expression kit (Invitrogen). Transformation of the supplied host strain GS115 was performed by electroporation following linearization of the vector with PmeI, SacI or DraI. To inhibit glycosylation expression was occasionally carried out in the presence of 10 μ g/ml tunicamycin. Recombinant TgMIC13 (aa 23-468) was either used directly in the form of culture supernatant or purified on NiNTA (Qiagen) and concentrated into 20mM NaH₂PO₄/Na₂HPO₄, pH 7.3. TgMCP3 (aa 54-565) and TgMCP4 (aa 526-1016) were not secreted, but soluble α -factor fusion-protein was obtained by lysis of cells in 50mM NaH₂PO₄/Na₂HPO₄, pH 7.4, 5% Glycerol, 1mM PMSF.

Preparation of Excreted Secreted Antigen fraction (ESA)

T. gondii tachyzoites freshly lysed from their host cells were harvested by centrifugation at 240g for 10 minutes and washed twice in IM (DMEM / 3% FBS / 10mM HEPES) prewarmed to 37°C. A pellet of 2,0 - 4,88 x 10⁸ parasites was resuspended in 1 ml IM, an aliquot of 50 μ l was taken as reference standard to allow estimation of the degree of secretion, and microneme secretion was stimulated by adding 10 μ l of 100% EtOH to the remaining 950 μ l. The sample was incubated for 10 minutes at RT, then 40 minutes at 37°C and finally cooled down to 0°C in an ice-water bath for 5min. Parasites were pelleted at 1000g, 4°C for 5 min. The supernatant was transferred to a new tube and centrifuged again for 5min, 4°C at 2000g. The supernatant (ESA) was collected and stored at -80°C.

Cell binding assays

These were performed as described previously (Blumenschein et al, 2007). GD1a di-sodium salt and GT1b tri-sodium salt were purchased from Alexis Biochemicals, L- α -phosphatidylcholine, cholesterol, NANA and glucuronic acid from Sigma, 3'Sia-LacNac and LacNac from Dextra, and α -2,3,6,8- *Vibrio cholerae* neuraminidase from Roche. Stock solutions of NANA and glucuronic acid in PBS were adjusted to pH 7.0.

Liposome inhibition assays were performed as follows. In glass vials, the lipid-linked oligosaccharide (here gangliosides GD1a and GT1b) were dispensed, together with carrier lipids, L- α -Phosphatidylcholine and cholesterol (all solutions in methanol), in a molar ratio of 1:5:5. The methanol was evaporated using a gentle stream of nitrogen and the samples dried completely by lyophilisation for 30 minutes. PBS was added to vials to give a nominal ganglioside concentration of 4 mM. Vials were sonicated in a water bath at 37°C for 30 min. Formation of liposomes was checked by dynamic light scattering.

Comparison of invasion efficiency of different strains

Comparison of different *T. gondii* strains for invasion efficiency was done using an RH-2YFP strain (Gubbels et al, 2003) as internal standard for parasite fitness. The details were described previously (Sawmynaden et al, 2008). Total number of parasite vacuoles and RH-2YFP parasite vacuoles were counted on 20 microscopic fields on each IFA-slide with a minimum of 440 vacuoles in total per slide. Only vacuoles containing at least two parasites were counted to be sure not to count extracellular parasites that resisted the washing. Each experiment was repeated at least three times. Statistical analysis was performed using PRISM.

Microarray analyses of recombinant MIC proteins

The sialyl oligosaccharide microarrays were generated with 88 lipid-linked oligosaccharide probes (**Supplementary Table 1**), which were arrayed in duplicate on nitrocellulose-coated

glass slides at 2 and 5 fmol per spot using a non-contact instrument (Palma et al, 2006). Analysis of carbohydrate-binding of the recombinant His-tagged MIC proteins was performed essentially as described (Blumenschein et al, 2007). In brief, each His-tagged MIC protein was precomplexed with mouse monoclonal anti-poly-histidine and biotinylated goat anti-mouse IgG antibodies (Sigma) in a ration of 1:2.5:2.5 (by weight), and overlaid onto the arrays at 40 µg/ml for TgMIC1-MARR and TgMIC13, and 20 µg/ml for NcMIC1-MARR. Binding was detected using Alexa 647 fluor conjugated streptavidin (Molecular Probes). Microarray data analysis and presentation were carried out using dedicated software (M. S. Stoll, unpublished). The binding to oligosaccharide probes was dose related, and results of 5 fmol per spot are shown.

Acknowledgements

We would like to thank C. Cherbuliez and N. Klages for excellent technical assistance, N. Jemmely for the generation of the anti-TgAMA-1 antibodies, W. Chai and Y. Zhang for mass spectrometric analyses of oligosaccharide probes, R. Childs for the microarrays, and M.P. Stoll for the microarray analyses software. We thank Dr. C. Rodriguez (University of Oviedo) and Dr. A. Hemphill (University of Bern) for kindly providing us with the C6 rat glioma cell line and NcMIC1 cDNA, respectively. We also thank M. Messer and T. Urashima for the 4-*O*-acetylated sialyllactose used in the microarray analyses. This project was supported by the Swiss National Foundation (DS), the UK Medical Research Council, and UK Research Councils' Basic Technology Grant (GR/S79268) and Translational Grant (EP/G037604/1) to TF. This work is part of the activities of the BioMalPar European Network of Excellence supported by a European grant (LSHP-CT-2004-503578) from the Priority 1 "Life Sciences, Genomics and Biotechnology for Health" in the 6th Framework Programme. DS is an International HHMI scholar. ASP is a fellow of the Fundação para a Ciência e Tecnologia

(SFRH/BPD/26515/2006, Portugal). JMS is a recipient of the EU-funded Marie Curie Action MalParTraining (MEST-CT-2005-020492), *The challenge of Malaria in the Post-Genomic Era*.

Author contributions

NF, DS and MJB designed the study. NF did sequence and phylogenetic analyses and performed invasion assays. NF and JDS generated parasite lines, produced recombinant proteins and performed cell binding assays. JDS conducted TgMIC13 knockout attempts, produced parasite ESA and generated TgMIC13 Abs. MK provided the synthetic glycolipids for the microarrays. YL and ASP designed and performed microarray experiments. YL analysed the microarray data. EL and SS produced recombinant NcMIC1. SM performed the modelling studies. NF and DS wrote the paper with contributions from YL, TF, SM and MJB.

Figure legends

Fig. 1

A cooperative function for the sialic acid binding TgMIC1-MAR domains and TgMIC4 in host cell invasion by *T. gondii*.

A. Schematic summarizing the domain organisation of the components of the TgMIC1-4-6 complex in the wt and *mic1ko*/MIC1myc complemented strain. **B.** Western blot analysis of parasite lines expressing TgMIC1 mutant proteins on the *mic1ko* background. **C, D and E.** TgMIC1 mutant proteins expressed on the *mic1ko* background were assessed for their ability to substitute for TgMIC1 and target the components of the TgMIC1-4-6 complex to the micronemes. IFA was performed on intracellular parasites multiplying in their vacuole. TgAMA-1 is used as a micronemal marker independent of the TgMIC1-4-6 complex. Scale bars, 1 μ m. A schematic summarizes the association/dissociation of the components of the

TgMIC1-4-6 complex in the different strains. An asterisk indicates a Thr to Ala substitution in the sialic acid binding site of the MAR domain. **F.** Comparison of host cell invasion efficiency by the various *T. gondii* mutant strains using an RH-2YFP strain as internal standard for parasite fitness. Error bars, standard deviation.

Fig. 2

Cell invasion assays using the *T. gondii mic1ko* strain demonstrate the existence of at least one more sialic acid parasite lectin. **A.** Invasion by *mic1ko* parasites in absence and presence of 10 mM and 20 mM free NANA or galactose. **B.** Assay comparing invasion of *mic1ko* parasites into host cells (HFFs) pre-treated or not with neuraminidase. Error bars, standard deviation.

Note that, compared to the parental strain the *mic1ko* shows a 50% reduced invasion phenotype (**Fig. 1F**) but its invasion efficiency was set to 100% here.

Fig 3

A family of MAR domain containing proteins in Apicomplexans

A. Schematic of the domain organisation of various MCPs. MAR domain type I (turquoise), MAR domain type II (dark green), MAR domain type II extension “the beta-finger” (light green), galectin-like domain (orange), repeat region (blue strips). The presence of a Thr in a MAR domain in an equivalent position to those critical for sialic acid binding in TgMIC1-MARR is indicated by a black T. The fourth MAR domain in TgMCP4 contains this Thr, however the sequence context does not fit with it being indicative of a potential sialic acid binding site; therefore the T is in parentheses.

B. Phylogenetic relationship of MCPs from *T. gondii*, *N. caninum* and *E. tenella*. All bootstrap values are >80. Since the domain structure of the different proteins varies, only sequence corresponding to the first two predicted MAR domains of each protein was used for

the analysis. Sequences were aligned in ClustalX and alignment positions containing gaps in >50% of the sequences were excluded from phylogenetic analyses. Phylogenetic analyses were carried out using POWER (<http://ca.expasy.org/>, neighbor-joining distance method, bootstrapping with 1000 replicates). Phylogenetic trees were generated using TREEVIEW.

Fig. 4

Characterisation of TgMIC13 (TgMCP2) in *T. gondii*.

A. Sub-cellular localisation by IFA in wild-type (top, confocal images) and in wt/TgMIC13ty parasites (bottom). **B.** Western blot analysis of endogenous and epitope-tagged TgMIC13 in wild-type and wt/TgMIC13ty parasites. An extract of HFF cells was also loaded as a control. **C.** Assessment of TgMIC13 solubility by fractionation (top). Cell binding assays using ESA from wt/TgMIC13ty parasites (bottom). S: supernatant, W: last of 4 washes, CB: cell bound fraction **D.** Transient transfection of TgMIC13ty into the *mic1ko* and the *mic6ko* show correct localisation of TgMIC13ty to the micronemes. Scale bars, 1µm.

Fig. 5

Binding characteristics of recombinant TgMIC13.

A. Cell binding assays using recombinant TgMIC1myc, TgMIC13myc and alpha-factor-TgMCP3myc fusion protein produced in *P. pastoris*; +tu/-tu: protein expressed in the presence/absence of tunicamycin. Top panel: Western blot indicating the relative concentration of the proteins in supernatants (S) used for the assay. Middle panel: Cell bound fractions (CB) probed for bound protein. Bottom panel: Control for use of equivalent amounts of cell material in each experiment. **B.** Pre-treatment of cells with neuraminidase abolishes binding by TgMIC1 and TgMIC13. S: supernatant, W: last of 4 washes, CB: cell bound fraction, C: background control. **C.** Binding of TgMIC13 is strongly reduced in cell binding assays by competition with free NANA but not glucuronic acid. **D.** Binding of TgMIC1 and

TgMIC13 to various CHO cell lines (K1: wt strain, 2: lec2 mutant with strong reduction in sialic acid surface expression, A and B: pgsA745 and pgsB618 mutants deficient in glycosaminoglycan synthesis, D: pgsD677 mutant deficient in two glycosyltransferase activities). Only the cell bound fractions of the assay are shown.

Fig. 6

Carbohydrate microarray analyses of recombinant TgMIC1-MARR expressed in *E. coli* (panel **A**), TgMIC13 expressed in *P. pastoris* (panel **B**) and NcMIC1-MARR expressed in *E. coli* (panel **C**) using 88 lipid-linked oligosaccharide probes printed on nitrocellulose-coated glass slides. Numerical scores of the binding signals are means of duplicate spots at 5 fmol/spot (with error bars). The complete list of probes, their sequences and binding scores are in Supplementary Table I. The binding signal for probe 88 was saturated and could not be accurately quantified (asterisk in B).

Fig. 7

A, B. Two orientations for the superposition of the homology model for the second MAR domain from TgMIC13 created using swiss-model (Bordoli et al, 2009) and the crystal structure of TgMIC1-MARR. TgMIC1 shown in cyan and TgMIC13 in violet; 3' sialyl-*N*-acetylactosamine is shown in an orange stick representation. Key side chains are drawn as stick representations minus attached protons and labelled with sequence position. **C.** Superposition of the homology model for the second MAR domain from NcMIC1 created using swiss-model (Bordoli et al, 2009) and the crystal structure of TgMIC1-MARR. TgMIC1 shown in cyan and NcMIC1 in violet; 3' sialyl-*N*-acetylactosamine is shown in an orange stick representation. Key side chains are drawn as stick representations minus attached protons and labelled with sequence position.

Fig. 8

Cell binding assays using recombinant TgMIC1myc and TgMIC13myc expressed in *P. pastoris*. I: input, W: last of 4 washes, CB: cell bound fraction, C: background control. Only the cell bound fractions are shown in panels A and C. **A.** Gangliosides GD1a and GT1b incorporated into liposomes compete with the host cell receptor(s) for binding to TgMIC1myc and TgMIC13myc. **B.** Cell binding assay on HFFs and C6 rat glioma cells using TgMIC1myc and TgMIC13myc. **C.** Free 3'SiaLacNac, LacNac and NANA used as competitors in cell binding assays with TgMIC1myc and TgMIC13myc.

Table I

Proposed orthologous relationships and Genebank (and toxoDB/geneDB) accession numbers for various MCPs.

<i>T. gondii</i>	<i>N. caninum</i>	<i>S. neurona</i>	<i>E. tenella</i>
MIC1 CAA96466 (80.m00012)	MIC1 AAL37729	Not id.	Not id.
MIC13/ MCP2 ABY81128 (55.m04865)	MCP2 (NC_LIV_092400)	Not id.	Not id.
MCP3 CAJ20583 (25.m00212)	MCP3 Chr 1b, contig 996	Not id.	Not id.
MCP4 CAJ20582 (25.m01822)	MCP4 Chr 1b, contig 996	Not id.	Not id.
Not present	MCP5 (NC_LIV_144780)	MCP5 EST4a79g11.y1 EST4a34d02.y1	Not id.
Not present	MCP6 (NC_LIV_131070)	Not id.	Not id.
Not present	MCP7 Contig 1031	Not id.	Not id.
Not present	Not present	Not id.	MIC3 ACJ11219
Not present	Not present	Not id.	MCP2 Contig 29262

Supplementary Materials and Methods

Immunofluorescence Analysis

IFA was carried out on intracellular parasites grown in HFFs seeded on glass coverslips. After inoculation of parasites, the cells were incubated for 24-36 hours at 37°C before fixation to let the parasites divide within their vacuoles and form rosettes. Fixation is routinely done with 4% paraformaldehyde for 20 min, followed by 3 min incubation with 0.1 M glycine in PBS to quench the reaction. For detection of dense granule proteins mild fixation conditions (10 min incubation in 4% paraformaldehyde) are used. Fixation and all further manipulations are carried out at room temperature. Fixed cells were permeabilised with 0.2% Triton-X100 in PBS for 20 min and blocked in 2% bovine serum albumin, 0.2% Triton-X100 in PBS for 20 min. The cells were then stained with primary antibodies followed by secondary antibodies Alexa488- or Alexa594-conjugated goat anti-mouse or goat anti-rabbit IgGs (Molecular Probes). After 4,6-diamidino-2-phenylindole (DAPI) staining, slides were mounted in Fluoromount G (Southern Biotech) and stored at 4°C in the dark. Micrographs were obtained on a Zeiss Axioskop 2 equipped with an Axiocam color charge-coupled device camera. Images were recorded and treated on computer through the AXIOVISION™ software. ADOBE PHOTOSHOP (Adobe Systems) was used for processing of images. Confocal images were collected with a Leica laser scanning confocal microscope (TCS-NT DM/IRB) using a x63 Plan-Apo objective with numerical aperture (NA) 1.40. Optical sections were recorded at 250 nm per vertical step and four times averaging.

Cell invasion assays in the presence of carbohydrate inhibitors

These were done as described previously (Blumenschein et al, 2007). Two independent experiments were performed in triplicate. Numbers of extracellular (attached) and

intracellular (invaded) parasites were counted in three microscopic fields respectively. For the untreated controls 75 parasites were counted per microscopic field on average. Invasion efficiency of the untreated parasites was set to 100%.

Cell invasion assays with neuraminidase treatment

These were done as described previously (Blumenschein et al, 2007) using *Vibrio cholerae* neuraminidase (Calbiochem). Three experiments were carried out in parallel and evaluated in blind manner. According to the staining, numbers of extracellular (attached) and intracellular (invaded) parasites were counted in ten microscopic fields respectively. A total of 80 parasites per microscopic field were counted on average and invasion of the untreated sample was set to 100%. Enzyme activity was monitored in parallel in a protein cell binding assay using recombinant TgMIC1myc.

Fractionation experiments

$2,0 - 4,88 \times 10^8$ *T. gondii* tachyzoites freshly egressed from their host cells were harvested by centrifugation at 240g , RT for 10 minutes and the pellet was resuspended in 1ml 1×PBS by rotating 1h at 4°C. After repeated freezing/thawing and sonication, the solution was centrifuged at maximum speed, for 30min, 4°C, and the supernatant was collected (fraction soluble in 1×PBS). The pellet was then resuspended in 1ml 1×PBS/0.3% Triton by rotating 1h, 4°C. After centrifugation, the supernatant was collected (fraction soluble in 1×PBS/0.3% Triton) and the pellet was resuspended in 1ml 1×PBS/0.2M Na₂CO₃, by rotating 1h, 4°C. After centrifugation, the supernatant was collected (fraction soluble in 1×PBS/0.2M Na₂CO₃). The supernatant and pellet fractions were run in SDS-PAGE and analyzed by western-blotting.

Cloning into expression vectors

Sequences corresponding to TgMIC13 (aa 23-468), TgMCP3 (aa 54-565) and TgMCP4 (aa 526-1016) were cloned into pET32-Xa/LIC (Novagen) *E. coli* and pPICZ α *Pichia* expression vectors. The coding region for NcMIC1 (aa 17-259) was amplified from pCR-XL-TOPO-NcMIC1 using primers NcMIC1-f and NcMIC1-r and inserted into pET32-Xa/LIC.

The TgMIC13 coding sequence was inserted into pET32-Xa/LIC according to the manufacturer's instructions. For expression from pPICZ α in *P. pastoris*, TgMIC13 was cloned in between EcoRI and XbaI restriction sites.

The TgMCP3 (aa 54-565) coding sequence was reamplified from pGEM-TgMCP3 using primer pair TgMIC1.3_2285 and TgMIC1.3_2286 and cloned into pPICZ α in between AvrII and NotI and into pET32-Xa/LIC-ROM6 in between NsiI and BamHI restriction sites.

TgMCP4 lacking the signal peptide was sub-cloned from pGEM-TgMCP4 into pPICZ α between EcoRI and NotI sites. Similar, the fragment coding for TgMCP4 (aa 526-1016) amplified from pGEM-MIC1.4 with primers TgMIC1/4-9_2287 and TgMIC1/4-4_1848 was digested with EcoRI and NotI and inserted into the pPICZ α *Pichia* expression vector. For expression in *E. coli*, the same primers were used and the fragment was cloned into a pET-Xa/LIC-derived vector (pET32-Xa/LIC-ROM6) which allows for cloning of sequences into the expression locus between EcoRI and PacI restriction sites.

Generation of Antibodies

Antibodies against the TgAMA1 C-terminus were raised in rabbits (EUROGENTEC, Belgium). The protein was expressed in *E. coli* BL21 strain as GST fusion from a pGEX-4T1 expression vector (pGEX-TgAMA-1tail) and purified on Glutathione-Sepharose (Amersham) according to the manufacturer's instructions. Polyclonal antibodies against TgMIC13 were

raised in rabbits (28 days immunization protocol, EUROGENTEC, Belgium) using full-length recombinant TgMIC13 expressed as described above in *P. pastoris*.

Supplementary figure S1

Stable expression of MIC1myc in the *mic1ko* background. The ability of the epitope-tagged protein to substitute for MIC1 and bring the components of the TgMIC1-4-6 complex to the micronemes is assessed by IFA on intracellular parasites. AMA-1 is used as a micronemal marker which is independent of the TgMIC1-4-6 complex. Most of TgMIC1myc is targeted to the micronemes; however, as previously reported overexpression causes some TgMIC1myc protein to be mistargeted. Scale bars, 1µm.

Supplementary figure S2

Alignment of MAR domain containing proteins (MCPs) from *T. gondii*, *N. caninum*, *S. neurona* and *E. tenella*. The position of the sialic acid binding motif relative to TgMIC1-MARR is boxed in red. Amino acid residues in TgMIC1_M2 critical for sugar binding and binding preference are marked with an asterisk (F169, A176, R178, K216, H218, T220, E206, E207). The aligned sequences correspond to the following amino acid stretches of the full-length sequences: TgMIC1_M1 (aa 35-143), M2 (aa 144-254); NcMIC1-M1 (aa 33-141), M2 (aa 142-252); TgMIC13_M1 (aa 51-154), M2 (aa 155-263), M3 (aa 274-375), M4 (aa 376-468); NcMCP2_M1 (aa 51-154), M2 (aa 155-263), M3 (aa 274-375), M4 (aa 376-468); TgMCP3_M1 (aa 118-222), M2 (aa 223-328), M3 (aa 339-439), M4 (aa 440-542); NcMCP3_M1 (aa 125-228), M2 (aa 229-334), M3 (aa 345-445), M4 (aa 446-550); TgMCP4_M1 (aa 554-658), M2 (aa 659-763), M3 (aa 774-873), M4 (aa 874-981); NcMCP4_M1 (aa 390-494), M2 (aa 495-600), M3 (aa 611-711), M4 (aa 712-819); NcMCP5_M1 (aa 54-157), M2 (aa 158-270), M3 (aa 348-451); NcMCP6_M1 (aa 51-154),

M2 (aa 155-266); NcMCP7_M1 (aa 53-156), M2 (aa 157-266); SnMCP5_M2, M3 and SnMCP6_M1 (derived from SnEST4a79g11.y1, SnEST4a34d02.y1 and SnESTbab01e09.y1, SnEST4a69h07.y1); EtMIC3_M1 (aa 45-148), M2 (aa 153-274), M3 (aa 294-425), M4: aa 445-576, M5: aa 596-727, M6 (aa 747-886), M7 (aa 887-980); EtMCP2_M1 (aa 37-138).

Supplementary figure S3

Alignment of the full-length amino acid sequences from TgMIC13 with NcMCP2 (top) and from TgMCP3 with NcMCP3 (bottom). The *N. caninum* genes were annotated *in silico* based on EST-data and the considerable conservation between the apicomplexan proteins.

Supplementary figure S4

Alignment of the full-length amino acid sequences from TgMCP4 with NcMCP4 (top) and from NcMCP5 with the partial sequence of SnMCP5 derived from ESTs SnEST4a79g11.y1 and SnEST4a34d02.y1 (bottom). The *N. caninum* genes were annotated *in silico* based on EST-data and the considerable conservation between the apicomplexan proteins.

Supplementary figure S5

Full-length amino acid sequences from NcMCP6, NcMCP7 and EtMCP2 as annotated *in silico* by comparison with the corresponding EST data set.

Supplementary Table I

Oligosaccharide probes^a investigated in carbohydrate microarray analyses, sorted by sialyl linkage and backbone sequence.

Pos ^b	Probe	Sequence	Binding intensity at ~5 fmol		
			TgMIC1 -MARR	TgMIC13	NcMIC1 -MARR
Neutral oligosaccharide probes					
1	LacNAc	Galβ-4GlcNAc-DH	36	- ^c	-
2	LNT	Galβ-3GlcNAcβ-3Galβ-4Glc-DH	16	-	-
3	LNnT	Galβ-4GlcNAcβ-3Galβ-4Glc-DH	-	-	-
4	LNFP-III	Galβ-4GlcNAcβ-3Galβ-4Glc-DH Fucα-3 Galβ-4GlcNAcβ-6	59	-	-
5	LNH	Galβ-4GlcNAcβ-6 Galβ-4Glc-DH Galβ-3GlcNAcβ-3 Galβ-4GlcNAcβ-2Manα-6	22	-	-
6	NA2	Manβ-4GlcNAcβ-4GlcNAc-DH Galβ-4GlcNAcβ-2Manα-3	-	-	-
2,3-linked sialyl oligosaccharide probes					
Gal, Lactose and N-acetylglucosamine-based					
7	GM4	NeuAcα-3Galβ-Cer	324	-	197
8	GM3	NeuAcα-3Galβ-4Glcβ-Cer	1386	-	298
9	GM3(Gc)	NeuGcα-3Galβ-4Glc-Cer	3594	1218	270
10	Haematoside	NeuAcα-3Galβ-4Glcβ-Cer	1556	-	771
11	NeuAcα-(3')Lac	NeuAcα-3Galβ-4Glc-DH	201	-	258
12	NeuAcα-(3')Lac-AO	NeuAcα-3Galβ-4Glc-AO	6286	1529	2995
13	NeuAcβ-(3')Lac-AO	NeuAcβ-3Galβ-4Glc-AO	8	-	-
14	Neuα-(3')Lac-AO	Neuα-3Galβ-4Glc-AO	12	258	81
15	Neu4,5Ac-(3')Lac	Neu4,5Acα-3Galβ-4Glc-DH	-	-	-
16	Neu4,5Ac-(3')Lac-AO	Neu4,5Acα-3Galβ-4Glc-AO	87	5149	-
17	NeuAcα-(3')LN ₁₋₃	NeuAcα-3Galβ-3GlcNAc-DH	2558	203	1133
18	NeuAcα-(3')LN ₁₋₃ -AO	NeuAcα-3Galβ-3GlcNAc-AO	20140	9228	6573
19	NeuAcα-(3')LN	NeuAcα-3Galβ-4GlcNAc-DH	1233	-	1036
20	NeuAcα-(3')LN-AO	NeuAcα-3Galβ-4GlcNAc-AO	19654	10441	8627

21	SA(3')-Lea-Tri-AO	NeuAc α -3Gal β -3GlcNAc-AO Fuc α -4	13233	2253	5403
22	SA(3')-Lex-Tri-AO	NeuAc α -3Gal β -4GlcNAc-AO Fuc α -3	17502	6205	5845
Lacto-N-neotetraose and Lacto-N-tetraose-based					
23	LSTa	NeuAc α -3Gal β -3GlcNAc β -3Gal β -4Glc-DH	5095	145	1429
24	Sialylparagloboside	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer	1619	-	706
25	GSC-31	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36	2367	-	483
26	GSC-516B	Neu α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 SU-6	42	-	-
27	C4U	NeuAc α -3Gal β -4GlcNAc β -3Gal β -3GlcNAc-DH SU-6 SU-6 SU-6	20762	6698	10514
28	SA(3')-LNFP-II	NeuAc α -3Gal β -3GlcNAc β -3Gal β -4Glc-DH Fuc α -4	1331	-	1406
29	SA(3')-LNFP-III	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4Glc-DH Fuc α -3	1060	-	440
30	GSC-105	NeuAc α -3Gal β -4GlcNAc β -3Gal β -Cer36 Fuc α -3	91	-	61
31	GSC-64	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 Fuc α -3	346	-	249
32	GSC-472	Neu α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 Fuc α -3 SU-6	-	-	-
33	GSC-268	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 Fuc α -3 SU-6	4814	4248	27309
34	GSC-268 deNAc	Neu α -3Gal β -4GlcNA β -3Gal β -4Glc β -Cer36 Fuc α -3 SU-6	13	-	-
35	GSC-269	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 Fuc α -3 SU-6	5754	3528	5398
36	GSC-406	Neu α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 Fuc α -3 SU-6 SU-6	40	-	56
37	GSC-270	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 Fuc α -3	8199	12666	22433
38	GSC-220	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 Fuc α -3 Fuc α -3	1950	15	1745
39	GSC-221	NeuAc α -3Gal β -4GlcNAc β -3Gal β -4GlcNAc β -3Gal β -4Glc β -Cer36 Fuc α -3	2291	68	838
Branched milk					
40	MSMFLNH	Gal β -4GlcNAc β -6 Fuc α -3 Gal β -4Glc-DH NeuAc α -3Gal β -3GlcNAc β -3	4005	217	989

<i>N</i> -glycan					
41	A2F(2-3)	$ \begin{array}{c} \text{NeuAc}\alpha\text{-3Gal}\beta\text{-4GlcNAc}\beta\text{-2Man}\alpha\text{-6} \quad \text{Fuc}\alpha\text{-6} \\ \quad \quad \quad \\ \text{Man}\beta\text{-4GlcNAc}\beta\text{-4GlcNAc} \\ \\ \text{NeuAc}\alpha\text{-3Gal}\beta\text{-4GlcNAc}\beta\text{-2Man}\alpha\text{-3} \end{array} $	6290	447	2259
Ganglioside-based					
42	GM2	$ \begin{array}{c} \text{GalNAc}\beta\text{-4Gal}\beta\text{-4Glc}\beta\text{-Cer} \\ \\ \text{NeuAc}\alpha\text{-3} \end{array} $	7724	1181	7106
43	GM1b	NeuAc α -3Gal β -3GalNAc β -4Gal β -4Glc β -Cer*	49	-	88
44	GM1	$ \begin{array}{c} \text{Gal}\beta\text{-3GalNAc}\beta\text{-4Gal}\beta\text{-4Glc}\beta\text{-Cer} \\ \\ \text{NeuAc}\alpha\text{-3} \end{array} $	32	-	155
45	GM1-penta	$ \begin{array}{c} \text{Gal}\beta\text{-3GalNAc}\beta\text{-4Gal}\beta\text{-4Glc-DH} \\ \\ \text{NeuAc}\alpha\text{-3} \end{array} $	93	-	656
46	GM1(Gc)	$ \begin{array}{c} \text{Gal}\beta\text{-3GalNAc}\beta\text{-4Gal}\beta\text{-4Glc}\beta\text{-Cer} \\ \\ \text{NeuGc}\alpha\text{-3} \end{array} $	75	94	117
47	GM1(Gc)-penta	$ \begin{array}{c} \text{Gal}\beta\text{-3GalNAc}\beta\text{-4Gal}\beta\text{-4Glc-DH} \\ \\ \text{NeuGc}\alpha\text{-3} \end{array} $	3683	1233	3051
48	GD1a	$ \begin{array}{c} \text{NeuAc}\alpha\text{-3Gal}\beta\text{-3GalNAc}\beta\text{-4Gal}\beta\text{-4Glc}\beta\text{-Cer} \\ \\ \text{NeuAc}\alpha\text{-3} \end{array} $	1814	1046	1184
49	GD1a-hexa	$ \begin{array}{c} \text{NeuAc}\alpha\text{-3Gal}\beta\text{-3GalNAc}\beta\text{-4Gal}\beta\text{-4Glc-DH} \\ \\ \text{NeuAc}\alpha\text{-3} \\ \\ \text{GalNAc}\beta\text{-4Gal}\beta\text{-3GalNAc}\beta\text{-4Gal}\beta\text{-4Glc}\beta\text{-Cer} \\ \quad \quad \\ \text{NeuGc}\alpha\text{-3} \quad \text{NeuAc}\alpha\text{-3} \end{array} $	10726	6123	5078
50	GalNAc-GD1a(Ac,Gc)	$ \begin{array}{c} \text{GalNAc}\beta\text{-4Gal}\beta\text{-3GalNAc}\beta\text{-4Gal}\beta\text{-4Glc}\beta\text{-Cer} \\ \quad \quad \\ \text{NeuAc}\alpha\text{-3} \quad \text{NeuGc}\alpha\text{-3} \end{array} $	2695	-	107

2,6-linked sialyl oligosaccharide probes

2,6-linked sialyl oligosaccharide probes					
51	NeuAc α -(6')Lac	NeuAc α -6Gal β -4Glc-DH	85	-	-
52	NeuAc α -(6')Lac-AO	NeuAc α -6Gal β -4Glc-AO	426	-	262
53	NeuAc β -(6')Lac-AO	NeuAc β -6Gal β -4Glc-AO	57	-	-
54	Neu α -(6')Lac-AO	Neu α -6Gal β -4Glc-AO	25	-	-
55	NeuAc α -(6')LN	NeuAc α -6Gal β -4GlcNAc	19	-	-
56	NeuAc α -(6')LN-AO	NeuAc α -6Gal β -4GlcNAc-AO	1352	122	388
57	Neu5,9Ac-(6')LN	Neu5,9Ac α -6Gal β -4GlcNAc-DH	-	-	55

Lacto-*N*-neotetraose and Lacto-*N*-tetraose-based

58	LSTb	$ \begin{array}{c} \text{Gal}\beta\text{-3GlcNAc}\beta\text{-3Gal}\beta\text{-4Glc-DH} \\ \\ \text{NeuAc}\alpha\text{-6} \end{array} $	195	-	-
59	LSTc	NeuAc α -6Gal β 4-GlcNAc β 3-Gal β 4-Glc-DH	1090	-	229
60	GSC-97	$ \begin{array}{c} \text{NeuAc}\alpha\text{-6Gal}\beta\text{-4GlcNAc}\beta\text{-3Gal}\beta\text{-4Glc}\beta\text{-Cer}^{36} \\ \\ \text{Fuc}\alpha\text{-3} \end{array} $	15	-	-

61	SA(6')-LNFP-VI	NeuAc α -6Gal β -4GlcNAc β -3Gal β -4Glc-DH Fuc α -3	158	-	106
Branched milk					
62	MSLNH	NeuAc α -6Gal β -4GlcNAc β -6 Gal β -4Glc-DH	94	-	-
63	MSLNnH-I	Gal β -3GlcNAc β -3 Gal β -4GlcNAc β -6 Gal β -4Glc-DH	268	-	43
64	DSLNNH	NeuAc α -6Gal β -3GlcNAc β -3 NeuAc α -6Gal β -4GlcNAc β -6 Gal β -4Glc-DH	333	-	161
65	MFMSLNnH	NeuAc α -6Gal β -4GlcNAc β -3 Gal β -4GlcNAc β -6 Fuc α -3 Gal β -4Glc-DH NeuAc α -6Gal β -3GlcNAc β -3	41	-	-
N-glycan					
66	A2(2-6)	NeuAc α -6Gal β -4GlcNAc β -2Man α -6 Man β -4GlcNAc β -4GlcNAc-DH NeuAc α -6Gal β -4GlcNAc β -2Man α -3	712	-	183

2,3:2,6-linked sialyl oligosaccharide probes

O-glycan					
67	DST	NeuAc α -3Gal β -3GalNAc-DH NeuAc α -6	2386	2525	4200
Lacto-N-neotetraose and Lacto-N-tetraose-based					
68	DSLNT	NeuAc α -3Gal β -3GlcNAc β -3Gal β -4Glc-DH NeuAc α -6	5230	1296	1617
N-glycan					
69	A3	NeuAc α -3Gal β -4GlcNAc β -2Man α -6 Man β -4GlcNAc β -4GlcNAc-DH NeuAc α -3Gal β -4GlcNAc β -4Man α -3 NeuAc α -6Gal β -4GlcNAc β -2	2345	26	855

2,8-linked sialyl oligosaccharide probes

Ganglioside-based					
70	GD3	NeuAc α -8NeuAc α -3Gal β -4Glc β -Cer	-	-	-
71	GD3-tetra	NeuAc α -8NeuAc α -3Gal β -4Glc-DH	-	-	112
72	GD3-tetra-AO	NeuAc α -8NeuAc α -3Gal β -4Glc-AO	99	-	15
73	GD2	GalNAc β -4Gal β -4Glc β -Cer NeuAc α -8NeuAc α -3	13	-	-
74	GD1b	Gal β -3GalNAc β -4Gal β -4Glc β -Cer NeuAc α -8NeuAc α -3	-	-	-
75	GQ1b	NeuAc α -8NeuAc α -3Gal β -3GalNAc β -4Gal β -4Glc β -Cer NeuAc α -8NeuAc α -3	6	-	-

Polysialyl					
76	SA2(α 8)	NeuAc α -8NeuAc-DH	25	-	-
77	SA3(α 8)	NeuAc α -8NeuAc α -8NeuAc-DH	113	-	13
78	SA4(α 8)	NeuAc α -8NeuAc α -8NeuAc α -8NeuAc-DH	454	-	2143
79	SA5(α 8)	NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc-DH ^d	2720	326	8011
80	SA6(α 8)	NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc-DH ^d	869	-	4714
81	SA7(α 8)	NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc-DH ^d	681	-	4100
82	SA8(α 8)	NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc-DH ^d	180	-	1366
83	SA9(α 8)	NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc-DH ^d	71	-	1152
84	SA10(α 8)	NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc-DH ^d	45	-	417
85	SA11(α 8)	NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc α -8NeuAc-DH ^d	82	-	630
2,3:2,8-linked sialyl oligosaccharide probes					
Ganglioside-based					
86	GT1a	NeuAc α -8NeuAc α -3Gal β -3GalNAc β -4Gal β -4Glc β -Cer NeuAc α -3	9	-	-
87	GT1b	NeuAc α -3Gal β -3GalNAc β -4Gal β -4Glc β -Cer NeuAc α -8NeuAc α -3	2742	2426	1821
2,9-linked sialyl oligosaccharide probe					
88	GSC-96	NeuAc α -9NeuAc α -3Gal β -4Glc β -Cer36	151	>>30000	1261

^a The oligosaccharide probes are all lipid-linked, and are from the collection assembled in the course of research in Glycosciences Laboratory. DH, designates NGLs prepared from reducing oligosaccharides by reductive amination with the amino lipid, 1,2-dihexadecyl-*sn*-glycero-3-phosphoethanolamine (DHPE; AO, NGLs prepared from reducing oligosaccharides by oxime ligation with an aminooxy (AO) functionalized DHPE (Liu et al., Chem. Biol. 14, 847–859, 2007); Cer, natural glycolipids with various ceramide moieties; Cer36, synthetic glycolipids with ceramide having a total of 36 carbon atoms. Neu, denotes de-N-acetylated neuraminic acid.

^bPos, position in screening microarray.

^c-, signal less than 1.

^dMajor component.

Supplementary Table II

Primer sequences

Primer name	Primer sequence	Gene to be amplified	Target vector
MIC1/2-5/1047 (sense)	5'-CGAATTCCCTTTTTCGACAAAATGGTGGGTGTTT TCCAGCTG-3'	TgMIC13	pTUB8TgMLCTy_HX
MIC1/2-6/1048 (antisense)	5'-CCCTGCAGGGCACTCTGTCTGAGGCGCTTTTCA TAGC-3'	TgMIC13	pTUB8TgMLCTy_HX
MIC1/2-1/1147 (sense)	5'-CCCGAATTCATCCAGCTTCGCCAGCATCG-3'	TgMIC13	pPICZ α
MIC1/2-2/1148 (antisense)	5'-GGGTCTAGAATGCACTCTGTCTGAGGCGCTT TTC-3'	TgMIC13	pPICZ α
TgMIC1.3_2285 (sense)	5'-cgatgcatcctaggGATCGTTTTGTCAACAAGGACC GAG-3'	TgMAR-CP3	pET32-Xa/LIC and pPICZ α
TgMIC1.3_2286 (reverse)	5'-ccggatccgcgcccgcGTATCCCCGTAGCCTTCGA TCctc-3'	TgMAR-CP3	pET32-Xa/LIC and pPICZ α
MIC1/3-7_1981 (sense)	5'-gatatcctctaccaatATGGCGAAAGCC-3'	TgMAR-CP3	pTUB8TgMLCTy_HX
MIC1/3-7_1982 (reverse)	5'-ctgcagAGTATCCCCGTAGCCTTCGATCCTCG-3'	TgMAR-CP3	pTUB8TgMLCTy_HX
TgMIC1/4-7_1983 (sense)	5'-gaattcgaccacgATGGTGCCCTTGTCGAGAATGGC-3'	TgMAR-CP4	pTUB8TgMLCTy_HX
TgMIC1/4-8_1984 (reverse)	5'-atgcatATTTACTGGTAGGATTTGCCGGGCC-3'	TgMAR-CP4	pTUB8TgMLCTy_HX
TgMIC1/4-9_2287 (sense)	5'-cgatgcatgaattcCCAGTGCCAACCCGGCCTGTGC-3'	TgMAR-CP4 (partial)	pPICZ α and pET32Xa/LIC
TgMIC1/4-3_1847 (sense)	5'-gaattcatgcatAATGTGTTTGGGGAGATGCAACC GCG-3'	TgMAR-CP4	pPICZ α and pROP1mycMIC1-GLD
TgMIC1/4-4_1848 (reverse)	5'-TTAATTAAGCGGCCGCTTTACTGGTAGGATTTC GCCGGGCC-3'	TgMAR-CP4	pPICZ α , pET32Xa/LIC and pROP1mycMIC1-GLD
TgMIC1-20_1731	5'-GACAAGCGGCATTATGCAGAAGAGGAAGGAA TTCG-3'	TgMIC1	pM2MIC1myc
TgMIC1-21_1732	5'CGAATTCCTTCCTCTTCTGCATAATGCCGCTTG TC-3'	TgMIC1	pM2MIC1myc
NcMIC1-f	5'GGTATTGAGGGTCGCGCTGGGGCAAAAACATACG GAGAAGCGTCG 3'	NcMIC1	pET32Xa/LIC

NcMIC1-r	5'AGAGGAGAGTTAGAGCCTTATTCTGGGACGTTGC CGTTATCATCTAAACA 3'	NcMIC1	pET32Xa/LIC
----------	---	--------	-------------

References

Blumenschein TM, 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(11): 2808-2820

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 4(1): 1-13

Brecht S, Carruthers VB, Ferguson DJ, Giddings OK, Wang G, Jakle U, Harper JM, Sibley LD, Soldati D (2001) The *Toxoplasma* micronemal protein MIC4 is an adhesin composed of six conserved apple domains. *J Biol Chem* 276(6): 4119-4127

Carruthers V, Boothroyd JC (2007) Pulling together: an integrated model of *Toxoplasma* cell invasion. *Curr Opin Microbiol* 10(1): 83-89

Carruthers VB, Hakansson S, Giddings OK, Sibley LD (2000) *Toxoplasma gondii* uses sulfated proteoglycans for substrate and host cell attachment. *Infect Immun* 68(7): 4005-4011

Carruthers VB, Sibley LD (1997) Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *European Journal of Cell Biology* 73(2): 114-123

Carruthers VB, Tomley FM (2008) Microneme proteins in apicomplexans. *Subcell Biochem* 47: 33-45

Cerede O, Dubremetz JF, Bout D, Lebrun M (2002) The *Toxoplasma gondii* protein MIC3 requires pro-peptide cleavage and dimerization to function as adhesin. *EMBO J* 21(11): 2526-2536

Cerede O, Dubremetz JF, Soete M, Deslee D, Vial H, Bout D, Lebrun M (2005) Synergistic role of micronemal proteins in *Toxoplasma gondii* virulence. *J Exp Med* 201(3): 453-463

Deutscher SL, Nuwayhid N, Stanley P, Briles EI, Hirschberg CB (1984) Translocation across Golgi vesicle membranes: a CHO glycosylation mutant deficient in CMP-sialic acid transport. *Cell* 39(2 Pt 1): 295-299

El Hajj H, Papoin J, Cerede O, Garcia-Reguet N, Soete M, Dubremetz JF, Lebrun M (2008) Molecular signals in the trafficking of *Toxoplasma gondii* protein MIC3 to the micronemes. *Eukaryot Cell* 7(6): 1019-1028

Esko JD, Rostand KS, Weinke JL (1988) Tumor formation dependent on proteoglycan biosynthesis. *Science* 241(4869): 1092-1096

Feizi T, Chai W (2004) Oligosaccharide microarrays to decipher the glyco code. *Nat Rev Mol Cell Biol* 5(7): 582-588

Fox BA, Ristuccia JG, Gigley JP, Bzik DJ (2009) Efficient gene replacements in *Toxoplasma gondii* strains deficient for nonhomologous end joining. *Eukaryot Cell* 8(4): 520-529

Fukuda MN, Dell A, Oates JE, Fukuda M (1985) Embryonal lactosaminoglycan. The structure of branched lactosaminoglycans with novel disialosyl (sialyl alpha 2----9 sialyl) terminals isolated from PA1 human embryonal carcinoma cells. *J Biol Chem* 260(11): 6623-6631

Garnett JA, Liu Y, Leon E, Allman S, Friedrich N, Saouros S, Curry S, Soldati-Favre D, Davis B, Feizi T, Matthews S (2009) Detailed insights from microarray and crystallographic studies into carbohydrate recognition by microneme protein 1 (MIC1) of *Toxoplasma gondii*. *Protein Sci*

Gubbels MJ, Li C, Striepen B (2003) High-throughput growth assay for *Toxoplasma gondii* using yellow fluorescent protein. *Antimicrob Agents Chemother* 47(1): 309-316

Harper JM, Hoff EF, Carruthers VB (2004) Multimerization of the *Toxoplasma gondii* MIC2 integrin-like A-domain is required for binding to heparin and human cells. *Mol Biochem Parasitol* 134(2): 201-212

Huynh MH, Carruthers VB (2009) Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking Ku80. *Eukaryot Cell* 8(4): 530-539

Huynh MH, Opitz C, Kwok LY, Tomley FM, Carruthers VB, Soldati D (2004) Trans-genera reconstitution and complementation of an adhesion complex in *Toxoplasma gondii*. *Cell Microbiol* 6(8): 771-782

Inoue S, Poongodi GL, Suresh N, Jennings HJ, Inoue Y (2003) Discovery of an alpha 2,9-PolyNeu5Ac glycoprotein in C-1300 murine neuroblastoma (clone NB41A3). *J Biol Chem* 278(10): 8541-8546

Iwersen M, Dora H, Kohla G, Gasa S, Schauer R (2003) Solubilisation and properties of the sialate-4-O-acetyltransferase from guinea pig liver. *Biol Chem* 384(7): 1035-1047

Keller N, Naguleswaran A, Cannas A, Vonlaufen N, Bienz M, Bjorkman C, Bohne W, Hemphill A (2002) Identification of a *Neospora caninum* microneme protein (NcMIC1) which interacts with sulfated host cell surface glycosaminoglycans. *Infect Immun* 70(6): 3187-3198

Kessler H, Herm-Gotz A, Hegge S, Rauch M, Soldati-Favre D, Frischknecht F, Meissner M (2008) Microneme protein 8--a new essential invasion factor in *Toxoplasma gondii*. *J Cell Sci* 121(Pt 7): 947-956

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(1): 43-53

Liu Y, Feizi T, Campanero-Rhodes MA, Childs RA, Zhang Y, Mulloy B, Evans PG, Osborn HM, Otto D, Crocker PR, Chai W (2007) Neoglycolipid probes prepared via oxime ligation for microarray analysis of oligosaccharide-protein interactions. *Chem Biol* 14(7): 847-859

Martin MJ, Rayner JC, Gagneux P, Barnwell JW, Varki A (2005) Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of N-glycolylneuraminic acid. *Proc Natl Acad Sci U S A* 102(36): 12819-12824

Meissner M, Reiss M, Viebig N, Carruthers VB, Toursel C, Tomavo S, Ajioka JW, Soldati D (2002) A family of transmembrane microneme proteins of Toxoplasma gondii contain EGF-like domains and function as escorters. *J Cell Sci* 115(Pt 3): 563-574

Monteiro VG, Soares CP, de Souza W (1998) Host cell surface sialic acid residues are involved on the process of penetration of Toxoplasma gondii into mammalian cells. *FEMS Microbiol Lett* 164(2): 323-327

Montoya JG, Liesenfeld O (2004) Toxoplasmosis. *Lancet* 363(9425): 1965-1976

Orlandi PA, Klotz FW, Haynes JD (1992) A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of Plasmodium falciparum recognizes the terminal Neu5Ac(alpha 2-3)Gal- sequences of glycophorin A. *J Cell Biol* 116(4): 901-909

Ortega-Barria E, Boothroyd JC (1999) A Toxoplasma lectin-like activity specific for sulfated polysaccharides is involved in host cell infection. *J Biol Chem* 274(3): 1267-1276

Palma AS, Feizi T, Zhang YB, Stoll MS, Lawson AM, Diaz-Rodriguez E, Campanero-Rhodes MA, Costa J, Gordon S, Brown GD, Chai WG (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(9): 5771-5779

Persson KE, McCallum FJ, Reiling L, Lister NA, Stubbs J, Cowman AF, Marsh K, Beeson JG (2008) Variation in use of erythrocyte invasion pathways by Plasmodium falciparum mediates evasion of human inhibitory antibodies. *J Clin Invest* 118(1): 342-351

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(3): 563-578

Rinninger A, Richet C, Pons A, Kohla G, Schauer R, Bauer HC, Zanetta JP, Vlasak R (2006) Localisation and distribution of O-acetylated N-acetylneuraminic acids, the endogenous substrates of the hemagglutinin-esterases of murine coronaviruses, in mouse tissue. *Glycoconj J* 23(1-2): 73-84

Saouros S, Blumenschein TM, Sawmynaden K, Marchant J, Koutroukides T, Liu B, Simpson P, Carpenter EP, Matthews SJ (2007) High-level bacterial expression and purification of apicomplexan micronemal proteins for structural studies. *Protein Pept Lett* 14(5): 411-415

Saouros S, Edwards-Jones B, Reiss M, Sawmynaden K, Cota E, Simpson P, Dowse TJ, Jakle 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(46): 38583-38591

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(11): 1149-1155

Sim BK, Orlandi PA, Haynes JD, Klotz FW, Carter JM, Camus D, Zegans ME, Chulay JD (1990) Primary structure of the 175K *Plasmodium falciparum* erythrocyte binding antigen and identification of a peptide which elicits antibodies that inhibit malaria merozoite invasion. *J Cell Biol* 111(5 Pt 1): 1877-1884

Soldati-Favre D (2008) Molecular dissection of host cell invasion by the apicomplexans: the glideosome. *Parasite* 15(3): 197-205

Takabatake N, Okamura M, Yokoyama N, Ikehara Y, Akimitsu N, Arimitsu N, Hamamoto H, Sekimizu K, Suzuki H, Igarashi I (2007) Glycophorin A-knockout mice, which lost sialoglycoproteins from the red blood cell membrane, are resistant to lethal infection of *Babesia rodhaini*. *Vet Parasitol* 148(2): 93-101

Tolia NH, Enemark EJ, Sim BK, Joshua-Tor L (2005) Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. *Cell* 122(2): 183-193

Xia D, Sanderson SJ, Jones AR, Prieto JH, Yates JR, Bromley E, Tomley FM, Lal K, Sinden RE, Brunk BP, Roos DS, Wastling JM (2008) The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. *Genome Biol* 9(7): R116

Figure 1

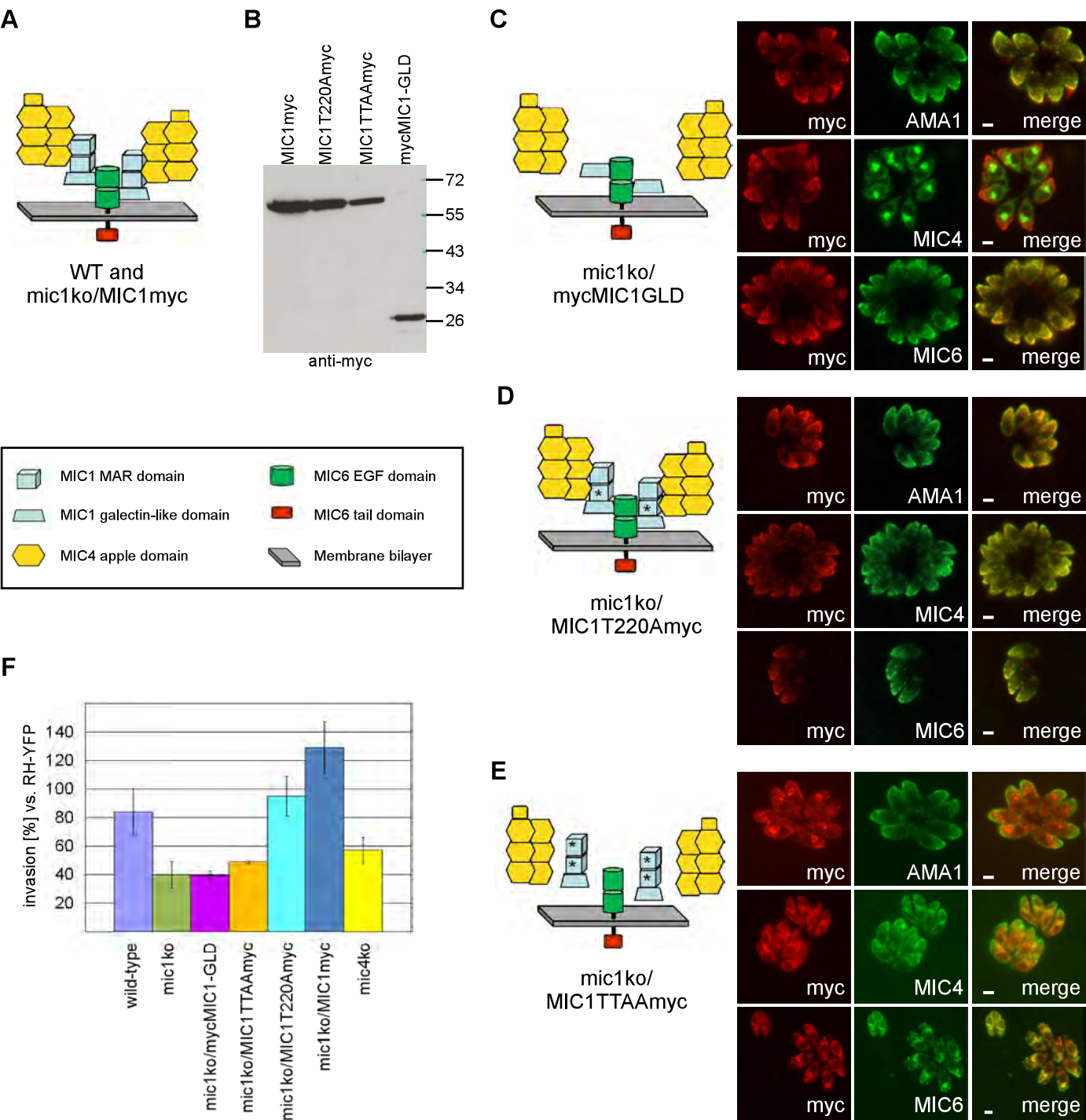


Figure 2

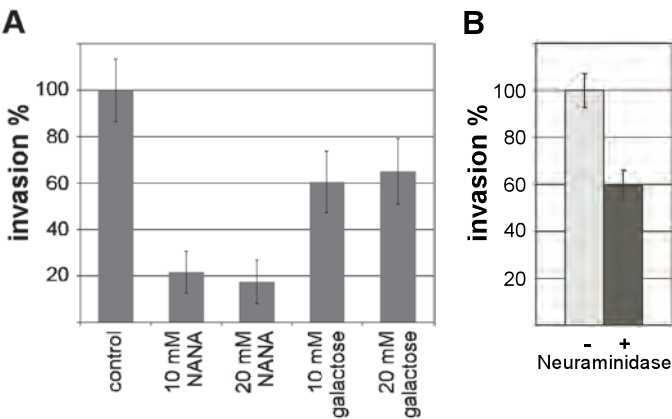


Figure 3

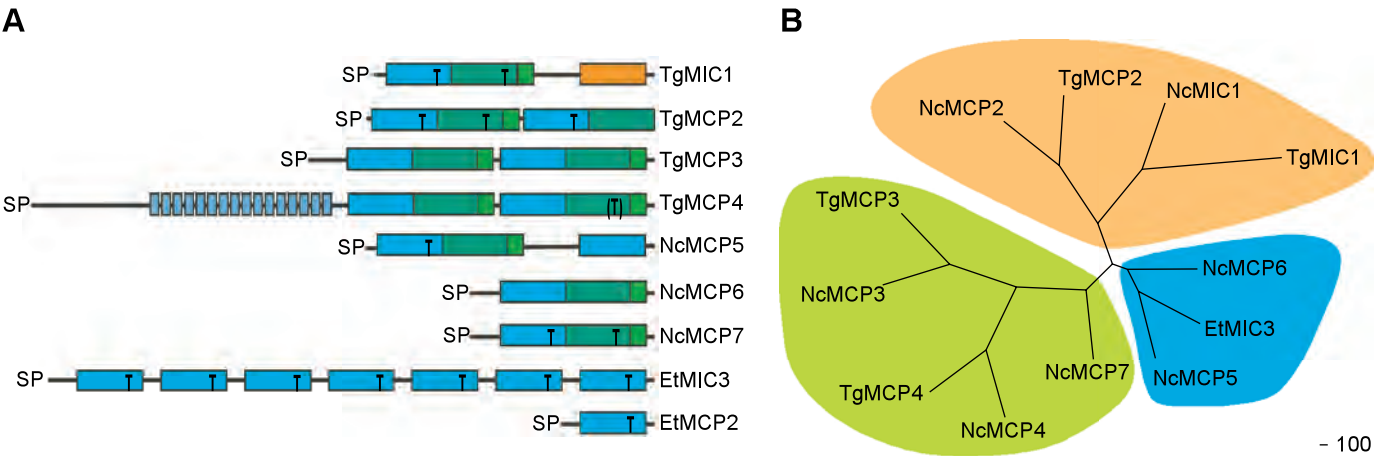


Figure 4

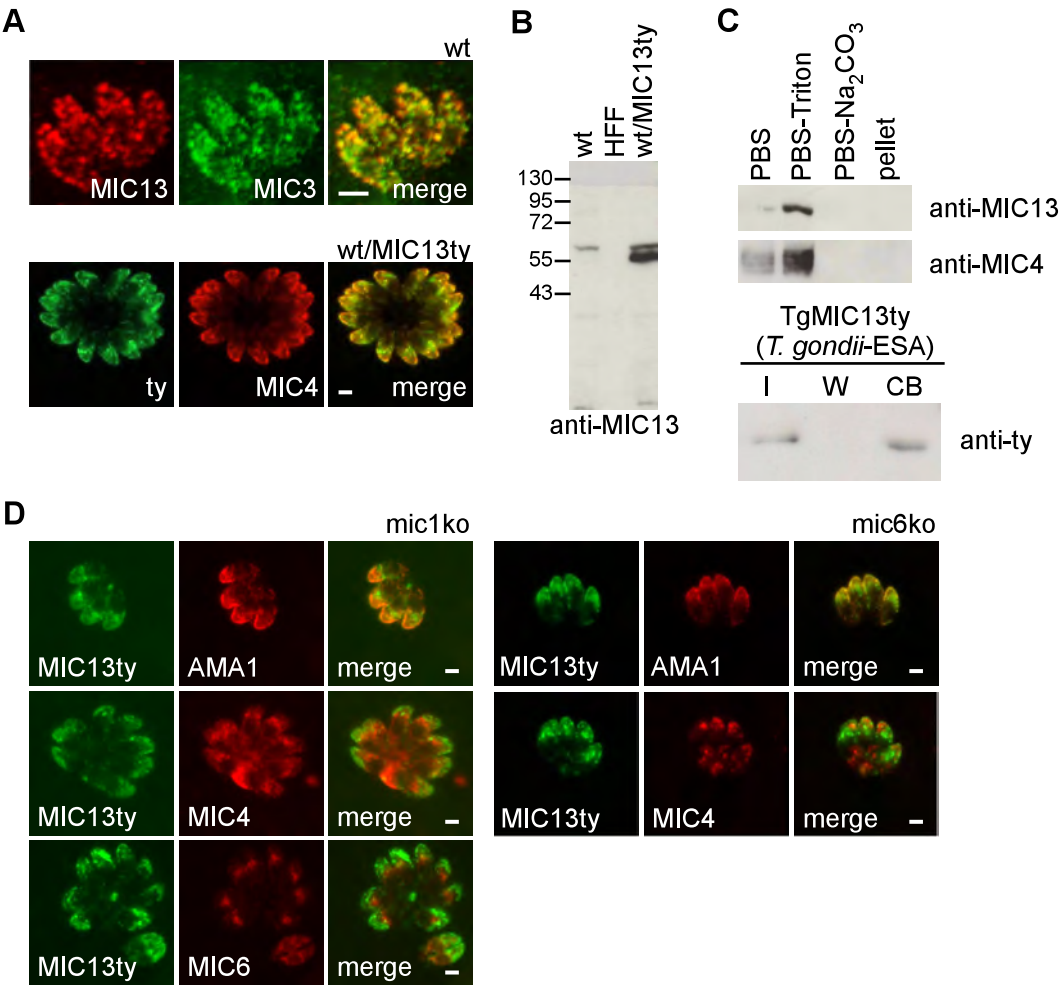


Figure 5

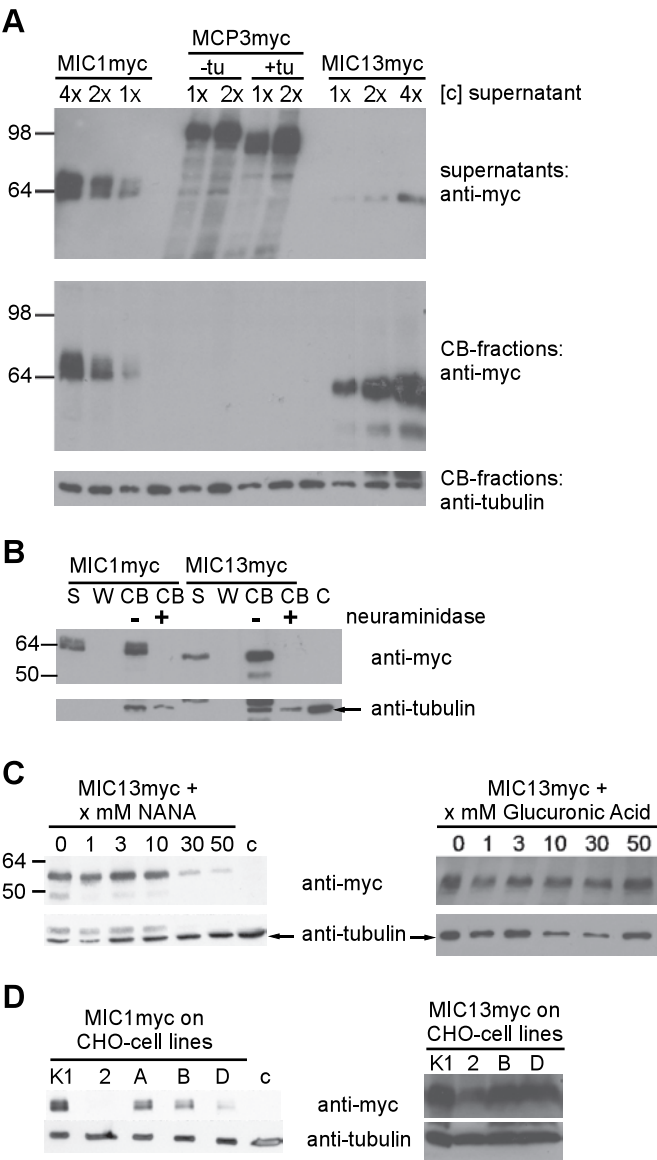


Figure 6

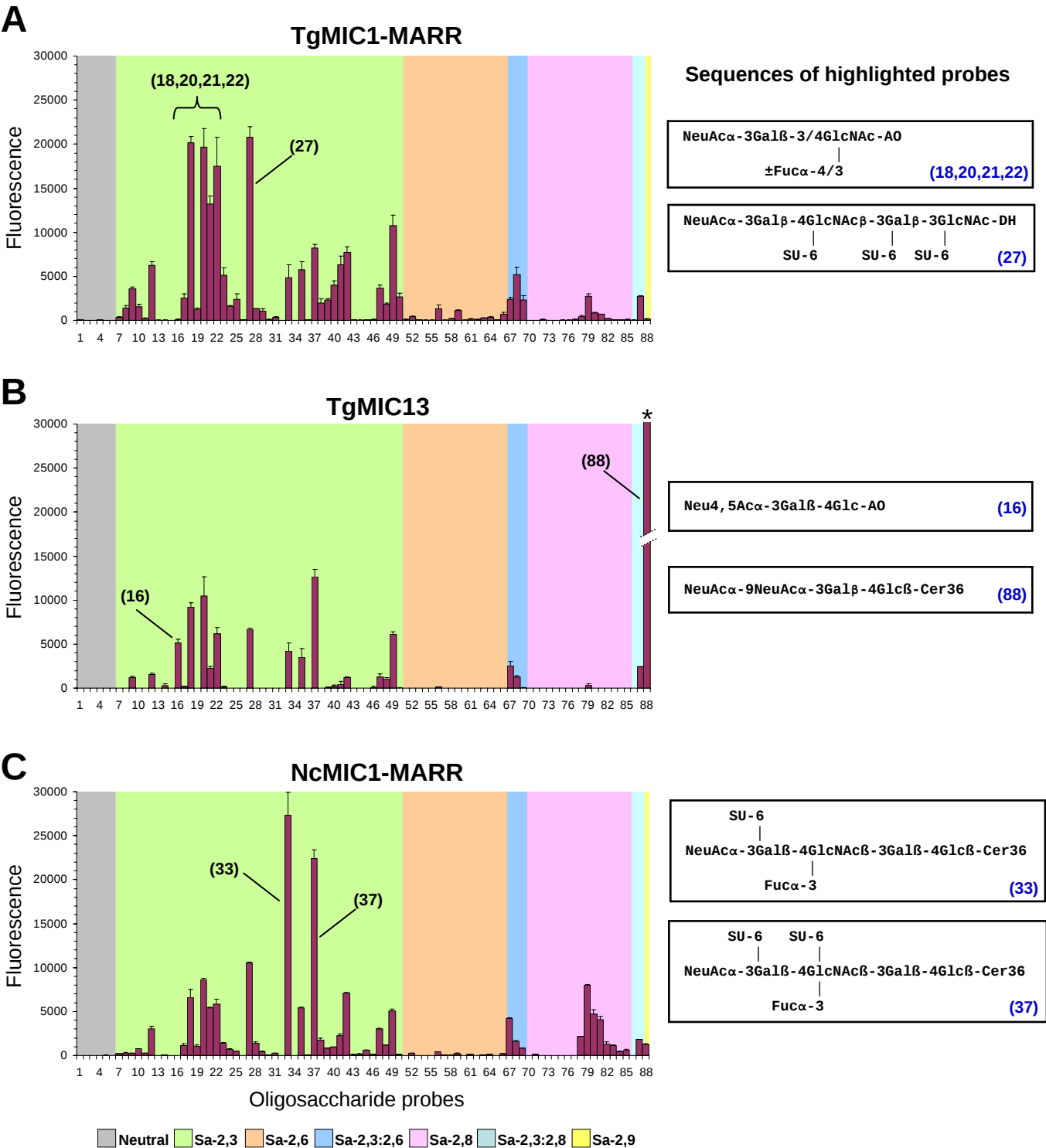


Figure 7

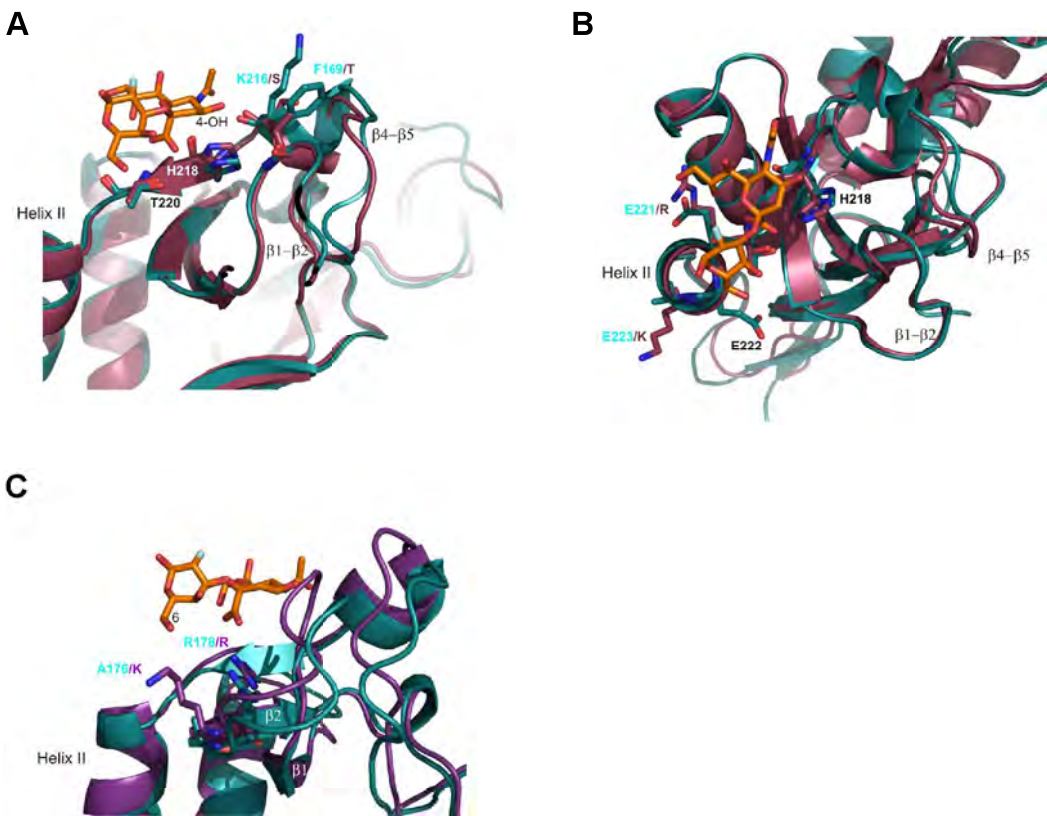
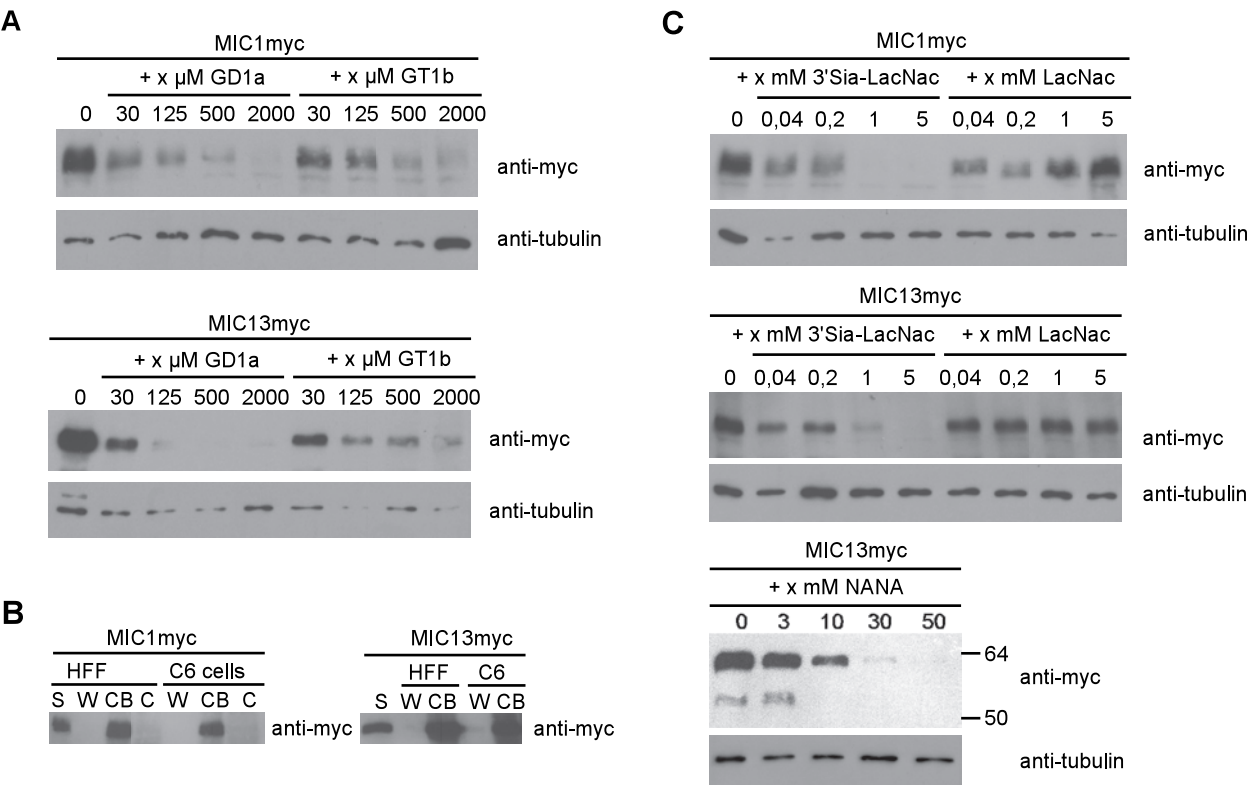
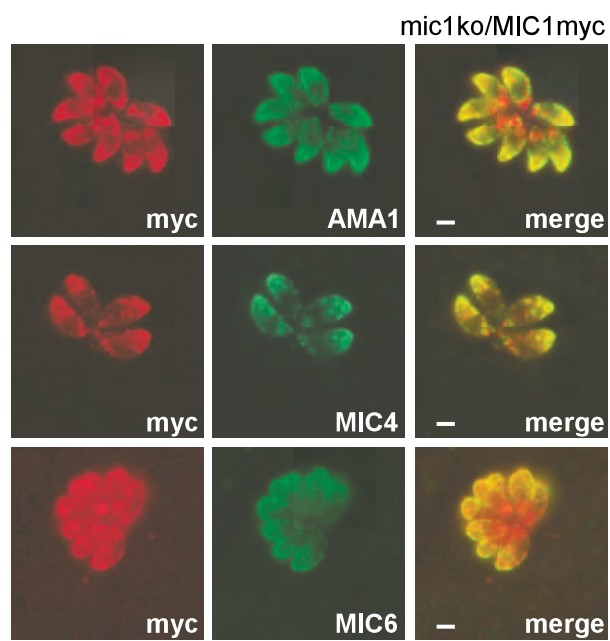


Figure 8



Supplementary figure 1



Supplementary figure 2

TgMIC1_M1	RYQQMQRQQLAAELCQ-SGL-RKMCVP--SSRI--RNAVL--ITHQNTLQPCID--TASL--ESNOENNGVNC--DLCHHTIP--PAG--	HRQENHATRHBI--SK--EGVQRFC--
NcMIC1_M1	RSQCTTCQVFKKLCDDGY-SKMCIP--ANQV--RQGLA--RKDQKLV--RCID--SAAF--AEGDENNVSLC--DDCAVSIPIPGC--	DRNETHARHDE--SQ--EGGVVRC--
TgMIC13_M1	ESQAEBSCHNRKQLCD-QGN-KKPCD--KTG--RYGTG--IQSQQTAR--PCF--AEV--KPS--GRKVQC--DDCNHYFFPLG--	DPNDFFHAHANQ--LNF--LDGVEKHC--
NcMCP2_M1	ETQAE--SACNNR--NQLCH--GGR--PD--CD--KIA--RYGTG--LQSQQTAR--PCF--AEV--KPS--GRKVQC--DDCNHYFFPLG--	EPNKEFFAHAAQ--LQF--AGVEKHC--
TgMCP3_M1	EQQD--DYCAKR--EQCCAGVLGKT--CTDG--LV--GRYVSQ--GSKL--EAG--HC--K--GA--LAHS--QRQVQC--RMCNLMKCFD--	NSQTKAKVMPG--DDW--DRKKWTVC--
NcMCP3_M1	EQQD--EDCAQ--EQLCAKAVQRK--CTKD--PV--GRYDPR--ARSV--EAG--QCHI--STENDHR--GRKVQC--RMCNLMKCFD--	RSETIKVAMPR--DDW--RRKKWTVC--
TgMCP4_M1	RIQQD--RACCTQ--QKLCSETIDGR--CGVS--GA--PFGPI--GHDS--DSQ--RCIE--ESI--INTT--TKGVQC--DDCNQMQLVPA--	SVKVTATPPLRA--SAF--SGRRWLYC--
NcMCP4_M1	RIQQD--RACCTAQ--QKLCSETIDAGSS--CRVH--SA--PFGPI--GHDS--DSQ--RCIE--ESI--INTT--TKGVQC--DDCNHMQLTET--	SAQVATLPQAA--SAF--SKRRWLYC--
NcMCP5_M1	PQLBETVCQK--KALCE--EGK--DA--CN--YELA--RYGAI--KETQDDMR--RCID--DRF--GTD--SSNVQC--DDCNNAIPLGE--	NPSTVEHTM--KD--ETE--VNRVPERC--
NcMCP6_M1	DTQEV--EFCQK--EALCA--TNL--PT--CN--HKA--PFGK--NKSQKNDV--RCIE--FSK--QST--KREVCQ--DDCNHYVPLCSG--	KSTIVVHTRAE--QS--ARESARLC--
NcMCP7_M1	EALQK--DTCTE--EALCA--TNL--PT--CN--HKA--PFGK--NKSQKNDV--RCIE--FSK--QST--KREVCQ--DDCNHYVPLCSG--	NPSTTIHARHSE--LNF--QOVSHFC--
EtMIC3_M1	RIQAV--EXLCGIF--NDLCT--LGV--OP--CN--GDA--RYGAI--KADTKDEAR--CA--GKT--GNS--ETKMQC--DDCYTFPEGNA--	NGDELTLTERP--QE--PAKITKSTC--
EtMIC3_M2	ESQD--DECSQAASVACR--SGL--QD--CN--EAVF--RYDGI--KSPSSQK--RCIT--EGA--DLE--AAGPGC--MCGAATMCLGAF--	NSRSVALTRNELTS--LRTQEEFC--
EtMIC3_M3/4/5	PTQED--ERKCAQFKESCN--SGL--AS--PCG--AAIY--RYDGI--NRSQTKR--RCII--EKE--DFS--KSGDGC--DDCSMTSRES--	SGPSTSHLTRPSE--QA--DSDKSNYC--
EtMIC3_M6	EALQK--EACVSE--GEKACS--NGL--RA--CN--SAMY--EKDVQ--NTTQTKR--RCIA--ASE--DFA--VSADVCAN--MGNITS--SGA--	NGVSTSHLTRNPQ--EE--MEKEVYC--
EtMIC3_M7	DCQV--SARCIYA--AQICI--TDN--SMCD--VA--RKVG----TS--DGP--QSA--DLS--RSGECC--BACNVVLCRGP--	NGVSTSHLTRNPQ--EE--MEKEVYC--
EtMCP2_M1	IIHR--DGLCLMT--TRACIEN--PH--CV--EA--RGVY--GSSQREEA--RCIS--VKE--DFS--LSRREC--DDCDIIECQAFS--	TGTAPSGAEDVG--DK--BELKEATC--
TgMIC1_M2	SPYQASAKKCNDA--P--GT--PRSK--*--FGNVVEVA--RCIE--KAS--LYS--VVAECASNC--TTWY--PGRRG--TSTELDKRHT--TEESG--	ROA--GSV--DSPCS--EVEVCLPKDENPPPLCLD--
NcMIC1_M2	EGFOANANS--CN--RP--GT--RKSK--FGHKEPVK--RCIE--K--PES--LFS--VFESE--SNCTTWS--PG--GRGLG--TATNLDKKHF--DESG--	LOA--TSV--PKACP--VGLVCLPRDQNPPLCLD--
TgMIC13_M2	EPFQMGNN--CQGVLA--DT--KDDTS--TASQAKA--RCIE--K--KDS--TYE--ARSIC--ONCEIE--PG--GRSE--TSGELL--SQHY--REKE--	QK--SQ--SAPCH--DACVATPINPPICAD--
NcMCP2_M2	EPFQD--GKK--CQGVLA--NS--RKDAE--TSSQQRK--RCIE--K--KDS--TYE--ARSIC--ONCEIE--PG--GRSE--TSGELL--SQHY--REKE--	QK--FQSG--AKPCS--G--ACVPTPVNPPVPCAD--
TgMCP3_M2	RYQQALALCNLAH--HGD--RYDHH--GQR--MCLD--LGD--ALD--RISYCAN--CGSPT--SVGEG--LPGLPR--SVRLVRH--	ESAF--AI--QPCPNAGD--CVSAVNPPLCLP--
NcMCP3_M2	RYQQALALCNLAH--HGD--RYDHH--GQR--MCLD--LGD--ALD--RISYCAN--CGSPT--SVGEG--LPGLPR--SVRLVRH--	ESAF--AI--QPCPNAGD--CVSAVNPPLCLP--
TgMCP4_M2	SPQF--ADALCNFR--TGN--RYDP--RTR--MCN--PVD--SLE--GOSYCAN--CGFT--PCAG--SPEIPT--RV--LIQSDTA--FAAL--	HPCRH--DEICIPSTVYPPACV--
NcMCP4_M2	ISFOR--EALALCNFR--TGT--RYDPV--RTR--MCN--AVD--SLE--GOSYCAN--CGFT--PCAG--SPEIPT--RV--LIQSDTA--FAAL--	HPCRH--DEICIPSTVYPPACV--
NcMCP5_M2	SPQQT--MECNLREGWVA--RG--LTSSAE--DQ--RCID--VAK--EYH--LQSQC--MCSRLRHP--GREGPDRNLTEND--	DIASDPCE--AALVCLISTPKNPAKCVS--
NcMCP5_M2	SPFQAL--RDCNNLKEGWVA--RKGL--EEER--SKT--RCID--VAV--EYH--LQSQC--MCSRLRHP--GREGPDRNLTEND--	DIASDPCE--AALVCLISTPKNPAKCVS--
NcMCP6_M2	PSSON--ANAKCTQQA--GS--RKDAE--LTSVNQGI--RCID--PSS--AYE--QKSTCASN--CEEVK--CHGRIDMYSEAVSAKH--	TERDKE--QDA--NKN--KTDC--GKICVNTVASPAQCVD--
NcMCP7_M2	SPQF--ADDCNRF--TKGSVA--RKA--HAGQS--AGGNEN--TGN--TFE--TVGLCA--MCKDEECV--GPNK--KHGT--DNPHGE--	DSD--NTF--PQO--SNGCS--SGGGCLSTPVNPPPLCLD--
TgMIC13_M3	RAQHF--DACQL--NCHCK--KDK--K--CT--TVVARE--DRGR--AGSQNDA--RCIS--LEE--DFQ--KTNPS--CDECNELP--QQA--	PEESTHHITWTK--PSK--DEATQKFC--
NcMCP2_M3	RAQHF--DACQOE--DRCLK--RDK--K--CK--TVVARE--DRGR--AGSQDTE--RCIA--LEE--NFQ--SKGSAC--DECNELP--QQG--	PEASTHHITWQD--PSK--DEAQKFC--
TgMCP3_M3	PAQPF--ETWCQAMDECR--KGKQTEDCF--GLW--RYELE--SPGGAGAR--RCIP--ESR--DFT--QLS--SC--GCNPNIS--GRQP--	AEVSPADIDFA--DR--AGD--MSYC--
NcMCP3_M3	VEQOE--ZAACQAMDECR--KGKQTEDCF--GLW--RYDI--RTPRETTA--RCIP--ESR--DFT--RLS--KC--GCNPNIS--GRQA--	AEVSPADIDFA--DR--AGD--MSYC--
TgMCP4_M3	SKQEE--ENTCQ--MDQCR--VGKWTTCY--QLW--RMVST--EDEE--EP--RCIP--FAL--NFS--DLS--TC--GCNPNIS--GRQA--	SPSTLTTEFSK--KT--AIN--IAPC--
NcMCP4_M3	KQOQ--GNICQ--QMDQLCR--KGQWTTCY--QLW--RMVGS--QKANGP--RCIP--FAL--NFS--DLS--TC--GCNPNIS--GRQA--	LTSVLTAAFPK--AI--AIN--PAPC--
NcMCP5_M3	QS--QGL--HARCAE--NYLCT--SENLA--CT--QPVARE--DYGV--YTGFEESI--RCIS--FAN--DQE--RKS--IC--MCAEYV--CDGG--	EPTFVVH--QOWSA--GK--TORKNDRC--
NcMCP5_M3	GT--SAL--ESICLA--DYLCE--TQQLDH--CLDDGDVA--VARE--DLGR--QDGAESPVR--RCVE--KQK--DKS--RSS--IC--MCAEYV--CDGG--	EPTFVVH--QOWSA--GK--TORKNDRC--
TgMIC13_M4	TERQF--AND--ICASEYA--DS--RYGPI--AGDQTSQ--RCIK--KAA--RAD--AEGQCA--CKGTET--CGGRSLR--DGLPSV--DDV--	NAHSG--SEAF--KSA--STEC--
NcMCP2_M4	ARQF--AND--ICESIP--DS--RYGS--AGDHTSQ--RCIK--KDA--RAD--AEGQCA--CKGTET--CGGRSLR--DGLPSV--DDV--	NAHSG--SEAF--KSA--STEC--
TgMCP3_M4	SAYQF--GUA--CAQHG--PRW--QONCN--TYK--ACIN--LDN--PPN--GWVGAGHC--EFVWCP--SAGT--GAVSD--SVQ--	DED--EST--FLAV--PDPCL--AGEKCEPTAQEPAYCSV--
NcMCP3_M4	SYQF--GUA--CAQHG--SGW--QONCN--TYK--ACIN--LDN--PPN--GWVGAGHC--EFVWCP--SAGT--GAVSD--SVQ--	DED--EST--FLAV--PDPCL--AGEKCEPTAQEPAYCSV--
TgMCP4_M4	SPQF--ANDHCNFR--G--EGW--QDWT--TDD--AC--Q--TTK--PTY--RWTC--MNCNTSL--CAAG--PEQLHSA--AAEAFVHP--DFA--ER--	SV--PSPCL--TGESCPTTAMNPPYCSE--
NcMCP4_M4	SPYQANAND--CAK--ING--EGW--QDWT--TDD--AC--Q--TTK--PTY--RWTC--MNCNTSL--CAAG--PEQLHSA--AAEAFVHP--DFA--ER--	SV--PSPCL--TGESCPTTAMNPPYCSE--

type I

MAR-domain

15

type II

MAR-domain

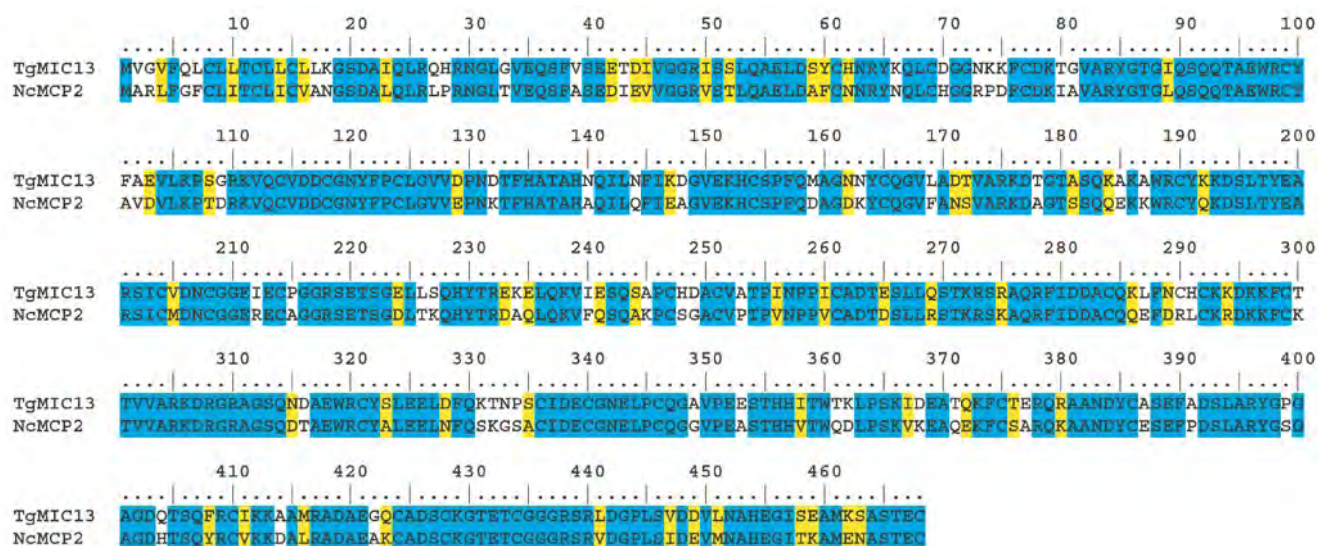
type I

MAR-domain

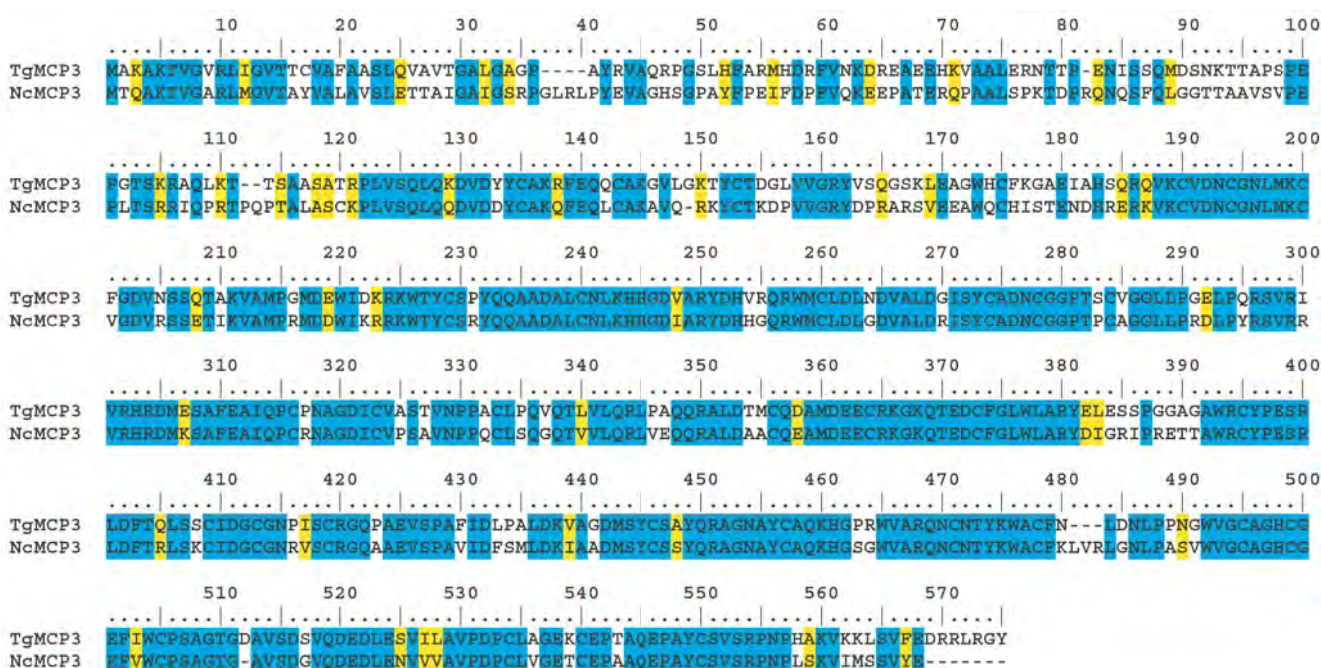
type II

MAR-domain

Supplementary figure 3

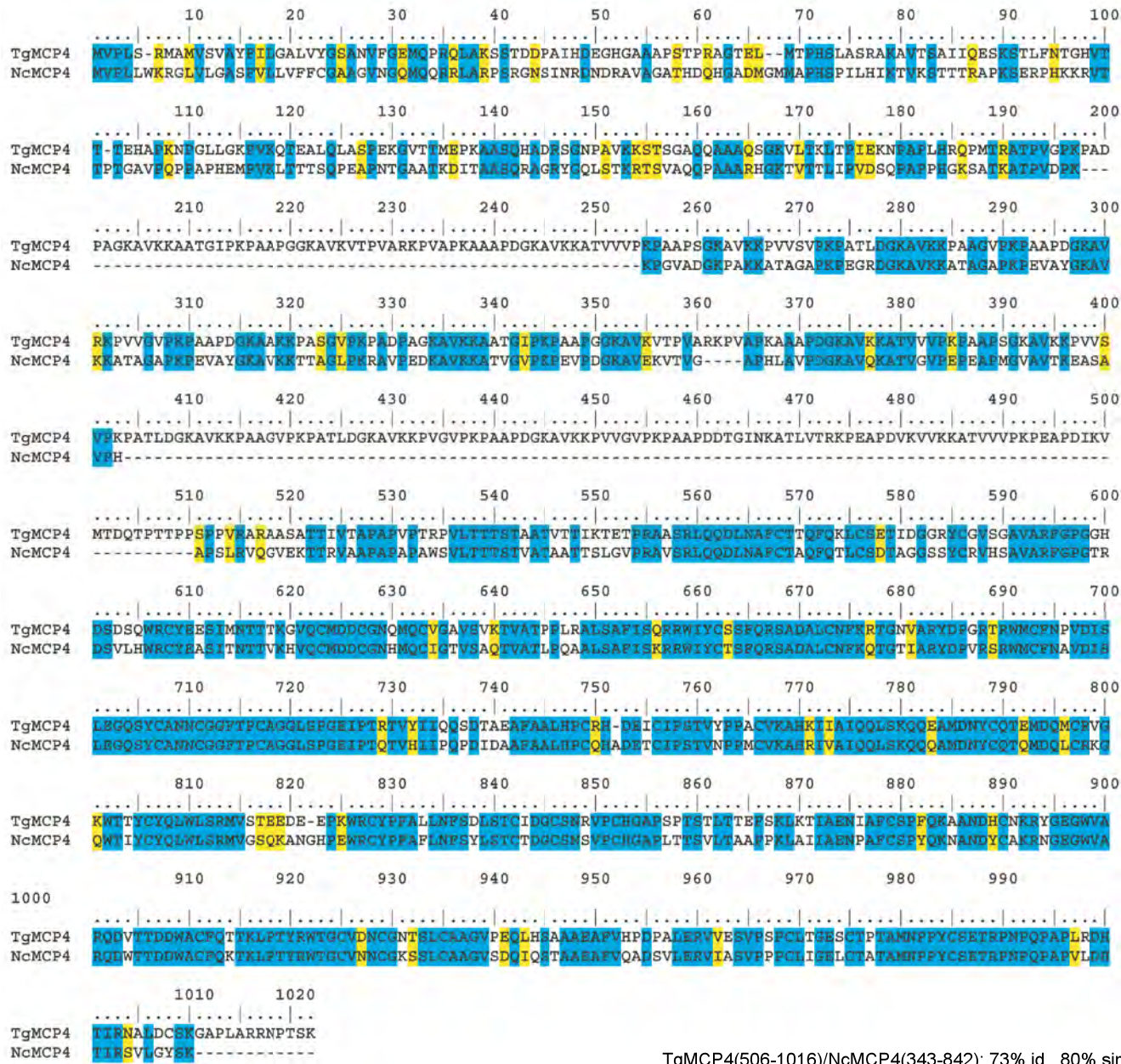


TgMIC13/NcMCP2: 73% id., 86% sim.

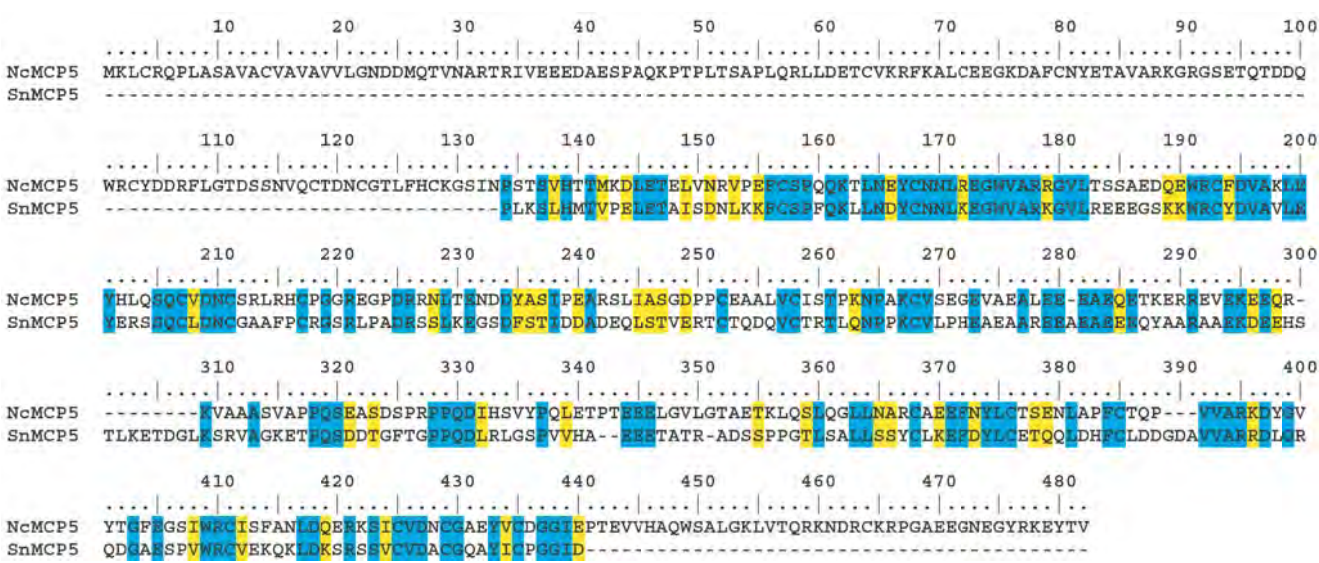


TgMCP3/NcMCP3: 67% id., 75% sim.

Supplementary figure 4



TgMCP4(506-1016)/NcMCP4(343-842): 73% id., 80% sim.



NcMCP5(134-427)/SnMCP5(partial): 37% id., 54% sim.

Supplementary figure 5

10 20 30 40 50 60 70 80 90 100
NcMCP6 MAFNFFCLGSTIAAVLASSQPAHAFAFERRLFGKIEGAQPVSPQLTLGVDDTLQEVLD EY CQKKFKALCASGDKVFCNRLAVARYAKGKENQDADMWRCF
110 120 130 140 150 160 170 180 190 200
NcMCP6 DDRSIVSSNRSVSCVDDCGNAIPCLGEVKSITVVHTTRHAEIQSLVARESARLCPSSQKAANAYCKTQQAGSVARKDAGLTSVNQGIIEWRCYDPSSLAYE
210 220 230 240 250 260
NcMCP6 QKSTCASNCGEEVKCHGGRIDMVSEAVSAKHVTRDKEIQDAINKNKTDCCKGKICVNTVASPAQCVD

10 20 30 40 50 60 70 80 90 100
NcMCP7 MTFNFLCIGSAVLAVATSSQAVHALERNLAGSIENDAMTSASLLSEDQMTVTLSALQLKLDTWCTEEFARLCATNLPTFCNHKAVARFGKGNKSQKNDVWR
110 120 130 140 150 160 170 180 190 200
NcMCP7 CYEPSKLQSTKREVQCVDNCGNYVPCLGSVNPSTTIHASRHSEILNFLQQQVSHFCSPIQKAADDFCNRKTKGSVARKALAHAGQSAGGWNCENTGNLTF
210 220 230 240 250 260 270 280 290 300
NcMCP7 ETVGLCADNCGKDEECVGGPNSKHGTDTNPHGEDSDLNTFVKQQSNGCSSGGGCLSTPVPNPPLCLDGVESGTHAESATESDHDRIGEIEFLCRGDNEPSC
310 320 330 340
NcMCP7 SQITKQKMLCPASSGTGGFQCLAAAEPGSMMLHTAKLKDCILCKR

10 20 30 40 50 60 70 80 90 100
EtMCP2 MKFFASRSLGAAAIALAVSCTASANCGTQGGVQEPPSILHRAMDGLCLETFTRACIENPHFCVEAVARRGVGGSSQREEAWRCYSVKELDFSLSRRECVD
110 120 130 140 150
EtMCP2 DCGDIIECQGAVSSENSLEHLSVTDILVKLLEDARHGTCKMQKRSIGNVALQR

4. Discussion

A wide spectrum of pathogens including bacteria and protozoa has adapted to an intracellular lifestyle which presents several advantages, including accessibility to nutrients from the host cell and partial protection from the host immune system. These pathogens have developed numerous strategies to enter their respective host cells. Sialic acid is abundantly found on cell surface glycoproteins and glycolipids and is widely exploited by viruses and bacteria to enter host cells. The most prominent example is probably the binding of influenza viruses to airway epithelium (Suzuki, 2005). Carbohydrate-bearing receptors are also important for host cell invasion by protozoan parasites of the phylum Apicomplexa and in particular sialic acid is now recognized to critically contribute to host cell recognition by a variety of species. Glycophorins A, B, and C are sialoglycoproteins of human erythrocytes that serve as important and highly specific receptors for the *P. falciparum* DBL-EBP family ligands. Sialoglycophorins on RBCs seem also to play an important role in *Babesia* infections (Takabatake et al, 2007a; Takabatake et al, 2007b). In this study we have shown that sialic acid constitutes a major determinant for host cell invasion by *T. gondii* and have characterized TgMIC1 and TgMIC13 as the key molecular players involved in Sia-dependent host cell recognition.

4.1 Function of MAR containing proteins (MCPs) in *T. gondii*

The TgMIC1-4-6 complex has been formally demonstrated to function in host cell invasion (Cerede et al, 2005; Reiss et al, 2001) and all available data supports a role for a hypothetical TgMIC13 containing complex in invasion as well. Despite several studies on the TgMIC1-4-6 complex and our efforts on the characterization of TgMIC13 it appears still difficult to pinpoint the exact role of these adhesins. In particular the timing and coordination of the individual parasite ligand-host cell receptor interactions during the invasion process is so far poorly understood.

4.1.1 Connection of adhesive complexes to the actin-myosin motor

Previous studies on the characterization of the *mic1ko* failed to detect any significant defect in parasite attachment to host cells (Cerede et al, 2005) which is in agreement with our own observations. This might be due to the experimental conditions used; alternatively the adhesive function of the TgMIC1-4-6 complex could be critical during the penetration process. So far it was impossible to localize this complex or TgMIC13 during invasion. Motility and invasion in apicomplexan parasites is driven by the glideosome (Soldati-Favre, 2008). Similar to TgMIC2 (Jewett & Sibley, 2003), *in vitro* experiments show that TgMIC6 is able to connect to this motor machinery via interaction with the F-actin binding protein aldolase ((Zheng et al, 2009) and Sheiner *et al.*, in revision). Therefore the TgMIC1-4-6 complex potentially provides a link between the host cell surface and the parasite motor. Whether this link is needed for reorientation, penetration or other mechanistic aspects of invasion is not yet known. However, our own unpublished results failed to detect an association of TgMIC6 with the motor machinery in the parasite during invasion. The interaction of TgMIC2 with aldolase relies on a tryptophan residue in the extreme C-terminus of the cytoplasmic tail (Buscaglia et al, 2003; Jewett & Sibley, 2003; Starnes et al, 2006) which is conserved in TgMIC12 and TgMIC6. A *mic6ko* strain was complemented with a mutant carrying a tryptophan to alanine substitution in this position. Subsequent invasion assays did not show any defect of the strain complemented with the mutant compared to a control strain complemented with a wild-type copy (data not shown). This result suggests that interaction of MIC6 with aldolase might not be important for the function of the TgMIC1-4-6 complex. Alternatively rearward translocation of the mutant complex over the parasite surface might be achieved through association with other TM-MICs capable of interacting with the glideosome making the interaction of TgMIC6 with aldolase redundant.

4.1.2 Function of TgMCP3 and TgMCP4

As we could not detect any carbohydrate binding for TgMCP3 and TgMCP4, their function remains so far a mystery. Parasite lines expressing epitope-tagged TgMCP3 and TgMCP4 were generated (Joana Santos, unpublished results). Both C-terminally tagged proteins were detected as a single species on Western Blot. In contrast to the majority of microneme proteins that target to the organelle when expressed as second copy with an epitope tag,

TgMCP3ty localized primarily to the dense granules and accumulates in the PV when tagged at either termini and under the control of tubulin or rhoptry promoters. Similar observations were made for C-terminally tagged TgMCP4. Although we can not exclude that the presence of the tag interferes with targeting, this behaviour suggests that level or timing of TgMCP3 and TgMCP4 expression is important, possibly reflecting a limited accessibility of their escorter(s). Alternatively, TgMCP3 and TgMCP4 might be *bona fide* dense granule proteins but all attempts to localise the endogenous proteins remained inconclusive so far. Despite the technical hitches, the similarity between the MAR domains in these two proteins at the amino acid level points towards a similar function and therefore the subcellular localisation of TgMCP4 is likely to mimic TgMCP3.

4.1.3 Complex formation by MCPs

Our data demonstrates that TgMIC13 does not belong to the TgMIC1-4-6 complex. Since TgMIC13 lacks a membrane spanning domain, we reasoned that the protein is likely to be associated to a TM-MIC which would serve as an escorter and anchor TgMIC13 to the parasite surface during invasion. Concerning the architecture of the hypothetical TgMIC13 containing complex, it is interesting to note that a dozen PAN-domain containing proteins with unknown function are present in the *T. gondii* genome (Figure 4.1). PAN domains classically serve extracellular functions and indeed several of these proteins have been localized to apical secretory organelles (Chen et al, 2008). Given that the MAR domain of TgMIC1 is interacting with the PAN domains of TgMIC4 ((Saouros et al, 2005) and Prof. Stephen Matthews, unpublished), it is plausible that TgMIC13 as well as TgMCP3 and TgMCP4 are each forming complexes with homologues of TgMIC4 or with other PAN-domain containing proteins.

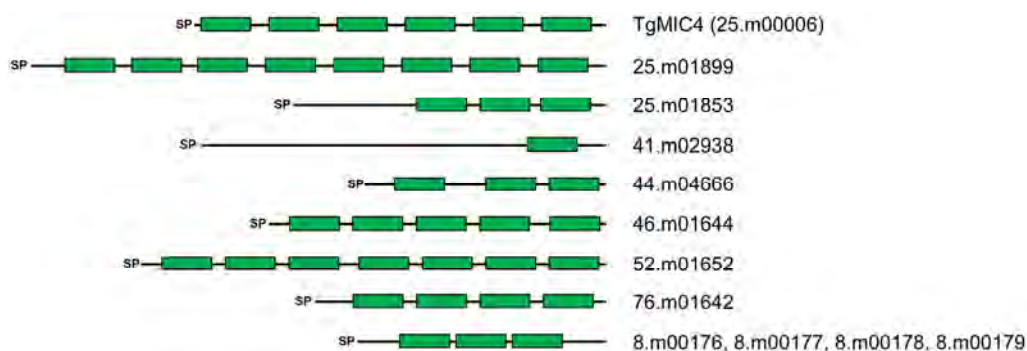


Figure 4.1: Schematic of PAN-domain containing proteins encoded in the genome of *T. gondii*. Accession numbers for toxoDB are indicated to the right.

For many lectins oligomerization was demonstrated to be important for interaction with their respective ligands. Oligomerization is a strategy to enhance the affinity and specificity of lectins toward carbohydrate ligands. For example multimerization of the major *Plasmodium* sporozoite surface protein, the circumsporozoite protein (PfCS), was suggested to be a requirement for sugar binding (Cerami et al, 1992b; Sinnis et al, 1994). This protein binds to heparin sulfate proteoglycans on the basolateral side of hepatocytes which is exposed to the blood stream (Cerami et al, 1992a; Frevert et al, 1993). In this way PfCS is believed to be involved in the selective attachment of sporozoites traveling in the blood circulation to these cells. Attachment is followed by invasion of hepatocytes, which initiates the liver stage development of the parasite. Structural studies have highlighted the necessity for PfEBA 175 to dimerize prior to interaction with the sialoglycoprotein glycophorin A during RBC invasion (Tolia et al, 2005). Oligomerization was also reported for a major micronemal antigen of *Sarcocystis muris* which is a galactose-specific lectin (Klein et al, 1998), as well as for several lectins in *T. gondii*, the best characterized being TgMIC2 and TgMIC3 (see introduction and (Cerede et al, 2002; Harper et al, 2004)). Analysis of recombinant TgMIC1 expressed in *P. pastoris* by native PAGE and gel filtration did not reveal any sign of oligomerization for the protein on its own (data not shown). Still the protein shows binding activity in cell binding assays showing that dimerization is not an absolute requirement. Instead it seems that dimerization has been replaced by the particular stoichiometry of the TgMIC1-4-6 complex which allows for the presentation of two molecules of TgMIC1 in close proximity. The results from our invasion assays suggest that although the presentation of a single TgMIC1 is functional in invasion, this arrangement of the complex indeed allows the complex to bind to host cell receptor(s) with increased affinity which in turn results in enhanced invasion efficiency. So far we have not addressed whether TgMIC13 is able to oligomerize on its own or as part of a hypothetical complex. Identification of the components of such a complex will be necessary to resolve this issue.

4.2 Is there a molecular basis for the broad host cell range of cyst-forming coccidians?

Among the Apicomplexa the potential to invade a large variety of cell types is unique to some coccidian parasites. Similar to *T. gondii* tachyzoites, *N. caninum* tachyzoites and *S. neurona* merozoites are able to infect a large variety of cells (Ellison et al, 2001; Hemphill et al, 2006). This is probably related to their common behavior of spreading through tissues during acute infection. We found that recombinant TgMIC1 and TgMIC13 are able to bind to HFFs, CHO cells and C6 rat glioma cells. Considering the potential spectrum of sialyl-oligosaccharides that can be bound by TgMIC1 and TgMIC13 as revealed by carbohydrate microarray analyses, and the wide distribution of sialic acid in vertebrate tissues, it is likely that these two lectins will bind to most cell types. Therefore, we hypothesize that TgMIC1 and TgMIC13 crucially contribute to the broad host cell range of *T. gondii*. However, among the multitude of proteins known to act on the surface of *T. gondii* others might as well play a role in this phenomenon.

TgAMA1 and TgRON proteins 2, 4, 5, and 8 form a complex at the moving junction (Straub et al, 2009). Based on the fact that the RONS were found to associate with the host cell PM (Besteiro et al, 2009), it has been proposed that the parasite provides its own anchor for invasion on the host cell surface. Furthermore it has been suggested that this could explain the broad range of cells invaded by *T. gondii* (Boothroyd & Dubremetz, 2008). Although the TgAMA1-RON2/4/5/8 complex is certainly essential for invasion of any cell type, it seems unlikely that it can provide an explanation for the considerable differences in host cell preferences and tissue tropism when comparing different apicomplexan parasites. Since the function of the complex is expected to be conserved across the phylum, it would imply that all apicomplexans have a similar host cell range, which is not true. Therefore other factors specific to subgroups of parasites should be responsible for the individual characteristics of apicomplexan parasites in discriminating between cells. How should these molecular factors look like? The domain composition of adhesive proteins is believed to reflect at least to some extent their functional characteristics. For this reason, comparing the level of conservation of protein domains across the phylum might provide indications on candidate proteins. Following these deliberations, which are the protein domains specific to coccidians? At least four different domains have to be considered: The MAR domain, the PAN domain, the chitin-binding-like (CBL) domain and the SRS domain.

4.2.1 MAR and apple domains in host cell recognition

Several lines of evidence support the potential role of TgMIC1 and TgMIC13, and more in general MCPs, in recognition of a broad range of cells. First, their binding characteristics are suited for engagement with a variety of cell surface sialyl-glycoconjugates which are widely distributed in vertebrate tissues. Second, a *mic1-3ko* strain displays decreased virulence *in vivo* (Cerede et al, 2005) and infection of mice with this attenuated strain results in reduced parasite load in a variety of tissues compared to the parental strain (Moire et al, 2009). Third, the MAR domain appears to be restricted to coccidians. Interestingly, the apple domain is specific to coccidians as well (Chen et al, 2008), although a divergent fold belonging to the PAN superfamily is present in PfAMA1 and this domain is conserved in AMA-1 homologues from all Apicomplexans. (Chesne-Seck et al, 2005; Pizarro et al, 2005)). As discussed above, it is plausible that similar to the interaction between TgMIC1 and TgMIC4, other apple domain containing proteins might associate with MCPs and form similar complexes.

N. caninum tachyzoites express homologues of TgMIC1 and TgMIC4, named NcMIC1 and NcMIC4. We also identified a putative TgMIC6 homologue (Sawmynaden et al, 2008); therefore we postulated the existence of a complex analogous to TgMIC1-4-6 in *Neospora*. Binding specificities of these two proteins have been characterized previously with NcMIC1 binding to sulphated GAGs (Keller et al, 2002) and NcMIC4 binding to lactose (Keller et al, 2004). We have revisited the characterization of binding preferences for NcMIC1 and found it to be specific for sialyl-oligosaccharides. These similarities to the TgMIC1-4-6 complex together with the identification of a TgMIC13 homologue in *N. caninum* corroborate the idea that these proteins fulfill conserved functions, possibly in the capacity of these two cyst-forming parasites to invade a broad range of host cells. In support of this view, expressed sequence tags (ESTs) coding for MCPs are present in another tissue-cyst forming coccidian, *Sarcocystis neurona*, although limited availability of genomic data precluded the complete annotation of the repertoire of MCPs. In addition, three apple-domain containing proteins have been characterized in *Sarcocystis muris*, among them the two major micronemal antigens. These proteins were named *S. muris* lectin 1 and 2 (SmL-1, SmL-2). The lectin activity of SmL-1 depends on dimerization of the protein. Haemagglutination assays suggest binding to N-acetyl-galactosamine (Klein et al, 1998). SmL-2 and another lectin, SmL-3, show binding to galactose (Klein et al, 2003; Muller et al, 2001). Antibodies directed against the two major micronemal antigens localized at least one of the two proteins to the moving junction during host cell invasion (Entzeroth et al, 1992; Klein et al, 2003).

E. tenella is a coccidian which does not spread through tissues and does not form tissue-cysts but nevertheless also expresses at least one MCP, EtMIC3 (Labbe et al, 2005)) and one apple domain containing protein, EtMIC5 (Brown et al, 2000; Brown et al, 2003). Furthermore EST and genomic data indicates that more MCPs and apple domain containing proteins might be expressed. EtMIC5 is known to form a multimeric complex with EtMIC4, a transmembrane protein composed of multiple TSP-1 and EGF-like domains (Periz et al, 2007). In this study, coimmunoprecipitation experiments failed to detect an interaction between EtMIC5 and EtMIC3 or any additional protein other than EtMIC4. However, our sequence and phylogenetic analyses indicate that *Eimeria* MCPs are divergent from those present in the cyst-forming coccidians. Therefore they might serve a different function.

4.2.2 Chitin-binding-like (CBL) domains in host cell recognition

Another domain which appears to be present only in coccidian parasites is the chitin-binding-like (CBL) domain found in TgMIC3, TgMIC8 as well as in two TgMIC8 paralogues named TgMIC8-2 and TgMIC8-3. BLAST searches identified a full set of homologues in *N. caninum* and indicated the presence of the CBL domain in *E. tenella* (Figure 4.2). Interestingly, a tryptophan residue critical for the adhesive function of TgMIC3 is conserved among all sequences (Cerede et al, 2005). In contrast, BLAST searches failed to identify a CBL domain in the genomes of *Plasmodium*, *Theileria* or *Cryptosporidium*. TgMIC8 clearly fulfils other functions in addition to adhesion (Cerede et al, 2005; Kessler et al, 2008). Also it has to be kept in mind that no attachment defect has been noticed in parasites depleted in TgMIC8 (Kessler et al, 2008). However, this is similar to observations on the *mic1ko* strain which does not exhibit any defect in parasite attachment either (Cerede et al, 2005). Consequently, the TgMIC8-MIC3 complex is another candidate potentially contributing to the capacity of *T. gondii* to invade such a broad range of cell types. This would be in agreement with our hypothesis since functional synergy has been demonstrated between TgMIC1 and TgMIC3 in *in vivo* infection (Cerede et al, 2005; Moire et al, 2009). Intriguingly, some chitin-binding domains can bind to N-acetylneuraminic acid (Raikhel-Lee et al, 1993), therefore it is possible that TgMIC8 and TgMIC3 form another sialyl-specific adhesive complex.

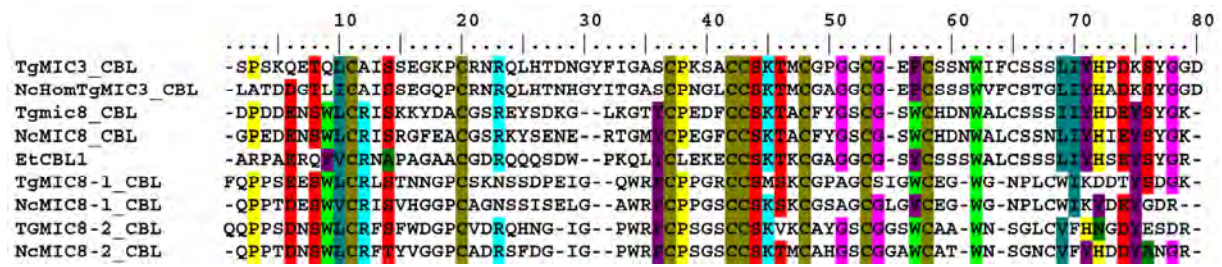


Figure 4.2: Alignment of CBL-domains. CBL domains were taken from TgMIC3, TgMIC8, TgMIC8.2, TgMIC8.3 and their homologues in *N. caninum* together with a CBL-domain identified by BLAST searches on contig 7300 of the *E. tenella* genome. Homologues in *N. caninum* were annotated *in silico* based on high similarity with the *T. gondii* proteins. In parentheses is indicated the position within the full-length sequence: TgMIC3_CBL (67-144), Tgmic8_CBL (52-126), TgMIC8-1_CBL (83-157), TGMIC8-2_CBL (108-181), NcHomTgMIC3_CBL (69-147), NcMIC8_CBL (52-127), NcMIC8-1_CBL (126-199), NcMIC8-2_CBL (87-160), EtCBL1 (contig7300).

4.2.3 Role of the surface antigen glycoprotein 1 (SAG1)-related superfamily in host cell recognition

Proteins of the SAG1-related superfamily (SRS) might also provide an explanation for the broad host cell range of *T. gondii*. The surface of *T. gondii* is populated by a family of GPI-anchored proteins of which surface antigen glycoprotein 1 (TgSAG1) is the prototypic member. The family comprises a collection of over 160 genes with stage specific expression pattern (Cleary et al, 2002; Jung et al, 2004). Among them TgSAG1 codes for the most abundantly expressed surface protein of the tachyzoite. Orthologous genes are present in the tissue-dwelling coccidians *Neospora* and *Sarcocystis* but not in more distantly related apicomplexans (Jung et al, 2004). However, recent structural modeling studies suggested that the 6-cys-domain present in *Theileria* and in a family of *Plasmodium* gamete surface coat proteins could be related to the SRS fold (Gerloff et al, 2005; He et al, 2002; Templeton, 2007). SRS proteins were previously proposed to be involved in the primary host cell attachment and to contribute to the broad range of cell types invaded by this group of parasites. This hypothesis is based on several observations. Earlier studies using monoclonal antibodies directed against TgSAG1 and TgSAG2A showed opposite effects of inhibition and

enhancement of tachyzoite attachment to host cells respectively (Grimwood & Smith, 1992; Grimwood & Smith, 1996; Mineo et al, 1993). Potential interaction with sulfated GAGs was proposed for most SRS proteins when the crystal structure of TgSAG1 was solved (He et al, 2002). TgSAG1 forms homodimers and the two subunits build a positively charged groove conserved among SRS proteins which could serve as interaction site with negatively charged GAGs. Binding to sulfated proteoglycans was detected for recombinant TgSAG3, but not TgSAG1 (Jacquet et al, 2001). Furthermore, a strain in which the *TgSAG3* gene has been disrupted revealed a decrease in adhesion of the parasite to host cells and a reduced virulence in mice (Dzierszinski et al, 2000). Conversely TgSAG1 deficient parasites showed increased host cell invasion and enhanced tissue burden in mice, but nevertheless were impaired in virulence (Rachinel et al, 2004). Taken together, these studies suggested that the different members of the SRS-family are functionally not identical.

Although these results do not exclude a role for SRS proteins in parasite attachment, their primary function appears to be related to the establishment of a chronic infection (Kim & Boothroyd, 2005; Kim et al, 2007; Saeij et al, 2008). SRS proteins expressed in the tachyzoite, the parasite form associated with acute infection, are highly immunogenic in mice and humans and TgSAG1 and TgSAG2A are known to be the immunodominant antigens (Bessieres et al, 1992; Partanen et al, 1984). TgSAG1 induces secretion of interferon (IFN)- γ by T cells, which is the major cytokine mediating resistance against *Toxoplasma* (Khan et al, 1988; Prigione et al, 2000; Suzuki et al, 1988). Stage conversion from the tachyzoite to the encysting bradyzoite form characteristic for chronic infection is accompanied by a switch in the expression of defined subsets of SRS-proteins (Cleary et al, 2002). Analysis of mutant parasites illegitimately expressing otherwise tachyzoite (TgSAG1) or bradyzoite (SRS9) specific proteins perturbed the establishment of a chronic infection (Kim & Boothroyd, 2005). From these experiments it was concluded that the switch of the surface coat enables the parasite to persist in face of a long-lasting immune response directed against tachyzoite specific antigens. In addition the switch is accompanied by a decrease in cytokine release thereby minimizing immunopathology to the host, an aspect that contributes to the achievement of a chronic infection by the parasite.

4.3 Benefits and intervention

Microneme proteins have been considered in the development of vaccines as well as diagnostic tests for parasitic diseases. The fact that they are secreted onto the parasite surface makes them accessible to antibodies and their important role in host cell invasion together with the discovery of invasion inhibitory antibodies indicates the possibility to block this step. For example, the micronemal antigens PfAMA1 and PfEBA175 are among the leading vaccine candidates to combat malaria. Vaccination strategies against neosporosis based on NcMIC1 alone provided partial protection in mice (Alaeddine et al, 2005) and the protection was improved when a combination of recombinant NcMIC1, NcMIC3 and NcROP2 was used for vaccination (Debache et al, 2009). This latter approach conferred complete protection from morbidity and led to a significant reduction in vertical transmission of neosporosis to pups.

The occurrence of MIC-specific antibodies in toxoplasma-positive individuals has been investigated. Antibodies specific to TgMIC1 were found in human patients and in particular, antibodies directed against the N-terminal part containing the MAR-domains were associated with acquired immunity (Holec et al, 2008). Antibody responses against TgMIC2, TgMIC3, TgMIC4, TgM2AP and TgAMA1 were detected in newborn with congenitally acquired toxoplasmosis, which is of importance for the development of reliable postnatal diagnostic tests (Buffolano et al, 2005). In a parallel study, phage display was used to identify antigenic regions in a variety of MICs. Peptide fragments from TgMIC2, TgM2AP, TgMIC3 and TgMIC4 were found to be recognized by T cells and antibodies of adults with acquired and children with congenital infection (Beghetto et al, 2005). A subsequent vaccination study in mice was performed using these antigenic MIC-regions that resulted in generation of TgMIC2, TgMIC4, TgM2AP and TgAMA1 IgG antibodies and in a strong decrease in numbers of brain cysts following oral challenge. Another study reported that a purified TgMIC1-TgMIC4 sub-complex is a potent antigen and acts as an effective vaccine in the mouse model (Lourenco et al, 2006). TgMIC3 is as well a potent vaccine candidate and antibodies directed against the CBL domain and the EGF domains were found to confer protection in mice (Ismael et al, 2009; Ismael et al, 2003).

Another approach to develop an effective vaccine is the use of genetically attenuated parasite strains as live vaccine. Such an attenuated *T. gondii* strain (S48) is already commercialized by Intervet, Schering-Plough Animal Health, as a live vaccine (Toxovax) for animals which is administered before mating and prevents abortion over a period of two

years. However, there have been concerns about the use of this strain, since the genetic basis for the attenuated phenotype appears to be not well defined, meaning that revertants might occur and could potentially be transmitted to humans in undercooked meat. The genetic lesion of the *mic1-3ko* strain with reduced virulence phenotype is well defined (Cerede et al, 2005) and was proven to be an effective vaccine against chronic and congenital toxoplasmosis in mice (Ismael et al, 2006). Vaccination almost completely prevented the generation of brain cysts after oral challenge, protected congenitally infected pups from death and again dramatically reduced the load of brain cysts in the pups. Further studies indicate that this live vaccine is also protecting from abortion in sheep (Moire et al, 2009).

4.4. Concluding remarks

In summary, our understanding of the molecular mechanisms underlying host cell invasion by apicomplexan parasites is not only of fundamental interest but also of practical relevance for the development of diagnostic tests and vaccines. Recognition of host cells through microneme proteins is an aspect that relates to parasite dissemination and disease progression. In this respect the results of this work contribute to our understanding of toxoplasma infection. The aims of the study have been partially fulfilled with the identification and characterization of two key molecular players, TgMIC1 and TgMIC13, in sialic acid-dependent host cell invasion by *T. gondii*. More work is needed to elucidate the precise function of the TgMIC1-4-6 complex and TgMIC13 as well as the significance of these and other parasite ligand-host cell receptor interactions such as those mediated through PAN and CBL domains. Identification of the relevant receptor(s) carrying the sialic acid determinants used by *T. gondii* to enter a host cell is the next challenging step in this study. The robust interactions of both lectins, TgMIC1 and TgMIC13, with several cell types offer a realistic chance to identify potential receptor(s) by biochemical approaches. Since substantial quantities of recombinant protein can be produced, pull-down experiments could be easily performed. Validation of potential receptor(s) would then require knockdown of the host cell receptor by RNAi or other approaches in combination with parasite invasion assays.

References

- Adams JH, Sim BK, Dolan SA, Fang X, Kaslow DC, Miller LH (1992) A family of erythrocyte binding proteins of malaria parasites. *Proc Natl Acad Sci U S A* **89**(15): 7085-7089
- Ajioka JW, Fitzpatrick JM, Reitter CP (2001) *Toxoplasma gondii* genomics: shedding light on pathogenesis and chemotherapy. *Expert Rev Mol Med* **2001**: 1-19
- Akhouri RR, Bhattacharyya A, Pattnaik P, Malhotra P, Sharma A (2004) Structural and functional dissection of the adhesive domains of *Plasmodium falciparum* thrombospondin-related anonymous protein (TRAP). *Biochem J* **379**(Pt 3): 815-822
- Alaeddine F, Keller N, Leepin A, Hemphill A (2005) Reduced infection and protection from clinical signs of cerebral neosporosis in C57BL/6 mice vaccinated with recombinant microneme antigen NcMIC1. *J Parasitol* **91**(3): 657-665
- Alexander DL, Mital J, Ward GE, Bradley P, Boothroyd JC (2005) Identification of the moving junction complex of *Toxoplasma gondii*: a collaboration between distinct secretory organelles. *PLoS Pathog* **1**(2): e17
- Amino R, Giovannini D, Thiberge S, Gueirard P, Boisson B, Dubremetz JF, Prevost MC, Ishino T, Yuda M, Menard R (2008) Host cell traversal is important for progression of the malaria parasite through the dermis to the liver. *Cell Host Microbe* **3**(2): 88-96
- Anantharaman V, Iyer LM, Balaji S, Aravind L (2007) Adhesion molecules and other secreted host-interaction determinants in Apicomplexa: insights from comparative genomics. *Int Rev Cytol* **262**: 1-74
- Aravind L, Iyer LM, Wellems TE, Miller LH (2003) *Plasmodium* biology: genomic gleanings. *Cell* **115**(7): 771-785
- Arrizabalaga G, Ruiz F, Moreno S, Boothroyd JC (2004) Ionophore-resistant mutant of *Toxoplasma gondii* reveals involvement of a sodium/hydrogen exchanger in calcium regulation. *J Cell Biol* **165**(5): 653-662
- Baker RP, Wijetilaka R, Urban S (2006) Two *Plasmodium* rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. *PLoS Pathog* **2**(10): e113
- Barragan A, Brossier F, Sibley LD (2005) Transepithelial migration of *Toxoplasma gondii* involves an interaction of intercellular adhesion molecule 1 (ICAM-1) with the parasite adhesin MIC2. *Cell Microbiol* **7**(4): 561-568
- Bastin P, Bagherzadeh Z, Matthews KR, Gull K (1996) A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. *Mol Biochem Parasitol* **77**(2): 235-239
- Baum J, Chen L, Healer J, Lopaticki S, Boyle M, Triglia T, Ehlgren F, Ralph SA, Beeson JG, Cowman AF (2009) Reticulocyte-binding protein homologue 5 - an essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum*. *Int J Parasitol* **39**(3): 371-380
- Baum J, Pinder M, Conway DJ (2003) Erythrocyte invasion phenotypes of *Plasmodium falciparum* in The Gambia. *Infect Immun* **71**(4): 1856-1863
- Beghetto E, Nielsen HV, Del Porto P, Buffolano W, Guglietta S, Felici F, Petersen E, Gargano N (2005) A combination of antigenic regions of *Toxoplasma gondii* microneme proteins induces protective immunity against oral infection with parasite cysts. *J Infect Dis* **191**(4): 637-645
- Bessieres MH, Le Breton S, Seguela JP (1992) Analysis by immunoblotting of *Toxoplasma gondii* exo-antigens and comparison with somatic antigens. *Parasitol Res* **78**(3): 222-228

- Besteiro S, Michelin A, Poncet J, Dubremetz JF, Lebrun M (2009) Export of a *Toxoplasma gondii* rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. *PLoS Pathog* **5**(2): e1000309
- Bhanot P, Frevert U, Nussenzweig V, Persson C (2003) Defective sorting of the thrombospondin-related anonymous protein (TRAP) inhibits *Plasmodium* infectivity. *Mol Biochem Parasitol* **126**(2): 263-273
- Binder EM, Lagal V, Kim K (2008) The prodomain of *Toxoplasma gondii* GPI-anchored subtilase TgSUB1 mediates its targeting to micronemes. *Traffic* **9**(9): 1485-1496
- Black MW, Boothroyd JC (2000) Lytic cycle of *Toxoplasma gondii*. *Microbiol Mol Biol Rev* **64**(3): 607-623
- Blumenschein TM, 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**(11): 2808-2820
- Boothroyd JC, Dubremetz JF (2008) Kiss and spit: the dual roles of *Toxoplasma* rhoptries. *Nat Rev Microbiol* **6**(1): 79-88
- Brecht S, Carruthers VB, Ferguson DJ, Giddings OK, Wang G, Jakle U, Harper JM, Sibley LD, Soldati D (2001) The *Toxoplasma* micronemal protein MIC4 is an adhesin composed of six conserved apple domains. *J Biol Chem* **276**(6): 4119-4127
- Brossier F, Jewett TJ, Lovett JL, Sibley LD (2003) C-terminal processing of the *Toxoplasma* protein MIC2 is essential for invasion into host cells. *J Biol Chem* **278**(8): 6229-6234
- Brossier F, Jewett TJ, Sibley LD, Urban S (2005) A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by *Toxoplasma*. *Proc Natl Acad Sci U S A* **102**(11): 4146-4151
- Brown PJ, Billington KJ, Bumstead JM, Clark JD, Tomley FM (2000) A microneme protein from *Eimeria tenella* with homology to the Apple domains of coagulation factor XI and plasma pre-kallikrein. *Mol Biochem Parasitol* **107**(1): 91-102
- Brown PJ, Mulvey D, Potts JR, Tomley FM, Campbell ID (2003) Solution structure of a PAN module from the apicomplexan parasite *Eimeria tenella*. *J Struct Funct Genomics* **4**(4): 227-234
- Brydges SD, Harper JM, Parussini F, Coppens I, Carruthers VB (2008) A transient forward-targeting element for microneme-regulated secretion in *Toxoplasma gondii*. *Biol Cell* **100**(4): 253-264
- Brydges SD, Zhou XW, Huynh MH, Harper JM, Mital J, Adjogble KD, Daubener W, Ward GE, Carruthers VB (2006) Targeted deletion of MIC5 enhances trimming proteolysis of *Toxoplasma* invasion proteins. *Eukaryot Cell* **5**(12): 2174-2183
- Buffolano W, Beghetto E, Del Pezzo M, Spadoni A, Di Cristina M, Petersen E, Gargano N (2005) Use of recombinant antigens for early postnatal diagnosis of congenital toxoplasmosis. *J Clin Microbiol* **43**(12): 5916-5924
- Bumstead J, Tomley F (2000) Induction of secretion and surface capping of microneme proteins in *Eimeria tenella*. *Mol Biochem Parasitol* **110**(2): 311-321
- Buscaglia CA, Coppens I, Hol WG, Nussenzweig V (2003) Sites of interaction between aldolase and thrombospondin-related anonymous protein in *plasmodium*. *Mol Biol Cell* **14**(12): 4947-4957
- Buxton D, Thomson K, Maley S, Wright S, Bos HJ (1991) Vaccination of sheep with a live incomplete strain (S48) of *Toxoplasma gondii* and their immunity to challenge when pregnant. *Vet Rec* **129**(5): 89-93
- Carruthers V, Boothroyd JC (2007) Pulling together: an integrated model of *Toxoplasma* cell invasion. *Curr Opin Microbiol* **10**(1): 83-89
- Carruthers VB, Giddings OK, Sibley LD (1999a) Secretion of micronemal proteins is associated with *Toxoplasma* invasion of host cells. *Cell Microbiol* **1**(3): 225-235

- Carruthers VB, Hakansson S, Giddings OK, Sibley LD (2000a) *Toxoplasma gondii* uses sulfated proteoglycans for substrate and host cell attachment. *Infect Immun* **68**(7): 4005-4011
- Carruthers VB, Moreno SN, Sibley LD (1999b) Ethanol and acetaldehyde elevate intracellular [Ca²⁺] and stimulate microneme discharge in *Toxoplasma gondii*. *Biochem J* **342** (Pt 2): 379-386
- Carruthers VB, Sherman GD, Sibley LD (2000b) The *Toxoplasma* adhesive protein MIC2 is proteolytically processed at multiple sites by two parasite-derived proteases. *J Biol Chem* **275**(19): 14346-14353
- Carruthers VB, Sibley LD (1997) Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *Eur J Cell Biol* **73**(2): 114-123
- Carruthers VB, Tomley FM (2008) Microneme proteins in apicomplexans. *Subcell Biochem* **47**: 33-45
- Cerami C, Frevert U, Sinnis P, Takacs B, Clavijo P, Santos MJ, Nussenzweig V (1992a) The basolateral domain of the hepatocyte plasma membrane bears receptors for the circumsporozoite protein of *Plasmodium falciparum* sporozoites. *Cell* **70**(6): 1021-1033
- Cerami C, Kwakye-Berko F, Nussenzweig V (1992b) Binding of malarial circumsporozoite protein to sulfatides [Gal(3-SO₄)β1-Cer] and cholesterol-3-sulfate and its dependence on disulfide bond formation between cysteines in region II. *Mol Biochem Parasitol* **54**(1): 1-12
- Cerede O, Dubremetz JF, Bout D, Lebrun M (2002) The *Toxoplasma gondii* protein MIC3 requires pro-peptide cleavage and dimerization to function as adhesin. *EMBO J* **21**(11): 2526-2536
- Cerede O, Dubremetz JF, Soete M, Deslee D, Vial H, Bout D, Lebrun M (2005) Synergistic role of micronemal proteins in *Toxoplasma gondii* virulence. *J Exp Med* **201**(3): 453-463
- Chaturvedi S, Qi H, Coleman D, Rodriguez A, Hanson PI, Striepen B, Roos DS, Joiner KA (1999) Constitutive calcium-independent release of *Toxoplasma gondii* dense granules occurs through the NSF/SNAP/SNARE/Rab machinery. *J Biol Chem* **274**(4): 2424-2431
- Chen XM, O'Hara SP, Huang BQ, Nelson JB, Lin JJ, Zhu G, Ward HD, LaRusso NF (2004) Apical organelle discharge by *Cryptosporidium parvum* is temperature, cytoskeleton, and intracellular calcium dependent and required for host cell invasion. *Infect Immun* **72**(12): 6806-6816
- Chen Z, Harb OS, Roos DS (2008) *In silico* identification of specialized secretory-organelle proteins in apicomplexan parasites and in vivo validation in *Toxoplasma gondii*. *PLoS One* **3**(10): e3611
- Chesne-Seck ML, Pizarro JC, Vulliez-Le Normand B, Collins CR, Blackman MJ, Faber BW, Remarque EJ, Kocken CH, Thomas AW, Bentley GA (2005) Structural comparison of apical membrane antigen 1 orthologues and paralogues in apicomplexan parasites. *Mol Biochem Parasitol* **144**(1): 55-67
- Chini EN, Nagamune K, Wetzel DM, Sibley LD (2005) Evidence that the cADPR signalling pathway controls calcium-mediated microneme secretion in *Toxoplasma gondii*. *Biochem J* **389**(Pt 2): 269-277
- Choe H, Moore MJ, Owens CM, Wright PL, Vasilieva N, Li W, Singh AP, Shakri R, Chitnis CE, Farzan M (2005) Sulphated tyrosines mediate association of chemokines and *Plasmodium vivax* Duffy binding protein with the Duffy antigen/receptor for chemokines (DARC). *Mol Microbiol* **55**(5): 1413-1422
- Cleary MD, Singh U, Blader IJ, Brewer JL, Boothroyd JC (2002) *Toxoplasma gondii* asexual development: identification of developmentally regulated genes and distinct patterns of gene expression. *Eukaryot Cell* **1**(3): 329-340
- Collins CR, Withers-Martinez C, Hackett F, Blackman MJ (2009) An inhibitory antibody blocks interactions between components of the malarial invasion machinery. *PLoS Pathog* **5**(1): e1000273
- Collins WE, Barnwell JW (2009) *Plasmodium knowlesi*: finally being recognized. *J Infect Dis* **199**(8): 1107-1108

- Coppens I, Dunn JD, Romano JD, Pypaert M, Zhang H, Boothroyd JC, Joiner KA (2006) *Toxoplasma gondii* sequesters lysosomes from mammalian hosts in the vacuolar space. *Cell* **125**(2): 261-274
- Daher W, Soldati-Favre D (2009) Mechanisms controlling glideosome function in apicomplexans. *Curr Opin Microbiol*
- Debache K, Alaeddine F, Guionaud C, Monney T, Muller J, Strohmusch M, Leib SL, Grandgirard D, Hemphill A (2009) Vaccination with recombinant NcROP2 combined with recombinant NcMIC1 and NcMIC3 reduces cerebral infection and vertical transmission in mice experimentally infected with *Neospora caninum* tachyzoites. *Int J Parasitol*
- Desimone TM, Jennings CV, Bei AK, Comeaux C, Coleman BI, Refour P, Triglia T, Stubbs J, Cowman AF, Duraisingh MT (2009) Cooperativity between *Plasmodium falciparum* adhesive proteins for invasion into erythrocytes. *Mol Microbiol*
- Di Cristina M, Spaccapelo R, Soldati D, Bistoni F, Crisanti A (2000) Two conserved amino acid motifs mediate protein targeting to the micronemes of the apicomplexan parasite *Toxoplasma gondii*. *Mol Cell Biol* **20**(19): 7332-7341
- Ding M, Clayton C, Soldati D (2000) *Toxoplasma gondii* catalase: are there peroxisomes in *Toxoplasma*? *J Cell Sci* **113** (Pt 13): 2409-2419
- Dobrowolski JM, Carruthers VB, Sibley LD (1997) Participation of myosin in gliding motility and host cell invasion by *Toxoplasma gondii*. *Mol Microbiol* **26**(1): 163-173
- Dobrowolski JM, Sibley LD (1996) *Toxoplasma* invasion of mammalian cells is powered by the actin cytoskeleton of the parasite. *Cell* **84**(6): 933-939
- Dolan SA, Miller LH, Wellems TE (1990) Evidence for a switching mechanism in the invasion of erythrocytes by *Plasmodium falciparum*. *J Clin Invest* **86**(2): 618-624
- Dolan SA, Proctor JL, Alling DW, Okubo Y, Wellems TE, Miller LH (1994) Glycophorin B as an EBA-175 independent *Plasmodium falciparum* receptor of human erythrocytes. *Mol Biochem Parasitol* **64**(1): 55-63
- Donald RG, Carter D, Ullman B, Roos DS (1996) Insertional tagging, cloning, and expression of the *Toxoplasma gondii* hypoxanthine-xanthine-guanine phosphoribosyltransferase gene. Use as a selectable marker for stable transformation. *J Biol Chem* **271**(24): 14010-14019
- Dowse TJ, Pascall JC, Brown KD, Soldati D (2005) Apicomplexan rhomboids have a potential role in microneme protein cleavage during host cell invasion. *Int J Parasitol* **35**(7): 747-756
- Dowse TJ, Soldati D (2005) Rhomboid-like proteins in Apicomplexa: phylogeny and nomenclature. *Trends Parasitol* **21**(6): 254-258
- Dubey JP, Lindsay DS, Speer CA (1998) Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clin Microbiol Rev* **11**(2): 267-299
- Duraisingh MT, Maier AG, Triglia T, Cowman AF (2003a) Erythrocyte-binding antigen 175 mediates invasion in *Plasmodium falciparum* utilizing sialic acid-dependent and -independent pathways. *Proc Natl Acad Sci U S A* **100**(8): 4796-4801
- Duraisingh MT, Triglia T, Ralph SA, Rayner JC, Barnwell JW, McFadden GI, Cowman AF (2003b) Phenotypic variation of *Plasmodium falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes. *EMBO J* **22**(5): 1047-1057
- Dzierszinski F, Mortuaire M, Cesbron-Delauw MF, Tomavo S (2000) Targeted disruption of the glycosylphosphatidylinositol-anchored surface antigen SAG3 gene in *Toxoplasma gondii* decreases host cell adhesion and drastically reduces virulence in mice. *Mol Microbiol* **37**(3): 574-582
- El Hajj H, Papoin J, Cerede O, Garcia-Reguet N, Soete M, Dubremetz JF, Lebrun M (2008) Molecular signals in the trafficking of *Toxoplasma gondii* protein MIC3 to the micronemes. *Eukaryot Cell* **7**(6): 1019-1028

- Ellison SP, Greiner E, Dame JB (2001) In vitro culture and synchronous release of *Sarcocystis neurona* merozoites from host cells. *Vet Parasitol* **95**(2-4): 251-261
- Entzeroth R, Kerckhoff H, Konig A (1992) Microneme secretion in Coccidia: confocal laser scanning and electron microscope study of *Sarcocystis muris* in cell culture. *Eur J Cell Biol* **59**(2): 405-413
- Fang J, Marchesini N, Moreno SN (2006) A *Toxoplasma gondii* phosphoinositide phospholipase C (TgPI-PLC) with high affinity for phosphatidylinositol. *Biochem J* **394**(Pt 2): 417-425
- Fang XD, Kaslow DC, Adams JH, Miller LH (1991) Cloning of the *Plasmodium vivax* Duffy receptor. *Mol Biochem Parasitol* **44**(1): 125-132
- Fourmaux MN, Achbarou A, Mercereau-Puijalon O, Biderre C, Briche I, Loyens A, Odberg-Ferragut C, Camus D, Dubremetz JF (1996) The MIC1 microneme protein of *Toxoplasma gondii* contains a duplicated receptor-like domain and binds to host cell surface. *Mol Biochem Parasitol* **83**(2): 201-210
- Frenkel JK, Smith DD (2003) Determination of the genera of cyst-forming coccidia. *Parasitol Res* **91**(5): 384-389
- Frevert U, Sinnis P, Cerami C, Shreffler W, Takacs B, Nussenzweig V (1993) Malaria circumsporozoite protein binds to heparan sulfate proteoglycans associated with the surface membrane of hepatocytes. *J Exp Med* **177**(5): 1287-1298
- Galinski MR, Medina CC, Ingravallo P, Barnwell JW (1992) A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. *Cell* **69**(7): 1213-1226
- Gantt S, Persson C, Rose K, Birkett AJ, Abagyan R, Nussenzweig V (2000) Antibodies against thrombospondin-related anonymous protein do not inhibit *Plasmodium* sporozoite infectivity in vivo. *Infect Immun* **68**(6): 3667-3673
- Garcia-Reguet N, Lebrun M, Fourmaux MN, Mercereau-Puijalon O, Mann T, Beckers CJ, Samyn B, Van Beeumen J, Bout D, Dubremetz JF (2000) The microneme protein MIC3 of *Toxoplasma gondii* is a secretory adhesin that binds to both the surface of the host cells and the surface of the parasite. *Cell Microbiol* **2**(4): 353-364
- Gaskins E, Gilk S, DeVore N, Mann T, Ward G, Beckers C (2004) Identification of the membrane receptor of a class XIV myosin in *Toxoplasma gondii*. *J Cell Biol* **165**(3): 383-393
- Gaur D, Singh S, Jiang L, Diouf A, Miller LH (2007) Recombinant *Plasmodium falciparum* reticulocyte homology protein 4 binds to erythrocytes and blocks invasion. *Proc Natl Acad Sci U S A* **104**(45): 17789-17794
- Gendrin C, Mercier C, Braun L, Musset K, Dubremetz JF, Cesbron-Delauw MF (2008) *Toxoplasma gondii* uses unusual sorting mechanisms to deliver transmembrane proteins into the host-cell vacuole. *Traffic* **9**(10): 1665-1680
- Gerloff DL, Creasey A, Maslau S, Carter R (2005) Structural models for the protein family characterized by gamete surface protein Pfs230 of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **102**(38): 13598-13603
- Green JL, Hinds L, Grainger M, Knuepfer E, Holder AA (2006) *Plasmodium* thrombospondin related apical merozoite protein (PTRAMP) is shed from the surface of merozoites by PfSUB2 upon invasion of erythrocytes. *Mol Biochem Parasitol* **150**(1): 114-117
- Green JL, Rees-Channer RR, Howell SA, Martin SR, Knuepfer E, Taylor HM, Grainger M, Holder AA (2008) The motor complex of *Plasmodium falciparum*: phosphorylation by a calcium-dependent protein kinase. *J Biol Chem* **283**(45): 30980-30989
- Grimwood J, Smith JE (1992) *Toxoplasma gondii*: the role of a 30-kDa surface protein in host cell invasion. *Exp Parasitol* **74**(1): 106-111

- Grimwood J, Smith JE (1996) *Toxoplasma gondii*: the role of parasite surface and secreted proteins in host cell invasion. *Int J Parasitol* **26**(2): 169-173
- Hadley TJ, Klotz FW, Pasvol G, Haynes JD, McGinniss MH, Okubo Y, Miller LH (1987) *Falciparum* malaria parasites invade erythrocytes that lack glycophorin A and B (MkMk). Strain differences indicate receptor heterogeneity and two pathways for invasion. *J Clin Invest* **80**(4): 1190-1193
- Hans D, Pattnaik P, Bhattacharyya A, Shakri AR, Yazdani SS, Sharma M, Choe H, Farzan M, Chitnis CE (2005) Mapping binding residues in the *Plasmodium vivax* domain that binds Duffy antigen during red cell invasion. *Mol Microbiol* **55**(5): 1423-1434
- Harper JM, Hoff EF, Carruthers VB (2004) Multimerization of the *Toxoplasma gondii* MIC2 integrin-like A-domain is required for binding to heparin and human cells. *Mol Biochem Parasitol* **134**(2): 201-212
- Harper JM, Huynh MH, Coppens I, Parussini F, Moreno S, Carruthers VB (2006) A cleavable propeptide influences *Toxoplasma* infection by facilitating the trafficking and secretion of the TgMIC2-M2AP invasion complex. *Mol Biol Cell* **17**(10): 4551-4563
- Harris PK, Yeoh S, Dluzewski AR, O'Donnell RA, Withers-Martinez C, Hackett F, Bannister LH, Mitchell GH, Blackman MJ (2005) Molecular identification of a malaria merozoite surface sheddase. *PLoS Pathog* **1**(3): 241-251
- He XL, Grigg ME, Boothroyd JC, Garcia KC (2002) Structure of the immunodominant surface antigen from the *Toxoplasma gondii* SRS superfamily. *Nat Struct Biol* **9**(8): 606-611
- Hemphill A, Vonlaufen N, Naguleswaran A (2006) Cellular and immunological basis of the host-parasite relationship during infection with *Neospora caninum*. *Parasitology* **133**(Pt 3): 261-278
- Herm-Gotz A, Weiss S, Stratmann R, Fujita-Becker S, Ruff C, Meyhofer E, Soldati T, Manstein DJ, Geeves MA, Soldati D (2002) *Toxoplasma gondii* myosin A and its light chain: a fast, single-headed, plus-end-directed motor. *EMBO J* **21**(9): 2149-2158
- Holec L, Gasior A, Brillowska-Dabrowska A, Kur J (2008) *Toxoplasma gondii*: enzyme-linked immunosorbent assay using different fragments of recombinant microneme protein 1 (MIC1) for detection of immunoglobulin G antibodies. *Exp Parasitol* **119**(1): 1-6
- Hoppe HC, Ngo HM, Yang M, Joiner KA (2000) Targeting to rhoptry organelles of *Toxoplasma gondii* involves evolutionarily conserved mechanisms. *Nat Cell Biol* **2**(7): 449-456
- Howell SA, Hackett F, Jongco AM, Withers-Martinez C, Kim K, Carruthers VB, Blackman MJ (2005) Distinct mechanisms govern proteolytic shedding of a key invasion protein in apicomplexan pathogens. *Mol Microbiol* **57**(5): 1342-1356
- Huynh MH, Carruthers VB (2006) *Toxoplasma* MIC2 is a major determinant of invasion and virulence. *PLoS Pathog* **2**(8): e84
- Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**(1): 23-28
- Ismael AB, Dimier-Poisson I, Lebrun M, Dubremetz JF, Bout D, Mevelec MN (2006) Mic1-3 knockout of *Toxoplasma gondii* is a successful vaccine against chronic and congenital toxoplasmosis in mice. *J Infect Dis* **194**(8): 1176-1183
- Ismael AB, Hedhli D, Cerede O, Lebrun M, Dimier-Poisson I, Mevelec MN (2009) Further analysis of protection induced by the MIC3 DNA vaccine against *T. gondii*: CD4 and CD8 T cells are the major effectors of the MIC3 DNA vaccine-induced protection, both Lectin-like and EGF-like domains of MIC3 conferred protection. *Vaccine* **27**(22): 2959-2966
- Ismael AB, Sekkai D, Collin C, Bout D, Mevelec MN (2003) The MIC3 gene of *Toxoplasma gondii* is a novel potent vaccine candidate against toxoplasmosis. *Infect Immun* **71**(11): 6222-6228

- Jacquet A, Coulon L, De Neve J, Daminet V, Haumont M, Garcia L, Bollen A, Jurado M, Biemans R (2001) The surface antigen SAG3 mediates the attachment of *Toxoplasma gondii* to cell-surface proteoglycans. *Mol Biochem Parasitol* **116**(1): 35-44
- Jethwaney D, Lepore T, Hassan S, Mello K, Rangarajan R, Jahnen-Dechent W, Wirth D, Sultan AA (2005) Fetuin-A, a hepatocyte-specific protein that binds *Plasmodium berghei* thrombospondin-related adhesive protein: a potential role in infectivity. *Infect Immun* **73**(9): 5883-5891
- Jewett TJ, Sibley LD (2003) Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. *Mol Cell* **11**(4): 885-894
- Jewett TJ, Sibley LD (2004) The *Toxoplasma* proteins MIC2 and M2AP form a hexameric complex necessary for intracellular survival. *J Biol Chem* **279**(10): 9362-9369
- Jung C, Lee CY, Grigg ME (2004) The SRS superfamily of *Toxoplasma* surface proteins. *Int J Parasitol* **34**(3): 285-296
- Kafsack BF, Pena JD, Coppens I, Ravindran S, Boothroyd JC, Carruthers VB (2009) Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells. *Science* **323**(5913): 530-533
- Karsten V, Qi H, Beckers CJ, Reddy A, Dubremetz JF, Webster P, Joiner KA (1998) The protozoan parasite *Toxoplasma gondii* targets proteins to dense granules and the vacuolar space using both conserved and unusual mechanisms. *J Cell Biol* **141**(6): 1323-1333
- Kasehagen LJ, Mueller I, Kiniboro B, Bockarie MJ, Reeder JC, Kazura JW, Kastens W, McNamara DT, King CH, Whalen CC, Zimmerman PA (2007) Reduced *Plasmodium vivax* erythrocyte infection in PNG Duffy-negative heterozygotes. *PLoS One* **2**(3): e336
- Kato N, Sakata T, Breton G, Le Roch KG, Nagle A, Andersen C, Bursulaya B, Henson K, Johnson J, Kumar KA, Marr F, Mason D, McNamara C, Plouffe D, Ramachandran V, Spooner M, Tuntland T, Zhou Y, Peters EC, Chatterjee A, Schultz PG, Ward GE, Gray N, Harper J, Winzeler EA (2008) Gene expression signatures and small-molecule compounds link a protein kinase to *Plasmodium falciparum* motility. *Nat Chem Biol* **4**(6): 347-356
- Keller N, Naguleswaran A, Cannas A, Vonlaufen N, Bienz M, Bjorkman C, Bohne W, Hemphill A (2002) Identification of a *Neospora caninum* microneme protein (NcMIC1) which interacts with sulfated host cell surface glycosaminoglycans. *Infect Immun* **70**(6): 3187-3198
- Keller N, Riesen M, Naguleswaran A, Vonlaufen N, Stettler R, Leepin A, Wastling JM, Hemphill A (2004) Identification and characterization of a *Neospora caninum* microneme-associated protein (NcMIC4) that exhibits unique lactose-binding properties. *Infect Immun* **72**(8): 4791-4800
- Kessler H, Herm-Gotz A, Hegge S, Rauch M, Soldati-Favre D, Frischknecht F, Meissner M (2008) Microneme protein 8--a new essential invasion factor in *Toxoplasma gondii*. *J Cell Sci* **121**(Pt 7): 947-956
- Khan IA, Eckel ME, Pfefferkorn ER, Kasper LH (1988) Production of gamma interferon by cultured human lymphocytes stimulated with a purified membrane protein (P30) from *Toxoplasma gondii*. *J Infect Dis* **157**(5): 979-984
- Kieschnick H, Wakefield T, Narducci CA, Beckers C (2001) *Toxoplasma gondii* attachment to host cells is regulated by a calmodulin-like domain protein kinase. *J Biol Chem* **276**(15): 12369-12377
- Kim SK, Boothroyd JC (2005) Stage-specific expression of surface antigens by *Toxoplasma gondii* as a mechanism to facilitate parasite persistence. *J Immunol* **174**(12): 8038-8048
- Kim SK, Karasov A, Boothroyd JC (2007) Bradyzoite-specific surface antigen SRS9 plays a role in maintaining *Toxoplasma gondii* persistence in the brain and in host control of parasite replication in the intestine. *Infect Immun* **75**(4): 1626-1634

- Klein H, Loschner B, Zyto N, Portner M, Montag T (1998) Expression, purification, and biochemical characterization of a recombinant lectin of *Sarcocystis muris* (Apicomplexa) cyst merozoites. *Glycoconj J* **15**(2): 147-153
- Klein H, Mueller S, Loeschner B, Toenjes RR, Braun G, Mueller EC, Otto A, Montag T (2003) Cloning, sequencing and recombinant expression of the open reading frame encoding a novel member of the *Sarcocystis muris* (Apicomplexa) microneme lectin family. *Parasitol Res* **90**(1): 84-86
- Kvicerova J, Ptackova P, Modry D (2007) Endogenous development, pathogenicity and host specificity of *Eimeria cahirinensis* Couch, Blaustein, Duszynski, Shenbrot and Nevo, 1997 (Apicomplexa: Eimeriidae) from *Acomys dimidiatus* (Cretzschmar 1826) (Rodentia: Muridae) from the Near East. *Parasitol Res* **100**(2): 219-226
- 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**(1): 43-53
- Larson ET, Parussini F, Huynh MH, Giebel JD, Kelley AM, Zhang L, Bogyo M, Merritt EA, Carruthers VB (2009) *Toxoplasma gondii* cathepsin L is the primary target of the invasion inhibitory compound LHSV. *J Biol Chem*
- Lavine MD, Knoll LJ, Rooney PJ, Arrizabalaga G (2007) A *Toxoplasma gondii* mutant defective in responding to calcium fluxes shows reduced in vivo pathogenicity. *Mol Biochem Parasitol* **155**(2): 113-122
- Levine ND, Ivens V (1988) Cross-transmission of *Eimeria* spp. (Protozoa, Apicomplexa) of rodents--a review. *J Protozool* **35**(3): 434-437
- Lourenco EV, Bernardes ES, Silva NA, Mineo JR, Panunto-Castelo A, Roque-Barreira MC (2006) Immunization with MIC1 and MIC4 induces protective immunity against *Toxoplasma gondii*. *Microbes and Infection* **8**(5): 1244-1251
- Lourenco EV, Pereira SR, Faca VM, Coelho-Castelo AA, Mineo JR, Roque-Barreira MC, Greene LJ, Panunto-Castelo A (2001) *Toxoplasma gondii* micronemal protein MIC1 is a lactose-binding lectin. *Glycobiology* **11**(7): 541-547
- Lovett JL, Marchesini N, Moreno SN, Sibley LD (2002) *Toxoplasma gondii* microneme secretion involves intracellular Ca²⁺ release from inositol 1,4,5-triphosphate (IP₃)/ryanodine-sensitive stores. *J Biol Chem* **277**(29): 25870-25876
- Lovett JL, Sibley LD (2003) Intracellular calcium stores in *Toxoplasma gondii* govern invasion of host cells. *J Cell Sci* **116**(Pt 14): 3009-3016
- Luo S, Ruiz FA, Moreno SN (2005) The acidocalcisome Ca²⁺-ATPase (TgA1) of *Toxoplasma gondii* is required for polyphosphate storage, intracellular calcium homeostasis and virulence. *Mol Microbiol* **55**(4): 1034-1045
- Maier AG, Baum J, Smith B, Conway DJ, Cowman AF (2009) Polymorphisms in erythrocyte binding antigens 140 and 181 affect function and binding but not receptor specificity in *Plasmodium falciparum*. *Infect Immun* **77**(4): 1689-1699
- Maier AG, Duraisingh MT, Reeder JC, Patel SS, Kazura JW, Zimmerman PA, Cowman AF (2003) *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nat Med* **9**(1): 87-92
- Martin MJ, Rayner JC, Gagneux P, Barnwell JW, Varki A (2005) Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of N-glycolylneuraminic acid. *Proc Natl Acad Sci U S A* **102**(36): 12819-12824
- Matuschewski K (2006) Getting infectious: formation and maturation of *Plasmodium* sporozoites in the Anopheles vector. *Cell Microbiol* **8**(10): 1547-1556
- Matuschewski K, Nunes AC, Nussenzweig V, Menard R (2002) *Plasmodium* sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. *EMBO J* **21**(7): 1597-1606

- Mayer DC, Cofie J, Jiang L, Hartl DL, Tracy E, Kabat J, Mendoza LH, Miller LH (2009) Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1. *Proc Natl Acad Sci U S A* **106**(13): 5348-5352
- Mayer DC, Jiang L, Achur RN, Kakizaki I, Gowda DC, Miller LH (2006) The glycophorin C N-linked glycan is a critical component of the ligand for the *Plasmodium falciparum* erythrocyte receptor BAEBL. *Proc Natl Acad Sci U S A* **103**(7): 2358-2362
- Mayer DC, Mu JB, Feng X, Su XZ, Miller LH (2002) Polymorphism in a *Plasmodium falciparum* erythrocyte-binding ligand changes its receptor specificity. *J Exp Med* **196**(11): 1523-1528
- McCormick CJ, Tuckwell DS, Crisanti A, Humphries MJ, Hollingdale MR (1999) Identification of heparin as a ligand for the A-domain of *Plasmodium falciparum* thrombospondin-related adhesion protein. *Mol Biochem Parasitol* **100**(1): 111-124
- Meissner M, Reiss M, Viebig N, Carruthers VB, Toursel C, Tomavo S, Ajioka JW, Soldati D (2002a) A family of transmembrane microneme proteins of *Toxoplasma gondii* contain EGF-like domains and function as escorts. *J Cell Sci* **115**(Pt 3): 563-574
- Meissner M, Schluter D, Soldati D (2002b) Role of *Toxoplasma gondii* myosin A in powering parasite gliding and host cell invasion. *Science* **298**(5594): 837-840
- Menard R (2005) Medicine: knockout malaria vaccine? *Nature* **433**(7022): 113-114
- Mercier C, Dubremetz JF, Rauscher B, Lecordier L, Sibley LD, Cesbron-Delauw MF (2002) Biogenesis of nanotubular network in *Toxoplasma* parasitophorous vacuole induced by parasite proteins. *Mol Biol Cell* **13**(7): 2397-2409
- Miller LH, Haynes JD, McAuliffe FM, Shiroishi T, Durocher JR, McGinniss MH (1977) Evidence for differences in erythrocyte surface receptors for the malarial parasites, *Plasmodium falciparum* and *Plasmodium knowlesi*. *J Exp Med* **146**(1): 277-281
- Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med* **295**(6): 302-304
- Miller LH, Mason SJ, Dvorak JA, McGinniss MH, Rothman IK (1975) Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science* **189**(4202): 561-563
- Mineo JR, McLeod R, Mack D, Smith J, Khan IA, Ely KH, Kasper LH (1993) Antibodies to *Toxoplasma gondii* major surface protein (SAG-1, P30) inhibit infection of host cells and are produced in murine intestine after peroral infection. *J Immunol* **150**(9): 3951-3964
- Mital J, Meissner M, Soldati D, Ward GE (2005) Conditional expression of *Toxoplasma gondii* apical membrane antigen-1 (TgAMA1) demonstrates that TgAMA1 plays a critical role in host cell invasion. *Mol Biol Cell* **16**(9): 4341-4349
- Moire N, Dion S, Lebrun M, Dubremetz JF, Dimier-Poisson I (2009) Mic1-3KO tachyzoite a live attenuated vaccine candidate against toxoplasmosis derived from a type I strain shows features of type II strain. *Exp Parasitol*
- Monteiro VG, Soares CP, de Souza W (1998) Host cell surface sialic acid residues are involved on the process of penetration of *Toxoplasma gondii* into mammalian cells. *FEMS Microbiol Lett* **164**(2): 323-327
- Montoya JG, Liesenfeld O (2004) Toxoplasmosis. *Lancet* **363**(9425): 1965-1976
- Morahan BJ, Wang L, Coppel RL (2009) No TRAP, no invasion. *Trends Parasitol* **25**(2): 77-84
- Moudy R, Manning TJ, Beckers CJ (2001) The loss of cytoplasmic potassium upon host cell breakdown triggers egress of *Toxoplasma gondii*. *J Biol Chem* **276**(44): 41492-41501

- Muller HM, Reckmann I, Hollingdale MR, Bujard H, Robson KJ, Crisanti A (1993) Thrombospondin related anonymous protein (TRAP) of *Plasmodium falciparum* binds specifically to sulfated glycoconjugates and to HepG2 hepatoma cells suggesting a role for this molecule in sporozoite invasion of hepatocytes. *EMBO J* **12**(7): 2881-2889
- Muller JJ, Muller EC, Montag T, Zyto N, Loschner B, Klein H, Heineman U, Otto A (2001) Characterization and crystallization of a novel *Sarcocystis muris* lectin, SML-2. *Acta Crystallogr D Biol Crystallogr* **57**(Pt 7): 1042-1045
- Nagamune K, Moreno SN, Chini EN, Sibley LD (2008) Calcium regulation and signaling in apicomplexan parasites. *Subcell Biochem* **47**: 70-81
- Nagamune K, Moreno SN, Sibley LD (2007) Artemisinin-resistant mutants of *Toxoplasma gondii* have altered calcium homeostasis. *Antimicrob Agents Chemother* **51**(11): 3816-3823
- Nagamune K, Sibley LD (2006) Comparative genomic and phylogenetic analyses of calcium ATPases and calcium-regulated proteins in the apicomplexa. *Mol Biol Evol* **23**(8): 1613-1627
- Ngo HM, Yang M, Paprotka K, Pypaert M, Hoppe H, Joiner KA (2003) AP-1 in *Toxoplasma gondii* mediates biogenesis of the rhoptry secretory organelle from a post-Golgi compartment. *J Biol Chem* **278**(7): 5343-5352
- O'Donnell RA, Hackett F, Howell SA, Treeck M, Struck N, Krnajski Z, Withers-Martinez C, Gilberger TW, Blackman MJ (2006) Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J Cell Biol* **174**(7): 1023-1033
- Okoyeh JN, Pillai CR, Chitnis CE (1999) *Plasmodium falciparum* field isolates commonly use erythrocyte invasion pathways that are independent of sialic acid residues of glycophorin A. *Infect Immun* **67**(11): 5784-5791
- Opitz C, Di Cristina M, Reiss M, Ruppert T, Crisanti A, Soldati D (2002) Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite *Toxoplasma gondii*. *EMBO J* **21**(7): 1577-1585
- Orlandi PA, Klotz FW, Haynes JD (1992) A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of *Plasmodium falciparum* recognizes the terminal Neu5Ac(alpha 2-3)Gal- sequences of glycophorin A. *J Cell Biol* **116**(4): 901-909
- Ortega-Barria E, Boothroyd JC (1999) A *Toxoplasma* lectin-like activity specific for sulfated polysaccharides is involved in host cell infection. *J Biol Chem* **274**(3): 1267-1276
- Partanen P, Turunen HJ, Paasivuo RT, Leinikki PO (1984) Immunoblot analysis of *Toxoplasma gondii* antigens by human immunoglobulins G, M, and A antibodies at different stages of infection. *J Clin Microbiol* **20**(1): 133-135
- Passos AP, Garcia CR (1998) Inositol 1,4,5-trisphosphate induced Ca²⁺ release from chloroquine-sensitive and -insensitive intracellular stores in the intraerythrocytic stage of the malaria parasite *P. chabaudi*. *Biochem Biophys Res Commun* **245**(1): 155-160
- Pasvol G, Jungery M (1983) Glycophorins and red cell invasion by *Plasmodium falciparum*. *Ciba Found Symp* **94**: 174-195
- Periz J, Gill AC, Hunt L, Brown P, Tomley FM (2007) The microneme proteins EtMIC4 and EtMIC5 of *Eimeria tenella* form a novel, ultra-high molecular mass protein complex that binds target host cells. *J Biol Chem* **282**(23): 16891-16898
- Perkins ME, Holt EH (1988) Erythrocyte receptor recognition varies in *Plasmodium falciparum* isolates. *Mol Biochem Parasitol* **27**(1): 23-34
- Persson KE, McCallum FJ, Reiling L, Lister NA, Stubbs J, Cowman AF, Marsh K, Beeson JG (2008) Variation in use of erythrocyte invasion pathways by *Plasmodium falciparum* mediates evasion of human inhibitory antibodies. *J Clin Invest* **118**(1): 342-351

- Pizarro JC, Vulliez-Le Normand B, Chesne-Seck ML, Collins CR, Withers-Martinez C, Hackett F, Blackman MJ, Faber BW, Remarque EJ, Kocken CH, Thomas AW, Bentley GA (2005) Crystal structure of the malaria vaccine candidate apical membrane antigen 1. *Science* **308**(5720): 408-411
- Preiser PR, Jarra W, Capiod T, Snounou G (1999) A rhoptry-protein-associated mechanism of clonal phenotypic variation in rodent malaria. *Nature* **398**(6728): 618-622
- Prigione I, Facchetti P, Lecordier L, Deslee D, Chiesa S, Cesbron-Delauw MF, Pistoia V (2000) T cell clones raised from chronically infected healthy humans by stimulation with *Toxoplasma gondii* excretory-secretory antigens cross-react with live tachyzoites: characterization of the fine antigenic specificity of the clones and implications for vaccine development. *J Immunol* **164**(7): 3741-3748
- Rabenau KE, Sohrabi A, Tripathy A, Reitter C, Ajioka JW, Tomley FM, Carruthers VB (2001) TgM2AP participates in *Toxoplasma gondii* invasion of host cells and is tightly associated with the adhesive protein TgMIC2. *Mol Microbiol* **41**(3): 537-547
- Rachinel N, Buzoni-Gatel D, Dutta C, Mennechet FJ, Luangsang S, Minns LA, Grigg ME, Tomavo S, Boothroyd JC, Kasper LH (2004) The induction of acute ileitis by a single microbial antigen of *Toxoplasma gondii*. *J Immunol* **173**(4): 2725-2735
- Rayner JC, Vargas-Serrato E, Huber CS, Galinski MR, Barnwell JW (2001) A *Plasmodium falciparum* homologue of *Plasmodium vivax* reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway. *J Exp Med* **194**(11): 1571-1581
- Reed MB, Caruana SR, Batchelor AH, Thompson JK, Crabb BS, Cowman AF (2000) Targeted disruption of an erythrocyte binding antigen in *Plasmodium falciparum* is associated with a switch toward a sialic acid-independent pathway of invasion. *Proc Natl Acad Sci U S A* **97**(13): 7509-7514
- 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**(3): 563-578
- Robson KJ, Frevert U, Reckmann I, Cowan G, Beier J, Scragg IG, Takehara K, Bishop DH, Pradel G, Sinden R, et al. (1995) Thrombospondin-related adhesive protein (TRAP) of *Plasmodium falciparum*: expression during sporozoite ontogeny and binding to human hepatocytes. *EMBO J* **14**(16): 3883-3894
- Rohrbach P, Friedrich O, Hentschel J, Plattner H, Fink RH, Lanzer M (2005) Quantitative calcium measurements in subcellular compartments of *Plasmodium falciparum*-infected erythrocytes. *J Biol Chem* **280**(30): 27960-27969
- Roos DS (2005) Genetics. Themes and variations in apicomplexan parasite biology. *Science* **309**(5731): 72-73
- Saeij JP, Arrizabalaga G, Boothroyd JC (2008) A cluster of four surface antigen genes specifically expressed in bradyzoites, SAG2CDXY, plays an important role in *Toxoplasma gondii* persistence. *Infect Immun* **76**(6): 2402-2410
- Saouros S, Edwards-Jones B, Reiss M, Sawmynaden K, Cota E, Simpson P, Dowse TJ, Jakle 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**(46): 38583-38591
- 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**(11): 1149-1155
- Scherf A, Lopez-Rubio JJ, Riviere L (2008) Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol* **62**: 445-470
- Sheiner L, Dowse TJ, Soldati-Favre D (2008) Identification of trafficking determinants for polytopic rhomboid proteases in *Toxoplasma gondii*. *Traffic* **9**(5): 665-677

- Shen Z, Jacobs-Lorena M (1999) Evolution of chitin-binding proteins in invertebrates. *J Mol Evol* **48**(3): 341-347
- Sim BK, Orlandi PA, Haynes JD, Klotz FW, Carter JM, Camus D, Zegans ME, Chulay JD (1990) Primary structure of the 175K *Plasmodium falciparum* erythrocyte binding antigen and identification of a peptide which elicits antibodies that inhibit malaria merozoite invasion. *J Cell Biol* **111**(5 Pt 1): 1877-1884
- Sim BK, Toyoshima T, Haynes JD, Aikawa M (1992) Localization of the 175-kilodalton erythrocyte binding antigen in micronemes of *Plasmodium falciparum* merozoites. *Mol Biochem Parasitol* **51**(1): 157-159
- Singh SK, Hora R, Belrhali H, Chitnis CE, Sharma A (2006) Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* **439**(7077): 741-744
- Sinnis P, Clavijo P, Fenyo D, Chait BT, Cerami C, Nussenzweig V (1994) Structural and functional properties of region II-plus of the malaria circumsporozoite protein. *J Exp Med* **180**(1): 297-306
- Soldati-Favre D (2008) Molecular dissection of host cell invasion by the apicomplexans: the glideosome. *Parasite* **15**(3): 197-205
- Soldati D, Boothroyd JC (1993) Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*. *Science* **260**(5106): 349-352
- Starnes GL, Coincon M, Sygusch J, Sibley LD (2009) Aldolase is essential for energy production and bridging adhesin-actin cytoskeletal interactions during parasite invasion of host cells. *Cell Host Microbe* **5**(4): 353-364
- Starnes GL, Jewett TJ, Carruthers VB, Sibley LD (2006) Two separate, conserved acidic amino acid domains within the *Toxoplasma gondii* MIC2 cytoplasmic tail are required for parasite survival. *J Biol Chem* **281**(41): 30745-30754
- Straub KW, Cheng SJ, Sohn CS, Bradley PJ (2009) Novel components of the Apicomplexan moving junction reveal conserved and coccidia-restricted elements. *Cell Microbiol* **11**(4): 590-603
- Striepen B, He CY, Matrajt M, Soldati D, Roos DS (1998) Expression, selection, and organellar targeting of the green fluorescent protein in *Toxoplasma gondii*. *Mol Biochem Parasitol* **92**(2): 325-338
- Striepen B, Jordan CN, Reiff S, van Dooren GG (2007) Building the perfect parasite: cell division in apicomplexa. *PLoS Pathog* **3**(6): e78
- Striepen B, Soldati D, Garcia-Reguet N, Dubremetz JF, Roos DS (2001) Targeting of soluble proteins to the rhoptries and micronemes in *Toxoplasma gondii*. *Mol Biochem Parasitol* **113**(1): 45-53
- Stubbs J, Simpson KM, Triglia T, Plouffe D, Tonkin CJ, Duraisingh MT, Maier AG, Winzeler EA, Cowman AF (2005) Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. *Science* **309**(5739): 1384-1387
- Sturm A, Graewe S, Franke-Fayard B, Retzlaff S, Bolte S, Roppenser B, Aepfelbacher M, Janse C, Heussler V (2009) Alteration of the parasite plasma membrane and the parasitophorous vacuole membrane during exo-erythrocytic development of malaria parasites. *Protist* **160**(1): 51-63
- Su C, Evans D, Cole RH, Kissinger JC, Ajioka JW, Sibley LD (2003) Recent expansion of *Toxoplasma* through enhanced oral transmission. *Science* **299**(5605): 414-416
- Suss-Toby E, Zimmerberg J, Ward GE (1996) *Toxoplasma* invasion: the parasitophorous vacuole is formed from host cell plasma membrane and pinches off via a fission pore. *Proc Natl Acad Sci U S A* **93**(16): 8413-8418
- Suzuki Y (2005) Sialobiology of influenza: molecular mechanism of host range variation of influenza viruses. *Biol Pharm Bull* **28**(3): 399-408
- Suzuki Y, Orellana MA, Schreiber RD, Remington JS (1988) Interferon-gamma: the major mediator of resistance against *Toxoplasma gondii*. *Science* **240**(4851): 516-518

- Takabatake N, Okamura M, Yokoyama N, Ikehara Y, Akimitsu N, Arimitsu N, Hamamoto H, Sekimizu K, Suzuki H, Igarashi I (2007a) Glycophorin A-knockout mice, which lost sialoglycoproteins from the red blood cell membrane, are resistant to lethal infection of *Babesia rodhaini*. *Vet Parasitol* **148**(2): 93-101
- Takabatake N, Okamura M, Yokoyama N, Okubo K, Ikehara Y, Igarashi I (2007b) Involvement of a host erythrocyte sialic acid content in *Babesia bovis* infection. *J Vet Med Sci* **69**(10): 999-1004
- Templeton TJ (2007) Whole-genome natural histories of apicomplexan surface proteins. *Trends Parasitol* **23**(5): 205-212
- Teo CF, Zhou XW, Bogyo M, Carruthers VB (2007) Cysteine protease inhibitors block *Toxoplasma gondii* microneme secretion and cell invasion. *Antimicrob Agents Chemother* **51**(2): 679-688
- Thompson JK, Triglia T, Reed MB, Cowman AF (2001) A novel ligand from *Plasmodium falciparum* that binds to a sialic acid-containing receptor on the surface of human erythrocytes. *Mol Microbiol* **41**(1): 47-58
- Tolia NH, Enemark EJ, Sim BK, Joshua-Tor L (2005) Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. *Cell* **122**(2): 183-193
- Tossavainen H, Pihlajamaa T, Huttunen TK, Raulo E, Rauvala H, Permi P, Kilpelainen I (2006) The layered fold of the TSR domain of *P. falciparum* TRAP contains a heparin binding site. *Protein Sci* **15**(7): 1760-1768
- Treeck M, Struck NS, Haase S, Langer C, Herrmann S, Healer J, Cowman AF, Gilberger TW (2006) A conserved region in the EBL proteins is implicated in microneme targeting of the malaria parasite *Plasmodium falciparum*. *J Biol Chem* **281**(42): 31995-32003
- Treeck M, Zacherl S, Herrmann S, Cabrera A, Kono M, Struck NS, Engelberg K, Haase S, Frischknecht F, Miura K, Spielmann T, Gilberger TW (2009) Functional analysis of the leading malaria vaccine candidate AMA-1 reveals an essential role for the cytoplasmic domain in the invasion process. *PLoS Pathog* **5**(3): e1000322
- Triglia T, Duraisingh MT, Good RT, Cowman AF (2005) Reticulocyte-binding protein homologue 1 is required for sialic acid-dependent invasion into human erythrocytes by *Plasmodium falciparum*. *Mol Microbiol* **55**(1): 162-174
- Triglia T, Thompson J, Caruana SR, Delorenzi M, Speed T, Cowman AF (2001) Identification of proteins from *Plasmodium falciparum* that are homologous to reticulocyte binding proteins in *Plasmodium vivax*. *Infect Immun* **69**(2): 1084-1092
- Urban S, Freeman M (2003) Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol Cell* **11**(6): 1425-1434
- Vaishnav S, Striepen B (2006) The cell biology of secondary endosymbiosis--how parasites build, divide and segregate the apicoplast. *Mol Microbiol* **61**(6): 1380-1387
- Varki A (1997) Sialic acids as ligands in recognition phenomena. *FASEB J* **11**(4): 248-255
- Varki A (2008) Sialic acids in human health and disease. *Trends Mol Med* **14**(8): 351-360
- Vetterling JM (1976) *Eimeria tenella*: host specificity in gallinaceous birds. *J Protozool* **23**(1): 155-158
- Weiss LM, Kim K (2000) The development and biology of bradyzoites of *Toxoplasma gondii*. *Front Biosci* **5**: D391-405
- Wertheimer SP, Barnwell JW (1989) *Plasmodium vivax* interaction with the human Duffy blood group glycoprotein: identification of a parasite receptor-like protein. *Exp Parasitol* **69**(4): 340-350
- Wetzel DM, Chen LA, Ruiz FA, Moreno SN, Sibley LD (2004) Calcium-mediated protein secretion potentiates motility in *Toxoplasma gondii*. *J Cell Sci* **117**(Pt 24): 5739-5748

- Wiersma HI, Galuska SE, Tomley FM, Sibley LD, Liberator PA, Donald RG (2004) A role for coccidian cGMP-dependent protein kinase in motility and invasion. *Int J Parasitol* **34**(3): 369-380
- Wright HT, Sandrasegaram G, Wright CS (1991) Evolution of a family of N-acetylglucosamine binding proteins containing the disulfide-rich domain of wheat germ agglutinin. *J Mol Evol* **33**(3): 283-294
- Zheng B, He A, Gan M, Li Z, He H, Zhan X (2009) MIC6 associates with aldolase in host cell invasion by *Toxoplasma gondii*. *Parasitol Res* **105**(2): 441-445
- Zhou XW, Blackman MJ, Howell SA, Carruthers VB (2004) Proteomic analysis of cleavage events reveals a dynamic two-step mechanism for proteolysis of a key parasite adhesive complex. *Mol Cell Proteomics* **3**(6): 565-576