Universität Bremen*
*Exzellent

Structure and Function Relationship of Trans-Sialidases from *Trypanosoma congolense*

Dissertation des Fachbereiches Biologie/Chemie
der Universität Bremen

September 2015

Zur Erlangung des Doktorgrades der Naturwissenschaften
- Dr. rer. nat. -

vorgelegt von
Mario Waespy

Die Vorliegende Arbeit wurde im Zeitraum von November 2011 bis Oktober 2015 im Arbeitskreis von Herrn Prof. Dr. Sørge Kelm am Zentrum für Biomolekulare Interaktionen des Fachbereiches Biologie/Chemie der Universität Bremen angefertigt.

1. Gutachter: Prof. Dr. Rita Gerardy-Schahn (Hannover Medical School, Germany)

2. Gutachter: Prof. Dr. Alvaro Acosta-Serrano (Liverpool School of Tropical Medicine, United Kingdom)

3. Gutachter: Prof. Dr. Oscar Campetella (University of General San Martín, Argentina)

Tag der Disputation: 13. November 2015

Für meine Eltern, meinen Bruder und meine leibliche Mutter

Es gibt zwei Arten sein
Leben zu leben:
Entweder so, als wäre
nichts ein Wunder, oder
so, als wäre
alles eines.

Albert Einstein

Im stillen Gedenken an Doris Schade und Gerhard Weichelt.

Echtheitserklärung

Ich versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.

Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Ich erkläre weiterhin, dass die vorliegende Arbeit weder in gleicher noch in ähnlicher Form bereits einem anderen Prüfungsverfahren vorliegt oder vorgelegen hat.

Mario Waespy Bremen, den 29.09.2015

Table of contents

I. Danksagung

Mein besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr. Sørge Kelm für die sehr interessante Aufgabenstellung, die exzellente Betreuung, die stetige Hilfs- und Diskussionsbereitschaft, sowie für die intensiven Gespräche und sein großes Vertrauen in mich und meine Arbeit. Vielen Dank auch für die ständige Hilfsbereitschaft, Unterstützung und wissenschaftliche Förderung meiner Person über den Rahmen dieser Arbeit hinaus und die Vorbereitung auf meine zukünftige Karriere als Wissenschaftler.

Frau Prof. Dr. Rita Gerardy-Schahn, Herrn Prof. Dr. Alvaro Acosta-Serrano und Herrn Prof. Dr. Oscar Campetella danke ich für die Übernahme der Gutachten dieser Arbeit.

Bei meinem Kollegen und Freund Dr. Thaddeus T. Gbem bedanke ich mich herzlich für die exzellente und nette Zusammenarbeit und die außerordentliche Hilfsbereitschaft, von technischen und persönlichen Problemen bis hin zur stetigen Mitarbeit an Manuskripten und die netten Kneipenabende. Dr. Frank Dietz danke ich für die angenehme und interessante Zusammenarbeit, sowie die Hilfestellung und Diskussion bei labortechnischen Problemen und anderen Anliegen, wie z.B. die kontinuierliche und essentielle Erörterung der Bundesliga-Ergebnisse.

Mein Dank gilt weiterhin allen Mitgliedern der Arbeitsgruppe Kelm, die mich immer unterstützt, mir bei Fragen und Anliegen stets zur Seite standen und für das stetige Ohr meiner kleinen, immer wiederkehrenden „Problemchen“. Ich danke Nazila Isakovic, Petra Berger, Sabine Limberg, Petra Seekamp, Hendrik Koliwer-Brandl, Tanja Dodenhof, Judith Weber, Jessica Nüsse und Veronika Kraaz für die schöne Zeit, die netten Konferenzen, Kneipen-, Weihnachtsmarkt- und Freimarkts-abende und natürlich Weihnachtsfeiern und Geburtstage.

Ich möchte mich ebenfalls ganz herzlich bei Paul Madge, Dr. Joe Tiralongo und Dr. Thomas Haselhorst für die exzellente und sehr erfolgreiche Zusammenarbeit bedanken und hoffe auch in naher Zukunft auf weitere, erfolgreiche gemeinsame Projekte.

Ganz herzlich möchte ich auch meiner Familie und meinen Freunden danken, in erster Linie meinen Eltern, die mir meine Ausbildung ermöglicht und mich während meines Studiums und meiner Dissertation in jeder Hinsicht unterstützt und immer wieder aufgebaut haben und meinem lieben Bruder Maik für die Unterstützung und Ermutigungen.

Schließlich gilt mein besonderer Dank meiner lieben Freundin Hava, meine Lebensgefährtin, meine beste Freundin und meine große Liebe. Danke für all deine unerschöpfliche Unterstützung, deine Fürsorge, deine oft nötigen Ermutigungen und deine Liebe, während meiner Dissertation und im Alltag. Ohne dich wäre diese Arbeit nie möglich gewesen.

„Ben seni seviyorum“

II. Structure of this thesis

The following paragraph provides general information about the structure of this work. This thesis generally consists of six main chapters, comprising Introduction, Objectives, Results, Summarising Discussion, Outlook and Appendix all written in English. Furthermore, a Summary is provided in English and German.

The aim of the introduction is to familiarise the reader with the field of carbohydrates, especially sialic acids but also oligosaccharides and the fundamental processes of protein glycosylation, as well as to the field of sialic acid binding and processing proteins, including siglecs and trans-sialidases (TS). Although the main focus directs on the latter, whereas associated aims of this thesis are elucidated in the subsequent objectives chapter.

The following results chapter is divided into five subchapters, two representing accepted publications (Chapter 3.1 and 3.2), one accepted manuscript (Chapter 3.4) and two manuscripts in preparation (Chapter 3.3 and 3.5). The first two chapters (3.1 and 3.2) describe the identification, cloning, expression and partial characterisation of 14 different, active TS from *Trypanosoma congolense* (*T. congolense*, TconTS) and their ability to transfer sialic acids from donor to acceptor glycoconjugates. Contents of chapter 3.3 comprise a detailed characterisation of TconTS regarding their pH optimum and enzymatic activities on blood glycoconjugates (serum glycoproteins) and fetuin. The following two chapters (3.4 and 3.5) focus on the lectin-like domain (LD) of TconTS and describe its carbohydrate binding activities and specificities (Chapter 3.4), as well as the influence on enzymatic catalysis (Chapter 3.5). Chapter 3.4 represents a manuscript, which has been recently accepted by the editorial board of PLOS-NTD and is currently in press. Chapter 3.3 and 3.5 are manuscripts under preparation, whereas latter has been prepared for future submission at PLOS-NTD, according to journal guidelines.

A summary of all publications and manuscripts is given in the fourth chapter termed summarising discussion, which generally represents an evaluation bringing all findings and conclusions of each publication/manuscript together, subsequently followed by chapter 5, representing a short outlook for further research and possible applications.

III.Summary

The study presented here addresses structural and functional relations of trans-sialidases (TS) from the African parasite *Trypanosoma congolense* and their biochemical characterisation. Parts of this work have already been published (Chapter 3.1, 3.2 and 3.4).

TS are unusual enzymes found in the flagellate protozoan parasite *Trypanosoma* and catalyse the stereo and region specific transfer of terminal sialic acid (Sia) from donor sialo-glycoconjugates to terminal galactose residues of suitable acceptor substrates, resulting in α2,3-sialylated glycoconjugates. Major research on trypanosomal TS has been done on *Trypanosoma cruzi*, the causative agent of Chagas' disease in Latin America. However, only little has been known about TS from the African *Trypanosoma congolense* (TconTS), the prevalent causative agent of African animal Trypanosomiasis (AAT) in livestock and domestic animals also termed nagana.

Preliminary search of the Welcome Trust Sanger Institute (WTSI) database using partial TconTS sequences revealed 11 closely related gene members termed TconTS1 sharing over 90 % amino acid sequence identity, as well as 3 additional, more distantly related TconTS family members sharing only 40 % sequence identity and therefore termed TconTS2, TconTS3 and TconTS4, respectively. Besides these TconTS genes, encoding active enzymes, 3 further gene members have been identified exhibiting only about 25 % amino acid sequence identity compared to TconTS1 and assumed to encode for inactive enzymes due to the lack of an tyrosine residue at the catalytic centre, known to be essential for catalysis and enzyme activity. All 11 TconTS1 variants as well as TconTS2, TconTS3 and TconTS4 were cloned, expressed as recombinant proteins in mammalian fibroblasts and subsequently biochemically analysed regarding their specific enzymatic activities. Significantly distinct catalytic activities and substrate affinities for all 14 active recombinant TconTS have been determined using a variety of several natural and synthetic substrates, as well as different reaction conditions (Chapter 3.1 – 3.3). Experimental results determined TconTS1 and TconTS2 as highly active TS, whereas TconTS3 and TconTS4 showed significant reduced enzymatic activities. One major exception constitutes TconTS1g, which exhibited a drastically lower Sia transfer activity compared to the other 10 TconTS1 variants, most likely due to the natural mutation R144C found at the catalytic centre in TconTS1g, proposed to be involved in substrate binding (Chapter 3.1).

In silico homology models of TconTS1, TconTS2, TconTS3 and TconTS4 have been generated to investigate a potential relation between the amino acid composition at the catalytic centre of each TconTS and the corresponding observed specific enzymatic activities,

also in respect to the catalytic mechanism published for other trypanosomal TS. However, no clear predictions regarding the differences in enzymatic properties of these four TconTS could be concluded, indicating the involvement of structural regions distant from the catalytic centre influencing the specific enzymatic activities. The most obvious candidate represents the C-terminal lectin-like domain (LD) of TconTS following the catalytic domain (CD). Therefore, a major aim of this study comprises the biochemically characterisation of TconTS-LDs including its carbohydrate binding ability and ligand specificity. LDs from all four TconTS were cloned and expressed as recombinant proteins. In cooperation with Dr. Joe Tiralongo and Dr. Thomas Haselhorst from the Institute for Glycomics (Griffith University Gold Coast, Australia) specific lectin activities for TconTS-LDs have been determined utilising several analytical methods specific to investigate protein-carbohydrate interactions, such as glycan array and saturation transfer difference nuclear magnetic resonance (STD NMR) spectroscopy (Chapter 3.4). In addition, a microtitre plate-based binding/inhibition assay was established to further characterise the lectin activity of TconTS-LD on glycoproteins (Chapter 3.4). In summary, specific binding activities of TconTS-LD to several galactose-, lactose- and mannose-containing glycans, as well as to high-mannose *N*-glycans of glycoproteins have been determined (Chapter 3.4). Furthermore, STD-NMR data provided strong evidence for simultaneous binding of both, lactose and α1-3,α1-6-mannotriose to two distinct binding sites on TconTS2-LD.

The fact that mannose is not a substrate for the catalytic domain of TconTS suggests a different, yet unknown function of TconTS. Experimental data obtained from size exclusion chromatography analysis of recombinant TconTS1 demonstrated its oligomerisation and that high-mannose *N*-glycans are essential for this process (Chapter 3.4). Although the glycosylation pattern of native TconTS still has remained unknown, this finding indicates a potential biological function of TconTS-LD in the organisation of cell surface glycoproteins on the parasites surface.

To further investigate a possible direct influence of TconTS-LD on enzymatic activity a strategy was established allowing the exchange and recombination of CDs and LDs from different recombinant TconTS. Chimeric, domain-swapped TconTS enzymes have been cloned and expressed in bacteria and fibroblasts and analysed for their specific catalytic activities (Chapter 3.5). Although clear Sia transfer activity has been determined for domain-swapped TconTS constructs, the overall enzymatic activities drastically differs relative to that determined for wild type TconTS. It has been demonstrated that if TconTS3-LD was recombined with TconTS1a-CD the corresponding domain swapped TconTS1a/TS3 enzyme

exhibit enhanced sialidase activity when expressed by *E. coli* and suppressed sialidase activity when expressed by fibroblasts relative to that observed for wild type TconTS1a, indicating the influence of LD on enzymatic activity (Chapter 3.5). Along this line, additional experimental results revealed strong evidence that *N*-glycosylation of TconTS-LD potentially influences enzymatic activities (Chapter 3.5). It will be of interest to investigate which precise structural factors are responsible for the observed modulation of enzyme activities and to decipher the underlying mechanisms.

IV. Zusammenfassung

Die hier vorliegende Arbeit befasst sich mit Untersuchungen zu Struktur- und Funktionsbeziehungen von Trans-Sialidasen (TS) des afrikanischen Parasiten *Trypanosoma congolense* und deren biochemischer Charakterisierung. Teile dieser Arbeit (Kapitel 3.1, 3.2 und 3.4) wurden bereits veröffentlich.

TS sind ungewöhnliche Enzyme, die in Flagellaten (Protozoen) Parasiten der Gattung *Trypanosoma* gefunden wurden und den stereo- und regiospezifischen Transfer von endständigen Sialinsäuren (Sia), sialylierter Donor-Glykokonjugate, auf terminale Galaktosereste entsprechender Akzeptormoleküle katalysieren. Produkte dieses Sia-Transfers sind α2,3-sialylierte Glykokonjugate. Der Großteil der trypanosomalen TS Forschung konzentrierte sich bislang hauptsächlich auf die TS des Parasiten *Trypanosoma cruzi* (TconTS), der Verursacher der lateinamerikanischen Chagas-Krankheit. Nur wenig ist hingegen über die TS des afrikanischen Parasiten *Trypanosoma congolense* bekannt, einer der Hauptverursacher der afrikanischen Schlafkrankheit (Animal African Trypanosomiasis AAT) bei Haus- und Nutztieren, lokal auch unter dem Namen Nagana bekannt.

Erste Untersuchungen der Welcome Trust Sanger Institute (WTSI) Datenbank, unter Verwendung von TconTS Teilsequenzen, ergaben insgesamt 14 verwandte Gensequenzen. Davon wurden 11 TconTS Gene mit einer Sequenzähnlichkeit von über 90 % und 3 weitere mit Sequenzähnlichkeiten von nur 40 %, relative zu ersteren, identifiziert und daher als TconTS1 und entsprechend TconTS2, TconTS3 und TconTS4 bezeichnet. Zusätzlich zu diesen 14 Gensequenzen, die für aktive TconTS Enzyme codieren, wurden 3 weitere gefunden, die jedoch nur etwa 25 % Sequenzähnlichkeit mit TconTS1 teilen und für inaktive Enzyme codieren, da diesen unter anderem ein für die Katalyse essentielles Tyrosin im aktiven Zentrum fehlt. Alle Gensequenzen der 11 TconTS1 Varianten sowie von TconTS2, TconTS3 und TconTS4 wurden kloniert, als rekombinante Proteine in Säuger-Fibroblasten exprimiert und anschließend die spezifischen Enzymaktivitäten biochemisch charakterisiert. Dabei wurden teilweise deutliche Unterschiede in den katalytischen Aktivitäten, sowie in den Substratspezifitäten, unter der Verwendung verschiedener, natürlicher und synthetischer Substrate und Reaktionsbedingungen, für alle 14 rekombinanten TconTS Enzyme beobachtet (Kapitel 3.1 – 3.3). Experimentelle Ergebnisse zeigten, dass TconTS1 und TconTS2 hochaktive TS sind, wohingegen sowie für TconTS3 als auch für TconTS4 relativ stark reduzierte Enzymaktivitäten im Vergleich zu ersteren bestimmt wurden. Eine Ausnahme bildet dabei die Variante TconTS1g, bei der eine deutlich geringere Sia Transfer-Aktivität, relative zu denen der anderen TconTS1 Varianten, gezeigt werden konnte. Die Ursache für

die reduzierte Enzymaktivität von TconTS1g ist vermutlich auf die natürliche Mutation R144C zurückzuführen, wobei angenommen wird, dass der durch Cystein ersetzte Argininrest, im aktiven Zentrum, direkt an der Substratbindung beteiligt ist und diese stabilisierende Wechselwirkung in TconTS1g fehlt (Kapitel 3.1).

In silico berechnete Homologiemodelle von TconTS1, TconTS2, TconTS3 und TconTS4 wurden erstellt, um mögliche Zusammenhänge zwischen der Zusammensetzung der katalytisch-essentiellen Aminosäuren im aktiven Zentrum der TconTS und den entsprechenden, beobachteten spezifischen Enzymaktivitäten zu erkennen, auch in Hinblick auf den, in der Literatur beschriebenen katalytischen Mechanismus anderer *Trypanosoma* TS. Allerdings konnten diesbezüglich keine klaren Aussagen zu den unterschiedlichen, enzymatischen Verhalten der vier TconTS Varianten gemacht werden. Diese Ergebnisse deuten jedoch darauf hin, dass noch andere Regionen, außer dem katalytischen Zentrum des Enzyms, scheinbar einen erheblichen Einfluss auf die katalytischen Aktivitäten der TS haben. Dabei bildet die C-terminale Lektin-ähnliche Domäne (LD), die der katalytischen Domäne (CD) direkt über einen α-Helix angeschlossen ist und der bisher keine funktionelle Bedeutung zugewiesen werden konnte, die plausibelste Region. Daher bestand ein wesentlicher Teil dieser Arbeit in der biochemischen Charakterisierung der TconTS-LD und dessen Potential als kohlenhydratbindende Domäne, sowie die damit verbundenen Ligandspezifitäten zu untersuchen. Dazu wurden die LDs aller vier TconTS kloniert und als rekombinante Proteine exprimiert. In Kooperation mit Dr. Joe Tiralongo und Dr. Thomas Haselhorst, vom Institute for Glycomics (Griffith Universität, Gold Coast, Australien), wurden mit Hilfe einer Reihe von analytischen Methoden wie z.B. Glycan-Array und Sättigungs-Transfer-Differenz(STD) Kernresonanz(NMR)-Spektroskopie, speziell für die Analyse von Protein-Kohlenhydrat Wechselwirkungen, die spezifischen Lektinaktvitäten der TconTS-LDs experimentell bestimmt (Kapitel 3.4). Zusätzlich dazu wurde ein Microtiter-Platten basierter Bindungs/Inhibitions-Test entwickelt, mit dessen Hilfe die Lektinaktivitäten der TconTS-LD mit immobilisierten Glykoproteinen weiter im Detail untersucht wurden (Kapitel 3.4). Zusammenfassend konnten die spezifischen Bindungsaktivitäten der rekombinanten TconTS-LD an verschiedene Galaktose-, Laktose- und Mannose-haltige Glykane sowie mit oligomannosidischen *N*-Glykanen von Glykoproteinen erfolgreich nachgewiesen werden (Kapitel 3.4). Zusätzlich ergaben experimentelle Ergebnisse der STD NMR Analysen einen deutlichen Hinweis auf eine mögliche simultane Bindung von Laktose und α1-3,β1-6-Mannotriose an zwei unterschiedliche Bindungsstellen in der TconTS-LD.

Die Tatsache, dass Mannose kein Substrat der TconTS-CD ist, deutet auf eine neue, bisher unbekannte Funktion der TconTS hin. Experimentelle Daten von Größenausschluss-chromatographie-Untersuchungen mit rekombinanten TconTS demonstrierten die *N*-Glykan-abhängige Oligomerisierung von TconTS1a (Kapitel 3.4). Auch wenn das Glyko-sylierungsmuster der nativen TconTS bis jetzt nicht bekannt ist, weisen diese Ergebnisse bereits auf eine mögliche biologische Funktion der TconTS-LD bei der Organisation der Zelloberflächen-Glykoproteine des Parasiten hin.

Um einen potentiellen, direkten Einfluss der TconTS-LD auf die Enzymaktivitäten zu untersuchen, wurde eine Strategie entwickelt, die den Austausch von CDs und LDs der verschiedenen TconTS ermöglicht. Enzym-Chimären bestehend aus den Domänen unterschiedlicher TconTS wurden kloniert, in Bakterien und Fibroblasten exprimiert und bezüglich ihrer katalytischen Aktivitäten analysiert (Kapitel 3.5). Eindeutige Sia-Transfer-aktivitäten der TconTS-Chimären wurden detektiert, wobei sich das generelle, katalytische Verhalten der Chimären deutlich von denen der Wildtyp TconTS unterscheidet. Dabei konnte gezeigt werden, dass, wenn TconTS3-LD mit TconTS1a-CD rekombiniert wurde, die resultierende TconTS1a/TS3-Chimäre relativ zu TconTS1a, eine erhöhte Sialidase-Aktivität aufwies, wenn das Enzym von Bakterien exprimiert wurde und eine deutlich verringerte Sialidase-Aktivität aufwies, wenn es von Fibroblasten exprimiert wurde. Dies war bereits ein Hinweis auf einen möglichen Einfluss der LD auf die Enzymaktiviäten (Kapitel 3.5). Weitere experimentelle Ergebnisse führten zu der Annahme, dass auch die *N*-Glykosylierung der TconTS-LD einen möglichen Einfluss auf die katalytische Aktivität der Enzyme hat (Kapitel 3.5). Daher besteht großes Interesse daran, die genauen, strukturellen Faktoren zu bestimmen, die für die beobachteten regulatorischen Effekte der Enzymaktivitäten verantwortlich sind und damit die zugrundeliegenden Mechanismen aufzuklären.

V. Abbreviations

aa	amino acids
AAT	Animal African Trypanosomiasis (Nagana)
bp	base pair
BSA	bovine serum albumin
BSF	bloodstream form
CD	catalytic domain
CMP-Neu5Ac	citidine-5'-monophosphate *N*-acetylneuraminic acid
Da	dalton
DANA	2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide 5'-triphosphate
DTT	dithiotreitol
E.coli	*Escherichia coli*
FCS	fetal calf serum
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GARP	glutamic acid-alanine-rich protein
Glc	glucose
GlcNAc	*N*-acetylglucosamine
GPI	glycosylphosphatidylinositol
HAT	Human African Trypanosomiasis (sleeping sickness)
HPAEC-PAD	high performance anion exchange chromatography – pulsed amperometric detection
IC_{50}	half maximal inhibitor concentration
Ig	immunoglobulin
IPTG	isopropylthiogalactoside
K_M	Michaelis-Menten constant
Lac	lactose (β-D-galactopyranosyl(1,4)-D-glucose)
LD	lectin domain
mAb	monoclonal antibody

Man	mannose
MU	4-methylumbelliferone
MUGal	2'(4-methylumbelliferyl)galactoside
MUNeu5Ac	2'(4-methylumbelliferyl)-α-D-*N*-acetylneuraminic acid
Neu5Ac	*N*-acetylneuraminic acid
Neu5Gc	*N*-glycolylneuraminic acid
NMR	nuclear magnetic resonance
ORF	open reading frame
PARP	procyclic acidic repetitive protein
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RT	room temperature
SA	sialidase
SAPA	shed acute phase antigen
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
Sia	sialic acid
Siglec	sialic acid binding immunoglobulin-like lectin
SPR	surface plasmon resonance
STD	saturation transfer difference
3'SL	3'-sialyllactose (Neu5Acα2,3-lactose)
6'SL	6'-sialyllactose (Neu5Acα2,6-lactose)
TBS	tris-buffered-saline
TS	trans-sialidase
TbruTS	*Trypanosoma brucei* trans-sialidase
TconTS	*Trypanosoma congolense* trans-sialidase
TcruTS	*Trypanosoma cruzi* trans-sialidase
TranTS	*Trypanosoma rangeli* trans-sialidase
TvivTS	*Trypanosoma vivax* trans-sialidase
Tris	2-amino-2(hydroxymethyl)-1,3-propanediol
v_{max}	maximum velocity
VSG	variable surface protein

Abbreviations for amino acids

Amino acid	One-letter symbol	Three-letter abbreviation	Amino acid	One-letter symbol	Three-letter abbreviation
Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparagine	N	Asn	Methionine	M	Met
Aspartic acid	D	Asp	Phenylalanine	F	Phe
Cystein	C	Cys	Proline	P	Pro
Glutamine	Q	Gln	Serine	S	Ser
Glutamic acid	E	Glu	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile	Valine	V	Val

1. Introduction

1 Introduction

1.1 Carbohydrates

1.1.1 Structure

Approximately 100 years ago, naturally occurring substances with the empirical formula $C_n(H_2O)_n$ ($n \geq 3$ to 9) were termed “carbon of hydrates” or “carbohydrates” due to the suggesting that the carbon atoms are in some way covalently attached to water [1,2]. The major group of carbohydrates comprise the family of saccharides or sugars, in which hexoses, besides pentoses, determine the most common and relevant constituents with related chemical structures. All hexoses in general consist of a six carbon atom chain as backbone, including a core of four hydroxymethylen groups, as well as a hydroxymethyl group on one end and either an aldehyde group (aldose) or an α-hydroxy ketone (ketose) on the other (Figure 1). Each carbon of the four hydroxylmethylene groups is bound to four chemically distinct substituents and therefore represents a chiral centre. However, the substituents around each optical active carbon can be arranged in two different ways, generating a total of 16 (2^4, m^k: m = configurations, k = chiral centres) possible hexoses (stereoisomers). If two hexoses differ in the stereochemical configuration of only one chiral centre, they are also referred to as epimers, for example in the case of D-glucose (Glc) and D-galactose (Gal, C-4 epimer to Glc). According to the rules of nomenclature for organic molecules, the numbering starts at the aldehyde carbon, which is the highest oxidised carbon in the structure and also referred to as the anomeric C-1. The configuration at the stereogenic centre furthest from the C-1 (C-5 for hexoses) determines the overall configuration of each sugar to either L or D. In solution, hexoses are usually found as an equilibrium mixture of linear and cyclic forms (Figure 1). The common cyclic six-ring conformation (pyranose) is created by the nucleophilic addition of the C-5 hydroxyl group to the aldehyde group (C-1) forming a hemiacetal structure with two possible configurations at the C-1, thus inducing a new asymmetric centre. The addition at C-2 leads to the appropriate cyclic five-ring hemiacetal structure (furanose). Depending on the direction of the nucleophilic attack, the hydroxyl group at C-1 in the cyclic form can either be oriented axial or equatorial.

These two possible configurations at the C-1 are distinguishable by the designations α (axial) and β (equatorial) anomers, as shown for D-glucose (Figure 1). Furthermore, in contrast to the other stereogenic centres on the cyclic hexose, the anomeric C-1 can undergo an interconversion of the stereoisomer, due to the permanent transition between the linear and cyclic form, termed mutarotation (fast swap in optical rotation, Figure 1).

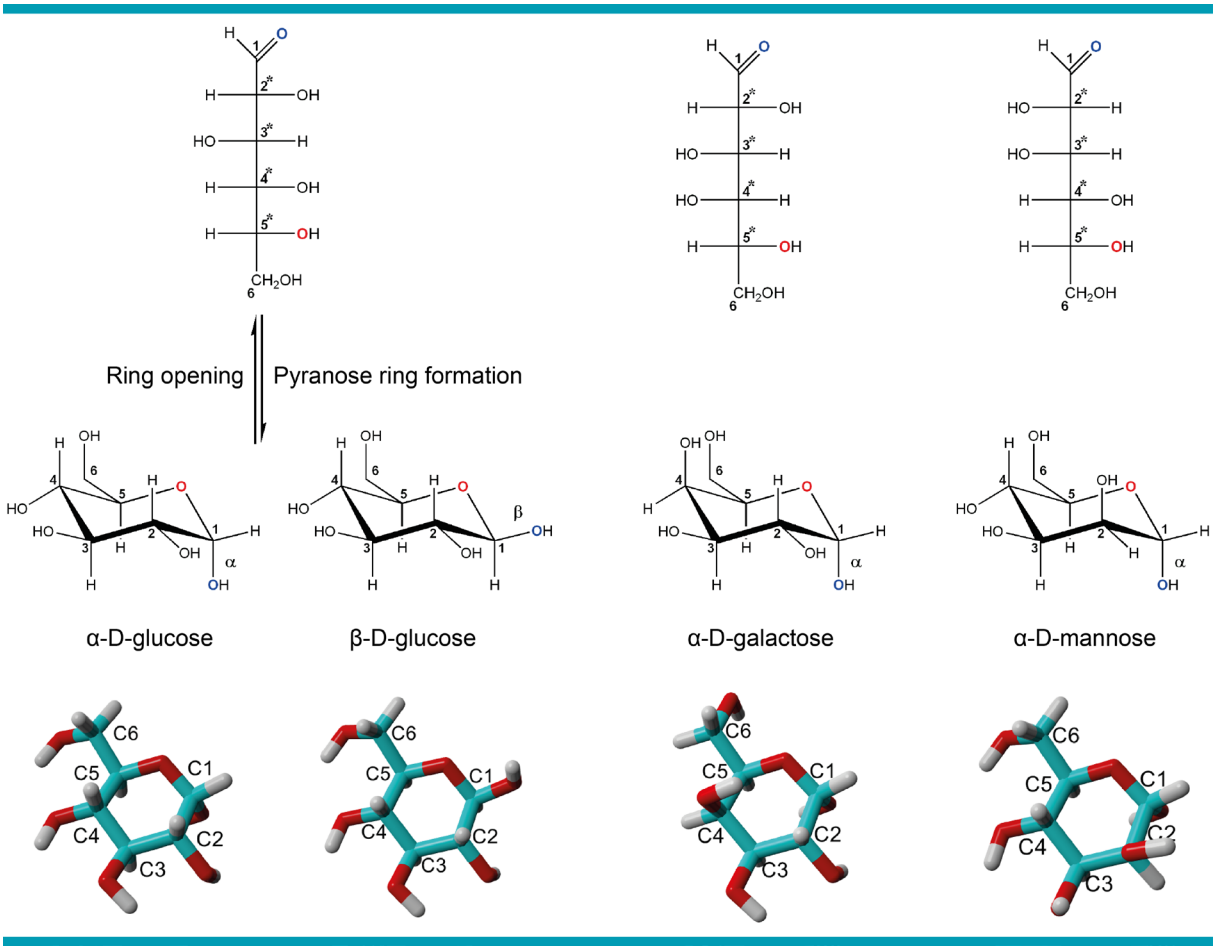


Figure 1. Stereochemistry of common hexoses. Monosaccharides are shown in Fischer projection (top row), Haworth projection (middle row) and as sterical stick-structures (bottom row). The reversible cyclisation of the linear hexose to the respective α or β pyranose ring is termed mutarotation and exemplarily shown for Glc. Stereogenic centres are indicated (*).

The distribution between the open chain and the two anomeric cyclic forms depends on the chemical structure/stability of each sugar, in which former is only present in minor amounts less then 0.01 % [1].

Only a few of the theoretically possible hexoses are commonly found in nature, whereas many of them can be derived from glucose, the most frequently occurring hexose, with only a single epimerisation and/or substitution. Substitutions often comprise modifications at C-6, such as oxidation to a carboxyl group (sugar acid, glucuronic acid, GlcA) or even the removal of C-6 to the resulting pentose, e.g. xylose (Xyl). Furthermore substitution of the C-2 hydroxyl group of glucose and galactose with an acetylated amino group leads to the frequently occurring *N*-acetylated hexoses *N*-acetylglucosamine (GlcNAc) and *N*-acetylgalactosamine (GalNAc). However, all of these common hexoses are normally found in D configuration.

Monosaccharides are able to form polymeric oligosaccharide structures through the bond formation between the anomeric carbon of one monosaccharide and the hydroxyl group

of another sugar (condensation). This fundamental *glycosidic bond*, which formally involves the reaction of a hemiacetal and an alcohol group to an acetal, represents the basis for a variety of different oligosaccharides formed from the broad monosaccharide building block. The combinatorial aspect of this molecular diversity can, in context to glycan structure, exceed those for peptides by orders of magnitude [3]. As an example for the vast increase in complexity of oligosaccharides, a simple disaccharide made of two identical glucose units will be compared to a dipeptide comprising two glycerine molecules. In principle, 11 different disaccharides are possible (5 reactive hydroxyl groups and 1 carbonyl group for each Glc) in contrast to only a single dipeptide. However, the increase in heterogeneity is even more drastically illustrated when going to lager scale. For example, four different amino acids lead to 24 different tetrapeptides but four different hexose monosaccharides can theoretically form 35,560 possible tetrasaccharides [4-6]. This high diversity derives from the variety in glycosidic linkages, including α and β configuration of the glycosidic bond resulting in two stereoisomers, the amount of hydroxyl groups on each monosaccharide leading to several possible regioisomers and the fact that one monosaccharide can exhibit more than two glycosidic bonds, therefore representing a branching point in the oligosaccharide as seen in glycoproteins [7,8]. An intriguing fact is that such a huge repertoire of biologically essential information is not encoded by the genome [9]. Two of the most common and extensively studied oligosaccharides are starch and glycogen, comprising of numerous glucose units linked to form a large biopolymer, which is utilised as an "energy-reservoir" in plants and animals, respectively. Cellulose serves as a constituent of the cell wall of plants and is composed of multiple β-1,4-linked glucose units. In contrast to that, glycosidic bonds of glucose monosaccharides in starch and glycogen are mainly formed by α-1,4-linkages, whereas also α-1,6-linkages are present in glycogen [2].

1.1.2 Carbohydrates, not only just an energy supplier

In general, four major classes of repeating biopolymers, being part of all living things on the planet, can be differentiated in oligonucleotides (DNA and RNA), lipids, proteins and carbohydrates. Latter present the most predominant component of biomasses formed on earth, since they are integrant of animals, plants and bacteria [1]. In contrast to the other three, on the molecular level well understood biopolymers mentioned above, biological functions of carbohydrates mainly have remained relative poorly understood, not at least because of their large and complex structural diversity in nature. Main important biological functions of carbohydrates, besides their role as energy supplier for cellular processes such as protein

synthesis, movement and transport, are particularly the involvement in cell-cell and cell-matrix interactions of complex cell-systems (organs) and organisms, also as a potent communication system.

All cells in nature are covered with a dense layer of a variety of carbohydrates on their plasma membrane termed glycocalyx, which can reach an intraluminal thickness of up to 300 – 500 nm [10]. It was primarily discovered in the early 1940s [11] and subsequently further investigated in the following years [12-14], accompanied with an increasing importance as an essential factor in vascular physiology, homeostasis and pathology [15-17]. This supramembrane carbohydrate coat comprises of single sugars molecules (monosaccharides) and extended sugar chains (oligosaccharides), which are in general specified as glycans, covalently linked (glycosidic linkage) to amino acid residues of soluble free or membrane bound polypeptides (glycoproteins). In addition it was estimated that half of all proteins in nature, known so far, are glycosylated [18].

However, the composition of proteoglycans, with their associated glycosaminoglycan (GAG) side chains forming the major constituent on the plasma membrane, and glycoproteins on the cell surface cannot be considered as a static overall picture, rather as a highly dynamic system exhibiting a permanent exchange of glycoconjugates [15]. Proteoglycans are considered to be the most important functional glycoconjugate species of the glycocalyx, comprising of a high number of different glycosaminoglycans attached, varying in length and type. The main five types of GAGs are: chondroitin/dermatan sulphate [19], heparan sulphate [20], keratan sulphate [21] and hyaluronic acid (hyaluronan) [22]. Structurally GAGs are linear dispersed hetero-oligosaccharides containing a repetitive core disaccharide, made of a hexosamine and an uronic acid (hexoses in which the C-6 is oxidised to a carboxyl group), which can vary between 50 and 150 units [23,24]. Further specifications are made regarding the type of hexosamine incorporated and other modifications, such as sulfation and acetylation.

Besides the class of proteoglycans, with their linear hetero-polysaccharide side chains, a large number of glycoproteins are present on the cell surface comprising of smaller but highly branched oligosaccharides in contrast to the former (further discussed in the following chapter).

The glycosylation pattern of glycoconjugates of a cell is determined by the activity levels of the different glycan modulating enzymes such as glycosyltransferases and glycosidases in cytosol, endoplasmic reticulum (ER), the Golgi apparatus and on the cell surface [25-27]. Significant changes of the glycoconjugate composition and glycosylation

pattern are associated with fundamental processes like embryogenesis, tissue development, stem-cell differentiation, growth, cell contact inhibition, cell-cell recognition (trans-interaction), host-pathogen interaction, cell signalling, host immune response, inflammation, pathogenesis, metastasis, intracellular trafficking, membrane stability and several fatal diseases such as cancer and diabetes [1,4,7,25-39]. The involvement in such a huge variety of different biological processes, clearly demonstrate the importance of carbohydrates in nature.

1.1.3 N- and O-linked glycosylation

The connection of two monosaccharides via formation of a glycosidic linkage requires energy, since it is an enthalpic and entropically unfavourable process [40]. This free energy needed for sugar oligomerisation is obtained from the hydrolysis of high-energy phosphate anhydride bonds in additional reactions, which hence are coupled to the former. As seen for many other energy consuming biological processes, adenosine triphosphate (ATP) is used to drive the formation of a nucleotide sugar donor, such as uridine diphosphate (UDP)-monosaccharide, although other nucleotide sugar donors are also present. The activated "high-energy" nucleotide sugar is then transported via specific nucleotide sugar transporter (NST, antiporter) into the ER or Golgi [41], where it can be transferred to another mono- or oligosaccharide. This reaction is catalysed via a substrate specific glycosyl-transferase, which hydrolyses the phosphoester bond of the activated nucleotide sugar donor and forms the new glycosidic linkage between the monosaccharides [42]. Different glycan modifying enzymes such as glycosyl-transferases and glycosidases exhibit substrate affinity specific to a certain sugar donor or acceptor, thus contributing to a broad regulatory glycosylation system [1]

As discussed in the previous chapter, glycans exhibit a tremendous structural heterogeneity, which is transmitted when glycans are for example attached to proteins. Thus forming a system of gylcoconjugates, modulated with a very high amount of information, providing the contribution to a wide range of biological processes. In general, two possible types of glycosidic linkages between a glycan and the polypeptide chain in glycoproteins can be differentiated. The first involves a covalent bond between the *N*-acetylglucosamine residue of the oligosaccharides reducing end, and the amide side chain of an asparagine residue (*N*-linked) in the polypeptide backbone. This *N*-linked glycosylation incorporates the essential amino acid sequence motif Asn-X-Ser/Thr (*N*-glycosylation site), whereas X represents any amino acid except proline [8]. The second type of linkage is determined by the glycosidic bond of *N*-acetylgalactosamine to the hydroxyl group of serine or threonine side chains (*O*-linked) of the polypeptide backbone [38,43]. Most glycoproteins containing *O*-linked glycans

are often also substituted with one or more *N*-linked oligosaccharides and the other way around [44].

N-glycosylation in eukaryotes starts simultaneously with the translocation of the newly synthesised polypeptide chain into the ER, thus determining it as a co-translational event. The oligosaccharide moiety of the lipid (Dolichol, Dol)-linked *N*-glycan precursor with the constitution $Glc_3Man_9GlcNAc_2$-PP(pyrophosphate)-Dol is transferred to the unfolded, nascent polypeptide chain, via a membrane-associated oligosaccharyl-transferase (OST). *In vivo* studies on cultured cells with depleting $Glc_3Man_9GlcNAc_2$-PP-Dol, revealed loss of cell viability as a consequence of disabled *N*-linked oligosaccharide synthesis [45,46]. This precursor also defines the basis of three types of *N*-glycans, namely high-mannose type oligosaccharides, hybrid oligosaccharides and complex oligosaccharides [47]. Further modifications of the oligosaccharide, including primal removal of three Glc and one Man residue to generate high-mannose type *N*-glycans ($Man_8GlcNAc_2$), occur in the ER via certain glucosidases and mannosidases, respectively. After transfer of the preliminary glycoprotein to the Golgi apparatus (*cis*-, *medium*- and *trans*-Golgi), subsequent modifications of the high-mannose type *N*-glycan through several different glycosidases and transferases lead to either hybrid or complex *N*-linked glycans [33]. This processing comprise the addition of terminal fucose (Fuc), Glc, GlcNAc, Gal, GalNAc and sialic acids (Sia), exhibiting different glycosidic linkages as well as the introduction of further branching and bisecting structures [48-55]. However, efficient and controlled *N*-linked glycan synthesis requires a strict and defined order, in which the participating glycosidases and glycosyl-transferases act. This is provided by the localisation of enzyme activities and the expression levels of the appropriate enzymes [42].

In addition, *N*-linked oligosaccharide structures processed in the ER generate substrates, which support enzymatic reglucosylation in the case of improper protein folding (unfolded protein response, UPR) [56,57]. Appropriate improperly or partially folded proteins will not be allowed to continue molecular trafficking and remain in the ER until proper folding is completed (or the improperly folded protein is degraded) and the glycoprotein may proceed to the next step along secretory pathways [58,59]. Several studies have demonstrated a causal role of deficient *N*-glycosylation in human congenital disorders of glycosylation (CDG) [60,61].

In contrast to *N*-glycans, *O*-linked oligosaccharides appear to be much more diverse in both, structure and function, since the former share at least a common central glycan-protein core and can be classified into a few structurally related types, whereas a different situation is

found for latter [34]. *O*-glycans provide glycosidic linkages involving Gal, GalNAc, Fuc, GlcNAc, Man or Xyl and serine, threonine or hydroxylysine residues of the polypeptide backbone [38,43]. In addition, there are numerous transferases available, which catalyse the glycosidic bond formation of GalNAc to serine or threonine residues [62] compared to only a single oligosaccharyl-transferase for *N*-glycosylation (OST).

O-glycosylation is a posttranslational modification catalysed by a variety of processing enzymes in the Golgi apparatus [43,63], including the addition of monosaccharides such as Gal, GlcNAc, Fuc and Sia after the addition of *O*-linked GalNAc [64,65]. These modifications are similar or identical to that of *N*-glycans, providing evidence for a possible overlap between both types of glycosylation, in which corresponding enzymes might be shared, also indicating their possible co-evolution [42]. However, *O*-glycans exhibit a very high heterogeneity and were demonstrated to effectively modulate enzyme activities by regulating transcription, protein-protein interactions, protein degradation and protein localization [34,66,67]. In addition, Zachara and Hart demonstrated that *O*-phosphorylation (another posttranslational modification involved in regulating protein function) and *O*-glycosylation compete for the same threonine or serine residues in the polypeptide chain, thus modifying the biological function of the mature protein by varying its phosphorylation pattern [34]. Furthermore, the fact that *O*-GlcNAc modified glycoproteins were shuttled from the cytoplasm to the nucleus of see slug *Aplysia* neurons, either indicates a potential role of the posttranslational modification as a alternative nuclear localizing signal (NLS) or as a retention signal, respectively [34,67].

Human congenital disorders of glycosylation (CDGs) for 12 defects in *N*-linked and 4 in *O*-linked glycosylation have been described so far [33,60,61,68]. For example latter include the Walker-Warburg syndrome, characterized by an congenital muscle dystrophy and complex brain and eye abnormalities (induced by an *O*-mannosylation defect [69]) and the hereditary multiple exostosis, which is characterised by bone and generalised somatic overgrowth (caused by an O-xylosylation defect [70]). In summary, *N*-linked and *O*-linked glycans have been associated with different unique functions and were demonstrated to be involved in a variety of essential biological processes and diseases, however many still remain to be explored.

1.2 Sialic acids

Glycoconjugates of all vertebrates commonly consist of hexoses and their derivatives. The only known fundamental exception constitutes the group of sialic acids (Sia), which

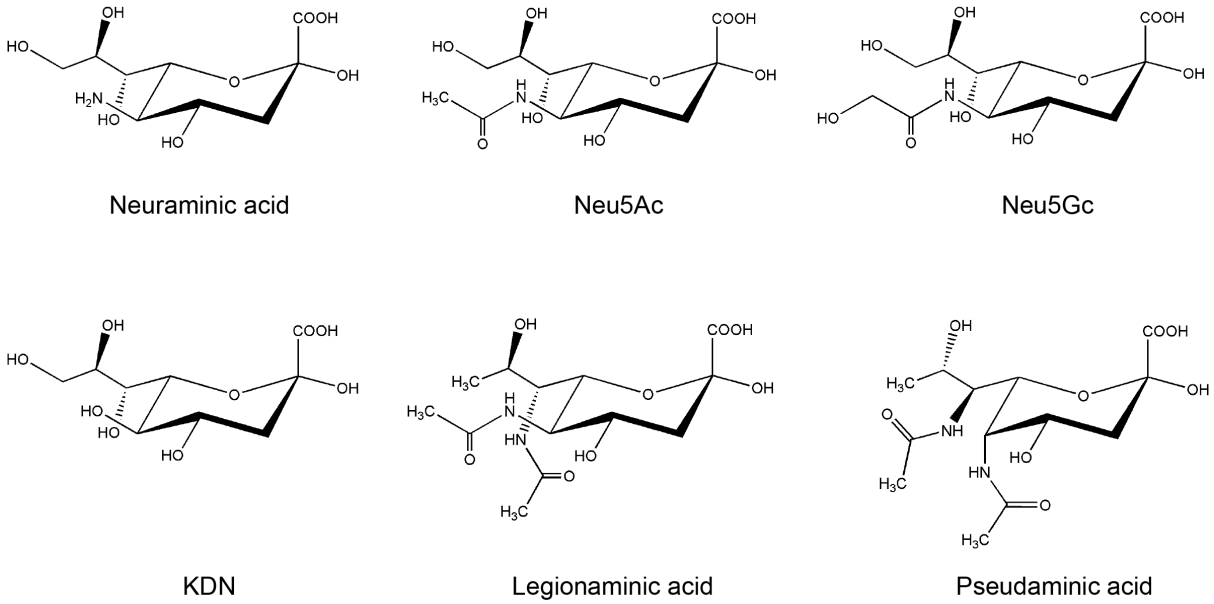


Figure 2. Structure of neuraminic acid and its most common derivatives. N-acetylneuraminic acid (Neu5Ac); N-glycolylneuraminic acid (Neu5Gc); 2-keto-3-deoxy-nonulosonic acid (KDN); 5,7-diamino-3,5,7,9-tetradeoxy-D-glycero-D-galacto-nonulosonic acid (legionaminic acid); 5,7-diamino-3,5,7,9-tetradeoxy-L-glycero-L-manno- nonulosonic acid (pseudaminic acid).

incorporate a number of unusual features compared to other common monosaccharides. Sia describes a family of 9-carbon α-keto acidic amino-monosaccharides [71] mainly found glycosidically linked to terminal Gal, GalNAc or Sia residues of complex oligosaccharide chains of glycoproteins or glycolipids. These are predominantly found on eukaryotic cell surfaces of vertebrates and *Deuterostomia* (e.g. *Asterias rubens*, starfish [72]) or secreted to extracellular medium [73,74]. Sia is also found as capsular polysaccharides or lipooligosaccharides of some pathogenic bacteria, such as *Escherichia coli* (*E. coli*) K1 [75].

The name **sialic** acid/**neuraminic** acid derived from the fundamental discovery of Sia in brain matter (**neuramine**, Blix in 1936) and on bovine submaxillary mucins (BSM, Klenk in 1941), a sialoglycoprotein from **salivary** glands of cows (salios: Greek word for saliva) [75,76]. The first conclusive structure of Neu5Ac was proposed by Gottschalk [77] and clearly illustrated the main specifications of these unique class of monosaccharides. So far over 62 naturally existing Sia forms have been identified, including the two most abundant occurring in eukaryotes, *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) (Figure 2) [76,78-80]. Interestingly the hydroxylase catalysing the reaction which generates the *N*-glycolyl goup in Neu5Gc is absent in humans, due to a mutation in the appropriate gene coding for the enzyme, after the evolutionary divergence of human from the great apes [75]. Similar gene defects occur in other species, including chicken [81,82].

Finally, one last unusual feature related exclusively to neuraminic acids is determined by their synthesis in eukaryotes, which incorporates the condensation of *N*-

acetylmannosamine-6-phosphate (C-4 to C-9 of Neu5Ac) and pyruvate (C-1 to C-3 of Neu5Ac) followed by dephosphorylation to Neu5Ac catalysed by Neu5Ac-9-phosphate synthase (NANS) and Neu5Ac-9-phosphate phosphatase (NANP), respectively [3,42]. Both reactions occur in the cytosol, whereas the activation of Neu5Ac to cytosine monophosphate (CMP)-Neu5Ac via CMP-sialic acid synthetase (CSS) occurs in the nucleus [83,84]. After transport of CMP-Neu5Ac into the Golgi apparatus, further modifications as well as the transfer of Neu5Ac to terminal monosaccharide units, catalysed by numerous different sialyl-transferases (ST), are made, followed by subsequent shuttling to the plasma membrane [83,85]. These STs represent key enzymes in the biosynthesis of sialylated oligosaccharides, glycoproteins and glycolipids [86], transferring Sia from the activated nucleotide sugar donor to terminal Gal, GalNAc or Sia residues of glycoconjugates [74]. The nomenclature of STs refers to the stereochemistry of the glycosidic linkage formed. For example the ST, which catalyses the transfer of Sia to another terminal Sia residue (polysialic acids, PSA), forming a new α2,8 glycosidic linkage is termed ST8Sia [74]. In addition, these polysialic acids (polySia) oligosaccharides, commonly found in vertebrates and *Deuterostomia* [87,88], can exceed 60 Sia units and were found to be involved in cell adhesion of eukaryotic cells, regeneration of the adult nervous system (neural cell adhesion molecule, NCAM [89]), cell trafficking and development of several essential organs, such as hart, brain, kidney, pancreas and respiratory tract [90-97]. Weinhold *et al.* demonstrated that double knockout of ST8SiaII and ST8SiaIV genes in mice, essential for PSA synthesis, resulted in a phenotype exhibiting several tissue defects accompanied by precocious death [98]. This clearly demonstrates the importance of such sialyl-transferases and their crucial function in cell development.

Besides common sialyl-linkages found in eukaryotes including α2,3 linkage to Gal, α2,6-linkage to Gal and GalNAc as well as α2,8 linkage to additional Sia, also more rarely occuring linkages such as α2,3 linkage to GalNAc, α2,6 linkage to Glc and GlcNAc, α2,4 linkage to Gal and GlcNAc and α2,9 linkage to additional Sia and Neu5Gc have been identified [72,99-107].

Furthermore, the nonulosonic acid family also comprise the common 2-keto-3-deoxy-nonulosonic acid (KDN), 5,7-diamino-3,5,7,9-tetradeoxy-D-glycero-D-galacto-nonulosonic acid (legionaminic acid) and 5,7-diamino-3,5,7,9-tetradeoxy-L-glycero-L-manno-nonulosonic acid (pseudaminic acid), whereas latter two are C-4-epimers (Figure 2) [76]. Besides these neuraminic acid derivates, a variety of substitutions mainly at the C-5 position (e.g. *N*-acetyl, *N*-glycolyl) [108] but also at the C-4, C-7, C-8 and C-9 hydroxyl groups are known, incorporating modifications such as *O*-methylation, *O*-acetylation, *O*-sulfation, as well as *O*-

acylation (formation of lactyl-esters) [76,109]. Furthermore, Bulai *et al*. demonstrated that 20% of the sialic acids bound to surface molecules of human red blood cells (RBC) are others than Neu5Ac or the unacetylated neuraminic acid [110]. These findings clearly demonstrate the wide chemical diversity of neuraminic acids, which goes beyond that of every other known monosaccharide [76,82]. The reasons for that still remain not entirely clarified, but some suggestions were made, regarding a potential result of an permanent evolutionary selection by host-pathogen interactions [74,109,111]. Sia are involved in a variety of biological processes, for example nervous system embryogenesis, cancer metastasis, immunological regulation, bacterial, viral and parasitic infections and the involvement in several serious diseases [39,76]. One of the well-known human disorders in Sia metabolism is called Sialuria (OMIM 269921). A extreme rare inborn disease, caused by a mutation in the gene encoding for the bifunctional enzyme GNE (uridine diphosphate (UDP)-GlcNAc 2-epimerase/ ManNAc kinase) [112-114], which catalyses formation of the precursor *N*-acetylmannos-amine-6-phosphate. Due to this defect an excessive synthesis and accumulation of free Sia in body fluids is observed, caused by defects in allosteric feedback inhibition of GNE [115,116]. Features of Sialuria incorporate hepatosplenomegaly (unusual enlargement of liver and spleen) and varying degrees of developmental delays [117]. Other disorders comprise the lysosomal storage disorder sialidosis (OMIM 256550), a neuraminidase 1 (NEU1) deficiency [118,119] and the xenosialitis, causing chronic inflammation, which is due to the generation of antibodies against Neu5Gc and Neu5Gc-containing glycans in human tissue (xeno-autoantigens), induced by the uptake of nonhuman Neu5Gc from dietary sources (red meat) [73,120].

1.2.1 Biological function and relevance

Together with the negatively charged GAGs, sialic acids mainly contribute to the overall negative charge of surface glycoproteins of the glycocalyx. This property is known to protect cells, such as the highly sialylated RBCs (more than 10 million Sia molecules per erythrocyte [1]), against immune response, therefore considered to be members of the innate immune system [82], and underlying glycoproteins from proteolytic degradation [121]. Furthermore, sialylation generates repulsive interactions between two negatively charged cells or sialylated molecules such as mucins [122], indicating the role of Sia in cell-cell interactions (adhesion) and involvement in cell signalling (e.g. hormone induced cell activation) [81,82]. Also its terminal positioning supported several hypotheses that Sia participate in modulating interactions with the environment. Along this line, Sia was reported to effectively mask

receptors, antigenic sites and not at least the directly underlying Gal, GalNAc residues [123]. This effect can be abolished by sialidases (unmasking of Gal), in which the subsequent binding of molecules or cells, such as macrophages, to exposed Gal residues can occur, for example via Gal-recognising receptors (Hepatocytes-asialoglycropotein receptor [121]). On the other hand, Sia itself can also serve as target molecule for hormones, antibodies and sialic acid specific carbohydrate binding proteins (lectins), expressed on several cell surfaces and also known as siglecs (sialic acid-binding immunoglobulin (Ig)-like lectin) [111]. They were found in human and animals [124] and demonstrated to be involved in many physiological and pathological processes [82,111,125-127].

1.3 Lectins

Lectins are defined as carbohydrate recognising and binding proteins [7,8,128]. Derived from the Latin word legere, synonymous for the English word “select”, lectins have been generalised to comprise all non-immune carbohydrate binding agglutinins without enzymatic functions, also termed hemagglutinins or phytoagglutinins due to their discovery from plant extracts, irrespective of blood-type specificity or source [129]. One intensively investigated example is ricin, a plant lectin isolated from seeds of the castor oil plant (*Ricinus communis*). It was suggested that ricin was the very first characterised hemagglutinin, described by Peter Hermann Stillmark (German-Baltic pharmacologist) 127 years ago (reviewed by Hartmut Franz 1988 [130]). Ricin is a highly toxic lectin, recognising terminal galactosyl residues of cell surface glycoconjugates. After attachment and following vesicular transport into the cell and to the ER, where it is able to avoid proteolytic destruction, ricin folds into an active conformation, which inactivates localised ribosomes. Consequences are the collapse of protein biosynthesis and subsequent cell death by initiated apoptosis [131]. The United States military developed a ricin bomb during World War II, which was tested by the British military, but fortunately never used as a mass destruction weapon [129].

With increasing knowledge about agglutination and hemagglutinins, multiple lectins were isolated and characterised from several organisms, such as plants, microorganisms and animals [129]. Interestingly, even though lectins from plants and animals exhibit similar carbohydrate binding specificities, they do not show any primary structural homology, which might be due to potential co-evolution [8], indicating a fundamental role of carbohydrate-lectin interactions. With the use of x-ray crystallographic methods, tertiary structures of numerous lectins could be solved over the last decades. It was observed that despite several differences in primary sequence of many lectins, significant similarities in their tertiary

structures exist [129]. Furthermore, additional structural studies regarding carbohydrate-lectin interaction suggested the involvement of only particular amino acids at a well-defined location in the protein, termed as carbohydrate-recognition domain (CRD). These domains commonly recognise the terminal sugar residues of an oligosaccharide, but also lectins with extended or even multiple thermodynamically distinct CRDs have been described, in which more than one monosaccharide of the same glycoconjugate interacts simultaneously with the lectin [129,132].

Lectins are classified in families and superfamilies according to structural relations and type of carbohydrate recognition [47,133]. L-type lectins represent the first carbohydrate-binding protein family discovered from seeds of leguminous plants [134], sharing structural conserved motifs which were also found in several animal lectins [47]. The family of L-type lectins has been intensively investigated, whereas many different members have been isolated, characterised and employed as practical tools in a variety of biochemical, analytical and biomedical processes, such as affinity purification of glycoconjugates, for specific glycan detection, or as diagnostic and therapeutic tools [7,47,135,136]. Other lectin families comprise M-type lectins, which are closely related to α-mannosidases of the ER and Golgi, recognising high-mannose-type glycans (Man_8); P-type lectins exhibiting unique β-rich structures and selective binding to phosphorylated Man-6-P; C-type lectins, which require Ca^{2+}-ions for carbohydrate recognition, include among others collectins and selectins (mannose-binding protein, MBP); S-type lectins (later termed galectins due to preferred βGal binding), which require free thiols for structure stability and binding activity (β-sandwich structure) and R-type lectins, exhibiting a CRD similar to that of ricin and binding affinities to various carbohydrate ligands [132,137-145]. Due to various carbohydrate-binding specificities of lectins listed above, they are involved in a variety of biological processes, such as ER associated protein degradation, protein sorting in the ER and post-Golgi, cell adhesion, rolling of lymphocytes on endothelial cell surface, B-cell activation (innate immune system), phagocytosis, glycan cross-linking in the extracellular matrix, enzyme targeting (glycoprotein hormone turnover) and are involved in human immunodeficiency virus (HIV) infection of T-cells [146-154]. However, lectins cannot be exclusively classified by their carbohydrate-binding properties, since some CRDs share similar features, for example C-type and L-type lectins both facilitate divalent Ca^{2+}-ions for carbohydrate recognition, respectively. However, they significantly differ in their protein overall structure, in which they are clearly distinguishable. In addition, in some cases also the oligosaccharide-ligands recognised by a specific type of CRD are remarkable similar in structure, whereas in other cases relatively

high structure diversity was observed [42,47]. These various and complex lectin-carbohydrate interactions generate a tremendous biological diversity of receptor-ligand recognition, also in respect to the heterogeneity of glycans itself. To discuss all these very interesting phenomena in detail goes beyond the scope of this thesis, but at this point it should be referred to some excellent publications and reviews available, for further reading [8,42,47,111,126,129,132,133,145,155-158].

1.3.1 Siglecs

Besides the previous discussed lectin families another very important group of carbohydrate binding proteins comprise the I-type lectins. This family belongs to the immunoglobulin (Ig)-superfamily, comprising over 500 mammalian members besides antibodies, in which a define set of Sia-binding proteins were identified forming a structural closely related group of CRDs termed siglecs (Sia-binding immunoglobulin-like lectins) [124,159]. They represent the most intensively and well-characterised group of I-type lectins regarding their structural and functional properties [42]. Siglecs are transmembrane proteins, constitute of multiple extracellular immunoglobulin (1 unusual V-set and 1 to 16 C2-set domains [124]) domains and one immunoglobulin constant (C-set) domain, as well as an inter-membrane domain and for some siglecs also several tyrosine containing, cytoplasmic immunotyrosine-based inhibition motifs (ITIMs) [111,160,161]. Latter can be phosphorylated by specific kinases and lead to modulation of cell-signalling cascades.

The three siglecs, sialoadhesin (Sn), CD22 (Siglec-2) and the myelin-associated glycoprotein (MAG) expressed on macrophages, B-cells and oligodendrocytes respectively, represented the first three family members, which were shown to specifically recognise terminal sialic acids of cell surface glycoconjugates, as target ligands in cell adhesion processes [160,162-164].

In addition, CD22 was cloned and expressed as recombinant protein, C-terminal fused to the hinge region and constant Fc-part of human IgG [163]. Subsequently it was demonstrated that the first two extracellular Ig-like domains (V-set and C-set) of CD22 are involved in Sia recognition [165-167], whereas only the first (V-set) was required in case of Sn [163,168]. Furthermore, also sialic acid binding for the siglecs CD33 (expressed by different cell types of the immune system), human myelin-associated glycoprotein (MAG), expressed on oligodendrocytes, and its avian ortholog Schwann cell myelin protein (SMP) (grouped together as Siglec-4a and b), could be demonstrated [163,169]. Not surprisingly, the Sia binding-site of MAG was also determined to be located in the N-terminal V-set Ig domain

[170]. The fact that all the four siglecs mentioned above share a homologue V-set domain, which has been identified as Sia binding domain, in addition to their function as cell adhesion molecules, lead to the designation “Sialoadhesin family” or “Sialoadhesins” [171], which was changed to siglecs in the following years [124]. Furthermore, additional siglecs have been identified, characterised and termed CD33-related Siglecs, due to their high degree in structural relation to CD33 [161].

Interestingly, all siglecs are expressed in a highly cell type specific fashion [111], leading to the assumption that they are involved in a variety of different cell processes. Indeed, it was demonstrated that Sn (restricted to macrophages [164]) and CD33 are involved in regulation of myeloid cell interactions [169,172], MAG was shown to play a role in myelin maintenance and to inhibit axonal growth [163,173-175], whereas CD22 was demonstrated to be a negative regulator of B-cell activation, and plays a potential role in B-cell guidance to bone marrow [176-179]. The diverse expression profile of CD33-related siglecs in the hematopoietic and immune cells [161], together with the presence of cytosolic ITIM motifs, suggests a potential role as inhibitory receptors, down-regulating innate immune cell reactivity [111,161]. Recently, Jellusova and Nitschke demonstrated that CD22 and Siglec-10 (CD33-related siglec) negatively modulate B-cell antigen receptor signalling [179]. Besides that, a potential involvement of CD33-related Siglec-5 in bacterial uptake has been reported [180]. All these functional related findings strongly suggest potential divergent Sia binding specificities of siglecs, which was approved in several studies. Remarkable different binding preferences were found for Sn, CD22 and MAG. In contrast to CD22, Sn and MAG preferably bind Neu5Ac [181,182], whereas former binds either Neu5Gc or Neu5Ac [183]. Regarding the Sia-linkage to the underlying Gal, CD22 is highly specific for α2,6-linked Sia [184], whereas Sn, MAG and CD33 exhibit binding specificities to α2,3-linked Sia with different affinities, respectively [111,163]. The presence of siglecs on different cell types of the immune and nervous system clearly demonstrates their fundamental importance in regulating both systems. However, with respect to the high structural diversity of Sia in nature as mentioned above, numerous unknown functions of siglecs may remain to be discovered.

1.4 Trypanosomes and trypanosomiasis

Plasmodium, *Leishmania* and *Trypanosoma* are all protozoan parasites, causing fatal diseases in humans as well as in domestic and wild animals [185]. Unicellular flagellate trypanosomes (Figure 3) are the major causative agent of Latin American Chagas’ disease [186] and African Trypanosomiasis, whereas latter is further divided into Human African

Trypanosomiasis (HAT, Sleeping sickness in human) [187] and Animal African Trypanosomiasis (AAT, Nagana, Sleeping sickness in animal) [188].

Kinetoplastid parasites evolve in a variety of hosts and vectors, depending on their presence and natural habitats respectively. For example, the Latin American parasite *Trypanosoma cruzi* (*T. cruzi*) is transmitted by blood sucking triatomine bugs through their defecation. In contrast, African trypanosomes are mainly taken up and transmitted by tsetse flies, during their blood meal [189]. Both, African and Latin American parasites undergo different developmental stages during their life cycles, in which they change their morphology and biochemical composition.

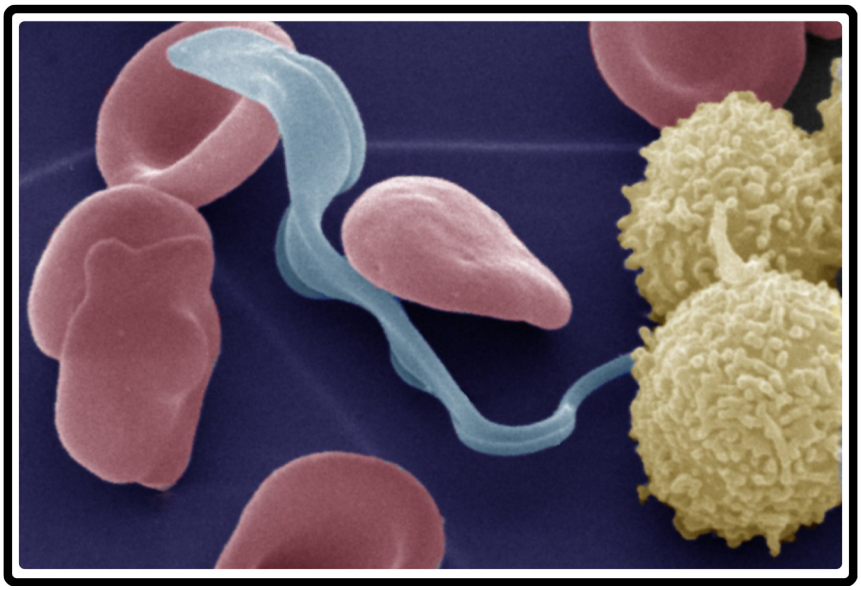


Figure 3. Coloured electron microscopic image of *Trypanosoma brucei* (light blue) in the presence of erythrocytes (red) and lymphocytes (yellow). The image was downloaded from the Deutsches Elektronen-Synchrotron (DESY) homepage. http://www.desy.de/information__services/press/press_releases/2012/pr_291112/index_eng.html. The original image was provided by Prof. Dr. Duszenko (permission was granted).

1.4.1 Chagas' disease, HAT and Nagana

Chagas' disease, caused by *T. cruzi*, was first discovered in 1909 by the Brazilian Scientist Carlos Chagas, who is the eponym of this disease and reported primary data about parasites transmission cycle and clinical manifestations [190]. Chagas' disease is categorised as one of the world's 13 most tropical neglected diseases [191]. First symptoms of the acute phase of this disease are in some cases relatively mild [186] and therefore often neglected, also because for most of the infected patients experience a spontaneous decay of symptoms after a short time, leading to fatal misdiagnosis. Symptoms of the chronic phase are more

lethal and will be only observable 25-30 years after infection [192,193]. Besides this it should be mentioned here that paleoparasitologists isolated *T. cruzi* DNA from 9000 years old human mummies, indicating that people had already suffered from Chagas' disease long ago [194]. Also Charles Darwin potentially suffered from this disease. In his records (*The Voyage of the Beagle* [195]), he mentioned that he got in contact with Triatomine bugs during his expedition to Latin America in 1835. As a possible consequence, he suffered from symptoms for the rest of his life, similar to that reported for patients infected with *T. cruzi* [196]. However it still remains unclear whether these symptoms are actually related to Trypanosomiasis, since Darwin died 1882 and Carlos Chagas (1879-1934) discovered the disease 27 years later. Recently, it was reported that Chagas' disease is not only restricted to Latin America anymore, but has also spread to non-endemic regions such as North America, Canada, Europe and the western Pacific regions, most likely due to infections by blood transfusion, organ transplantation, directly from mother to infant or even orally through food contamination by vector's feces [197-203].

HAT is a fatal, tropical neglected disease (WHO) that is caused by two subspecies of the African *Trypanosoma brucei* (*T. brucei*), namely *T. brucei gambiense* and *T. brucei rhodesiense* [204]. In contrast, *T. b. brucei*, which is only infective for livestock [205], together with the related *Trypanosoma congolense* (*T. congolense*) and *Trypanosoma vivax* (*T. vivax*), represent the major causative agents of the wasting Animal African Trypanosomiasis (AAT), also termed Nagana in livestock and wild animals [188]. *T. brucei spp.* and *T. congolense* are transmitted by several tsetse fly sub-species (*Glossina* spp.), which are exclusively found in sub-Saharan Africa (tsetse belt, covering an area grater than the entire Australian continent), therefore also localising HAT and AAT in that region. Furthermore, two additional species have been reported, *Trypanosoma evansi* (*T. evansi*), causing surra [206], and the non-pathogenic *Trypanosoma lewisi* (*T. lewisi*), which are widely spread over more then three continents. This might be due to the wide distribution of their correlated vectors. In addition, further trypanosomal subspecies and some cases of atypical Human Trypanosomiasis (aHT) have been reported and were recently reviewed [207].

Without appropriate treatment, HAT is a fatal disease, estimated to affect 10-30 thousand individuals every year, representing a recent decline compared to 2006 with about 60 thousand new cases [208]. This drop is most likely related to several reasons, such as the cessation of civil wars, the increased engagement of different organisations (e.g. WHO), production and free supply of anti-trypanosomal drugs, efficient strategies for vector control and active surveillance of the population in endemic countries [209]. However, many cases

remain undiagnosed, due to a massive lack of clinical representation and surveillance in unstable areas. In addition, it has been reported that some tsetse fly subspecies have already adapted to areas with high population densities as in West Africa [210], thus possibly spreading this disease.

Pathogenesis evolves in two distinct stages, including the haemolymphatic stage characterised by presence of parasites only in the blood and lymphatic system, followed by a more chronic phase, in which trypanosomes start invading the central nervous system and even cross the blood brain barrier into the brain and cerebrospinal fluid. Several stage specific clinical symptoms were observed. Finally, untreated sleeping sickness will lead to coma or death [187].

Nagana, the wasting disease in domestic and wild animals is not affecting humans directly, but indirectly by annual losses in agricultural productivity of about 4.5 billion USD. This is mainly due to death of about 3 million cattle, whereas 40 million are threatened, and the low productivity of those who survived but suffered from wasting effects of the disease [211]. Thus animal trypanosomiasis impedes agriculture development in tsetse fly afflicted regions. Common domestic animals suffering from AAT, besides cattle, include sheep, goats, pigs, horses and camels. Also Nagana is fatal without proper treatment, whereas typical symptoms comprise anemia, drastic weight loss, infertility and immunosuppression.

1.4.2 Life cycle of African trypanosomes

African trypanosomes are mainly transmitted by tsetse flies, and require both, mammalian host and tsetse vector for efficient development. During their life cycle *T. brucei* and *T. congolense* undergo complex developmental changes, immediately initiated after ingestion of an infectious blood meal by the tsetse fly [212-214]. A schematic illustration of the general life cycle of African trypanosomes is shown in Figure 4. Trypanosomal bloodstream forms (BSF) are taken up during blood meal and settle in the midgut of the tsetse fly, where they differentiate into the non-infectious, long procyclic trypomastigotes. One fundamental feature of this form is a drastically change in composition of their surface glycoproteins to evade flies digestive system, which comprises several antimicrobial peptides and lectins as defence system [215-217]. After cease dividing they become mesocyclic trypomastigotes and start migrating to proboscis lumen (*T. congolense*) or salivary glands (*T. brucei ssp.*) of the insect. By reaching the proboscis or salivary glands, parasites differentiate into epimastigotes, which are able to adhere to salivary gland epithelium cells (*T. brucei ssp.*) or fly's proboscis and mouth-parts (*T. congolense*) via their flagellum.

In this stage, a second amplification occurs, after which they detach and differentiate into infective metacyclic trypomastigotes. The parasitic developmental cycle in the fly requires 2-5 weeks [218,219]. After injection of metacyclic trypomastigotes into the mammalian host during a blood meal, parasites keep multiplying at the site of injection for a

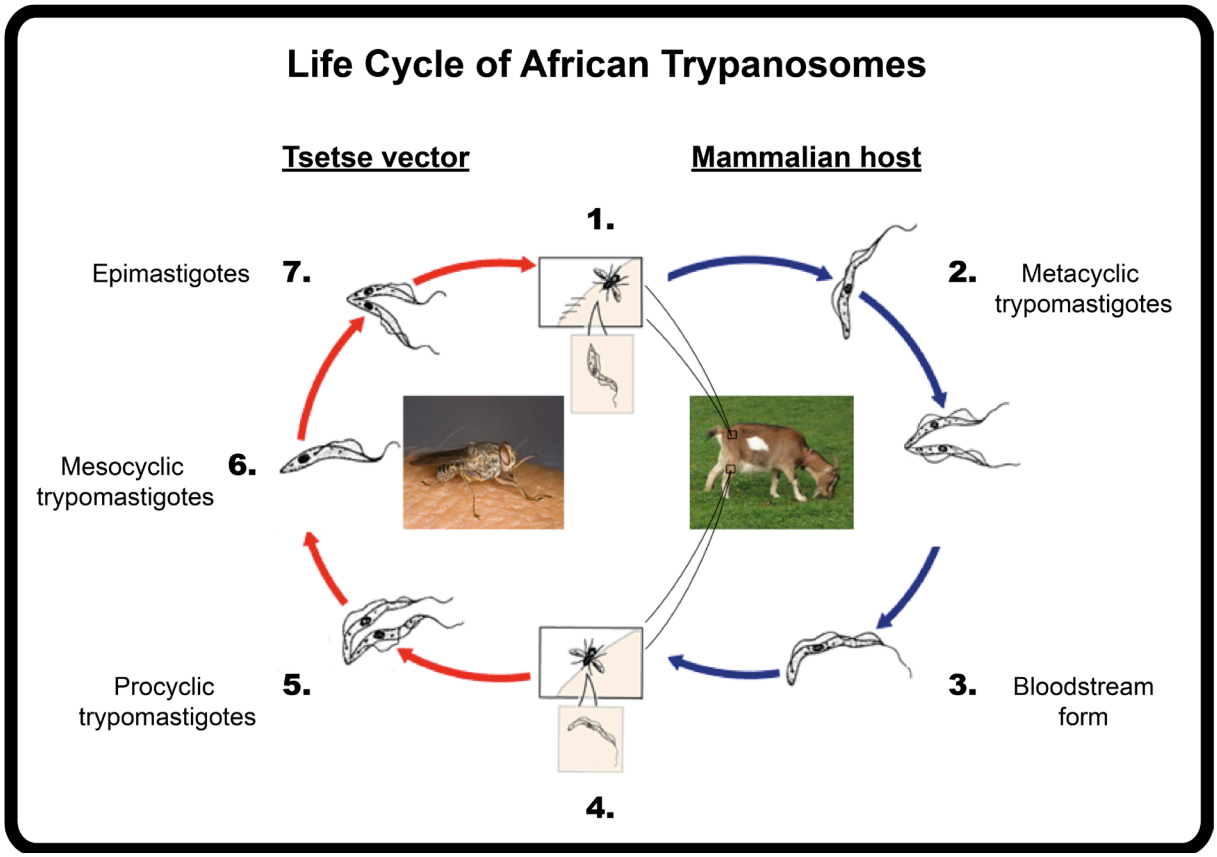


Figure 4. Schematic scheme showing the general life cycle of African trypanosomes.**1**) Infected tsetse fly taking a blood meal transmits infectious traypanosomes into mammalian host. **2**) Injected metacyclic trypomastigotes transform into bloodstream trypomastigotes, multiply in blood and establish an infection. **3,4**) Bloodstream trypomastigotes are taken up by the tsetse fly during blood meal on infected animals. **5**) Ingested bloodstream trypomastigotes transform into procyclic trypomastigotes and proliferate in the tsetse midgut. **6**) Mesocylic trypomastigotes migrate towards proboscis (*T. congolense*) or salivary glands (*T. brucei ssp*). **7**) Epimastigotes multiply and subsequent transform into infectious metacyclic trypomastigotes. Note that this is a general description of the trypanosomal life cycle and that slight differences between several species exist. The graphic is a modified version of the original downloaded from Centers for Disease Control and Prevention (CDC), (http://www.cdc.gov/para-sites/sleepingsickness/biology.html)(permission was granted).

few days, before they invade the lymphatic system followed by migration to the bloodstream, thus finally differentiating into BSF trypomastigotes and completing their life cycle [220].

1.4.3 Parasite's strategies to evade immune response in host and vector

In contrast to *T. cruzi*, which invades host cells to evade immune response [30], African trypanosomes are extracellular parasites, thus exhibiting different strategies to ensure its

survival in the mammalian host. BSF trypomastigotes are able to multiply by longitudinal binary fission at all sites in the mammalian host, in which a periodically fluctuating parasitaemia is observed, due to immune response. The reason for that is mainly related to the parasite surface, which consists of a various number of antigenic, but highly variable surface glycoproteins (VSG).

Trypanolytic factors (TLF1 and TLF2 [221]) found in human serum are able to lyse trypanosomes upon entry [222], therefore serving as a natural host innate parasite defence mechanism (immunity) besides the innate complex immune system. However, trypanosomes evolved multiple strategies to evade these innate defence mechanisms, in order to insure their survival in host and establishment of chronic infection. Besides evolved resistance against TLF, it was recently found that specific trypanosomal adenylate cyclases inhibit TNF (Tumor necrosis factor) production of myeloid-cells and therefore block the early host immune response [223,224]. Nevertheless this parasite-host interaction has to be well regulated, since a parasitical overgrowth would lead to early host death and therefore counterproductive for the establishment of pathogenesis and a subsequent vector infection. One possibility to overcome a potential overgrowth involves the differentiation from the proliferating to a non-dividing form, which is pre-adapted for a effective subsequent vector infection [220]. However, another more efficient system to evade the immune system and ensure permanent parasitaemia in the mammalian host is termed antigenic variation, a strategy which is also observed for other pathogens [225,226].

The surface of *T. brucei* and *T. congolense* metacyclic and BSF trypomastigotes is densely covered with glycosylphosphatidylinositol (GPI)-membrane bound surface glycoproteins (VSG) [227]. Each trypanosome has the potential to express hundreds of different VSGs, encoded by a repertoire of more than 1000 distinct VSG genes, whereas only one is expressed at a time [228]. These immunogenic VSGs induce host immune responses, which initially set the parasite under permanent immune pressure. Before the first generation of parasites, expressing the same VSG, are killed completely by the host's immune response, a second generation has evolved, presenting another type of VSG, which will not be recognised by currently generated antibodies [229]. These new antigenic distinct trypanosomes amplify and overgrow the first until they in turn are cleared by another set of antibodies. This process repeats multiple times, thus resulting in a periodically observable parasitaemia [230]. This efficient switching of the entire VSG coat occurs spontaneously, antibody independent and enables the parasite to continuously evade new antibody attacks and results in the development of a chronic infection in the host [230,231].

However, another scenario is observed in the fly vector. In contrast to the VSG coat of metacyclic trypomastigotes and BSF of *T. brucei* and *T. congolense*, both, their procyclic and epimastigote forms exhibit a more restricted composition of surface glycoproteins, termed procyclins [227,232]. After trypamosomes were taken up by tsetse fly during a blood meal and subsequent differentiation into procyclic trypomastigotes, parasites shed their VSG coats via involvement of GPI-phospholipase C as well as a metalloprotease [219], and express a new layer of procyclins. In *T. brucei* this layer consist of two related types of procyclins, comprising either multiple tandem repeats of glutamic acid and proline (EP, single-letter amino acid code) or protease-resistant pentapeptide repeats of the type GTEEP, respectively [218,233,234]. The appropriate analogues in *T. congolense* have also been identified, termed protease-resistant surface molecule (PRS), glutamic acid-alanine rich protein (GARP) [235] as well as *T. congolense* procyclin [236]. These surface glycoproteins exhibit a specific expression profile during development in tsetse fly. They are glycosylated with polymannose and galactose containing oligosaccharides [237], for which the exact composition still has remained unclear. Procyclins are membrane bound via a unique, modified GPI-anchor [238], consisting of an unusual lipid structure and a polydisperse carbohydrate unit comprising of a branched unique poly-*N*-acetyllactosamine side chain and a mannose core [236,238,239]. In addition, it has been demonstrated by Engstler and Schauer [240,241], as well as Schenkman and co-worker's [242] that terminal galactose residues of this poly-*N*-acetyllactosamine side chains are substituted with α2,3-linked sialic acids. Interestingly trypanosomes are unable to synthesise Sia and utilise unusual enzymes termed trans-sialidases (TS), which appear to be unique to these parasites, to transfer terminal Sia residues from the environment to their own surface molecules, [241,243,244] (A more detailed description of these enzymes is provided in chapter 1.5). However, strong evidence suggested that polyanionic procyclins, in combination with the negatively charged Sia residues of their GPI-anchors, form a strong negatively charged glycocalyx, which protects the parasites against the fly's digestive system [241,244]. Ruepp *et al*. demonstrated that the introduction of GARP into procyclin-knock out *T. brucei* recovered their ability to establish infection in the flies midgut, indicating their fundamental role in parasite survival and development [245]. Nevertheless, these suggestions were supported by findings, demonstrating the absence of procyclins or related structures in *T. vivax*, a trypanosomal species, which apparently develops exclusively in the flies proboscis, but seems to be unable to colonise the flies midgut [236] and therefore do not require a protection against digestive enzymes.

However, these strategies of antigenic variation, incorporating the expression of VSGs in mammalian host, as well as the expression of procyclins and their subsequent sialylation in the vector represent very effective defensive systems, which seem to be unique to trypanosomes, and ensures survival and development at all stages during their life cycles.

1.4.4 Therapies and pharmaceuticals against trypanosomiasis

In order to fight Chagas' disease, HAT and Nagana, several vaccination strategies have been generated over the past years, but not yet resulted in a single promising therapy with a 100 % sterile immunity against the disease or even found its way to a serious field application [246]. With respect to the antigenic variation strategy of the parasite to evade host immune system discussed above, it is obvious that a potential vaccination against VSG is impossible. Another major challenge is the ability of some trypanosomes to cross the blood-brain barrier (BBB) and invade the brain and central nervous system, because most pharmaceuticals cannot pass and therefore are ineffective. Thus, questionable chemotherapies were used to treat HAT infections, including the medicaments pentamidine and suramin for the early phase, as well as eflornithine, nifurtimox and melarsoprol against the chronic phase of the disease [247]. Besides an increasing resistance against these drugs, a series of adverse and even toxic effects were observed but endured due to the absence of more effective therapeutic alternatives [247]. In addition, increasing treatment failure rates for suramin (polyanionic naphthalene bearing compound with unknown mechanism of action) [248] and melarsoprol (highly toxic arsenical diamidine prodrug, which irreversibly and unspecifically inactivate enzymes) [249] have already been reported and represent a current, fundamental growing problem in Africa. However, the mechanisms of action for most of the trypanocidal-compounds (anti-trypanosomiasis medicaments) remain still unclear. A recent publication regarding the identification of several trypanosomal factors involved in anti-trypanosomal drug interactions, revealed advances for a better understanding of drug efficacy and resistance and may provide a further step towards novel drug design and more efficient therapeutic applications [250].

1.5 Trypanosomal trans-sialidase

The biological importance of Sia and its modifying enzymes, such as sialyl-transferases and sialidases (neuraminidases) have been discussed in the previous chapters (1.2). On one hand, these enzymes were found in several catabolic and regulatory processes in eukaryotic cells [251], as well as virulence factors in some bacteria and viruses, such as *Vibiro cholerae* (VC-neuraminidase) [252] and Influenza virus (IV-neuraminidase) [253], on the other hand.

Biological functions related to these sialidases comprise unmasking and generation of receptor sites, enhancing affinities to host cells, protein degradation and the involvement in pathogenesis [123,254]. However, first reports concerning sialidase activities in trypanosomes were made by Esievo in 1979 and Pereira in 1983 who detected free Sia production by *T. vivax* and *T. cruzi* [255], respectively. Surprisingly, Sia (Neu5Ac and Neu5Gc) have also been detected in glycoconjugates of *T. cruzi*, which were not synthesised by the parasite itself due to a lack of the appropriate Sia processing machinery [256]. Only a few years later in 1987, it was observed that *T. cruzi* sialidase does not only exhibit hydrolysing activity, but surprisingly, was also capable of transferring terminal Sia form host glycoconjugates onto trypanosomal surface molecules, and was therefore termed trans-sialidase (TS) [30,257,258]. In addition, this novel TS was found to be involved in the generation of surface receptor molecules for mammalian cell invasion by *T. cruzi*, thus already indicating its essential role in parasite survival and pathogenesis in the host [30,259]. With respect to these fundamental findings, in the following years, enhanced attention was also paid to African trypanosomes regarding a possible TS occurrence. Not surprisingly, developmental regulated TS from procyclic *T. brucei* [241,258] and *T. congolense* [241,260] were detected and subsequently partial characterised.

Trypanosomal TS were generally found to be located on the cell surface of the parasite attached to the plasma membrane via a GPI-anchor similar to those of procyclins described in the previous chapter (1.4.3). However, *T. cruzi* TS (TcruTS) not only acts on the cell surface but is also shed into the blood during host infection [261], whereas the same is assumed for *T. congolense* TS (TconTS). Interestingly, comparisons between TcruTS and the African trypanosomal TS, *T. congolense* TS (TconTS) and T. brucei TS (TbruTS) regarding structure, enzyme activity, substrate specificity and other properties, revealed significant similarities. For example both, *T. congolense* and *T. cruzi* express a number of different TS, encoded by several genes, exhibiting different sialyl-transferase and sialidase activities [260,262-267]. However, in addition only few active TS were found in *T. brucei* [268,269]. Interestingly a large TS gene family comprising over 1000 members were identified in *T. cruzi*, of which the majority encodes for inactive TcruTS [266,270]. A different situation exists in the African trypanosomes, where 4 TS genes were identified for *T. vivax*, 9 for *T. brucei* and 14 for *T. congolense* so far [267-269,271] including those described during the course of this thesis [272,273].

One unique feature of trypanosomal TS is the transfer of Sia to terminal galactose residues of oligosaccharides or glycoconjugates without the requirement of the activated

form, CMP-Sia, as donor substrate [274]. In the presence of suitable acceptor molecules, TS preferentially transfer Sia from one glycoconjugate to another, instead of hydrolysing the glycosidic bond [30]. For example, from a biological point of view, the enzymatic behaviour of TS, in the absence of suitable Sia acceptor molecules, is the same as that of common sialidases [275], determined by the release of free Sia from sialylated glycoconjugates. In addition, by the current state of knowledge regarding the catalytic mechanism of TS, the reaction would chemically rather be described as a transfer of Sia to a water molecule (detailed explanations are provided in following chapters). In this context, it should also be mentioned that trypanosomal TS share several structural homologies with common viral and bacterial sialidases [252,275], which will be discussed in the following chapters. Nevertheless, trypanosomes utilise TS to sialylate their own surface molecules, hence generating a negatively charged glycocalyx, which is known to play a fundamental role in several biological processes during the life cycle of trypanosomes in host and vector (chapter 1.4.2). In summary, TS modulate the presence, localisation and quantity of Sia on parasites' surface molecules and their environment. Also an essential role for TS in parasite development was suggested, since its expression is stage specifically regulated in the mammalian and insect forms. This hypothesis has been subsequently confirmed by several following studies [267,276-279], clearly representing TS as a distinct virulence factor in Chagas' disease and African Nagana not only in procyclic insect forms but also in BSF. For example Nok and Balogun demonstrated TconTS involvement in anaemia in animals, caused by the parasitical desialylation of erythrocytes [267,277].

1.5.1 TS substrate specificities

Trypanosomal TS catalyse the reversible transfer of preferentially α2,3-linked Sia to terminal βGal residues of glycoconjugates to specifically form a new α2,3 glycosidic linkage [30]. This enzymatic transfer activity is temperature, pH and acceptor substrate concentration dependent [280,281]. A variety of different donor and acceptor substrates have been tested with live trypanosomes and later with recombinant TS to further investigate substrate specificities [241,243,244,258,260]. Using *T. brucei* TS (TbruTS), Sia donor substrates, such as Neu5Ac-α(2,3)-lactose (3'SL), Neu5Ac-α(2,3)-lactitol, fetuin (sialo-glycoprotein), methylumbelliferone-Neu5Ac (MU-Neu5Ac), as well as human and bovine erythrocytes were found to be suitable donors for sialyl-transferase activity, whereas only negligible transfer rates have been observed for Neu5Ac-α(2,6)-lactose (6'SL) and Neu5Ac-α(2,8)-Neu5Ac, respectively [240,243]. In contrast, neither CMP-Neu5Ac, the well-known substrate for

eukaryotic and bacterial STs, nor the monosaccharide Neu5Ac were utilised as substrates [243]. Similar enzyme activities and substrate specificities have been observed for TconTS, for which Sia-transfer has additionally been detected using procyclic GARP or PARP as Sia acceptor substrates [241]. Interestingly, one exception regarding substrate specificity was observed by Tiralongo *et al.*, who demonstrated the efficient Sia transfer using 6'SL as donor substrate, catalysed by two different isolated TS-forms from *T. congolense* procyclic trypomastigotes, respectively [260]. Furthermore, substrate specificities for several Sia acceptor molecules and consequent TS activities have been determined for both, TconTS and TbruTS. Although significant transfer activities were detected using galactose (weak acceptor), lactose, lactitol, lactobionic acid, β-methyl-galactopyranoside, asialofetuin and asialo-PARP [241,282]. Asialo-GARP has also been demonstrated to be a suitable acceptor substrate for TconTS [241]. In contrast, mannose, glucose, galactose, methylumbelliferone and α-methyl-galactopyranoside turned out to be non-suitable Sia acceptor substrates for TconTS and TbruTS, respectively [241,243]. In addition, Sia donor and acceptor specificities have been demonstrated for TcruTS, including for example the transfer of α2,3-linked Neu5Gc or even its *O*-acetylated derivate, but with less efficiency, besides Neu5Ac [242,283-286]. Other suitable donor substrates comprise pNP(*p*-nitrophenole)-αNeu5Ac and Neu5Ac-α(2-3)Galβ(1-4)GlcNAc (3'-sialyl-*N*-acetyllactosamin) [281,287]. Furthermore, several Sia acceptor substrates, such as *N*- and *O*-glycans of asialofetuin, some asialo-gangliosides and mucins were trans-sialylated by TcruTS [242,288]. On the other hand, Man, Gal, GalNAc and terminal α-linked Gal were identified as non-suitable acceptor substrates for the transfer of Sia by TcruTS [284]. These findings clearly demonstrated the general regio specific substrate preferences of trypanosomal TS towards α2,3-linked Sia donors and terminal βGal-containing oligosaccharides, glycoproteins and glycolipids as Sia acceptor substrates, as well as the close relation between the two African trypanosomal species *T. brucei* and *T. congolense*.

Another interesting observation regarding TS-siglec interactions was made by Neubacher *et al.* [289]. They utilised recombinant *T. cruzi* TS to specifically synthesise a set of novel, unnatural sialyl-oligosaccharides, such as pNP-Neu5Ac and several other derivatives. These Neu5Ac analogues were used to study substrate specificity and binding affinity of MAG, since latter specifically recognises α2,3-linked Sia [163], the product of TS trans-sialylation. This also reflects the potential of trypanosomal TS to generate sialyl-glycoconjugates *in vivo*, which may represent a set of suitable ligands for siglecs, preferentially recognising α2,3-linked Sia ligands.

Reasons for the region and stereo selectivity of TS to specific donor and acceptor substrates are determined by the defined architecture of the catalytic centre, which will be discussed in the following chapters.

1.5.2 TS mediated interactions between parasite and siglecs from host cells

TS mediated trans-sialylation to specifically generate terminal α2,3-linked Sia containing oligosaccharides on parasites surface, will not only enhance the repulsive electrochemical properties but also provides a set of structures which can serve as potential ligands for different host cell lectins. The myelin associated glycoprotein (MAG), which is a member of a superfamily of sialic acid binding proteins called siglecs [163], represents such an example. Several family members are expressed on the plasma membrane of different cell types of the immune and nervous system, which has been already discussed in previous chapter (1.3.1). Sialoadhesin (Siglec-1, Sn), a cell adhesion molecule expressed on the cell surface of macrophages, has been demonstrated to be involved in Sia mediated association of *T. cruzi* procyclic trypomastigotes [290,291]. The parasites' association with macrophages observed was drastically reduced, when anti-Sn antibodies or desialylated macrophages were used [291], indicating a Sia-siglec mediated interaction, probably initiated by TcruTS.

Another interesting finding comprises the interaction of *T. cruzi* with the inhibitory murine Siglec-E of dendritic cells. Erdmann *et al.* demonstrated that *T. cruzi* trypomastigotes suppressed the production of inflammatory cytokines and subsequent T-cell activation due to Sia mediated binding of the parasite to murine Siglec-E on host cell surface [292]. It was proposed that this effect might be induced by TS-dependent transfer of Sia from dendritic sialo-glycoconjugates, which serve as ligands for *cis*-interactions with Siglec-E to trypomastigote mucins, hence generating sialylated ligands on the parasite surface for *trans*-interactions with Siglec-E. Such *cis*-interaction mimicry maintain inhibitory effects of Siglec-E and therefore impair the immune response [291-295]. Thus, Siglec specific interactions with sialo-glycoconjugates of the parasite may represent the basic mechanism, which enables trypanosomes to sabotage and modulate host immune system. This weakening effect of TS was even more obvious, when transfecting TS into *Leishmania major* (transgenic), resulting in an significant increase in virulence, thus declaring TS as a kind of inflammatory cytokine mimicry, termed parasitokine [293].

In summary, these findings illustrate the importance of TS in Sia mediated association to siglec positive immune-competent host cells, as a potential target for trypanosomes in the early phase of infection, as well as the importance in establishment of chronic infection.

1.5.3 The structure of trypanosomal TS

In the early 1990s genes encoding for several TcruTS enzymes haven been identified [30,296-299]. The appropriate amino acid sequence for enzymatic active TS proteins were predicted and indicate significant structural similarities with known viral and bacterial sialidases [30,296,300,301].

Affranchino *et al*. reported about an antigen from *T. cruzi* trypomastigotes, which was shed into cell culture supernatant from cultured procyclic trypomastigotes and blood of infected patients and animals recently after infection, therefore being a major secreted antigen in the acute phase of the disease and termed Shed-Acute-Phase-Antigen (SAPA) [302]. SAPA was declared as a highly antigenic protein comprising a number of C-terminal repeats, as well as some N-terminal motifs, which were subsequently found to be conserved in bacterial and viral sialidases [297]. Only two years later, in 1991, Pereira *et al*. surprisingly investigated a high gene sequence homology of SAPA with *T. cruzi* sialidases [296]. Finally, in 1992 Parodi *et al*. identified SAPA not only as the major sialidase but also the major TS of *T. cruzi*, and suggested the hydrolytic as well as the sialyl-transferase activity to be located in the N-terminal domain [299] of the two domain comprising enzyme [303]. This suggestion was subsequently confirmed, although it has been demonstrated that SAPA repeats in the C-terminus were not required for enzymatic activity of TcruTS [304,305], but rather increase TS half-life in blood due to its highly antigenic feature [261,306]. Furthermore, it was observed that the N-terminal catalytic domain (CD) of TcruTS contains all the amino acid residues, which were known to interact with terminal Sia residues of glycoconjugates in the catalytic centre of sialidases from *Vibrio cholera* and *Salmonella* [307,308]. Primary structure alignments, revealed further insights into the potential tertiary structure of TcruTS, suggesting that CD is followed by a large domain of unknown function, exhibiting fibronectin type III (Fn3) similarities, the segment containing SAPA repeats and finally a GPI-anchor [296].

To further characterise the region in TcruTS, which is required for sialyl-transferase activity, Smith and Eichinger expressed several chimeric recombinant proteins, containing amino acid segments of TcruTS and *T. rangeli* sialidase (TranSA), respectively [309]. In addition, they also performed point mutation experiments to identify amino acids essential for Sia transferase activity. Impressively, without the employment of any structural resolving methods, such as nuclear magnetic resonance (NMR)-spectroscopy or X-ray crystallography, they were able to relative precisely predict a couple of essential tertiary structure elements. Besides several others, one interesting finding was that the C-terminal Fn3-like domain does

not induce sialyl-transferase activity to TranSA, since latter also exhibit this domain and its exchange from TcruTS to TranSA did not add Sia transfer activity to the TranSA [309].

In addition to that, however, the Fn3-like domain was demonstrated to be required for the expression of enzymatically active recombinant TcruTS and TranSA, respectively [305,310]. This suggests a crucial role of the Fn3-like domain in enzyme folding or indicates other, so far, unknown functions.

The breakthrough regarding tertiary structure and catalytic mechanism of TcruTS, succeeded with the resolved crystal structure of TcruTS [311], as well as TranSA [312] reported by Buschiazzo *et al*. in 2002 and 2000, respectively. Both, TcruTS and TranSA exhibit a similar overall structure. Although, the predicted six-bladed β-propeller topology at the N-terminus, as seen for microbial sialidases [300], was confirmed. The architecture of the catalytic centre exhibits a similar arrangement and orientation of amino acids, which were suggested to be involved in catalysis. Amino acids of the TcruTS hydrophobic Sia binding pocket comprise an essential arginine triad (R35, R245 and R314), which stabilise the carboxyl group of Sia, a glutamic acid (E357) residue, which in turn stabilise the R35, a tyrosine (Y342) and another glutamic acid (E230) residue, which are involved in the formation of the covalent transition state, and finally an aspartic acid (D59), which catalysis the cleavage and subsequent formation of the Sia-Gal glycosidic linkage [311,313] (Figure 5).

Most of these and other residues of the catalytic centre are well conserved among the trypanosomal TS family [260,267,269,313]. One major finding includes the natural occurring mutation of Y342 to histidine (Y342H), observed for inactive TS [270]. This consequently indicates the essential role of Y342 in enzyme activity and confirms it as a main factor to differentiate between active and inactive TS.

Furthermore, the C-terminal Fn3-like domain, following the CD, exhibit a high structural similarity to the carbohydrate recognition domain (CRD) of common plant lectins, typically characterised by a conserved β-barrel topology [314-317]. Hence suggesting that the C-terminal domain may provide a potential carbohydrate-binding-site and consequently termed “lectin-like” domain (LD) due to its structural homology to plant lectins [312].

TcruTS CD is approximately 371 and LD 237 amino acids long, whereas both domains are connected via a 22 amino acid long α-helix (generating a 70 kDa globular core). These defined arrangements result in a relatively close contact between the two domains (Figure 5). Interestingly, the surface area of this extended interface of TcruTS CD and LD was found to be significantly larger compared to that of other microbial sialidases, such as *Vibrio cholera*

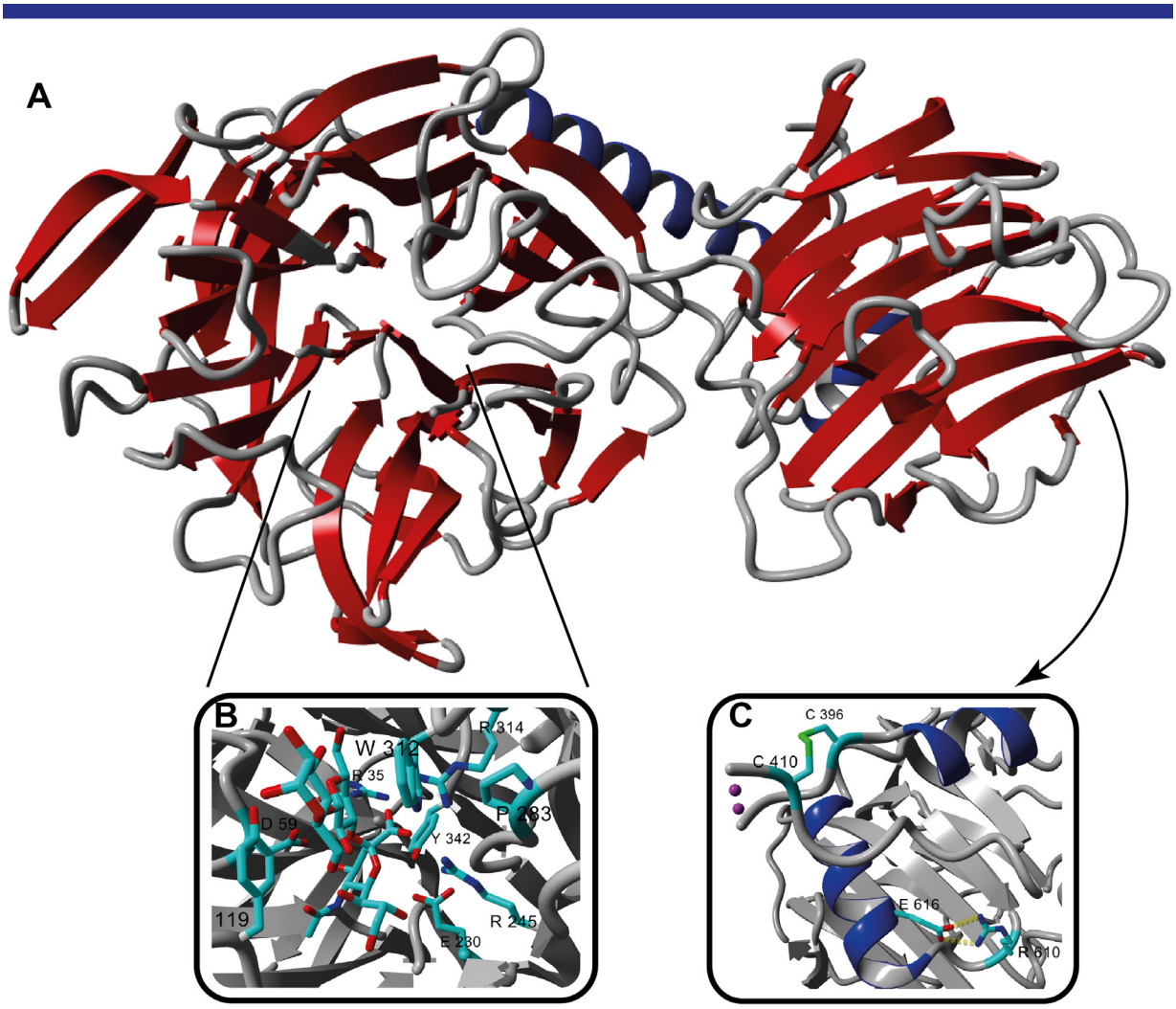


Figure 5. Crystal structure of TcruTS (PDB code: 1S0I) [311]. **A**) Overall structure, comprising the six-bladed β-propeller topology of the catalytic domain (CD) on the left, the fibronectin type III (Fn3) topology of the lectin-like domain (LD) on the right and the α-helix connecting both domains. **B**) Catalytic center of CD in complex with α-2,3-sialyllactose. Critical amino acids are shown and labelled with their respective numbers. **C**) Rear view of LD relative to A. The well-conserved disulfide-bridge (C396-C410), as well as the functional relevant salt-bridge (R610-E616) are illustrated. Purple spheres indicate the non-resolved loop of TcruTS crystal structure.

or *Macrobdella decora* neuraminidase [301,318], thus illustrating a structural topology distinct from those of classical sialidases, but comprising similar structural elements.

So far no crystal structures of any African trypanosomal TS or sialidases are available. However, several *in silico* data, including amino acid alignments and homology modelling of TbruTS and TconTS revealed high overall structure homology to TcruTS and TranSA [260,263,267,268,319] (Further data are part of this thesis). Another structural element shared by all trypanosomal TS comprise a single disulfide bridge (formed between C396 and C410 in TcruTS), which is located right after the connecting α-helix and constricts a small hairpin loop, which includes a short non-conserved amino acid sequence (Figure 5). This loop has remained unresolved in the crystal structures of both, TcruTS and TranSA [311,312],

indicating its relatively high steric flexibility. Besides the α-helix connecting CD and LD, a second α-helix was observed from both crystal structures, located at the end of the C-terminal LD, sterically close to the first α-helix and the following unresolved loop (Figure 5).

Interestingly, point mutation experiments on TcruTS, carried out by Smith *et al.*, suggested the relevance of a specific salt-bridge, formed by arginine (R642) and glutamine (E648) side chains [309] (Figure 5). They demonstrated that its disruption by site-directed mutagenesis of any of the two amino acids involved resulted in increased TcruTS Sia transfer activity [309]. Surprisingly, this crucial structural element, which is well conserved in the trypanosomal TS family, is located far away from the catalytic centre and therefore, not likely to be directly involved in Sia transfer [320]. So far, there are still no explanations available for this functional effect. However, these findings clearly suggest a potential role for TS LD in enzymatic activity, although its mechanism still remains unclear. Therefore, biochemical characterisation of TconTS-LD, including its carbohydrate recognition potential and ligand specificity, has represented an integral part of this thesis.

1.5.4 Catalytic mechanism

Several approaches have clearly demonstrated the fundamental role of trypanosomal TS in the survival and development of the parasite in host and vector, as well as in pathogenesis (detailed discussed in the previous chapters 1.4 – 1.5). With respect to this, it seems more than obvious to elucidate the highly stereo and region specific mechanism of TS Sia transfer activities, in order to achieve enhanced insight into this unique class of enzymes.

The general catalytic mechanism, by which common glycosidases hydrolyse the glycosidic bond of oligosaccharides, has been extensively studied and revealed the presence of both, inverting (chiral inversion of the anomeric carbon) as well as retaining (retain configuration at the anomeric carbon) glycosidases [321]. In contrast to that, it was found that trypanosomal TS exclusively act as retaining enzymes [241,243,260,283], similar as classical *exo*-sialidases, suggesting an analogues overall mechanism.

However, compared to glycosidases, which utilize a pair of carboxylic acids in the catalytic centre, acting as proton donor (acid-catalysed hydrolysis) and nucleophile to form an oxoniumion-like transition state, sialidases do not contain such negatively charged groups as their central catalyst, most likely due to appearance of potential repulsive interactions with the carboxyl group of Sia. Hence a different catalytic mechanism exists for sialidases, proposed by several studies [251,322,323], involving a distinct set of amino acid residues in the active centre, catalysing the reaction.

The crystal structures of TcruTS and TranSA, representing the most intensive studied TSs, resolved by Buschiazzo *et al.* revealed several highly conserved amino acid residues in the catalytic centre of TS and SA (as mentioned in the previous chapter 1.5.3), which were also found in other microbial and human sialidases and known to be involved in catalytic activity. This new structural data together with subsequent mutagenesis results on selected amino acid residues of the active centre [312], has provided evidence for a potential catalytic mechanism analogue to that for common *exo*-sialidases [275,322] as follows: 1: Effective binding of the Sia donor substrate to the active site (Michaelis complex) is mainly stabilised by three arginine residues (arginine triad), forming several hydrogen bonds to the carboxyl group at C-1, as well as to the C-8 hydroxyl group of the glycerol side chain of Sia, which is accompanied by conformational change from a chair ($^2C_5$) into a boat ($^{2,5}B$) configuration [313]; 2: An aspartate residue, located close to the glycosidic linkage between Sia and the underling oligosaccharide, directly acts as an acid/base catalyst and leads to the cleavage of the glycosidic bond, subsequently releasing the reducing end glycan structure; 3: The resulting Sia oxocarbonium-intermediate is then stabilised by a nucleophilic tyrosine residue, which is in turn formed by deprotonation via a neighbouring glutamate residue; 4: The same aspartate residue, which catalysed the cleavage of the initial glycosidic bond, deprotonates the C-3 hydroxyl group of the terminal βGal moiety of the acceptor substrate (or a water molecule in terms of hydrolysis), which is followed by a nucleophilic attacking at the C-2 of Sia, finally forming a new α2,3-glycosidic linkage accompanied by a conformational change of Sia back into $^2C_5$ configuration; 5: The formed sialo-glycoconjugate is then released from the active centre [312].

This proposed conventional double-displacement or “ping-pong” mechanism is typical for common retaining sialidases and generally characterised by a long-lived reaction intermediate [259,280,312]. In contrast, from a chemical point of view, such oxocarbonium ion-like transition states, as proposed for TcruTS, are energetically unfavourable and in order to that relatively unstable and very short-lived, hence suggesting a different mechanism for trans-sialylation. Indeed, Watts *et al*. demonstrated that trans-sialylation catalysed by TcruTS proceeds via a covalent sialyl-enzyme-intermediate [324]. Using deoxy-2,3-difluorosialic acid as substrate, they were able to trap this covalent intermediate, as it has been reported for other glycosidases [325]. Furthermore, they could identify tyrosine 342 (Y342) as the respective nucleophile of TcruTS forming a covalent bond between the oxygen of Y342 side chain and the anomeric C-2 of Sia, before its transfer to an appropriate acceptor substrate occurs.

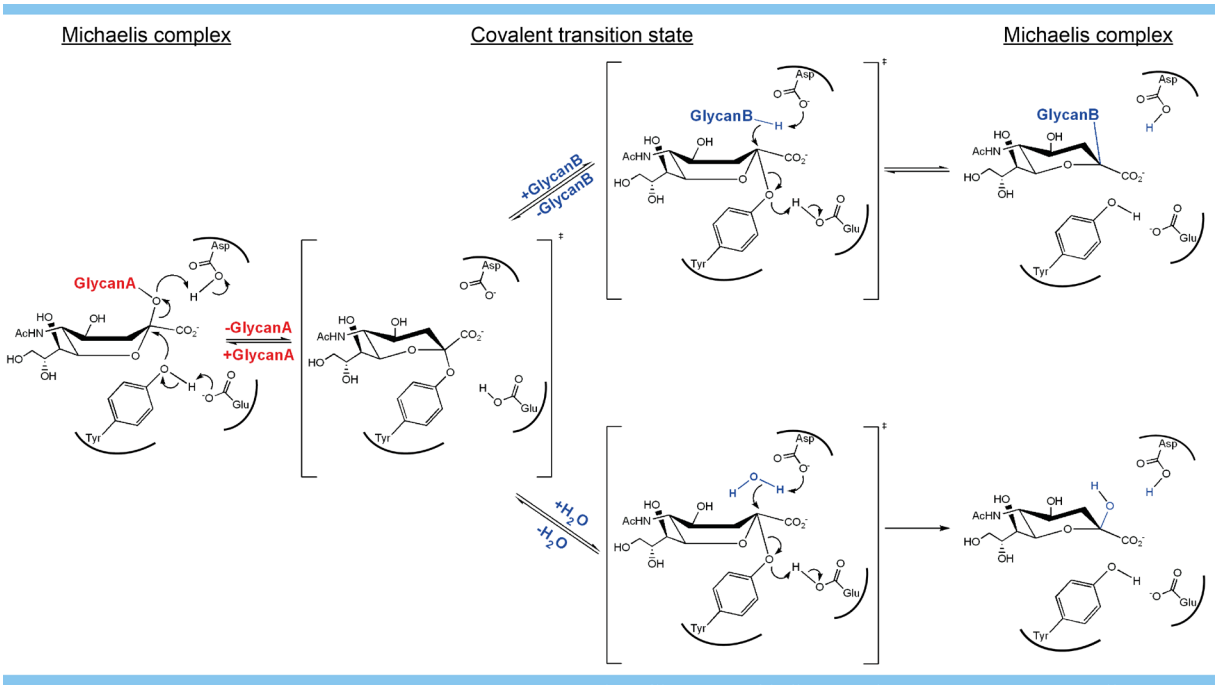


Figure 6. Proposed catalytic mechanism for trypanosomal TS. The Mechaelis complex is formed upon Sia donor substrate binding to the catalytic centre of TcruTS. The formation of the covalent sialyl-enzyme intermediate (shown in brackets) and the release of the aglycon moiety are catalysed by aspartate (acid/base catalyst), glutamate and tyrosine (nucleophile) residues of the catalytic centre. After binding of the galactosyl-substrate to the acceptor binding-site, Sia is transferred (or hydrolysis) to the terminal galactose residue, followed by the release of the new sialoside (or Neu5Ac) from the binding pocket of TcruTS. The figure was generated based on the catalytic mechanism proposed by Amaya *et al.* [313].

Interestingly, such a covalent aryl-sialyl-intermediate is unique for trypanosomal TS in respect to retaining sialidases reported so far. In respect to these findings, Amaya *et al.* proposed an adapted catalytic mechanism for TcruTS (Figure 6), also considering the glutamate residue 230 (E230) as the corresponding acid/base catalyst responsible for deprotonation of Y342 [313]. In principle, the mechanism follows the same scheme as described above, although difference comprise the formation of the covalent intermediate before a suitable acceptor substrate is bound, which is accompanied by an electrophilic migration of the C-2 towards the nucleophile via sugar ring relaxation of Sia to its $^{2}C_{5}$ conformation [313] (Figure 6). Acceptor substrate binding to the active centre of TcruTS (Michaelis complex) and the nucleophilic attack of the βGal C-3 hydroxyl group at the anomeric carbon of Sia occur in a reverse process as mentioned above.

However, since the catalytic centre of trypanosomal TS is structurally very closely related to that of viral sialidases and also TranSA, the question still remains why trypanosomal TS act so efficiently in transferring compared to just hydrolysing the glycosidic bond with subsequent release of free Sia. Investigations regarding the comparison of kinetic reaction parameters and substrate binding properties of trypanosomal and viral enzymes revealed similar turnover rates but significant different ligand interactions [276,307],

suggesting a potential rearrangement of TS active centre and/or additional substrate interactions. Although the presence of distinct acceptor and donor substrate binding-sites in TS would provide a potential ternary complex mechanism contributing to a preferred Sia transfer activity.

One primary observation was made in correlation with the crystal structure of TcruTS, identifying several hydrophobic residues at the active binding groove, which significantly contribute to an overall hydrophobic character and therefore suggested to contribute to exclude water molecules from the catalytic centre [311]. Furthermore, two of these hydrophobic amino acids, a tyrosine (Y119) and a tryptophan (W312) of TcruTS, which are located at opposite positions close to the active site cleft (Figure 5), were suggested to play a crucial role in enzyme activity by generating stacking interactions (π-π^*) with cyclohexane rings of the donor and acceptor substrate, respectively [311-313]. Interestingly, Y119 is substituted by a serine residue in TranSA (S141), whereas the W312 is present in both enzymes (Figure 7). Key mutagenesis experiments on these two amino acids in TcruTS, carried out by Paris *et al.*, clearly demonstrated their direct involvement in substrate binding [326]. When mutating the W312 to alanine, a complete loss of Sia transfer activity for TcruTS was observed, independent of the donor substrate used, although the sialidase activity was also significantly decreased [326]. These finding demonstrated the involvement of W312 in donor substrate binding. Interestingly, this conserved tryptophan residue was found in a completely different orientation in TranSA (W334, Figure 7), due to a adjacent glutamine residue Q305 (P283 in TcruTS), which sterically hinders W334 to occupy the same configuration as in TcruTS [326]. Substitution of P283 to glutamine in TcruTS revealed mutants with abolished Sia transfer as well as sialidase activities, clearly demonstrating the requirement of an accurate orientation of W312 for efficient substrate binding and transfer activities.

In addition, site directed mutagenesis of Y119 to serine in TcruTS, drastically reduced Sia transfer but almost retained sialidase activities [326], indicating the crucial role of Y119 in acceptor substrate binding, potentially being a part of the acceptor binding-site [312], and obviously representing an essential, characteristic feature of trypanosomal TS. These suggestions were further supported by a mutant TranSA, in which Sia transfer activities has been induced by five different point mutations, also including those residues discussed above [327]. Furthermore, it has been proposed that the lactose moiety of the ligand is bound between W312 and Y119 [312,326]. Numerous subsequent molecular dynamics simulations

on TcruTS in complex with several ligands revealed another striking feature of the catalytic mechanism, which has not been observed from crystallographic data.

Demir *et al*. demonstrated a high plasticity of the catalytic grove of TcruTS as a potential consequence for the different catalytic behaviour of TcruTS and TranSA, respectively [328]. They showed that the high flexibility of the W312 comprising loop in the apo (without ligand bound)-enzyme (open form), and its motion upon donor substrate binding (holo-enzyme), leads to a more closed conformation of the catalytic centre with significantly decreased solvent exposure, probably enhancing sialyl-enzyme intermediate maintenance and facilitating subsequent acceptor substrate binding [328]. In addition, also the flexibility of Y119 provides its motion to a more outward position relative to the catalytic centre, which allows stacking interactions with the acceptor ligand, and an inward conformation proposed to be essential for blocking the Sia binding-site [313,328,329]. In addition, the orientation of

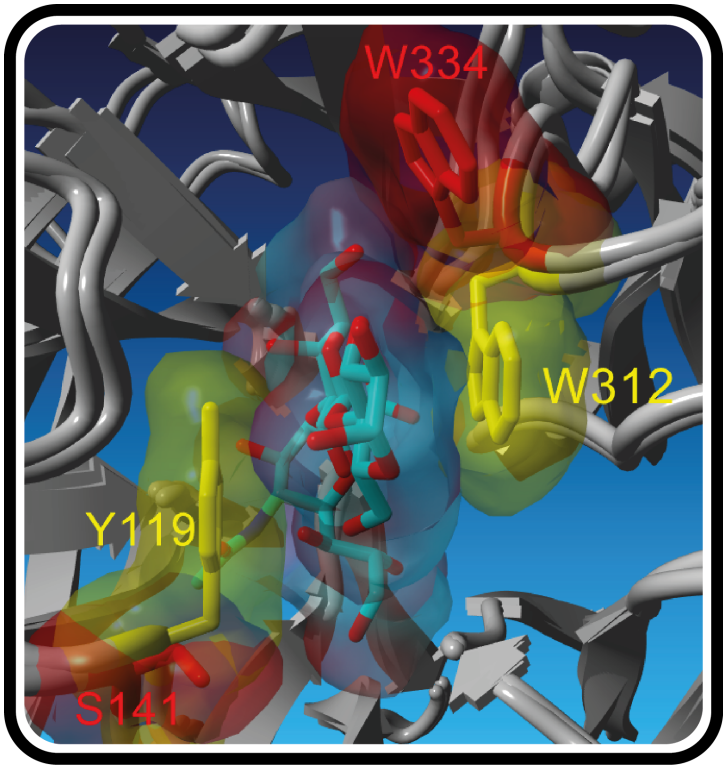


Figure 7. Structural comparison of *T. cruzi* TS (TcruTS) and *T. rangeli* SA (TranSA). The superimposed catalytic centres of TcruTS (PDB code: 1S0I) and TranSA (PDB code: 1N1S) in complex with 3'-sialyllactose are illustrated. Conserved tryptophan residues W312 in TcruTS (yellow) and W334 in TranSA (red) of the flexible loop, which defines the open and closed form of the catalytic pocket, exhibit distinct orientations, respectively. Together with the tyrosine 119 (yellow), W312 stabilise the ligand via stacking interactions with the cyclohexane ring of the lactose moiety. Due to the substitution of Y to S141 in TranSA together with the altered orientation of W334, in contrast to that of TcruTS, a lack of interactions to the lactose moiety can be observed. Van der Waals surfaces of the illustrated residues and 3'-sialyllactose are shown in transparent spheres, respectively.

Y119 was found to be correlated with the W312 loop motion [328]. In contrast to TcruTS, TranSA generally exhibits a more open conformation of the catalytic grove before and upon substrate binding [328], representing a possible explanation for the different catalytic behaviour of both enzymes. Recent *in silico* free-energy calculations of the sialyl-enzyme intermediate of TcruTS and TranSA, revealed a significant higher stability for that of TranSA

compared to TcruTS, therefore indicating a subsequent Sia-transfer to a suitable donor substrate as rather unfavourable [330].

In summary, these highly dynamic structural regulatory processes, controlling the substrate entry and binding, represent a much more complex situation for trypanosomal TS activity as it was previously assumed and might be also responsible for the preferred Sia transfer activities compared to sialidases. A recent study has demonstrated the experimental existence of an acceptor substrate binding-site in TcruTS, which is generated upon donor substrate binding, accompanied by subsequent conformational changes and clearly distinct from that reported for the donor substrate [331]. From their novel experimental and *in silico* data, they concluded that the Sia transfer mechanism catalysed by trypanosomal TS cannot simply be described as a classical sequential or conventional double-displacement (ping-pong) mechanism, but rather as a hybrid of both systems characterised by the formation of a ternary complex after the sialyl-enzyme intermediate is formed [331]. In conclusion, this new insights into the catalytic mechanism of trypanosomal TS provide a better understanding of this unique Sia transfer and opens new strategies for synthetic anti-TS drug development.

1.5.5 TS inhibitors, a strategy to fight trypanosomiasis

As discussed in previous chapters, trypanosomes provide several strategies to evade host immune response and vectors digestive system to ensure their survival and development during their life cycles. However, potential strategies against the parasite in controlling the vector have been used effectively but also implicate several striking challenges [187,191,332]. Furthermore, the fact that eradication of the entire trypanosomal reservoir is obviously impossible, another more logical strategy in controlling the parasite includes anti-trypanosome vaccination. Although, a major problem constitutes the surface coat of the parasite during host infection, which is made of densely packed VSGs and permanently changed due to parasites antigenic variation strategy. Thus representing vaccination against these targets as totally ineffective (see previous chapter 1.4.3). Besides some established drugs, such as pentamidine, nifurtimox and melarsoprol currently used against trypanosomiasis with only moderate success (see previous chapter 1.4.4), the interest in identifying new potential targets and developing novel strategies towards more efficient anti-trypanosomal drug-design has significantly increased over the past decade [333]. One promising drug target against trypanosomiasis comprise the family of TS, since they have been demonstrated to play a key role for parasite survival and pathogenesis (see previous chapters 1.5), they are expressed in the different life stages of the parasite during its

development in both, vector and host and do not occur naturally in either of them. Several different potential inhibitors have been tested, but unfortunately with only minor inhibitory efficiency as expected [334,335]. Generally they can be differentiated in two major classes of synthetic structures, the Sia mimicking derivatives, also comprising the Sia donor substrates, and the acceptor substrate analogue compounds [334]. One example constitutes 2,3-didehydro-2-deoxy-*N*-acetylneuraminic acid (DANA), the well-studied and potent inhibitor of the influenza virus neuraminidase [253,336]. However, this effective sialidase inhibitor and even other derivatives with further modifications at several positions in the cyclohexane ring as well as additional functional groups, revealed only extremely weak inhibition potential for TcruTS in the higher mM range [288,327,337,338]. In addition also deoxy-2,3-difluorosialic acid, which has been used to trap and characterise the sialyl-TS intermediate (discussed in the previous chapter 1.5.5), only exhibits reasonable inhibitory effects at relatively high mM concentrations [339].

On the other hand, acceptor substrate analogous, such as lactitol, lactobionic acid or galactosyl-arabinose have been also identified as only very weak TcruTS inhibitors [282,340]. Nevertheless, a modified derivate of the ganglioside GM_3 (A membrane-bound glycosphingolipid with the general composition Neu5Ac-Gal-Glc-ceramide) has been identified as one of the most potent TcruTS inhibitors known so far (IC_{50} value in the lower µM range) [283], most likely due to its ability to bind both, donor and acceptor substrate binding-sites in TcruTS. Using this potent TcruTS inhibitor as a model structure and new strategy towards more efficient derivatives, novel potentially inhibitory compounds have been synthesised, also in respect to maintain their multivalent interaction to both binding-sites in TS [338]. Surprisingly, these new class of inhibitors turned out to be only efficient at low mM concentrations just as some other compounds mentioned above. Nevertheless, following the strategy to develop trypanosomal TS inhibitors targeting both, acceptor and donor binding-sites but with additional substitutions to improve binding affinities, several other compounds could be generated during the past years.

Utilizing the “click chemistry” [341] approach, Carvalho *et al*. recently synthesised a library of triazole-substituted galactosides, a group that has been explored as potent inhibitors for different glycosyl-transferases [342], glycosidases [343,344] and influenza neuraminidase [345,346], as novel potent TcruTS inhibitors, bearing additional substitutions at C-1 and C-6, such as aryls, azoles, aliphatic alcohols and others [347]. Certainly, also these compounds were proven to be moderate to weak TcruTS inhibitors (low mM range) *in vitro*, but surprisingly showed significant trypanocidal activity against cultured *T. cruzi* trypomastigotes

[347]. On that basis a set of further differential modified 1,2,3-triazole sialyl-mimetic neoglycoconjugates have been recently synthesised by Campo *et al.* and verified to inhibit TcruTS even more effectively (around 1 mM) compared to previously reported synthetic compounds [348]. However, even if these compounds do not exhibit a desirable inhibitory activity in the low μM to nM range, a clear improvement towards more effective multivalent substances has been observed. Currently, a series of different relatively potent TcruTS inhibitors exist, comprising modified gangliosides [283], sulfonamide chalcones [349], modified octyl-galactosides [350], flavonoids (Myricetin) and anthraquinones [351], but no 100 % specific and sufficient effective trypanosomal TS inhibitor has been established so far, possibly due to the fact that trypanosomal TS have still not been fully characterised and understood well enough in detail. Therefore, further characterisation of these unique enzymes represents a fundamental necessity for the synthesis of more effective TS inhibitors to efficiently fight trypanosomiasis in Latin America, Africa and all the other threatened countries around the planet.

1.6 References

1. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al. Discovery and Classification of Glycan-Binding Proteins. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.
2. Wade LG. Organic Chemistry. Prentice Hall; 2010.
3. Varki A. Essentials of Glycobiology. CSHL Press; 1999.
4. Sharon N, Lis H. Lectins as cell recognition molecules. Ullstein Mosby. 1989 Oct 13;246(4927):227–34.
5. Sharon N, Lis H. Carbohydrates in cell recognition. Sci Am. 1993 Jan;268(1):82–9.
6. Gabius HJ. Biological information transfer beyond the genetic code: the sugar code. Naturwissenschaften. 2000 Mar;87(3):108–21.
7. Mody R, Joshi S, Chaney W. Use of lectins as diagnostic and therapeutic tools for cancer. J Pharmacol Toxicol Methods. 1995 Feb;33(1):1–10.
8. Gorelik E, Galili U, Raz A. On the role of cell surface carbohydrates and their binding proteins (lectins) in tumor metastasis. Cancer Metastasis Rev. 2001;20(3-4):245–77.
9. Feizi T, Mulloy B. Carbohydrates and glycoconjugates. Glycomics: the new era of carbohydrate biology. Curr Opin Struct Biol. 2003 Oct;13(5):602–4.
10. van den Berg BM, Vink H, Spaan JAE. The endothelial glycocalyx protects against myocardial edema. Circ Res. Lippincott Williams & Wilkins; 2003 Apr 4;92(6):592–4.
11. Danielli JF. Capillary permeability and oedema in the perfused frog. J Physiol (Lond). 1940 Mar 14;98(1):109–29.
12. Chambers R, Zweifach BW. Intercellular cement and capillary permeability. Physiol Rev. 1947 Jul;27(3):436–63.
13. Copley AL. Hemorheological aspects of the endothelium-plasma interface. Microvasc Res. 1974 Sep;8(2):192–212.
14. Luft JH. Fine structures of capillary and endocapillary layer as revealed by ruthenium red. Fed Proc. 1966 Nov;25(6):1773–83.
15. Pries AR, Secomb TW, Gaehtgens P. The endothelial surface layer. Pflugers Arch - Eur J Physiol. 2000 Sep;440(5):653–66.
16. Ballinger ML, Nigro J, Frontanilla KV, Dart AM, Little PJ. Regulation of glycosaminoglycan structure and atherogenesis. Cell Mol Life Sci. Birkhäuser-Verlag; 2004 Jun;61(11):1296–306.
17. Nieuwdorp M, Meuwese MC, Vink H, Hoekstra JBL, Kastelein JJP, Stroes ESG. The endothelial glycocalyx: a potential barrier between health and vascular disease. Curr Opin Lipidol. 2005 Oct;16(5):507–11.
18. Taniguchi N, Miyoshi E, Gu J, Jianguo G, Honke K, Matsumoto A. Decoding sugar functions by identifying target glycoproteins. Curr Opin Struct Biol. 2006 Oct;16(5):561–6.
19. Sugahara K, Mikami T, Uyama T, Mizuguchi S, Nomura K, Kitagawa H. Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. Curr Opin Struct Biol. 2003 Oct;13(5):612–20.
20. Esko JD, Selleck SB. Order out of chaos: assembly of ligand binding sites in heparan sulfate. Annu Rev Biochem. 2002;71:435–71.
21. Funderburgh JL. Keratan sulfate: structure, biosynthesis, and function. Glycobiology. 2000 Oct;10(10):951–8.

22. Laurent TC, Fraser JR. Hyaluronan. FASEB J. 1992 Apr;6(7):2397–404.

23. Jackson RL, Busch SJ, Cardin AD. Glycosaminoglycans: molecular properties, protein interactions, and role in physiological processes. Physiol Rev. 1991 Apr;71(2):481–539.

24. Camejo G, Hurt-Camejo E, Wiklund O, Bondjers G. Association of apo B lipoproteins with arterial proteoglycans: pathological significance and molecular basis. Atherosclerosis. 1998 Aug;139(2):205–22.

25. Sharon N. Carbohydrates. Sci Am. 1980 Nov;243(5):90–116.

26. Opdenakker G, Rudd PM, Ponting CP, Dwek RA. Concepts and principles of glycobiology. FASEB J. 1993 Nov;7(14):1330–7.

27. Zheng M, Fang H, Tsuruoka T, Tsuji T, Sasaki T, Hakomori S. Regulatory role of G_{M3} ganglioside in $\alpha 5\beta 1$integrin receptor for fibronectin-mediated adhesion of FUA169 cells. J Biol Chem. 1993 Jan 25;268(3):2217–22.

28. Akiyama SK, Yamada SS, Yamada KM. Analysis of the role of glycosylation of the human fibronectin receptor. J Biol Chem. 1989 Oct 25;264(30):18011–8.

29. Chammas R, Veiga SS, Line S, Potocnjak P, Brentani RR. Asn-linked oligosaccharide-dependent interaction between laminin and gp120/140. An $\alpha 6/\beta 1$ integrin. J Biol Chem. 1991 Feb 15;266(5):3349–55.

30. Schenkman S, Jiang MS, Hart GW, Nussenzweig V. A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. Cell. 1991 Jun 28;65(7):1117–25.

31. Disney MD, Seeberger PH. The use of carbohydrate microarrays to study carbohydrate-cell interactions and to detect pathogens. Chem Biol. Elsevier; 2004 Dec;11(12):1701–7.

32. Nimrichter L, Gargir A, Gortler M, Altstock RT, Shtevi A, Weisshaus O, et al. Intact cell adhesion to glycan microarrays. Glycobiology. Oxford University Press; 2004 Feb;14(2):197–203.

33. Haltiwanger RS, Lowe JB. Role of glycosylation in development. Annu Rev Biochem. 2004;73:491–537.

34. Zachara NE, Hart GW. Cell signaling, the essential role of O-GlcNAc! Biochim Biophys Acta. 2006 May;1761(5-6):599–617.

35. Lanctot PM, Gage FH, Varki AP. The glycans of stem cells. Curr Opin Chem Biol. 2007 Aug;11(4):373–80.

36. Sharon N, Lis H. Lectins. Dordrecht: Springer Science & Business Media; 2007. 1 p.

37. Blomme B, Van Steenkiste C, Callewaert N, Van Vlierberghe H. Alteration of protein glycosylation in liver diseases. J Hepatol. Elsevier; 2009 Mar;50(3):592–603.

38. Patsos G, Hebbe-Viton V, Robbe-Masselot C, Masselot D, San Martin R, Greenwood R, et al. *O*-glycan inhibitors generate aryl-glycans, induce apoptosis and lead to growth inhibition in colorectal cancer cell lines. Glycobiology. Oxford University Press; 2009 Apr;19(4):382–98.

39. Freire-de-Lima L, Oliveira IA, Neves JL, Penha LL, Alisson-Silva F, Dias WB, et al. Sialic acid: a sweet swing between mammalian host and *Trypanosoma cruzi*. Front Immunol. Frontiers; 2012;3:356.

40. Atkins P, de Paula J. Physical chemistry for the life sciences. Oxford University Press, USA; 2011. 1 p.

41. Gerardy-Schahn R, Oelmann S, Bakker H. Nucleotide sugar transporters: biological and functional aspects. Biochimie. 2001 Aug;83(8):775–82.

42. Taylor ME, Drickamer K. Introduction to glycobiology. Oxford University Press; 2011. 1 p.

43. Röttger S, White J, Wandall HH, Olivo JC, Stark A, Bennett EP, et al. Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus. J Cell Sci. 1998 Jan;111 (Pt 1):45–60.

44. Kornfeld R, Kornfeld S. Comparative aspects of glycoprotein structure. Annu Rev Biochem. 1976;45:217–37.

45. Takatsuki A, Tamura G. Effect of tunicamycin on the synthesis of macromolecules in cultures of chick embryo fibroblasts infected with Newcastle disease virus. J Antibiot. 1971 Nov;24(11):785–94.

46. Elbein AD. Inhibitors of the biosynthesis and processing of *N*-linked oligosaccharides. CRC Crit Rev Biochem. 1984;16(1):21–49.

47. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al. C-type Lectins. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

48. Kumar R, Stanley P. Transfection of a human gene that corrects the Lec1 glycosylation defect: evidence for transfer of the structural gene for *N*-acetylglucosaminyltransferase I. Mol Cell Biol. 1989 Dec;9(12):5713–7.

49. Kumar R, Yang J, Larsen RD, Stanley P. Cloning and expression of *N*-acetylglucosaminyl-transferase I, the medial Golgi transferase that initiates complex *N*-linked carbohydrate formation. Proc Natl Acad Sci USA. 1990 Dec;87(24):9948–52.

50. Sarkar M, Hull E, Nishikawa Y, Simpson RJ, Moritz RL, Dunn R, et al. Molecular cloning and expression of cDNA encoding the enzyme that controls conversion of high-mannose to hybrid and complex *N*-glycans: UDP-*N*-acetylglucosamine: α-3-D-mannoside β-1,2-*N*-acetylglucosaminyltransferase I. Proc Natl Acad Sci USA. 1991 Jan 1;88(1):234–8.

51. Ioffe E, Stanley P. Mice lacking *N*-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid *N*-linked carbohydrates. Proc Natl Acad Sci USA. 1994 Jan 18;91(2):728–32.

52. Ioffe E, Liu Y, Stanley P. Essential role for complex *N*-glycans in forming an organized layer of bronchial epithelium. Proc Natl Acad Sci USA. 1996 Oct 1;93(20):11041–6.

53. Ioffe E, Liu Y, Stanley P. Complex *N*-glycans in *Mgat1* null preimplantation embryos arise from maternal *Mgat1* RNA. Glycobiology. 1997 Oct;7(7):913–9.

54. Metzler M, Gertz A, Sarkar M, Schachter H, Schrader JW, Marth JD. Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantation development. EMBO J. 1994 May 1;13(9):2056–65.

55. Campbell RM, Metzler M, Granovsky M, Dennis JW, Marth JD. Complex asparagine-linked oligosaccharides in *Mgat1*-null embryos. Glycobiology. 1995 Jul;5(5):535–43.

56. Kaufman RJ. Orchestrating the unfolded protein response in health and disease. J Clin Invest. American Society for Clinical Investigation; 2002 Nov;110(10):1389–98.

57. Kaufman RJ, Scheuner D, Schröder M, Shen X, Lee K, Liu CY, et al. The unfolded protein response in nutrient sensing and differentiation. Nat Rev Mol Cell Biol. 2002 Jun;3(6):411–21.

58. Ellgaard L, Molinari M, Helenius A. Setting the standards: quality control in the secretory pathway. Ullstein Mosby. 1999 Dec 3;286(5446):1882–8.

59. Trombetta ES, Parodi AJ. *N*-glycan processing and glycoprotein folding. Adv Protein Chem. 2001;59:303–44.

60. Marquardt T, Denecke J. Congenital disorders of glycosylation: review of their molecular bases, clinical presentations and specific therapies. Eur J Pediatr. 2003 Jun;162(6):359–79.

61. Jaeken J, Carchon H. Congenital disorders of glycosylation: the rapidly growing tip of the iceberg. Curr Opin Neurol. 2001 Dec;14(6):811–5.

62. Marth JD. Complexity in *O*-linked oligosaccharide biosynthesis engendered by multiple polypeptide *N*-acetylgalactosaminyltransferases. Glycobiology. 1996 Oct;6(7):701–5.

63. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of *O*-linked glycosylation. Crit Rev Biochem Mol Biol. 1998;33(3):151–208.

64. Hanisch FG, Reis CA, Clausen H, Paulsen H. Evidence for glycosylation-dependent activities of polypeptide *N*-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on mucin glycopeptides. Glycobiology. 2001 Sep;11(9):731–40.

65. Mitra N, Sinha S, Ramya TNC, Surolia A. *N*-linked oligosaccharides as outfitters for glycoprotein folding, form and function. Trends Biochem Sci. 2006 Mar;31(3):156–63.

66. Hanover JA. Glycan-dependent signaling: *O*-linked *N*-acetylglucosamine. FASEB J. Federation of American Societies for Experimental Biology; 2001 Sep;15(11):1865–76.

67. Hart GW, Housley MP, Slawson C. Cycling of *O*-linked β-*N*-acetylglucosamine on nucleocytoplasmic proteins. Nature. 2007 Apr 26;446(7139):1017–22.

68. Freeze HH. Disorders in protein glycosylation and potential therapy: tip of an iceberg? J Pediatr. 1998 Nov;133(5):593–600.

69. Beltrán-Valero de Bernabé D, Currier S, Steinbrecher A, Celli J, van Beusekom E, van der Zwaag B, et al. Mutations in the *O*-mannosyltransferase gene POMT1 give rise to the severe neuronal migration disorder Walker-Warburg syndrome. Am J Hum Genet. 2002 Nov;71(5):1033–43.

70. Bernfield M, Götte M, Park PW, Reizes O, Fitzgerald ML, Lincecum J, et al. Functions of cell surface heparan sulfate proteoglycans. Annu Rev Biochem. 1999;68:729–77.

71. Blix FG, Gottschalk A, Klenk E. Proposed nomenclature in the field of neuraminic and sialic acids. Nature. 1957 May 25;179(4569):1088.

72. Muralikrishna G, Reuter G, Peter-Katalinić J, Egge H, Hanisch FG, Siebert HC, et al. Identification of a new ganglioside from the starfish *Asterias rubens*. Carbohydr Res. 1992 Dec 15;236:321–6.

73. Varki NM, Strobert E, Dick EJ, Benirschke K, Varki A. Biomedical differences between human and nonhuman hominids: potential roles for uniquely human aspects of sialic acid biology. Annu Rev Pathol. 2011;6:365–93.

74. Chen X, Varki A. Advances in the biology and chemistry of sialic acids. ACS Chem Biol. 2010 Feb 19;5(2):163–76.

75. Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM. Diversity of microbial sialic acid metabolism. Microbiol Mol Biol Rev. 2004 Mar;68(1):132–53.

76. Angata T, Varki A. Chemical diversity in the sialic acids and related α-keto acids: an evolutionary perspective. Chem Rev. 2002 Feb;102(2):439–69.

77. Gottschalk A. The chemistry and biology of sialic acids and related substances. 1960. 1 p.

78. Varki A. Loss of *N*-glycolylneuraminic acid in humans: mechanisms, consequences, and implications for hominid evolution. Am J Phys Anthropol. 2001;Suppl 33(S33):54–69.

79. Varki A. *N*-glycolylneuraminic acid deficiency in humans. Biochimie. 2001 Jul;83(7):615–22.

80. Kamerling JP, Gerwig GJ. Structural analysis of naturally occurring sialic acids. Methods Mol Biol. New Jersey: Humana Press; 2006;347:69–91.

81. Varki A. Diversity in the sialic acids. Glycobiology. 1992 Feb;2(1):25–40.

82. Schauer R. Sialic acids as regulators of molecular and cellular interactions. Curr Opin Struct Biol. 2009 Oct;19(5):507–14.

83. Kean EL, Münster-Kühnel AK, Gerardy-Schahn R. CMP-sialic acid synthetase of the

nucleus. Biochim Biophys Acta. 2004 Jul 6;1673(1-2):56–65.

84. Altheide TK, Hayakawa T, Mikkelsen TS, Diaz S, Varki N, Varki A. System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents: Evidence for two modes of rapid evolution. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2006 Sep 1;281(35):25689–702.

85. Li Y, Chen X. Sialic acid metabolism and sialyltransferases: natural functions and applications. Appl Microbiol Biotechnol. Springer-Verlag; 2012 May;94(4):887–905.

86. Harduin-Lepers A, Mollicone R, Delannoy P, Oriol R. The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. Glycobiology. 2005 Aug;15(8):805–17.

87. Troy FA. Polysialylation: from bacteria to brains. Glycobiology. 1992 Feb;2(1):5–23.

88. Miyata S, Sato C, Kumita H, Toriyama M, Vacquier VD, Kitajima K. Flagellasialin: a novel sulfated α2,9-linked polysialic acid glycoprotein of sea urchin sperm flagella. Glycobiology. 2006 Dec;16(12):1229–41.

89. Schachner M, Martini R. Glycans and the modulation of neural-recognition molecule function. Trends Neurosci. 1995 Apr;18(4):183–91.

90. Edelman GM, Crossin KL. Cell adhesion molecules: implications for a molecular histology. Annu Rev Biochem. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1991;60(1):155–90.

91. Rutishauser U. NCAM and its polysialic acid moiety: a mechanism for pull/push regulation of cell interactions during development? Dev Suppl. 1992;:99–104.

92. Tang J, Landmesser L, Rutishauser U. Polysialic acid influences specific pathfinding by avian motoneurons. Neuron. 1992 Jun;8(6):1031–44.

93. Zuber C, Lackie PM, Catterall WA, Roth J. Polysialic acid is associated with sodium channels and the neural cell adhesion molecule N-CAM in adult rat brain. J Biol Chem. 1992 May 15;267(14):9965–71.

94. Rutishauser U, Landmesser L. Polysialic acid in the vertebrate nervous system: a promoter of plasticity in cell-cell interactions. Trends Neurosci. 1996 Oct;19(10):422–7.

95. Walsh FS, Doherty P. Neural cell adhesion molecules of the immunoglobulin superfamily: role in axon growth and guidance. Annu Rev Cell Dev Biol. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1997;13(1):425–56.

96. Brusés JL, Rutishauser U. Roles, regulation, and mechanism of polysialic acid function during neural development. Biochimie. 2001 Jul;83(7):635–43.

97. Simon P, Bäumner S, Busch O, Röhrich R, Kaese M, Richterich P, et al. Polysialic acid is present in mammalian semen as a post-translational modification of the neural cell adhesion molecule NCAM and the polysialyltransferase ST8SiaII. Journal of Biological Chemistry. American Society for Biochemistry and Molecular Biology; 2013 Jun 28;288(26):18825–33.

98. Weinhold B, Seidenfaden R, Röckle I, Mühlenhoff M, Schertzinger F, Conzelmann S, et al. Genetic ablation of polysialic acid causes severe neurodevelopmental defects rescued by deletion of the neural cell adhesion molecule. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2005 Dec 30;280(52):42971–7.

99. Tsuji S, Datta AK, Paulson JC. Systematic nomenclature for sialyltransferases. Glycobiology. 1996 Oct;6(7):v–vii.

100. Iwasaki M, Inoue S, Kitajima K, Nomoto H, Inoue Y. Novel oligosaccharide chains on polysialoglycoproteins isolated from rainbow trout eggs. A unique carbohydrate sequence with a sialidase-resistant sialyl group, DGalNAcβ1 leads to 4 (NeuGc2 leads to 3)DGalNAc. Biochemistry. 1984 Jan 17;23(2):305–10.

101. Green ED, Adelt G, Baenziger JU, Wilson S, van Halbeek H. The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of *N*-glycanase-released oligosaccharides by 500-megahertz ^{1}H NMR spectroscopy. J Biol Chem. 1988 Dec 5;263(34):18253–68.

102. Kitagawa H, Takaoka M, Nakada H, Fukui S, Funakoshi I, Kawasaki T, et al. Isolation and structural studies of human milk oligosaccharides that are reactive with a monoclonal antibody MSW 113. J Biochem. 1991 Oct;110(4):598–604.

103. Kubo H, Irie A, Inagaki F, Hoshi M. Gangliosides from the eggs of the sea urchin, *Anthocidaris crassispina*. J Biochem. 1990 Aug;108(2):185–92.

104. Takasaki S, Yamashita K, Suzuki K, Iwanaga S, Kobata A. The sugar chains of cold-insoluble globulin. A protein related to fibronectin. J Biol Chem. 1979 Sep 10;254(17):8548–53.

105. Slomiany A, Slomiany BL. Structures of the acidic oligosaccharides isolated from rat sublingual glycoprotein. J Biol Chem. 1978 Oct 25;253(20):7301–6.

106. Fukuda MN, Dell A, Oates JE, Fukuda M. Embryonal lactosaminoglycan. The structure of branched lactosaminoglycans with novel disialosyl (sialyl α2-9 sialyl) terminals isolated from PA1 human embryonal carcinoma cells. J Biol Chem. 1985 Jun 10;260(11):6623–31.

107. Kitazume S, Kitajima K, Inoue S, Troy FA, Cho JW, Lennarz WJ, et al. Identification of polysialic acid-containing glycoprotein in the jelly coat of sea urchin eggs. Occurrence of a novel type of polysialic acid structure. J Biol Chem. 1994 Sep 9;269(36):22712–8.

108. Comb DG, Roseman S. The sialic acids. I. The structure and enzymatic synthesis of *N*-acetylneuraminic acid. J Biol Chem. 1960 Sep;235:2529–37.

109. Traving C, Schauer R. Structure, function and metabolism of sialic acids. Cell Mol Life Sci. 1998 Dec;54(12):1330–49.

110. Bulai T, Bratosin D, Pons A, Montreuil J, Zanetta JP. Diversity of the human erythrocyte membrane sialic acids in relation with blood groups. FEBS Lett. 2003 Jan 16;534(1-3):185–9.

111. Varki A, Angata T. Siglecs--the major subfamily of I-type lectins. Glycobiology. Oxford University Press; 2006 Jan;16(1):1R–27R.

112. Montreuil J, Biserte G, Strecker G, Spik G, Fontaine G, Farriaux JP. Description of a new type of melituria, called sialuria. Clin Chim Acta. 1968 Jul;21(1):61–9.

113. Fontaine G, Biserte G, Montreuil J, Dupont A, Farriaux JP. Sialuria: an original metabolic disorder. Helv Paediatr Acta. 1968;:Suppl17:1–32.

114. Nishino I, Carrillo-Carrasco N, Argov Z. GNE myopathy: current update and future therapy. J Neurol Neurosurg Psychiatr. BMJ Publishing Group Ltd; 2015 Apr;86(4):385–92.

115. Kamerling JP, Strecker G, Farriaux JP, Dorland L, Haverkamp J, Vliegenthart JF. 2-Acetamidoglucal, a new metabolite isolated from the urine of a patient with sialuria. Biochim Biophys Acta. 1979 Mar 22;583(3):403–8.

116. Thomas GH, Reynolds LW, Miller CS. Overproduction of *N*-acetylneuraminic acid (sialic acid) by sialuria fibroblasts. Pediatr Res. 1985 May;19(5):451–5.

117. Enns GM, Seppala R, Musci TJ, Weisiger K, Ferrell LD, Wenger DA, et al. Clinical course and biochemistry of sialuria. J Inherit Metab Dis. 2001 Jun;24(3):328–36.

118. Bonten E, van der Spoel A, Fornerod M, Grosveld G, d'Azzo A. Characterization of human lysosomal neuraminidase defines the molecular basis of the metabolic storage disorder sialidosis. Genes Dev. 1996 Dec 15;10(24):3156–69.

119. Pshezhetsky AV, Richard C, Michaud L, Igdoura S, Wang S, Elsliger MA, et al. Cloning, expression and chromosomal mapping of human lysosomal sialidase and characterization of

mutations in sialidosis. Nat Genet. 1997 Mar;15(3):316–20.

120. Samraj AN, Läubli H, Varki N, Varki A. Involvement of a non-human sialic acid in human cancer. Front Oncol. Frontiers; 2014;4:33.

121. Kelm S, Schauer R. Sialic acids in molecular and cellular interactions. Int Rev Cytol. 1997;175:137–240.

122. Schauer R. Chemistry, metabolism, and biological functions of sialic acids. Adv Carbohydr Chem Biochem. 1982;40:131–234.

123. Schauer R. Sialic acids and their role as biological masks. Trends in Biochemical Sciences. 1985;10(9):357-360

124. Crocker PR, Clark EA, Filbin M, Gordon S, Jones Y, Kehrl JH, et al. Siglecs: a family of sialic-acid binding lectins. Glycobiology. 1998 Feb;8(2):v.

125. Koliwer-Brandl H, Siegert N, Umnus K, Kelm A, Tolkach A, Kulozik U, et al. Lectin inhibition assays for the analysis of bioactive milk sialoglycoconjugates. International Dairy Journal. Elsevier Ltd; 2011 Jun 1;21(6):413–20.

126. Kelm S, Madge P, Islam T, Bennett R, Koliwer-Brandl H, Waespy M, et al. C-4 modified sialosides enhance binding to Siglec-2 (CD22): towards potent Siglec inhibitors for immunoglycotherapy. Angew Chem Int Ed Engl. 2013 Mar 25;52(13):3616–20.

127. Nitschke L. CD22 and Siglec-G regulate inhibition of B-cell signaling by sialic acid ligand binding and control B-cell tolerance. Glycobiology. Oxford University Press; 2014 Sep;24(9):807–17.

128. Minko T. Drug targeting to the colon with lectins and neoglycoconjugates. Adv Drug Deliv Rev. 2004 Mar 3;56(4):491–509.

129. Sharon N, Lis H. History of lectins: from hemagglutinins to biological recognition molecules. Glycobiology. Oxford University Press; 2004 Nov;14(11):53R–62R.

130. Franz H. Advances in lectin research. Band 1. Ullstein Mosby; 1988. 187 p.

131. Spooner RA, Lord JM. Ricin trafficking in cells. Toxins (Basel). Multidisciplinary Digital Publishing Institute; 2015 Jan;7(1):49–65.

132. Weis WI, Drickamer K. Structural basis of lectin-carbohydrate recognition. Annu Rev Biochem. 1996;65:441–73.

133. Drickamer K. Two distinct classes of carbohydrate-recognition domains in animal lectins. J Biol Chem. 1988 Jul 15;263(20):9557–60.

134. Fiedler K, Simons K. A putative novel class of animal lectins in the secretory pathway homologous to leguminous lectins. Cell. 1994 Jun 3;77(5):625–6.

135. Loris R, Hamelryck T, Bouckaert J, Wyns L. Legume lectin structure. Biochim Biophys Acta. 1998 Mar 3;1383(1):9–36.

136. Hamelryck TW, Loris R, Bouckaert J, Dao-Thi MH, Strecker G, Imberty A, et al. Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. Journal of Molecular Biology. 1999 Mar 5;286(4):1161–77.

137. Barondes SH, Cooper DN, Gitt MA, Leffler H. Galectins. Structure and function of a large family of animal lectins. J Biol Chem. 1994 Aug 19;269(33):20807–10.

138. Epperson TK, Patel KD, McEver RP, Cummings RD. Noncovalent association of P-selectin glycoprotein ligand-1 and minimal determinants for binding to P-selectin. J Biol Chem. 2000 Mar 17;275(11):7839–53.

139. Hemmerich S, Rosen SD. Carbohydrate sulfotransferases in lymphocyte homing. Glycobiology. 2000 Sep;10(9):849–56.

140. Hughes RC. Galectins as modulators of cell adhesion. Biochimie. 2001 Jul;83(7):667–76.

141. Rini JM. Lectin structure. Annu Rev Biophys Biomol Struct. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1995;24(1):551–77.

142. Somers WS, Tang J, Shaw GD, Camphausen RT. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe^X and PSGL-1. Cell. 2000 Oct 27;103(3):467–79.

143. Turner MW. Mannose-binding lectin: the pluripotent molecule of the innate immune system. Immunol Today. 1996 Nov;17(11):532–40.

144. Vestweber D, Blanks JE. Mechanisms that regulate the function of the selectins and their ligands. Physiol Rev. 1999 Jan;79(1):181–213.

145. Weis WI, Taylor ME, Drickamer K. The C-type lectin superfamily in the immune system. Immunol Rev. 1998 Jun;163:19–34.

146. Bleijs DA, Geijtenbeek TB, Figdor CG, van Kooyk Y. DC-SIGN and LFA-1: a battle for ligand. Trends Immunol. 2001 Aug;22(8):457–63.

147. Dwir O, Kansas GS, Alon R. Cytoplasmic anchorage of L-selectin controls leukocyte capture and rolling by increasing the mechanical stability of the selectin tether. J Cell Biol. Rockefeller Univ Press; 2001 Oct 1;155(1):145–56.

148. Hernandez JD, Baum LG. Ah, sweet mystery of death! Galectins and control of cell fate. Glycobiology. 2002 Oct;12(10):127R–36R.

149. Hu Y, Szente B, Kiely JM, Gimbrone MA. Molecular events in transmembrane signaling via E-selectin. SHP2 association, adaptor protein complex formation and ERK1/2 activation. J Biol Chem. 2001 Dec 21;276(51):48549–53.

150. Lübke T, Marquardt T, Etzioni A, Hartmann E, Figura von K, Körner C. Complementation cloning identifies CDG-IIc, a new type of congenital disorders of glycosylation, as a GDP-fucose transporter deficiency. Nat Genet. 2001 May;28(1):73–6.

151. Pöhlmann S, Baribaud F, Doms RW. DC-SIGN and DC-SIGNR: helping hands for HIV. Trends in immunology. 2001 Dec;:643–6.

152. Stahl PD, Ezekowitz RA. The mannose receptor is a pattern recognition receptor involved in host defense. Curr Opin Immunol. 1998 Feb;10(1):50–5.

153. van der Merwe PA. Leukocyte adhesion: high-speed cells with ABS. Curr Biol. 1999 Jun 3;9(11):R419–22.

154. Wallis R, Shaw JM, Uitdehaag J, Chen C-B, Torgersen D, Drickamer K. Localization of the serine protease-binding sites in the collagen-like domain of mannose-binding protein: indirect effects of naturally occurring mutations on protease binding and activation. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2004 Apr 2;279(14):14065–73.

155. Weis WI, Quesenberry MS, Taylor ME, Bezouska K, Hendrickson WA, Drickamer K. Molecular mechanisms of complex carbohydrate recognition at the cell surface. Cold Spring Harb Symp Quant Biol. 1992;57:281–9.

156. Weis WI, Drickamer K, Hendrickson WA. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. Nature. Nature Publishing Group; 1992 Nov 12;360(6400):127–34.

157. Kerrigan AM, Brown GD. C-type lectins and phagocytosis. Immunobiology. 2009;214(7):562–75.

158. Ghazarian H, Idoni B, Oppenheimer SB. A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. Acta Histochem. 2011 May;113(3):236–47.

159. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al. I-type Lectins. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

160. Crocker PR, Varki A. Siglecs, sialic acids and innate immunity. Trends Immunol. 2001 Jun;22(6):337–42.

161. Crocker PR, Varki A. Siglecs in the immune system. Immunology. 2001 Jun;103(2):137–45.

162. Crocker PR, Gordon S. Mouse macrophage hemagglutinin (sheep erythrocyte receptor) with specificity for sialylated glycoconjugates characterized by a monoclonal antibody. J Exp Med. 1989 Apr 1;169(4):1333–46.

163. Kelm S, Pelz A, Schauer R, Filbin MT, Tang S, de Bellard ME, et al. Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. Curr Biol. 1994 Nov 1;4(11):965–72.

164. Crocker PR, Kelm S, Dubois C, Martin B, McWilliam AS, Shotton DM, et al. Purification and properties of sialoadhesin, a sialic acid-binding receptor of murine tissue macrophages. EMBO J. 1991 Jul;10(7):1661–9.

165. Engel P, Wagner N, Miller AS, Tedder TF. Identification of the ligand-binding domains of CD22, a member of the immunoglobulin superfamily that uniquely binds a sialic acid-dependent ligand. J Exp Med. 1995 Apr 1;181(4):1581–6.

166. Law CL, Aruffo A, Chandran KA, Doty RT, Clark EA. Ig domains 1 and 2 of murine CD22 constitute the ligand-binding domain and bind multiple sialylated ligands expressed on B and T cells. J Immunol. 1995 Oct 1;155(7):3368–76.

167. Nath D, van der Merwe PA, Kelm S, Bradfield P, Crocker PR. The amino-terminal immunoglobulin-like domain of sialoadhesin contains the sialic acid binding site. Comparison with CD22. J Biol Chem. 1995 Nov 3;270(44):26184–91.

168. Crocker PR, Mucklow S, Bouckson V, McWilliam A, Willis AC, Gordon S, et al. Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains. EMBO J. 1994 Oct 3;13(19):4490–503.

169. Freeman SD, Kelm S, Barber EK, Crocker PR. Characterization of CD33 as a new member of the sialoadhesin family of cellular interaction molecules. Blood. 1995 Apr 15;85(8):2005–12.

170. Tang S, Shen YJ, DeBellard ME, Mukhopadhyay G, Salzer JL, Crocker PR, et al. Myelin-associated glycoprotein interacts with neurons via a sialic acid binding site at ARG118 and a distinct neurite inhibition site. J Cell Biol. 1997 Sep 22;138(6):1355–66.

171. Kelm S, Schauer R, Crocker PR. The Sialoadhesins--a family of sialic acid-dependent cellular recognition molecules within the immunoglobulin superfamily. Glycoconj J. 1996 Dec;13(6):913–26.

172. Crocker PR, Hartnell A, Munday J, Nath D. The potential role of sialoadhesin as a macrophage recognition molecule in health and disease. Glycoconj J. 1997 Aug;14(5):601–9.

173. Filbin MT. Myelin-associated glycoprotein: a role in myelination and in the inhibition of axonal regeneration? Curr Opin Neurobiol. 1995 Oct;5(5):588–95.

174. Lopez PHH, Ahmad AS, Mehta NR, Toner M, Rowland EA, Zhang J, et al. Myelin-associated glycoprotein protects neurons from excitotoxicity. Journal of Neurochemistry. 2011 Jan 7;116(5):900–8.

175. Quarles RH. Myelin-associated glycoprotein (MAG): past, present and beyond. J Neurochem. 2007 Jan 4;100:1431–1448.

176. Powell LD, Sgroi D, Sjoberg ER, Stamenkovic I, Varki A. Natural ligands of the B cell adhesion molecule CD22 beta carry *N*-linked oligosaccharides with α-2,6-linked sialic acids that are required for recognition. J Biol Chem. 1993 Apr 5;268(10):7019–27.

177. Sgroi D, Varki A, Braesch-Andersen S, Stamenkovic I. CD22, a B cell-specific immunoglobulin superfamily member, is a sialic acid-binding lectin. J Biol Chem. 1993 Apr 5;268(10):7011–8.

178. Nitschke L, Floyd H, Ferguson DJ, Crocker PR. Identification of CD22 ligands on bone marrow sinusoidal endothelium implicated in CD22-dependent homing of recirculating B cells. J Exp Med. 1999 May 3;189(9):1513–8.

179. Jellusova J, Nitschke L. Regulation of B cell functions by the sialic acid-binding receptors siglec-G and CD22. Front Immunol. Frontiers; 2011;2:96.

180. Jones C, Virji M, Crocker PR. Recognition of sialylated meningococcal lipopolysaccharide by siglecs expressed on myeloid cells leads to enhanced bacterial uptake. Mol Microbiol. 2003 Sep;49(5):1213–25.

181. Collins BE, Yang LJ, Mukhopadhyay G, Filbin MT, Kiso M, Hasegawa A, et al. Sialic acid specificity of myelin-associated glycoprotein binding. J Biol Chem. 1997 Jan 10;272(2):1248–55.

182. Collins BE, Kiso M, Hasegawa A, Tropak MB, Roder JC, Crocker PR, et al. Binding specificities of the sialoadhesin family of I-type lectins. Sialic acid linkage and substructure requirements for binding of myelin-associated glycoprotein, Schwann cell myelin protein, and sialoadhesin. J Biol Chem. 1997 Jul 4;272(27):16889–95.

183. Kelm S, Schauer R, Manuguerra JC, Gross HJ, Crocker PR. Modifications of cell surface sialic acids modulate cell adhesion mediated by sialoadhesin and CD22. Glycoconj J. 1994 Dec;11(6):576–85.

184. Powell LD, Varki A. I-type lectins. J Biol Chem. 1995 Jun 16;270(24):14243–6.

185. WHO. Research Priorities for Chagas Disease, Human African Trypanosomiasis and Leishmaniasis. Technical Report Series. 2012 Feb 17;975:1–116.

186. WHO. Control of Chagas disease. World Health Organization technical report series. 2002 pp. 1–109.

187. Malvy D, Chappuis F. Sleeping sickness. Clin Microbiol Infect. 2011 Jul;17(7):986–95.

188. Connor RJ. The impact of nagana. Onderstepoort J Vet Res. 1994 Dec;61(4):379–83.

189. Duszenko M. Die Erreger der Schlafkrankheit. Biologie in unserer Zeit. 1998;2:72–81.

190. Chagas C. Nova tripanozomiase humana. Mem Inst Oswaldo Cruz. 1909 Jun 7;1:159–218.

191. Hotez PJ, Molyneux DH, Fenwick A, Kumaresan J, Sachs SE, Sachs JD, et al. Control of neglected tropical diseases. N Engl J Med. 2007 Sep 6;357(10):1018–27.

192. Clayton J. Chagas disease 101. Nature. 2010 Jun 24;465(7301):S4–5.

193. Rassi A, Marin-Neto JA. Chagas disease. Lancet. 2010 Apr 17;375(9723):1388–402.

194. Aufderheide AC, Salo W, Madden M, Streitz J, Buikstra J, Guhl F, et al. A 9,000-year record of Chagas' disease. Proc Natl Acad Sci USA. National Acad Sciences; 2004 Feb 17;101(7):2034–9.

195. Darwin C. The Origin of Species and The Voyage of the “Beagle.” Everyman's Library; 2012.

196. Bernstein RE. Darwin’s illness: Chagas’ disease resurgens. Journal of the Royal Society of Medicine. 1984 July;77(7):608–609.

197. Villalba R, Fornés G, Alvarez MA, Román J, Rubio V, Fernández M, et al. Acute Chagas' disease in a recipient of a bone marrow transplant in Spain: case report. Clin Infect Dis. 1992 Feb;14(2):594–5.

198. Cimo PL, Luper WE, Scouros MA. Transfusion-associated Chagas' disease in Texas: report of a case. Tex Med. 1993 Dec;89(12):48–50.

199. Leiby DA, Lenes BA, Tibbals MA, Tames-Olmedo MT. Prospective evaluation of a patient with *Trypanosoma cruzi* infection transmitted by transfusion. N Engl J Med. 1999 Oct 14;341(16):1237–9.

200. Kun H, Moore A, Mascola L, Steurer F, Lawrence G, Kubak B, et al. Transmission of *Trypanosoma cruzi* by heart transplantation. Clin Infect Dis. Oxford University Press; 2009 Jun 1;48(11):1534–40.

201. Alarcón de Noya B, Díaz-Bello Z, Colmenares C, Ruiz-Guevara R, Mauriello L, Zavala-Jaspe R, et al. Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela. J Infect Dis. Oxford University Press; 2010 May 1;201(9):1308–15.

202. Chappuis F, Mauris A, Holst M, Albajar-Vinas P, Jannin J, Luquetti AO, et al. Validation of a rapid immunochromatographic assay for diagnosis of *Trypanosoma cruzi* infection among Latin-American Migrants in Geneva, Switzerland. J Clin Microbiol. American Society for Microbiology; 2010 Aug;48(8):2948–52.

203. Clayton J. Chagas' disease: pushing through the pipeline. Nature. 2010 Jun 24;465(7301):S12–5.

204. World Health Organization. Control and surveillance of human African trypanosomiasis. World Health Organ Tech Rep Ser. 2013;(984):1–237.

205. Jackson AP, Sanders M, Berry A, McQuillan J, Aslett MA, Quail MA, et al. The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human african trypanosomiasis. PLoS Negl Trop Dis. Public Library of Science; 2010;4(4):e658.

206. Hoare CA. The trypanosomes of mammals. J samll Anim Pract. 1972 May 16;13:671–2.

207. Truc P, Büscher P, Cuny G, Gonzatti MI, Jannin J, Joshi P, et al. Atypical human infections by animal trypanosomes. Ndung'u JM, editor. PLoS Negl Trop Dis. 2013 Sep 12;7(9):e2256.

208. WHO. Human African trypanosomiasis (sleeping sickness): epidemiological update. Wkly Epidemiol Rec. 2006 Feb 24;81(8):71–80.

209. Chappuis F, Lima MA, Flevaud L, Ritmeijer K. Human African trypanosomiasis in areas without surveillance. Emerging Infect Dis. 2010 Feb;16(2):354–6.

210. Courtin F, Sidibé I, Rouamba J, Jamonneau V, Gouro A, Solano P. Population growth and global warming: impacts on tsetse and trypanosomoses in West Africa. Parasite. 2009 Mar;16(1):3–10.

211. Hursey BS. The programme against African trypanosomiasis: aims, objectives and achievements. Trends Parasitol. 2001 Jan;17(1):2–3.

212. Vickerman K, Tetley L, Hendry KA, Turner CM. Biology of African trypanosomes in the tsetse fly. Biol Cell. 1988;64(2):109–19.

213. Aksoy S, Gibson WC, Lehane MJ. Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis. Adv Parasitol. 2003;53:1–83.

214. Gibson W, Bailey M. The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes. Kinetoplastid Biol Dis. 2003;2(1):1.

215. Welburn SC, Maudlin I, Ellis DS. Rate of trypanosome killing by lectins in midguts of different species and strains of Glossina. Med Vet Entomol. 1989 Jan;3(1):77–82.

216. Hao Z, Kasumba I, Lehane MJ, Gibson WC, Kwon J, Aksoy S. Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. Proc Natl Acad Sci USA. National Acad Sciences; 2001 Oct 23;98(22):12648–53.

217. Abubakar LU, Bulimo WD, Mulaa FJ, Osir EO. Molecular characterization of a tsetse fly midgut proteolytic lectin that mediates differentiation of African trypanosomes. Insect Biochem Mol Biol. 2006 Apr;36(4):344–52.

218. Acosta-Serrano A, Vassella E, Liniger M, Kunz Renggli C, Brun R, Roditi I, et al. The surface coat of procyclic *Trypanosoma brucei*: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. Proc Natl Acad Sci USA. National Acad Sciences;

2001 Feb 13;98(4):1513–8.

219. Urwyler S, Studer E, Renggli CK, Roditi I. A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of *Trypanosoma brucei*. Mol Microbiol. Blackwell Publishing Ltd; 2007 Jan;63(1):218–28.

220. Vickerman K. Developmental cycles and biology of pathogenic trypanosomes. Br Med Bull. 1985 Apr;41(2):105–14.

221. Rifkin MR. Identification of the trypanocidal factor in normal human serum: high density lipoprotein. Proc Natl Acad Sci USA. 1978 Jul;75(7):3450–4.

222. Alsford S, Currier RB, Guerra-Assunção JA, Clark TG, Horn D. Cathepsin-L can resist lysis by human serum in *Trypanosoma brucei brucei*. PLoS Pathog. Public Library of Science; 2014 May;10(5):e1004130.

223. Salmon D, Vanwalleghem G, Morias Y, Denoeud J, Krumbholz C, Lhommé F, et al. Adenylate cyclases of *Trypanosoma brucei* inhibit the innate immune response of the host. Ullstein Mosby. American Association for the Advancement of Science; 2012 Jul 27;337(6093):463–6.

224. Salmon D, Bachmaier S, Krumbholz C, Kador M, Gossmann JA, Uzureau P, et al. Cytokinesis of *Trypanosoma brucei* bloodstream forms depends on expression of adenylyl cyclases of the ESAG4 or ESAG4-like subfamily. Mol Microbiol. Blackwell Publishing Ltd; 2012 Apr;84(2):225–42.

225. Hensley SE, Das SR, Gibbs JS, Bailey AL, Schmidt LM, Bennink JR, et al. Influenza A virus hemagglutinin antibody escape promotes neuraminidase antigenic variation and drug resistance. PLoS ONE. Public Library of Science; 2011;6(2):e15190.

226. Nowak MA, May RM. Mathematical biology of HIV infections: antigenic variation and diversity threshold. Math Biosci. 1991 Sep;106(1):1–21.

227. Pays E, Nolan DP. Expression and function of surface proteins in *Trypanosoma brucei*. Mol Biochem Parasitol. 1998 Mar 1;91(1):3–36.

228. De Lange T, Borst P. Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. Nature. 1982 Sep 30;299(5882):451–3.

229. Robinson NP, Burman N, Melville SE, Barry JD. Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. Mol Cell Biol. 1999 Sep;19(9):5839–46.

230. Hall JPJ, Wang H, Barry JD. Mosaic VSGs and the scale of *Trypanosoma brucei* antigenic variation. PLoS Pathog. Public Library of Science; 2013;9(7):e1003502.

231. Baral TN. Immunobiology of African trypanosomes: need of alternative interventions. J Biomed Biotechnol. Hindawi Publishing Corporation; 2010;2010:389153.

232. Roditi I, Furger A, Ruepp S, Schürch N, Bütikofer P. Unravelling the procyclin coat of *Trypanosoma brucei*. Mol Biochem Parasitol. 1998 Mar 1;91(1):117–30.

233. Mehlert A, Zitzmann N, Richardson JM, Treumann A, Ferguson MA. The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of *Trypanosoma brucei*. Mol Biochem Parasitol. 1998 Mar 1;91(1):145–52.

234. Roditi I, Clayton C. An unambiguous nomenclature for the major surface glycoproteins of the procyclic form of *Trypanosoma brucei*. Mol Biochem Parasitol. 1999 Sep 20;103(1):99–100.

235. Bütikofer P, Vassella E, Boschung M, Renggli CK, Brun R, Pearson TW, et al. Glycosylphosphatidylinositol-anchored surface molecules of *Trypanosoma congolense* insect forms are developmentally regulated in the tsetse fly. Mol Biochem Parasitol. 2002 Jan;119(1):7–16.

236. Utz S, Roditi I, Kunz Renggli C, Almeida IC, Acosta-Serrano A, Bütikofer P. *Trypanosoma congolense* procyclins: unmasking cryptic major surface glycoproteins in procyclic forms. Eukaryotic Cell. 2006 Aug;5(8):1430–40.

237. Treumann A, Zitzmann N, Hülsmeier A, Prescott AR, Almond A, Sheehan J, et al. Structural characterisation of two forms of procyclic acidic repetitive protein expressed by procyclic forms of *Trypanosoma brucei*. Journal of Molecular Biology. 1997 Jun 20;269(4):529–47.

238. Ferguson MA, Murray P, Rutherford H, McConville MJ. A simple purification of procyclic acidic repetitive protein and demonstration of a sialylated glycosyl-phosphatidylinositol membrane anchor. Biochem J. 1993 Apr 1;291 (Pt 1):51–5.

239. Field MC, Menon AK, Cross GA. A glycosylphosphatidylinositol protein anchor from procyclic stage *Trypanosoma brucei*: lipid structure and biosynthesis. EMBO J. 1991 Oct;10(10):2731–9.

240. Engstler M, Schauer R. Sialidases from African trypanosomes. Parasitol Today (Regul Ed). 1993 Jun;9(6):222–5.

241. Engstler M, Schauer R, Brun R. Distribution of developmentally regulated trans-sialidases in the Kinetoplastida and characterization of a shed trans-sialidase activity from procyclic *Trypanosoma congolense*. Acta Tropica. 1995 May;59(2):117–29.

242. Schenkman S, Eichinger D, Pereira ME, Nussenzweig V. Structural and functional properties of *Trypanosoma* trans-sialidase. Annu Rev Microbiol. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1994;48(1):499–523.

243. Engstler M, Reuter G, Schauer R. The developmentally regulated trans-sialidase from *Trypanosoma brucei* sialylates the procyclic acidic repetitive protein. Mol Biochem Parasitol. 1993 Sep;61(1):1–13.

244. Pontes de Carvalho LC, Tomlinson S, Vandekerckhove F, Bienen EJ, Clarkson AB, Jiang MS, et al. Characterization of a novel trans-sialidase of *Trypanosoma brucei* procyclic trypomastigotes and identification of procyclin as the main sialic acid acceptor. J Exp Med. 1993 Feb 1;177(2):465–74.

245. Ruepp S, Kurath U, Renggli CK, Brun R, Roditi I. Glutamic acid/alanine-rich protein from *Trypanosoma congolense* is the functional equivalent of "EP" procyclin from *Trypanosoma brucei*. Mol Biochem Parasitol. 1999 Jan 5;98(1):151–6.

246. La Greca F, Magez S. Vaccination against trypanosomiasis: can it be done or is the trypanosome truly the ultimate immune destroyer and escape artist? Hum Vaccin. 2011 Nov;7(11):1225–33.

247. Fairlamb AH. Chemotherapy of human African trypanosomiasis: current and future prospects. Trends Parasitol. 2003 Nov;19(11):488–94.

248. Pépin J, Milord F. The treatment of human African trypanosomiasis. Adv Parasitol. 1994;33:1–47.

249. de Koning HP. Ever-increasing complexities of diamidine and arsenical crossresistance in African trypanosomes. Trends Parasitol. Elsevier; 2008 Aug;24(8):345–9.

250. Alsford S, Eckert S, Baker N, Glover L, Sanchez-Flores A, Leung KF, et al. High-throughput decoding of antitrypanosomal drug efficacy and resistance. Nature. Nature Publishing Group; 2012 Jan 25;:1–6.

251. Monti E, Miyagi T. Structure and function of mammalian sialidases. Topics in Current Chemistry. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012.

252. Haselhorst T, Wilson JC, Thomson RJ, McAtamney S, Menting JG, Coppel RL, et al. Saturation transfer difference (STD) ^{1}H-NMR experiments and in silico docking experiments to probe the binding of *N*-acetylneuraminic acid and derivatives to *Vibrio cholerae* sialidase. Proteins. 2004 Apr 28;56(2):346–53.

253. Taylor NR, Itzstein von M. A structural and energetics analysis of the binding of a series of *N*-acetylneuraminic-acid-based inhibitors to influenza virus sialidase. J Comput Aided Mol Des. 1996 Jun;10(3):233–46.

254. Stewart-Tull DE, Ollar RA, Scobie TS. Studies on the *Vibrio cholerae* mucinase complex. I. Enzymic activities associated with the complex. J Med Microbiol. 1986 Dec;22(4):325–33.

255. Pereira ME. A developmentally regulated neuraminidase activity in *Trypanosoma cruzi*. Ullstein Mosby. 1983 Mar 25;219(4591):1444–6.

256. Schauer R, Reuter G, Mühlpfordt H, Andrade AF, Pereira ME. The occurrence of *N*-acetyl- and *N*-glycoloylneuraminic acid in *Trypanosoma cruzi*. Hoppe-Seyler's Z Physiol Chem. 1983 Aug;364(8):1053–7.

257. Zingales B, Carniol C, de Lederkremer RM, Colli W. Direct sialic acid transfer from a protein donor to glycolipids of trypomastigote forms of *Trypanosoma cruzi*. Mol Biochem Parasitol. 1987 Nov;26(1-2):135–44.

258. Engstler M, Reuter G, Schauer R. Purification and characterization of a novel sialidase found in procyclic culture forms of *Trypanosoma brucei*. Mol Biochem Parasitol. 1992 Aug;54(1):21–30.

259. Schenkman S, Eichinger D. *Trypanosoma cruzi* trans-sialidase and cell invasion. Parasitol Today (Regul Ed). 1993 Jun;9(6):218–22.

260. Tiralongo E, Schrader S, Lange H, Lemke H, Tiralongo J, Schauer R. Two trans-sialidase forms with different sialic acid transfer and sialidase activities from *Trypanosoma congolense*. J Biol Chem. 2003 Jun 27;278(26):23301–10.

261. Buscaglia CA, Alfonso J, Campetella O, Frasch AC. Tandem amino acid repeats from *Trypanosoma cruzi* shed antigens increase the half-life of proteins in blood. Blood. 1999 Mar 15;93(6):2025–32.

262. Frasch AC. Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. Parasitol Today (Regul Ed). 2000 Jul;16(7):282–6.

263. Tiralongo E, Martensen I, Grötzinger J, Tiralongo J, Schauer R. Trans-sialidase-like sequences from *Trypanosoma congolense* conserve most of the critical active site residues found in other trans-sialidases. Biol Chem. 2003 Aug;384(8):1203–13.

264. Risso MG, Garbarino GB, Mocetti E, Campetella O, Gonzalez Cappa SM, Buscaglia CA, et al. Differential expression of a virulence factor, the trans-sialidase, by the main *Trypanosoma cruzi* phylogenetic lineages. J Infect Dis. Oxford University Press; 2004 Jun 15;189(12):2250–9.

265. Jäger AV, Muiá RP, Campetella O. Stage-specific expression of *Trypanosoma cruzi* trans-sialidase involves highly conserved 3' untranslated regions. FEMS Microbiol Lett. 2008 Jun;283(2):182–8.

266. Freitas LM, Santos dos SL, Rodrigues-Luiz GF, Mendes TAO, Rodrigues TS, Gazzinelli RT, et al. Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of *Trypanosoma cruzi* reveal an undetected level of complexity. Rodrigues MM, editor. PLoS ONE. 2011 Oct 19;6(10):e25914.

267. Coustou V, Plazolles N, Guegan F, Baltz T. Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. Cellular Microbiology. 2012 Feb 13;14(3):431–45.

268. Montagna G, Cremona ML, Paris G, Amaya MF, Buschiazzo A, Alzari PM, et al. The trans-sialidase from the african trypanosome *Trypanosoma brucei*. Eur J Biochem. 2002 Jun;269(12):2941–50.

269. Nakatani F, Morita YS, Ashida H, Nagamune K, Maeda Y, Kinoshita T. Identification of a second catalytically active trans-sialidase in *Trypanosoma brucei*. Biochemical and

Biophysical Research Communications. Elsevier Inc; 2011 Nov 18;415(2):421–5.

270. Cremona ML, Sánchez DO, Frasch AC, Campetella O. A single tyrosine differentiates active and inactive *Trypanosoma cruzi* trans-sialidases. Gene. 1995 Jul 4;160(1):123–8.

271. Jackson AP, Allison HC, Barry JD, Field MC, Hertz-Fowler C, Berriman M. A cell-surface phylome for African trypanosomes. Tschudi C, editor. PLoS Negl Trop Dis. 2013 Mar 21;7(3):e2121.

272. Koliwer-Brandl H, Gbem TT, Waespy M, Reichert O, Mandel P, Drebitz E, et al. Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*. BMC Biochem. 2011;12(1):39.

273. Gbem TT, Waespy M, Hesse B, Dietz F, Smith J, Chechet GD, et al. Biochemical diversity in the *Trypanosoma congolense* trans-sialidase family. PLoS Negl Trop Dis. 2013;7(12):e2549.

274. Schauer R, Kamerling JP. The chemistry and biology of trypanosomal trans-sialidases: virulence factors in Chagas' disease and sleeping sickness. Chembiochem. WILEY-VCH Verlag; 2011 Oct 17;12(15):2246–64.

275. Chong AK, Pegg MS, Taylor NR, Itzstein von M. Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. Eur J Biochem. 1992 Jul 1;207(1):335–43.

276. Buschiazzo A, Campetella O, Frasch AC. *Trypanosoma rangeli* sialidase: cloning, expression and similarity to *T. cruzi* trans-sialidase. Glycobiology. 1997 Dec;7(8):1167–73.

277. Nok AJ, Balogun EO. A bloodstream *Trypanosoma congolense* sialidase could be involved in anemia during experimental trypanosomiasis. J Biochem. 2003 Jun;133(6):725–30.

278. Nagamune K, Acosta-Serrano A, Uemura H, Brun R, Kunz Renggli C, Maeda Y, et al. Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors. J Exp Med. Rockefeller Univ Press; 2004 May 17;199(10):1445–50.

279. Guegan F, Plazolles N, Baltz T, Coustou V. Erythrophagocytosis of desialylated red blood cells is responsible for anaemia during *Trypanosoma vivax* infection. Cellular Microbiology. 2013 Aug;15(8):1285–303.

280. Ribeirão M, Pereira-Chioccola VL, Eichinger D, Rodrigues MM, Schenkman S. Temperature differences for trans-glycosylation and hydrolysis reaction reveal an acceptor binding site in the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. Glycobiology. 1997 Dec;7(8):1237–46.

281. Neubacher BR, Schmidt D, Ziegelm ller P, Thiem J. Preparation of sialylated oligosaccharides employing recombinant trans-sialidase from *Trypanosoma cruzi*. Org Biomol Chem. 2005;3(8):1551.

282. Agusti R. Lactose derivatives are inhibitors of *Trypanosoma cruzi* trans-sialidase activity toward conventional substrates in vitro and in vivo. Glycobiology. 2004 Apr 14;14(7):659–70.

283. Vandekerckhove F, Schenkman S, Pontes de Carvalho L, Tomlinson S, Kiso M, Yoshida M, et al. Substrate specificity of the *Trypanosoma cruzi* trans-sialidase. Glycobiology. 1992 Dec;2(6):541–8.

284. Scudder P, Doom JP, Chuenkova M, Manger ID, Pereira ME. Enzymatic characterization of β-D-galactoside α2,3-trans-sialidase from *Trypanosoma cruzi*. J Biol Chem. 1993 May 5;268(13):9886–91.

285. Acosta-Serrano A, Almeida IC, Freitas-Junior LH, Yoshida N, Schenkman S. The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. Mol Biochem Parasitol. 2001 May;114(2):143–50.

286. Agustí R, Giorgi ME, de Lederkremer RM. The trans-sialidase from *Trypanosoma cruzi*

efficiently transfers α-(2-3)-linked *N*-glycolylneuraminic acid to terminal β-galactosyl units. Carbohydr Res. 2007 Nov 26;342(16):2465–9.

287. Vetere A, Paoletti S. Complete synthesis of 3'-sialyl-*N*-acetyllactosamine by regioselective transglycosylation. FEBS Lett. 1996 Dec 16;399(3):203–6.

288. Ferrero-García MA, Trombetta SE, Sánchez DO, Reglero A, Frasch AC, Parodi AJ. The action of *Trypanosoma cruzi* trans-sialidase on glycolipids and glycoproteins. Eur J Biochem. 1993 Apr 15;213(2):765–71.

289. Neubacher B, Scheid S, Kelm S, Frasch AC, Meyer B, Thiem J. Synthesis of Neu5Ac oligosaccharides and analogues by transglycosylation and their binding properties as ligands to MAG. Chembiochem. WILEY-VCH Verlag; 2006 Jun;7(6):896–9.

290. Monteiro VG, Lobato CSS, Silva AR, Medina DV, de Oliveira MA, Seabra SH, et al. Increased association of *Trypanosoma cruzi* with sialoadhesin positive mice macrophages. Parasitol Res. Springer-Verlag; 2005 Nov;97(5):380–5.

291. Jacobs T, Erdmann H, Fleischer B. Molecular interaction of Siglecs (sialic acid-binding Ig-like lectins) with sialylated ligands on *Trypanosoma cruzi*. Eur J Cell Biol. 2010 Jan;89(1):113–6.

292. Erdmann H, Steeg C, Koch-Nolte F, Fleischer B, Jacobs T. Sialylated ligands on pathogenic *Trypanosoma cruzi* interact with Siglec-E (sialic acid-binding Ig-like lectin-E). Cellular Microbiology. Blackwell Publishing Ltd; 2009 Nov;11(11):1600–11.

293. Gao W, Luquetti AO, Pereira MA. Immunological tolerance and its breakdown in Chagas' heart disease: role of parasitokines. Front Biosci. 2003 Jan 1;8:e218–27.

294. Gao W, Pereira MA. Interleukin-6 is required for parasite specific response and host resistance to *Trypanosoma cruzi*. Int J Parasitol. 2002 Feb;32(2):167–70.

295. Gao W, Wortis HH, Pereira MA. The *Trypanosoma cruzi* trans-sialidase is a T cell-independent B cell mitogen and an inducer of non-specific Ig secretion. Int Immunol. 2002 Mar;14(3):299–308.

296. Pereira ME, Mejia JS, Ortega-Barria E, Matzilevich D, Prioli RP. The *Trypanosoma cruzi* neuraminidase contains sequences similar to bacterial neuraminidases, YWTD repeats of the low density lipoprotein receptor, and type III modules of fibronectin. J Exp Med. 1991 Jul 1;174(1):179–91.

297. Pollevick GD, Affranchino JL, Frasch AC, Sánchez DO. The complete sequence of a shed acute-phase antigen of *Trypanosoma cruzi*. Mol Biochem Parasitol. 1991 Aug;47(2):247–50.

298. Uemura H, Schenkman S, Nussenzweig V, Eichinger D. Only some members of a gene family in *Trypanosoma cruzi* encode proteins that express both trans-sialidase and neuraminidase activities. EMBO J. European Molecular Biology Organization; 1992 Nov;11(11):3837–44.

299. Parodi AJ, Pollevick GD, Mautner M, Buschiazzo A, Sánchez DO, Frasch AC. Identification of the gene(s) coding for the trans-sialidase of *Trypanosoma cruzi*. EMBO J. European Molecular Biology Organization; 1992 May;11(5):1705–10.

300. Vimr ER. Microbial sialidases: does bigger always mean better? Trends Microbiol. 1994 Aug;2(8):271–7.

301. Crennell S, Garman E, Laver G, Vimr E, Taylor G. Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. Structure. 1994 Jun 15;2(6):535–44.

302. Affranchino JL, Ibañez CF, Luquetti AO, Rassi A, Reyes MB, Macina RA, et al. Identification of a *Trypanosoma cruzi* antigen that is shed during the acute phase of Chagas' disease. Mol Biochem Parasitol. 1989 May 15;34(3):221–8.

303. Cazzulo JJ, Frasch AC. SAPA/trans-sialidase and cruzipain: two antigens from *Trypanosoma*

cruzi contain immunodominant but enzymatically inactive domains. FASEB J. 1992 Nov;6(14):3259–64.

304. Campetella OE, Uttaro AD, Parodi AJ, Frasch AC. A recombinant *Trypanosoma cruzi* trans-sialidase lacking the amino acid repeats retains the enzymatic activity. Mol Biochem Parasitol. 1994 Apr;64(2):337–40.

305. Schenkman S, Chaves LB, Pontes de Carvalho LC, Eichinger D. A proteolytic fragment of *Trypanosoma cruzi* trans-sialidase lacking the carboxyl-terminal domain is active, monomeric, and generates antibodies that inhibit enzymatic activity. J Biol Chem. 1994 Mar 18;269(11):7970–5.

306. Buscaglia CA, Campetella O, Leguizamon MS, Frasch AC. The repetitive domain of *Trypanosoma cruzi* trans-sialidase enhances the immune response against the catalytic domain. J Infect Dis. 1998 Feb;177(2):431–6.

307. Crennell SJ, Garman EF, Laver WG, Vimr ER, Taylor GL. Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. Proc Natl Acad Sci USA. 1993 Nov 1;90(21):9852–6.

308. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure. 1995 Nov 15;3(11):1197–205.

309. Smith LE, Eichinger D. Directed mutagenesis of the *Trypanosoma cruzi* trans-sialidase enzyme identifies two domains involved in its sialyltransferase activity. Glycobiology. 1997 Apr;7(3):445–51.

310. Smith LE, Uemura H, Eichinger D. Isolation and expression of an open reading frame encoding sialidase from *Trypanosoma rangeli*. Mol Biochem Parasitol. 1996 Jul;79(1):21–33.

311. Buschiazzo A, Amaya MF, Cremona ML, Frasch AC, Alzari PM. The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. Mol Cell. 2002 Oct;10(4):757–68.

312. Buschiazzo A, Tavares GA, Campetella O, Spinelli S, Cremona ML, París G, et al. Structural basis of sialyltransferase activity in trypanosomal sialidases. EMBO J. 2000 Jan 4;19(1):16–24.

313. Amaya MF, Watts AG, Damager I, Wehenkel A, Nguyen T, Buschiazzo A, et al. Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. Structure. 2004 May;12(5):775–84.

314. Delbaere LT, Vandonselaar M, Prasad L, Quail JW, Wilson KS, Dauter Z. Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution. Journal of Molecular Biology. 1993 Apr 5;230(3):950–65.

315. Hester G, Kaku H, Goldstein IJ, Wright CS. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. Nat Struct Biol. 1995 Jun;2(6):472–9.

316. Bourne Y, Roussel A, Frey M, Rougé P, Fontecilla-Camps JC, Cambillau C. Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: fine specificity of the monosaccharide-binding site. Proteins. 1990;8(4):365–76.

317. Wright CS. 2.2 Å resolution structure analysis of two refined *N*-acetylneuraminyl-lactose--wheat germ agglutinin isolectin complexes. Journal of Molecular Biology. 1990 Oct 20;215(4):635–51.

318. Luo Y, Li SC, Li YT, Luo M. The 1.8 Å structures of leech intramolecular trans-sialidase complexes: evidence of its enzymatic mechanism. Journal of Molecular Biology. 1999 Jan 8;285(1):323–32.

319. Montagna GN, Donelson JE, Frasch ACC. Procyclic *Trypanosoma brucei* expresses separate sialidase and trans-sialidase enzymes on its surface membrane. Journal of Biological Chemistry. 2006 Nov 3;281(45):33949–58.

320. Amaya MF, Buschiazzo A, Nguyen T, Alzari PM. The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T.cruzi* trans-sialidase. Journal of Molecular Biology. 2003 Jan;325(4):773–84.

321. Ly HD, Withers SG. Mutagenesis of glycosidases. Annu Rev Biochem. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1999;68(1):487–522.

322. Burmeister WP, Henrissat B, Bosso C, Cusack S, Ruigrok RW. Influenza B virus neuraminidase can synthesize its own inhibitor. Structure. 1993 Sep 15;1(1):19–26.

323. Albouz-Abo S, Turton R, Wilson JC, Itzstein von M. An investigation of the activity of recombinant rat skeletal muscle cytosolic sialidase. FEBS Lett. Elsevier; 2005 Feb 14;579(5):1034–8.

324. Watts AG, Damager I, Amaya ML, Buschiazzo A, Alzari P, Frasch AC, et al. *Trypanosoma cruzi* trans-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. J Am Chem Soc. 2003 Jun;125(25):7532–3.

325. Wicki J, Rose DR, Withers SG. Trapping covalent intermediates on β-glycosidases. Meth Enzymol. 2002;354:84–105.

326. París G, Cremona ML, Amaya MF, Buschiazzo A, Giambiagi S, Frasch AC, et al. Probing molecular function of trypanosomal sialidases: single point mutations can change substrate specificity and increase hydrolytic activity. Glycobiology. 2001 Apr;11(4):305–11.

327. Paris G, Ratier L, Amaya MF, Nguyen T, Alzari PM, Frasch ACC. A sialidase mutant displaying trans-sialidase activity. Journal of Molecular Biology. 2005 Jan 28;345(4):923–34.

328. Demir O, Roitberg AE. Modulation of catalytic function by differential plasticity of the active site: case study of *Trypanosoma cruzi* trans-sialidase and *Trypanosoma rangeli* sialidase †. Biochemistry. 2009 Apr 21;48(15):3398–406.

329. Mitchell FL, Neres J, Ramraj A, Raju RK, Hillier IH, Vincent MA, et al. Insights into the activity and specificity of *Trypanosoma cruzi* trans-sialidase from molecular dynamics simulations. Biochemistry. 2013 May 28;52(21):3740–51.

330. Pierdominici-Sottile G, Palma J, Roitberg AE. Free-energy computations identify the mutations required to confer trans-sialidase activity into *Trypanosoma rangeli* sialidase. Proteins. 2013 Oct 17;82(3):424–35.

331. Oliveira IA, Goncalves AS, Neves JL, Itzstein von M, Todeschini AR. Evidence of ternary complex formation in *Trypanosoma cruzi* trans-sialidase catalysis. Journal of Biological Chemistry. 2014 Jan 3;289(1):423–36.

332. Barrett MP. The rise and fall of sleeping sickness. Lancet. Elsevier; 2006 Apr 29;367(9520):1377–8.

333. Schofield CJ, Jannin J, Salvatella R. The future of Chagas' disease control. Trends Parasitol. Elsevier; 2006 Dec;22(12):583–8.

334. Neres J, Bryce RA, Douglas KT. Rational drug design in parasitology: trans-sialidase as a case study for Chagas' disease. Drug Discov Today. 2008 Feb;13(3-4):110–7.

335. Neres J, Brewer ML, Ratier L, Botti H, Buschiazzo A, Edwards PN, et al. Discovery of novel inhibitors of *Trypanosoma cruzi* trans-sialidase from *in silico* screening. Bioorg Med Chem Lett. 2009 Feb 1;19(3):589–96.

336. Janakiraman MN, White CL, Laver WG, Air GM, Luo M. Structure of influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dehydro analog at 1.8-Å resolution: implications for the catalytic mechanism. Biochemistry. 1994 Jul

12;33(27):8172–9.

337. Neres J, Bonnet P, Edwards PN, Kotian PL, Buschiazzo A, Alzari PM, et al. Benzoic acid and pyridine derivatives as inhibitors of *Trypanosoma cruzi* trans-sialidase. Bioorganic & Medicinal Chemistry. 2007 Mar 1;15(5):2106–19.

338. Streicher H, Busse H. Building a successful structural motif into sialylmimetics-cyclohexenephosphonate monoesters as pseudo-sialosides with promising inhibitory properties. Bioorganic & Medicinal Chemistry. 2006 Feb 15;14(4):1047–57.

339. Buchini S, Buschiazzo A, Withers SG. A new generation of specific *Trypanosoma cruzi* trans-sialidase inhibitors. Angew Chem Int Ed Engl. WILEY-VCH Verlag; 2008;47(14):2700–3.

340. Agustí R, Giorgi ME, Mendoza VM, Gallo-Rodriguez C, de Lederkremer RM. Comparative rates of sialylation by recombinant trans-sialidase and inhibitor properties of synthetic oligosaccharides from *Trypanosoma cruzi* mucins-containing galactofuranose and galactopyranose. Bioorganic & Medicinal Chemistry. 2007 Apr 1;15(7):2611–6.

341. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. Angew Chem Int Ed Engl. WILEY-VCH Verlag GmbH; 2002 Jul 15;41(14):2596–9.

342. Lee LV, Mitchell ML, Huang S-J, Fokin VV, Sharpless KB, Wong C-H. A potent and highly selective inhibitor of human α-1,3-fucosyltransferase via click chemistry. J Am Chem Soc. American Chemical Society; 2003 Aug 13;125(32):9588–9.

343. Rossi LL, Basu A. Glycosidase inhibition by 1-glycosyl-4-phenyl triazoles. Bioorg Med Chem Lett. 2005 Aug 1;15(15):3596–9.

344. Zhou Y, Zhao Y, O' Boyle KM, Murphy PV. Hybrid angiogenesis inhibitors: synthesis and biological evaluation of bifunctional compounds based on 1-deoxynojirimycin and aryl-1,2,3-triazoles. Bioorg Med Chem Lett. 2008 Feb 1;18(3):954–8.

345. Li J, Zheng M, Tang W, He P-L, Zhu W, Li T, et al. Syntheses of triazole-modified zanamivir analogues via click chemistry and anti-AIV activities. Bioorg Med Chem Lett. 2006 Oct 1;16(19):5009–13.

346. Weïwer M, Chen C-C, Kemp MM, Linhardt RJ. Synthesis and biological evaluation of non-hydrolizable 1,2,3-triazole linked sialic acid derivatives as neuraminidase inhibitors. European J Org Chem. 2009 Jun;2009(16):1–20.

347. Carvalho I, Andrade P, Campo VL, Guedes PMM, Sesti-Costa R, Silva JS, et al. "Click chemistry" synthesis of a library of 1,2,3-triazole-substituted galactose derivatives and their evaluation against *Trypanosoma cruzi* and its cell surface trans-sialidase. Bioorganic & Medicinal Chemistry. 2010 Apr 1;18(7):2412–27.

348. Campo VL, Sesti-Costa R, Carneiro ZA, Silva JS, Schenkman S, Carvalho I. Design, synthesis and the effect of 1,2,3-triazole sialylmimetic neoglycoconjugates on *Trypanosoma cruzi* and its cell surface trans-sialidase. Bioorganic & Medicinal Chemistry. 2012 Jan 1;20(1):145–56.

349. Kim JH, Ryu HW, Shim JH, Park KH, Withers SG. Development of new and selective *Trypanosoma cruzi* trans-sialidase inhibitors from sulfonamide chalcones and their derivatives. Chembiochem. WILEY-VCH Verlag; 2009 Oct 12;10(15):2475–9.

350. Harrison JA, Kartha KPR, Fournier EJL, Lowary TL, Malet C, Nilsson UJ, et al. Probing the acceptor substrate binding site of *Trypanosoma cruzi* trans-sialidase with systematically modified substrates and glycoside libraries. Org Biomol Chem. The Royal Society of Chemistry; 2011 Mar 7;9(5):1653–60.

351. Arioka S, Sakagami M, Uematsu R, Yamaguchi H, Togame H, Takemoto H, et al. Potent inhibitor scaffold against *Trypanosoma cruzi* trans-sialidase. Bioorganic & Medicinal Chemistry. 2010 Feb 15;18(4):1633–40.

2.
Objectives

2 Objectives

Trans-sialidases are unique enzymes, expressed on trypanosomal cell surfaces during the entire life cycle, and utilised to protect the parasite from immune response, as well as to generate receptor molecules for host cell interactions. Due to its essential role in pathogenesis and parasite survival in mammalian host and insect vector, research on trypanosomal TS has increased over the past decades, mainly focusing on TcruTS from the Latin American parasite *T. cruzi*. However, only little is known about the orthologous TS from the African parasite *T. congolense*.

The main aim of this study was to identify and biochemically characterise trans-sialidases from *T. congolense* to get further insights into i) enzymatic activity, ii) structure-function relationships, iii) synthetic and natural substrate affinity and specificity, iv) as well as its biological function also in comparison to other already well-described trypanosomal TS. Furthermore, additional biochemical data of TconTS should provide new strategies for the synthesis of novel specific and highly potent trypanosomal TS inhibitors. The specific objectives in these four sections were:

i. Enzymatic activity of TconTS: To study enzyme activities in detail, it is necessary to express the proteins encoded by different TS-like genes from *T. congolense* genome available at the Welcome Trust Sanger Institute database. For this, all variants were cloned and expressed as recombinant proteins in fibroblasts. Using well-established standard reaction conditions and fetuin as highly sialylated donor substrate, reaction product analysis employing a carbohydrate optimised HPLC-based method, was used to provide comparable data on enzymatic activities. In addition, a strategy to swap and recombine catalytic (CDs) and lectin-like domains (LDs) from different TconTS was established and chimeric recombinant enzymes were expressed by fibroblasts and bacteria to investigate the influence of LD on enzymatic activities.

ii. Structure-function relations: Since there is no crystal structure or NMR resolved structure of TconTS available so far, homology modelling using the published crystal structure of TcruTS as template was used to generate TconTS structure models based on amino acid sequence alignment of all enzymes. Together with the data obtained on TconTS enzyme activities these models should not only provide further insights into structure-function relations and enhance understanding of TS function, but also be helpful in respect to future TconTS applications as catalysts for sialylation reactions.

iii. Substrate affinity and specificity: Besides the selected sialyl-donor substrates 3'SL and fetuin for TS reactions also more natural substrates available in the mammalian

host as well as the parasites' midgut environment, such as serum glycoproteins, have been chosen to investigate enzymatic activities of the different TconTS variants, mimicking a more native system. On the other hand, TconTS activities were also be investigated under conditions related to tsetse vector environment. Since *T. congolense* migrates through different tissues with different environmental conditions during its developmental life cycle in tsetse vector, which is e.g. characterised by a pH shift, it was important to investigate TconTS pH dependency. Results should provide new insight into enzyme catalysis and support synthesis of more effective, stage specific inhibitors and/or the establishment of other anti-trypanosomal therapeutic strategies.

iv. Biological function: Many studies have been focused on the structural topology of the catalytic domain (CD) and the active centre of TS and sialidase, in respect to biological function and catalytic mechanism. However, only very little is known about TS lectin-like domains (LD), for which no functional data and/or biological relevance have been described so far. Therefore, it was of fundamental interest to biochemically characterise TconTS-LD function by cloning and expressing recombinant proteins as model structures. In cooperation with Dr. Joe Tiralongo and Dr. Thomas Haselhorst (Institute for Glycomics, Griffith University, Goldcoast Australia) carbohydrate binding affinities and specificities of several TconTS-LDs were determined, utilising a number of techniques specific to investigate protein-ligand interactions, such as SPR, glycan array and STD NMR. Additionally, establishment of a microtitre-based TconTS-LD binding/inhibition assay, as well as other TS activity studies regarding the influence of LD on enzyme activity provided further essential data in respect to the biological function of these unusual enzymes.

3.
Results

3.1 Results

Biochemical characterisation of trans-sialidase TS1 variants from *Trypanosoma congolense*

Hendrik Koliwer-Brandl, Thaddeus T Gbem, Mario Waespy, Olga Reichert, Phillip Mandel, Eric Drebitz, Frank Dietz and Sørge Kelm (2011)

BMC Biochemistry, ***12****, 39.*

(doi:10.1093/bioinformatics/18.suppl_2.S153)

Contribution of Mario Waespy:

- Purification of TconTS1g
- TconTS1g enzyme assays
- Evaluation of TconTS1g enzymatic activity data obtained from HPAEC-PAD analysis

Koliwer-Brandl *et al. BMC Biochemistry* 2011, **12**:39
http://www.biomedcentral.com/1471-2091/12/39

BMC Biochemistry

RESEARCH ARTICLE Open Access

Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*

Hendrik Koliwer-Brandl, Thaddeus T Gbem, Mario Waespy, Olga Reichert, Philipp Mandel, Eric Drebitz, Frank Dietz and Sørge Kelm*

Abstract

Background: Animal African trypanosomiasis, sleeping sickness in humans and Nagana in cattle, is a resurgent disease in Africa caused by *Trypanosoma* parasites. Trans-sialidases expressed by trypanosomes play an important role in the infection cycle of insects and mammals. Whereas trans-sialidases of other trypanosomes like the American *T. cruzi* are well investigated, relatively little research has been done on these enzymes of *T. congolense*.

Results: Based on a partial sequence and an open reading frame in the WTSI database, DNA sequences encoding for eleven *T. congolense* trans-sialidase 1 variants with 96.3% overall amino acid identity were amplified. Trans-sialidase 1 variants were expressed as recombinant proteins, isolated and assayed for trans-sialylation activity. The purified proteins produced α2,3-sialyllactose from lactose by desialylating fetuin, clearly demonstrating their trans-sialidase activity. Using an HPLC-based assay, substrate specificities and kinetic parameters of two variants were characterized in detail indicating differences in substrate specificities for lactose, fetuin and synthetic substrates. Both enzymes were able to sialylate asialofetuin to an extent, which was sufficient to reconstitute binding sites for Siglec-4. A mass spectrometric analysis of the sialylation pattern of glycopeptides from fetuin revealed clear but generally similar changes in the sialylation pattern of the *N*-glycans on fetuin catalyzed by the trans-sialidases investigated.

Conclusions: The identification and characterization of a trans-sialidase gene family of the African parasite *T. congolense* has opened new perspectives for investigating the biological role of these enzymes in Nagana and sleeping sickness. Based on this study it will be interesting to address the expression pattern of these genes and their activities in the different stages of the parasite in its infection cycle. Furthermore, these trans-sialidases have the biotechnological potential to be used for enzymatic modification of sialylated glycoconjugates.

Background

Animal African trypanosomiasis, called sleeping sickness in humans and Nagana in cattle, is a resurgent disease in Africa. Nagana is caused by *Trypanosoma congolense* (*T. congolense*), *Trypanosoma vivax* (*T. vivax*) and *Trypanosoma brucei* (*T. brucei*) subspecies. Most research on African trypanosomes has focused on *T. brucei*, whereas only few studies have been done with other African trypanosomes including *T. congolense*. In wild animals, these parasites cause relatively mild infections while in domestic animals they cause a severe, often fatal disease. Because of Nagana, stock farming is very difficult within the tsetse belt of Africa [1].

Although of crucial importance for their survival, cyclical transmission and hence pathogenicity of trypanosomes, trypanosomes lack the biochemical metabolic machinery synthesizing sialic acids (Sia), but use a unique enzyme, trans-sialidase (TS) to transfer Sia onto the parasites surface molecules from the environment. Structurally TS belong to the family of sialidases (SA). In contrast to the usual sialyltransferases, TS does not utilize CMP-activated Sia as monosaccharide donors, but catalyzes the transfer of carbohydrate-linked Sia to another glycan forming a new α2,3-glycosidic linkage to galactose or *N*-acetylgalactosamine.

Whereas more detailed studies exist on the role of TS in the pathogenicity of *T. cruzi*, the etiologic agent of

* Correspondence: skelm@uni-bremen.de
Centre for Biomolecular Interactions Bremen, Department of Biology and Chemistry, University of Bremen, Leobener Str./NW2/B2235, 28359 Bremen, Germany

Chagas diseases in South America, where TS was first discovered [2], the current knowledge about the corresponding enzymes in the African trypanosomes is very limited. Of all the African trypanosomes, only *T. brucei* full length TS genes have been cloned and studied [3]. Sialylation of parasite surfaces is believed to protect the parasites from the action of glycolytic enzymes as well as from immunocompetent substances present in the tsetse gut and blood meal respectively, as well as influencing the interaction of parasites with the gut epithelial cells [4-6]. In the African trypanosomes where TS is thought to be expressed only in the procyclic insect stages [5,7], TS has been shown to increase the survival, maturation and hence establishment of the parasites in the vector midgut [8].

Two TS forms, named TS-form 1 and TS-form 2, have been purified from procyclic *T. congolense* cultures [6]. Interestingly, glutamic acid and alanine-rich protein (GARP) was co-purified with TS-form 1, suggesting that GARP may be a natural substrate for TS-form 1. Interestingly, TS-form 1 had significantly less SA activity and higher TS activity, whereas SA activity was predominately found in preparations of TS-form 2. An anti-*T. congolense* TS antibody (mAb 7/23) was developed using TS form 1 as antigen. This antibody is specific for *T. congolense* TS recognizing TS-form 1 and TS-form 2, but does not bind to *T. brucei* TS. Peptides micro sequencing revealed the amino acid sequence VVDPTVVAK in TS-form 1. Subsequently, fragments of two TS genes (TS1 and TS2) were sequenced, sharing about 50% sequence identity [9]. TS1 encoded this peptide sequence, whereas in TS2 this sequence ended in VVK. These data strongly suggested that the gene product of TS1 has been present in TS-form 1. Nevertheless, it has remained unclear whether only TS1 and/or TS2 gene products were present in TS-form 1 and TS-form 2 preparations. Due to the fact that the monoclonal antibody mAb 7/23 bound both TS preparations, it is quite possible that TS-form 2 contained at least some amounts of TS1 gene product, which might have been responsible for the TS activity of this preparation.

Here, we report the cloning of eleven trans-sialidase TS1 variants from *T. congolense* and their recombinant expression in CHO_{Lec1} cells. Furthermore, the enzymatic properties of two of these recombinant TS1 variants were compared with TS from *T. cruzi*.

Results

Diversity of TS1 genes and structural model

Based on the partial sequence of TS1 [GenBank: AJ535487.1] [9], an open reading frame in the Welcome Trust Sanger Institute (WTSI) database was identified. The full-length translation product consists of 750 amino acids extending the partial sequence of TS1 by 153 amino acids at the N-terminus and by 84 amino acids at the C-terminus. It contains a 16 amino acids N-terminal signal peptide and a catalytic domain (residues 17-467), which is connected through a long α helix (residues 468-491) to a lectin domain (residues 492-732) followed by a potential C-terminal GPI-anchor attachment site (residues 733-750, identified by big-PI predictor [10]). Furthermore, nine potential *N*-glycosylation sites were identified (Figure 1).

T. congolense TS1 shares about 57% sequence identity with *T. brucei* TS [EMBL: AAG32055.1] and 48% with *T. cruzi* TS [EMBL: BAA09334.1] (Figure 2). The *T. brucei* TS has a prolonged N-terminus of approx. 90 amino acids, which is conserved in *T. congolense* TS1 sharing 50% amino acids, but is absent in *T. cruzi* TS. The catalytic domain of both African proteins has 60% and the lectin domain 43% sequence similarities. *T. congolense* TS1, like *T. brucei* TS and *T. rangeli* SA, has no C-terminal SAPA domain typical for *T. cruzi* TS [11]. Almost all amino acid residues reported to be required for TS activity are identical in TS1 with the exception of A325 (corresponding to P231 in *T. cruzi* TS) [11], R127, G344-Q346 and Y408 (corresponding to Y248 and W312, respectively in *T. cruzi* TS) [12] (Figure 2).

To produce recombinant protein for enzyme characterization, the DNA encoding amino acids 17-732 was amplified using genomic *T. congolense* DNA as a template and inserted into a mammalian expression vector as described in Methods. 13 clones were picked from two independent cloning experiments and sequenced. Interestingly, not all the 13 clones had identical sequences and eleven different sequences were obtained (TS1a through TS1j), exhibiting an overall amino acid identity of 96.3%. A more detailed search of the WTSI database using these sequences as queries confirmed the presence of these TS1 genes in the *T. congolense* genome database. As shown in Figure 3, changes are not evenly distributed over the protein sequences. Eight were found in the lectin domain and 17 in the catalytic domain, some close to the predicted active site as shown in Figure 4A.

For a better understanding of how these differences may affect TS function, we calculated a model structure (Figure 4) for TS1 e-1 by homology modeling based on the crystal structure of *T. cruzi* TS [12], which was complexed with the Sia derivative 3-fluoro-5-*N*-acetyl-9-benzamido-2,9-dideoxy-neuraminic acid. The superimposed structures of *T. cruzi* TS and the *T. congolense* TS1 e-1 model had a root mean square deviation (RMSD) of 0.685 Å over 594 aligned residues.

In Figure 4A, amino acids of the active site are highlighted. Most of the amino acids reported to be relevant for TS activity are identical in all *T. congolense* TS1 variants (white labels). However, differences to *T. cruzi* TS

Figure 1 Primary sequence of TS1a. The full length coding domain sequence [EMBL: HE583283] with corresponding amino acid translations is shown. The recombinant protein was generated without the N-terminal signal peptide and without the C-terminal region predicted to be replaced by a GPI anchor in the native protein (framed boxes). Predicted *N*-glycosylation sites are highlighted by light grey boxes.

were identified at three positions (yellow labels in Figure 4A). (I) At position 325 all *T. congolense* TS1 variants have an alanine, like in *T. brucei* TS, replacing a proline occurring in *T. cruzi* TS (P231); (II) Y408 of all *T. congolense* TS1 variants corresponds to a tryptophan in *T. cruzi* TS (W321) and *T. brucei* TS; (III) the group of G342, G343 and Q344 replaces a tyrosine (Y248) in *T. cruzi* TS. In addition, near the catalytic site at position 407 (red label) in *T. congolense* TS1 variants, a serine or valine occurs instead of arginine (R311) in *T. cruzi* TS. Interestingly, similar differences occur also in *T. brucei* TS (Figure 2). Since these amino acids are close to the active site, they could influence the acceptor binding specificity. The arginine at position 144 (blue label) is conserved in all TS, with the exception of *T. congolense* TS1g, where it is a cysteine.

In Figure 4B the amino acid positions are highlighted, which have different side chains in TS1a-TS1j (Figure 3). It should be noted that these are all on the same side of the protein as the catalytic site. Striking is a cluster of amino acid variations in the lectin domain (position 599 to 602 and 643) suggesting that these changes may influence substrate binding of larger substrate molecules, such as glycoproteins.

Characterization of *T. congolense* TS1 enzyme activity

All eleven TS1 gene products (TS1a-TS1j) were expressed as recombinant proteins and were recognized by the anti- *T. congolense* TS antibody (mAb 7/23) [6] (data not shown). For all TS1 variants similar robust TS activity could be determined, except for TS1g. This variant, which carries cysteine instead of arginine at position 144, had only very low TS activity. However, in contrast to the other variants, TS1g released free Sia from fetuin at about 50% of the transfer to lactose. Two of the *T. congolense* TS1 variants, TS1b and TS1 e-1, were further characterized. They differ in eleven of the total 25 positions with amino acid variations listed in Figure 3, three in the catalytic domain and eight in the lectin domain.

The donor substrates fetuin 3'SL or pNP-Neu5Ac and the acceptor substrates lactose, galactose or Gal-MU were employed to determine sialidase and trans-sialidase activities. For this purpose, a new assay was established

Koliwer-Brandl *et al. BMC Biochemistry* 2011, **12**:39
http://www.biomedcentral.com/1471-2091/12/39

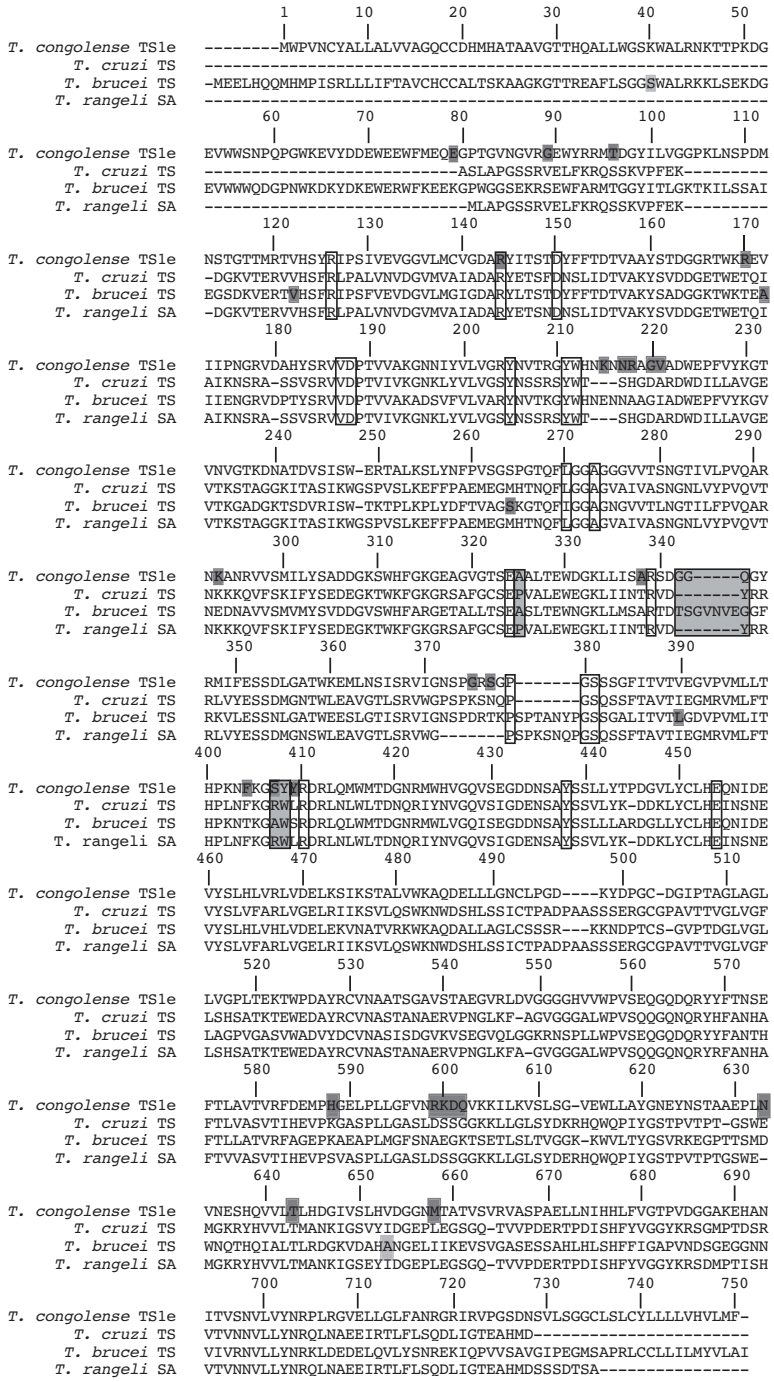


Figure 2 Primary sequence alignment of trypanosomal trans-sialidases. The amino acid sequence of *T. congolense* [EMBL: HE583287] TS1 e-1 was aligned with *T. cruzi* TS [EMBL: BAA09334.1, PDB: 3B69], *T. brucei* TS [EMBL: AAG32055.1] and *T. rangeli* SA [EMBL: AAC95493.1] based on a structural alignment of *T. congolense* TS1 e-1 with *T. cruzi* TS generated with Yasara during homology modeling. Amino acids, which have been proposed to be relevant for enzymatic activity are marked with black frames if conserved or with black frames and light grey background if not conserved. Positions with variations occurring in *T. congolense* TS1 or *T. brucei* TS [3] are highlighted by dark grey boxes.

Koliwer-Brandl *et al. BMC Biochemistry* 2011, **12**:39
http://www.biomedcentral.com/1471-2091/12/39

position	TS1a	TS1b	TS1c	TS1d	TS1e	TS1e'	TS1f	TS1g	TS1h	TS1i	TS1j
Catalytic domain											
79	E	K	E	E	E	E	E	E	E	E	E
89	G	T	G	G	G	G	G	G	A	G	G
96	K	K	T	T	T	K	K	K	K	K	T
144	R	R	R	R	R	R	R	C	R	R	R
170	R	R	R	S	R	R	S	R	R	R	R
215	K	K	R	Q	K	K	Q	K	K	Q	Q
217	N	N	D	D	N	N	D	N	N	D	D
218	R	R	K	E	R	R	E	R	R	E	E
220	G	G	A	G	G	G	A	G	G	G	G
221	V	V	I	V	V	V	I	V	V	V	V
293	K	K	K	K	K	K	K	K	K	K	T
338	A	A	A	A	A	A	T	A	A	A	A
375	G	G	K	G	G	G	P	K	K	K	S
377	S	S	N	S	S	S	S	N	N	N	S
404	L	L	L	L	F	L	L	L	L	L	L
407	S	S	S	V	S	S	S	S	S	S	S
409	Y	Y	Y	S	Y	Y	Y	Y	Y	Y	Y
Lectin domain											
588	Q	Q	Q	Q	H	H	H	Q	R	R	R
599	R	R	R	R	R	R	R	R	H	R	R
600	E	E	E	E	K	K	K	K	K	K	K
601	G	G	G	G	D	D	D	G	G	G	G
602	K	K	K	K	Q	Q	Q	K	K	K	K
633	D	D	D	D	N	N	N	N	E	D	D
643	A	A	A	A	T	T	T	T	A	A	A
658	T	T	T	T	M	M	M	T	T	M	T

Figure 3 Amino acid variations found in *T. congolense* TS1a-TS1j. TS1a [EMBL: HE583283], TS1b [EMBL: HE583284], TS1c [EMBL: HE583285], TS1d [EMBL: HE583286], TS1 e-1 [EMBL: HE583287] , TS1 e-2 [EMBL: HE583288], TS1f [EMBL: HE583289], TS1g [EMBL: HE583290], TS1h [EMBL: HE583291], TS1i [EMBL: HE583292], TS1j [EMBL: HE583293]. Differences in amino acids are highlighted (light grey: conservative; dark grey: modest; black: drastic change).

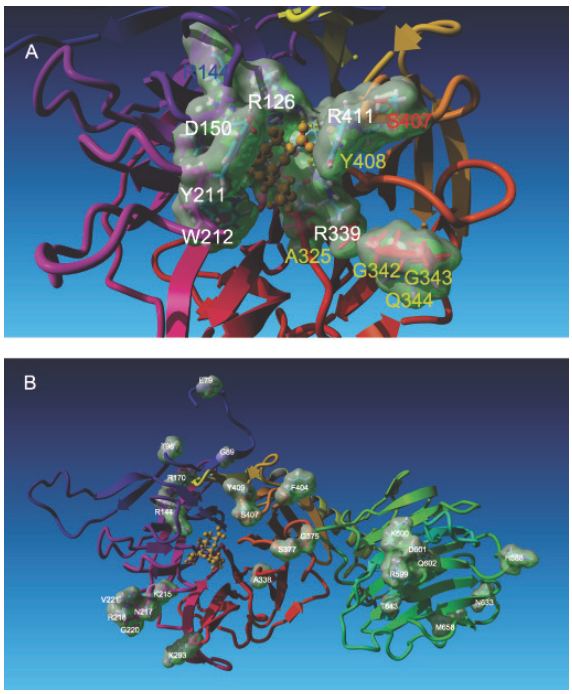


Figure 4 Homology model of *T. congolense* TS1. The crystal structure of *T. cruzi* TS [12] in complex with 3-fluoro-5-*N*-acetyl-9-benzamido-2,9-dideoxy-neuraminic acid was used as template to calculate a model structure for *T. congolense* TS1 e-1. Only the Neu5Ac part (orange) of the ligand in the binding site of the complex is illustrated. (A) Conserved amino acids of the active site are labeled in white. Amino acids at positions reported to be important for ligand binding in *T. cruzi* TS [12], which are not conserved in *T. congolense* are labeled in yellow. The red labeled position 407 is a serine or valine in *T. congolense* TS1 forms. R144, which is exchanged to a cysteine in TS1g, is labeled in blue. (B) Green clouds mark positions at which different amino acids occur in other *T. congolense* TS1 forms.

as described under Methods, using HPAEC-PAD to quantify sialylated oligosaccharide products with the detection limit of 20 pmol 3'SL corresponding to 0.5 μM in the reaction mixture. In standard assays, 50 μL TS reactions were set up with 50 ng TS1b or TS1 e-1, 100 μg fetuin (approx. 600 μM bound Sia) as donor substrate and 100 nmol acceptor substrate (2 mM e. g. lactose or galactose). Under these conditions, linear product formation was obtained for up to 2500 pmol corresponding to 50 μM 3'SL (Figure 5).

If lactose is used as a donor substrate under standard conditions, 3'SL concentration increases linearly for about 30 minutes before the reaction velocity started to decrease (Figure 5A). It should be noted that lactose was sialylated almost twice as fast by TS1b than by TS1 e-1. In contrast to lactose, galactose was sialylated at the same rate by both enzymes, but at about 20-fold lower velocity than lactose.

Different specific activities were obtained for *T. congolense* TS1b, TS1 e-1 and *T. cruzi* TS (Figure 6). The reaction velocity was linearly dependent on the amount of TS as long as the concentration of the product 3'SL was below 50 μM. Under standard conditions 50 μM 3'SL was produced in 30 minutes with 50 ng TS. If 200 ng TS or more were used, product formation was independent of the amount of TS, probably due to the increased use of 3'SL as a donor substrate in the reverse reaction, finally leading to an equilibrium between lactose, 3'SL, sialylated and desialylated glycans on fetuin. This equilibrium apparently was reached in 30 minutes with 500 ng TS (266 ± 4 μM 3'SL for *T. cruzi* TS, 194 ± 6 μM 3'SL for TS1b and 165 ± 7 μM 3'SL for TS1 e-1). After 20 h incubation, 50 ng TS was sufficient to reach the equilibrium. Independent of the amount of enzyme used, for all three TS applied similar final concentrations of 3'SL were obtained after 20 h incubation (Table 1).

The HPAEC-PAD method used allowed not only determining the TS, but also SA activity, since free Sia and 3' SL could be quantified from the same chromatogram. In standard reactions (50 ng TS, 30 min incubation time) no SA activity could be detected, both in the presence or absence of lactose as an acceptor substrate. This suggests that these TS1 variants usually need an acceptor substrate like lactose to cleave Sia from a donor substrate. However, after 20 h incubation, free Sia was detected. The quantity of Sia released was dependent on the amount of TS used (Table 1). Besides standard TS reactions with fetuin as donor and lactose as acceptor substrate, TS reactions with 2 mM 3'SL as donor and 2 mg/mL ASF as acceptor substrate were performed. In these reactions, free Sia was detected after short reaction times and after incubation for 24 h, 0.5-1 mM free Sia were produced (data not shown).

Koliwer-Brandl *et al. BMC Biochemistry* 2011, **12**:39
http://www.biomedcentral.com/1471-2091/12/39

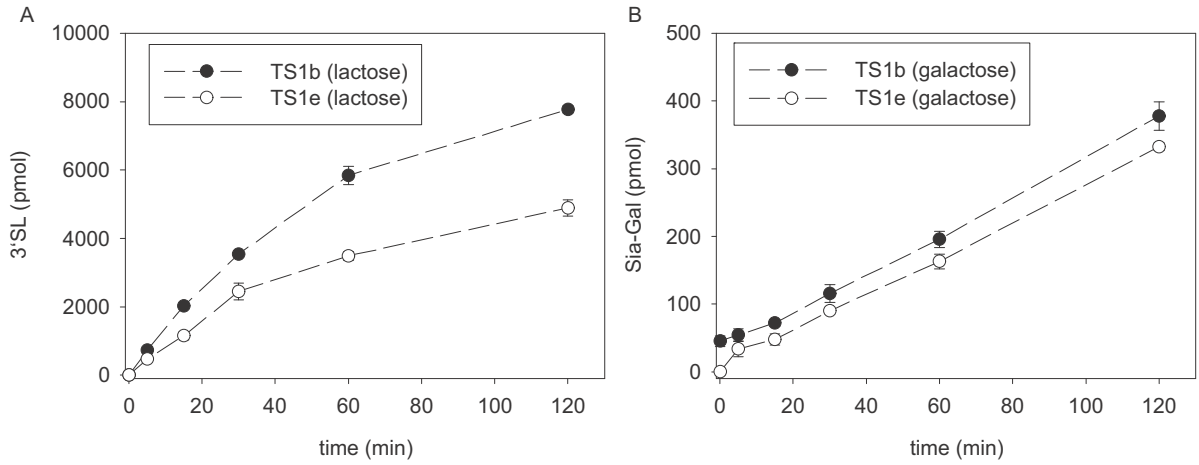


Figure 5 Time dependence of TS reactions. Reactions of 100 µg fetuin with 100 nmol lactose (A) or 100 nmol galactose (B) were started with 50 ng TS in 50 µL and incubated at 37°C. The amount of 3'SL produced was determined by HPAEC-PAD as described under Methods. Data points are means ± standard deviations of three replicates.

For kinetic experiments, assays were incubated for 30 minutes using 50 ng TS, since under these conditions 3'SL production was linear for all three TS. To determine the kinetic parameters for the acceptor substrates lactose (Figure 7A) or Gal-MU (Figure 7C), 100 µg fetuin (600 µM bound Sia) was used as donor substrate. The lowest K_M for lactose was found for *T. cruzi* TS with 327 µM compared to 1683 µM for TS1b and 727 µM for TS1 e-1 (Table 2). Furthermore, *T. cruzi* TS was able to produce twice more 3'SL than TS1b and fourfold more than TS1 e-1 under these conditions.

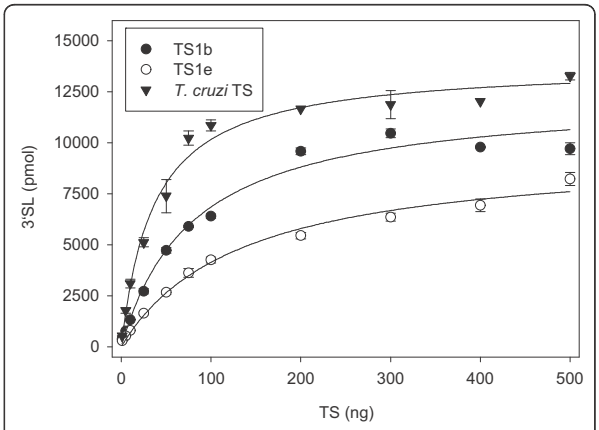


Figure 6 Concentration dependence of TS reactions. Reactions of 100 µg fetuin with 100 nmol lactose were started with varying amounts of TS in overall 50 µL 10 mM phosphate buffer, pH 7.4 and incubated at 37°C for 30 min. The amount of 3'SL produced was determined by HPAEC-PAD as described under Methods. Data points are means ± standard deviations of three determinations.

To determine the kinetic parameters for the donor substrate fetuin (Figure 7B), 2 mM lactose was used as acceptor substrate. Both *T. congolense* TS1 had similar v_{max}-values, whereas the v_{max} for *T. cruzi* TS was about fivefold higher. Different to the K_M of lactose, the lowest K_M for fetuin was exhibited by TS1b with 359 µM, which is about fivefold lower compared to TS1 e-1 with 1617 µM and about 12-fold lower compared to *T. cruzi* TS with 4124 µM.

Kinetic studies with TS were also performed for the acceptor substrate Gal-MU (Table 2) and the donor substrate pNP-Neu5Ac. Almost similar K_M and v_{max}-values were found for both *T. congolense* TS1. The substrate pNP-Neu5Ac was only weakly used as a donor substrate by all three TS species. Therefore, no reliable kinetic parameters could be determined.

Table 1 Free Sia and 3'SL production at equilibrium of the reaction

	Neu5Ac [pmol]	3'SL [pmol]
0 ng TS	262 ± 46	0
50 ng *T. cruzi* TS	580 ± 8	12597 ± 115
250 ng *T. cruzi* TS	1820 ± 13	11199 ± 288
50 ng TS1b	415 ± 7	15378 ± 117
250 ng TS1b	1095 ± 129	14435 ± 2226
50 ng TS1 e-1	348 ± 83	12719 ± 3057
250 ng TS1 e-1	1186 ± 9	13655 ± 491
only fetuin	116 ± 8	0

Neu5Ac and 3'SL produced by TS after 20 h incubation under standard conditions were determined as described under Material and Methods. Data points are means ± standard deviations of three replicates.

Koliwer-Brandl *et al. BMC Biochemistry* 2011, **12**:39
http://www.biomedcentral.com/1471-2091/12/39

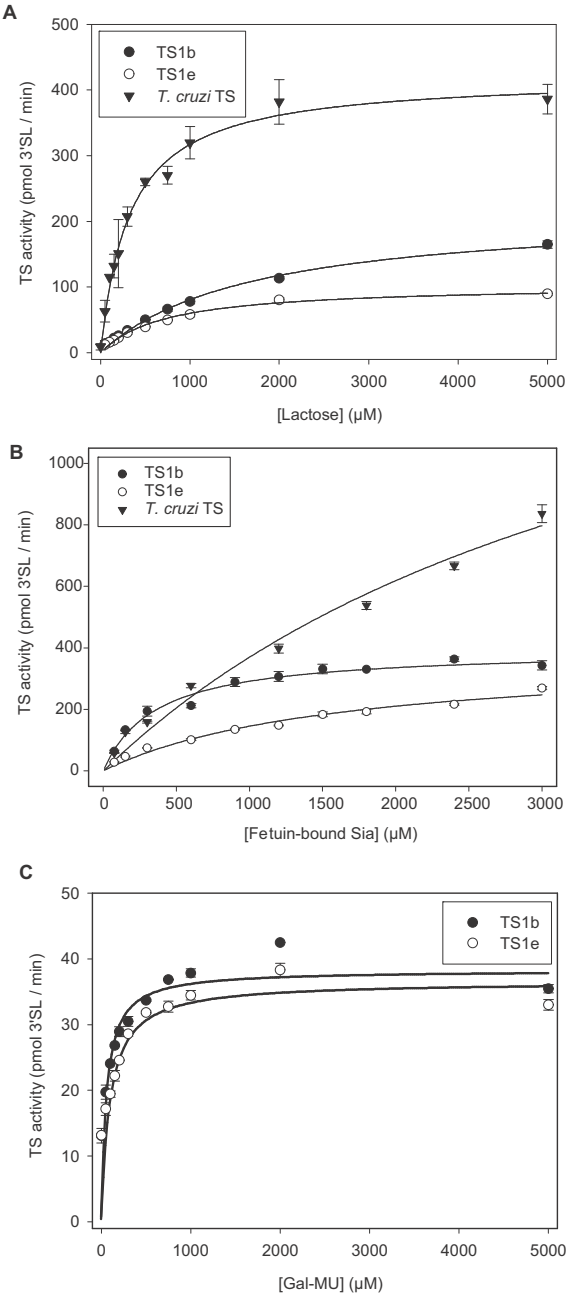


Figure 7 Michaelis-Menten kinetics of TS reactions. All reactions were started with 50 ng TS in a final volume of 50 μL and incubated for 30 min at 37°C. **(A)** 600 μM fetuin-bound Sia was used as donor substrate with varying concentrations of lactose as acceptor substrate. **(B)** Varying concentrations of fetuin-bound Sia were used as donor substrate with 2 mM lactose as acceptor substrate. **(C)** Varying amounts Gal-MU were used as acceptor substrate and 50 nmol 3'SL as donor substrate. The amount of 3'SL (in the case of **(A)** and **(B)**) or 3'Sia-Gal-MU (in the case of **(C)**) produced was determined by HPAEC-PAD as described under Methods. Data points are means ± standard deviations of three replicates.

Sialylation pattern of glycopeptides

Structural differences between TS1 variants may influence the enzymes preference for glycans on glycoproteins, such as fetuin. This could possibly result in different sialylation patterns on glycoproteins after incubation with TS. Fetuin contains 3 *N*-glycosylation sites and 3 *O*-glycans, which all can serve as Sia donors in TS reactions [13]. To investigate the specificity of TS towards different *N*-glycans on fetuin, we used MALDI-TOF-MS to determine the sialylation pattern of glycopeptides (GPs) from trypsin-digested fetuin after incubation with TS and lactose (Figure 8).

The sialylation patterns of three glycopeptides, GP 127-141 (dibranched or tribranched), and GP 54-85 (tribranched) were determined. All 14 potential sialylation variants of these glycopeptides could be identified unambiguously and quantified from the MALDI-TOF-MS spectra. In untreated fetuin most branches on the three *N*-glycans investigated were sialylated, whereas upon treatment with TS and lactose after 30 minutes a clear shift towards incompletely sialylated glycans was observed. After 24 h TS incubation the relative amounts of unsialylated glycans was further increased and monosialylated glycans represented the most prominent species on both, di- and tribranched glycans (Figure 9).

Sialylation of glycoproteins

As described above, *T. congolense* TS1b and TS1 e-1 readily used fetuin as donor substrate for the production of 3'SL. However, long-term TS reactions or experiments with higher amounts of TS had suggested that the reverse reaction also takes place. Therefore, we investigated whether *T. congolense* TS1b and TS1 e-1 can restore sialylated glycans on *Vibrio cholerae* sialidase-treated fetuin (ASF) as model glycoprotein. Resialylation experiments were performed with 100 μg ASF as acceptor and 100 nmol 3'SL as donor substrate as well as 50 ng TS1 in 50 μL to start the TS reaction and were incubated up to 24 h. This resialylation partially reversed the shift in electrophoretic mobility in SDS-PAGE observed for sialidase-treated fetuin (Figure 10). Also by MALDI-MS of glycopeptides, the sialylation of unsialylated glycans was confirmed (data not shown).

Furthermore, we addressed the question of whether through this reaction recognition sites for siglecs can be restored. For this purpose, TS-treated ASF was immobilized to a microtitre plate and used as target for Siglec-4, which preferentially binds α2,3-linked Sia. Under these conditions robust Siglec-4 binding was observed of ASF that had been treated with TS for 4 h. A prolonged (up to 24 h) TS reaction only led to little further increase reaching 40% of binding levels observed with native fetuin (Figure 11).

Koliwer-Brandl *et al. BMC Biochemistry* 2011, **12**:39
http://www.biomedcentral.com/1471-2091/12/39

Table 2 Kinetic parameters of *T. cruzi* TS, *T. congolense* TS1b and TS1 e-1

	Acceptor substrates				Donor substrates	
	Lactose		Gal-MU		Fetuin-bound Sia[x]	
	v_{max} (µmol/(min × mg TS))	K_M (µM)	v_{max} (µmol/(min × mg TS))	K_M (µM)	v_{max} (µmol/(min × mg TS))	K_M (µM)
TS1b	4.3 ± 0.1	1683 ± 101	0.77 ± 0.03	57 ± 14	7.9 ± 0.3	359 ± 45
TS1 e-1	2.1 ± 0.1	727 ± 48	0.72 ± 0.03	74 ± 17	7.6 ± 0.5	1617 ± 223
***T. cruzi* TS**	8.4 ± 0.3	327 ± 31	n.d.	n.d.	37.9 ± 6.0	4124 ± 985

K_M and v_{max} were calculated from Michaelis-Menten kinetics (Figure 7) by SigmaPlot. Data points are means ± standard deviations of three replicates. × Approximately 30 nmol Sia per 100 µg fetuin. Abbr.: n.d.: not determined.

Discussion

In 2003 Tiralongo et al. published a partial sequence for *T. congolense* TS1 [9]. Our approach to obtain the full length sequence of TS1 led to the discovery of 11 variants of this gene with an overall sequence identity of 96.3% in the genome of this parasite. The previously published partial TS1 sequence, which had been assembled from 47 independent PCR reactions, turned out to represent a mixture of fragments from the 11 TS1 forms identified in this study. Therefore, it is likely that that sequence doesn't exist in nature.

Similar TS-like gene families occur also in other trypanosomes. In *T. brucei* the situation appears to be less complex, since only 8 TS closely related genes have been identified [3] and these differences lead to 6 amino acid mutations. The largest TS gene family has been found in *T. cruzi*. Its 140 members fall into three different groups [14]. Blood stream trypomastigotes express two of these groups, one with TS activity and the other only with lectin activity. A third group has TS activity and is expressed by epimastigotes. At present it is unknown whether the expression of the different TS genes in African trypanosomes is also stage-dependent.

T. congolense TS1 shares about 57% identical amino acids with *T. brucei* TS and 48% sequence identity with *T. cruzi* TS. By comparison, the American *T. rangeli* SA and *T. cruzi* TS are more closely related with approximately 70% sequence identity [15].

The homology model for *T. congolense* TS1 based on crystal structures from *T.cruzi* TS and *T. rangeli* SA provided (I) insight in differences in the active site and its surrounding between TS from these parasites and (II) revealed the spatial distribution of the amino acid differences between the TS1 variants.

Compared to *T. cruzi* TS only three amino acids are changed in TS1 (A325, S407 and Y408). At position 325 a proline at the corresponding position of *T. cruzi* TS appears to be relevant for full TS activity [11]. However, in both, *T. brucei* TS and in *T. congolense* TS1, position 325 is an alanine. Tiralongo et al., (2003) postulated A325 might be common to African trypanosomes and does not seem to have an effect on enzymatic activity [9]. Our data have demonstrated that an alanine at this position is well compatible with TS activity, since all active TS1 variants have A325. The exchanges at 407 and 408 may be relevant for the different substrate

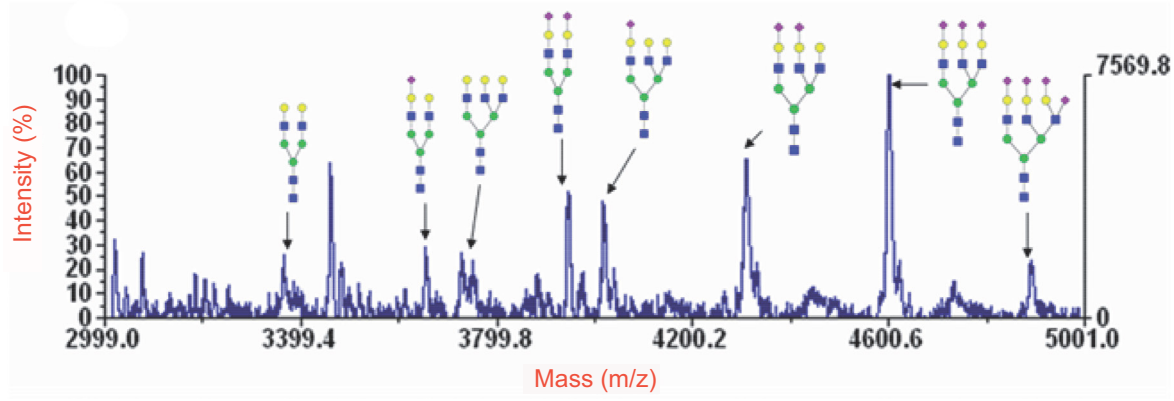


Figure 8 MALDI-TOF-MS of glycopeptide 127-141 sialylation variants. 100 µg fetuin were treated with trypsin and analyzed by MALDI-TOF-MS as described under Methods. The symbols code for the monosaccharide units is according to the nomenclature of the Consortium for Functional Glycomics (http://www.functionalglycomics.org): GlcNAc (blue squares), Man (green circles, Gal (yellow circles), Neu5Ac (purple diamonds).

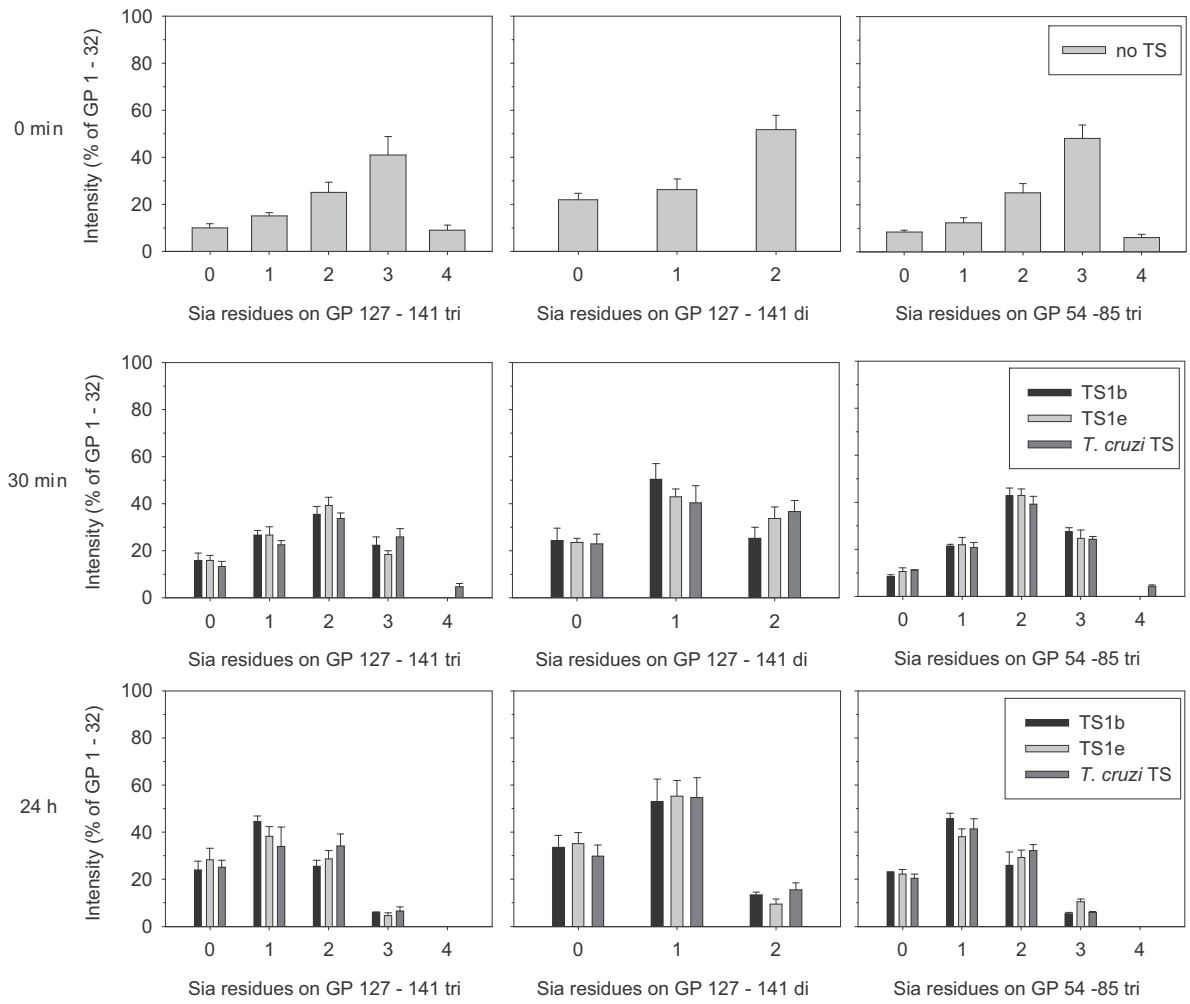


Figure 9 Relative changes of the sialylation pattern of fetuin *N*-glycans. Shown are the sialylation patterns of the glycopeptides GP 127-141 with dibranched (GP 127-141 di) or tribranched (GP 127-141 tri) glycans and of the glycopeptide GP 54-85 with tribranched glycans (GP 54-85 tri) relative to the non-sialylated GP 1 -32. 100 μg fetuin was incubated with 50 ng TS1b, TS1 e-1 or *T. cruzi* TS and 100 nmol lactose for the times indicated, trypsinized and glycopeptides analyzed by MALDI-MS as described under Methods. Data points are averages ± standard deviations of three to ten replicates.

specificities of TS species, since this area possibly participates in the interaction with the underlying galactose.

With the full length coding sequence of TS1 available, the enzymes were expressed and characterized as recombinant proteins to address the question of whether TS1 accounts for the TS activity in the two preparations (TS-form 1 and TS-form 2) from procyclic parasite cultures [6]. The amino acid sequences of the eleven TS1 variants described here contain the peptide VVDPTVVAK present in TS-form 1. Furthermore, all recombinant TS1 variants reacted with the anti-*T. congolense* TS antibody (mAb 7/23) [6]. Based on this information, it is now clear that TS1 was present in both, TS-form1 and TS-form 2. However, two observations made with recombinant TS1 suggest that the native enzyme preparations of TS-form 1 and even more for TS-form 2 contained additional TS1-like proteins. (I) The reaction velocities with the synthetic substrates pNP-Neu5Ac (as donor) and Gal-MU (as acceptor) were much lower than with fetuin and lactose for all three TS. Tiralongo et al., 2003 [6] determined a K_M of 500 μM for Gal-MU, which is approximately 10-fold higher than for TS1b (57 μM) and TS1 e-1 (74 μM). These differences could be due to the presence of other TS1-like enzymes in the preparations of TS-form 1 and TS-form 2. (II) Using the substrate Neu5Ac-MU no SA or TS activity could be detected for the *T. congolense* TS1 variants investigated here, whereas cleavage of

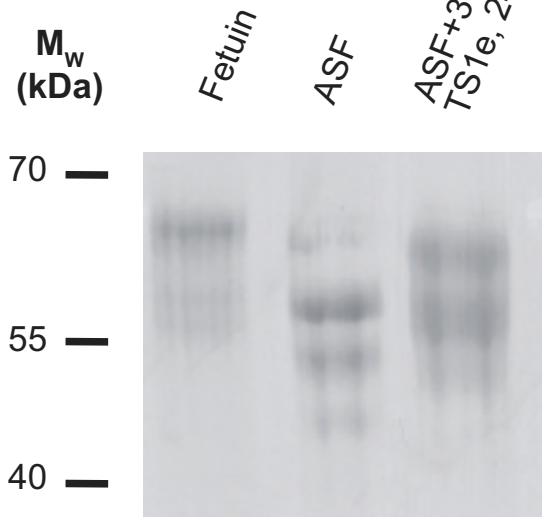


Figure 10 SDS-PAGE of resialylated fetuin. ASF (45 to 58 kDa): *Vibrio cholera* sialidase-treated fetuin (58 to 65 kDa). For resialylation ASF was treated with TS1 e-1 and 3'SL for 24 h. 2 µg protein per lane were run on a 12% polyacrylamide gel. Abbr.: M_W: molecular weight

this substrate indicated SA activity in preparations of TS-form 1 and TS-form 2 [6], as well as TS activity in procyclic parasites [16]. Also this apparent discrepancy can easily be explained by the presence of other TS1-like enzymes accepting Neu5Ac-MU as TS or SA substrate.

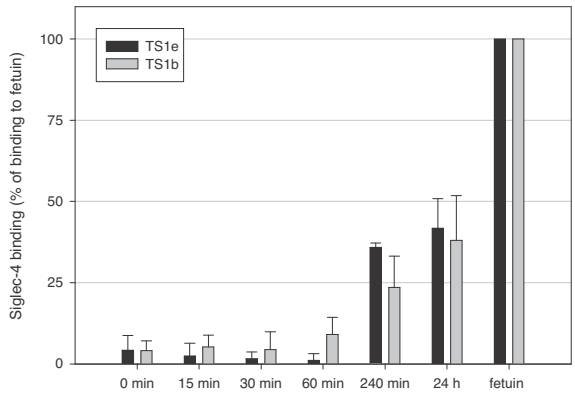


Figure 11 Binding of Siglec-4 to resialylated fetuin. *Vibrio cholera* sialidase-treated fetuin (ASF) was treated with TS1 e-1 and 3'-SL for the times indicated as described under Material and Methods. Siglec-4 binding relative to untreated fetuin is shown. Data points are means ± standard deviations of three independent determinations.

Ten of the eleven recombinant TS1 variants revealed TS activity, which is in agreement with the TS-form 1 purified by Tiralongo et al. [6]. Only TS1g, which carries a cysteine at position 144 instead of an arginine, showed very low TS but clear SA activity. This suggests that R144 is important but not essential for the transfer reaction and hinders SA activity in *T. congolense* TS1. In a previous study on *T. rangeli* SA [17] R144 has been postulated to form a hydrogen bond to the O4 of Sia. However, it remains unclear how this could relate to TS activity. Furthermore, the homology model of TS1 does not provide evidence for such a hydrogen bond for the binding site of *T. congolense* TS1.

The TS1 homology model (Figure 4B) displays also the position of amino acid exchanges in the eleven TS1 forms identified. It could be speculated that these positions are relevant determinants of antigenic sites and that these variations help to escape recognition by the immune system. However, if TS1 is expressed in the procyclic form in the insect vector, this is unlikely to be relevant. Interestingly, these exchanges are located on the same side of the protein where substrate binding occurs, which opens the possibility that these changes influence the binding specificities. The cluster of changes from position 599 to 602 and 643 could be involved in recognition of larger substrate molecules, such as glycoproteins. Positions, 407 and 408 (Figure 4B), as well as positions 375, 377 and 404 (Figure 4A) are closer to the active site in the catalytic domain, possibly contributing to interactions with substrates. It should be noted that the six amino acid differences occurring in *T. brucei* TS are not found in clusters like in *T. congolense* TS1, and furthermore, they are not even on the same side of the protein.

For a more detailed characterization of their differences in activity, we choose two TS1 variants, TS1b and TS1 e-1. The two amino acid sequences of TS1b and TS1 e-1 differ mainly in the lectin domain (Figure 4), whereby the exchange from E600 in TS1b to K600 in TS1 e-1 represents the most drastic change.

The observation that lactose is a much better acceptor substrate than Gal is in agreement with previously reported relative transfer activity data for the preparations TS-form 1 and TS-form 2 [9]. Interestingly, with fetuin as donor substrate *T. cruzi* TS, *T. congolense* TS1b and TS1 e-1 produced different amounts of 3'SL in 30 minutes under identical conditions. It is likely that these differences are due to distinct substrate specificities for the sialoglycans of fetuin. However, after 20 h, equilibriums for the reactions were attained and the three TS applied produced almost the same amounts of 3'SL. Most likely this reflects a similar equilibrium for these three TS.

Koliwer-Brandl *et al. BMC Biochemistry* 2011, **12**:39
http://www.biomedcentral.com/1471-2091/12/39

Clear differences were found in the kinetic parameters of TS1b and TS1 e-1 for lactose and fetuin. Reliable kinetic parameter for 3'SL as donor substrate could not be determined, because of the inaccurate quantification of the high concentrations of 3'SL as donor substrate. Whereas TS1b has a slightly higher K_M (1683 μM) for lactose compared to TS1 e-1 (727 μM), the opposite and more pronounced difference was observed for fetuin-bound Sia, where TS1 e-1 has approximately five-fold higher K_M (81 μM) compared to TS1b (17 μM). In combination, this implies that the ratio of K_M for lactose/K_M for fetuin-bound Sia is approximately 100 for TS1b, whereas it is only 10 for TS1 e-1. By comparison, for *T. cruzi* TS the K_M for both substrates is quite similar (326 over 206 μM) and its v_{max} was fourfold higher than for the TS1 isoforms. The differences in the kinetic parameters for fetuin observed for TS1b and TS1 e-1 are possibly related to altered affinities resulting from amino acid divergences in the lectin domain of TS1b and TS1 e-1. This would suggest a mechanism linking the lectin domain to the enzymatic properties of TS.

No release of free Sia could be detected after 30 minutes of TS reactions with fetuin as donor and lactose as acceptor substrates, demonstrating the absence of SA activity. However, after 20 h incubation, free Sia was detected clearly indicating SA activity. This activity correlated with the TS amount present, implying SA activity to be a side reaction observable only in extended reactions. Interestingly, in the reverse TS reactions with 3'SL as donor substrate for sialylation of galactose residues of ASF, free Sia is detected very early in the reaction (data not shown). This suggests that the free Sia detected in extended TS reactions times with fetuin as donor substrate is mainly the product of a SA side activity of the reaction using 3'SL as substrate. Therefore, the amount of free Sia could provide indirect information on the velocity of the reverse reaction. This assumption is further supported by the fact that lower amounts of TS can led to the same amount of final 3'SL but produce less free Sia as side product (Table 1). In this context it should be noted, that TS from *T. cruzi* clearly produced more free Sia than *T. congolense* TS1b or TS1 e-1. The structural basis for this phenomenon is unclear but may be related to the kinetics of the reaction. TS have been reported to follow ping-pong bi bi kinetics [12,18,19]. It will be interesting to investigate the structure-function relationship of this phenomenon and whether this is related to the SA activity of TS1g.

The TS substrate specificities for the glycans of the donor substrate fetuin were investigated by a MALDI-TOF analysis of TS treated glycopeptides from trypsin digested fetuin, since the glycosylation of fetuin is well established [13,20-23]. Three glycopeptides coming from two of the three *N*-glycosylation sites (di- and tribranched GP 127-141, and tribranched GP 54-85) could be analyzed reliably. The predicted masses of GP 142-169 di could not be identified in any spectra. GP 54-85 di and GP 142-169 tri differed only in one Da, which were not resolved by the equipment available and were excluded from the analysis. The peptide containing the three *O*-glycosylation sites could not be detected due to its high mass. But it is important to note, that TS1 clearly utilizes sialylated *O*-glycans as donor substrates as indicated by a rapid unmasking of peanut agglutinin recognition sites (data not shown).

Only minor differences in the sialylation pattern of fetuin GPs were observed using the different TS species. In summary, these were too small to draw a conclusion that these TS differ in their substrate specificities for fetuin glycans. In general, the TS applied cleaved Sia from *N*-glycans of glycopeptides investigated, but also transferred Sia back to these *N*-glycans. *N*-glycans that carried three and four Sia molecules in the case of the tribranched *N*-glycans as well as *N*-glycans that carried two Sia molecules in the case of dibranched *N*-glycans before TS incubation, were reduced to predominantly 0-2 Sia molecules in the case of tribranched and 0-1 Sia molecule in the case of dibranched *N*-glycans after 24 h TS incubation.

Whereas in the reactions discussed above Sia was transferred from fetuin to synthesize 3'SL, we could also show that *T. congolense* TS1 as well as *T. cruzi* TS transfer Sia in the reverse direction from 3'SL to glycoproteins. Furthermore, the TS reaction restores binding of Sia-binding proteins, such as Siglec-4. Due to the reversibility of the reaction, a complete resialylation of an acceptor substrate applying TS cannot be expected under these conditions. Nevertheless, differences in the kinetic parameters as shown for two of the eleven *T. congolense* TS1 variants could be used in kinetically controlled reactions to optimize the TS reaction to one or the other product, making the TS1 variants interesting tools for biotechnological applications. Thus, TS1 can be utilized to transfer Sia in α2,3-linkage on biologically relevant glycoproteins containing terminal galactose as Sia acceptor.

Conclusions

For the first time, full length TS from the African parasite *T. congolense* has been cloned and sequenced, opening new perspectives for investigations on the biological role of these enzymes in the pursuit of a cure for Nagana and sleeping sickness. Eleven *T. congolense* TS1 variants were identified and expressed as recombinant proteins. The eleven TS1 differ in 25 of 702 amino acid positions and a structural model revealed that these variations occur in three clusters on the side of the protein that is open to substrate binding. Ten of these TS1

Koliwer-Brandl *et al. BMC Biochemistry* 2011, **12**:39
http://www.biomedcentral.com/1471-2091/12/39

variants share predominantly TS and little SA activity. Only one, TS1g, has much lower TS but increased SA activity, probably due to an exchange of an arginine to a cysteine at position 144. Interestingly, the kinetic parameters of two characterized TS1 variants reveal subtle differences in substrate specificities. However, these did not lead to major differences in the sialylation pattern of *N*-glycans on fetuin after treatment with different TS variants. Finally a proof of principle has been provided that these TS can be used to sialylate glycoconjugates to create binding sites for Sia-binding proteins like Siglec-4.

It will be interesting to investigate the expression patterns of TS1 variants in the parasite's life cycle in future investigations addressing their importance for the manifestation of midgut colonization and maturation in the tsetse vector with possible implications for the transmission to the mammalian host.

Methods

Materials

Complete Mini, EDTA free protease-inhibitor tablets and *Vibrio cholerae* sialidase were purchased from Roche Diagnostics, Mannheim, Germany. *Pfu* DNA polymerase and restriction enzymes *Bam*HI and *Spe*I, isopropyl β-D-1-thiogalacto-pyranoside (IPTG), PageBlue and molecular weight marker (PageRuler) were from Fermentas, St. Leon-Rot, Germany. Trypsin was from Promega, Mannheim, Germany, 2,5-Dihydroxybenzoic acid from Bruker Daltonics, Billerica, USA. Ultrafiltration units Vivaspin6 and VivaCell250 were from Sartorius, Göttingen, Germany. BCA Protein Assay Kit was purchased from Thermo Scientific Pierce, Rockford, USA. Anti-SNAP-tag rabbit polyclonal antibody was from GenScript, Piscataway, USA. Anti-*Strep*-tag rabbit polyclonal antibody, StrepTactin beads and buffers were purchased from IBA, Göttingen, Germany and hygromycin from PAA, Pasching, Austria. 2'-(4-Methylumbelliferyl)-α-D-*N*-acetylneuraminic acid sodium salt hydrate (MU-Neu5Ac), 4-methylumbelliferyl ß-D-galactoside (MU-Gal), 4-methylumbelliferone (MU), glucoronic acid, Ex-cell® CD CHO media, fetuin and PEI transfection reagent were purchased from Sigma-Aldrich, Munich, Germany. X-ray film, enhanced chemiluminescence system, Ni-NTA and Q-Sepharose FF were purchased from GE Healthcare, Munich, polyvinylidene difluoride membranes and ZipTips from Millipore, Schwalbach, Germany.

Cloning and expression of recombinant TS1

The published partial sequences of *Trypanosoma congolense* TS1 [Genebank: AJ535487.1] [6,9] was used as starting query for searching the *T. congolense* genomic database for pathogen genomics at the WTSI (http://www.sanger.ac.uk/). Based on the obtained sequence fragments, an open reading frame encoding TS1 was assembled. Based on this, primers were designed to amplify TS1 from *T. congolense* (strain STIB 249) genomic DNA [9] using *Pfu* DNA polymerase in a nested PCR reaction leaving out the N-terminal signal peptide sequence and the C-terminal GPI anchor attachment site. Both outside primers (forward ATG CGG CCG GTG AAT TGT TAN and reverse CAT CAG CAC ATG CAC GAG CAN) were degenerate at the 3' end, whereas the internal primers (forward CGA CTA GTC AGT GCT GCG ACC ACA TGC AN and reverse CGG GAT CCG TCG CTC CCA GGC ACA CGA AN) were designed to introduce *Spe*1 and *Bam*HI restriction sites, respectively. These restriction sites were used to ligate the PCR-products in frame into a modified pDEF [24] vector (pDEF-T3C/SNAPstrep) providing in frame a Transin cleavable signal peptide [25], a 3C-protease cleavage site [26] followed by SNAP (Covalys, Witterswil, Switzerland) and *Strep* (IBA, Göttingen, Germany) tags. The pDEF-T3C/SNAPstrep was obtained as follows: The coding sequence for the Transin signal peptide has been introduced into pcDNA 3 Amp *Strep-tag* [27] using the *Hind*III/*Bam*HI-digested linker obtained by hybridization of the following oligonucleotides: sense 5'-CGAAGCTTATGAAAGGGCTCCCAGTCCTGC TGTGGCTGTGTACGGCTGTGTGCTC

ATCCTACCCATTGCATGGCAGTGAAGAAGAT GCTGGCATGGAGACTAGTGGATCCCG

and antisense primer 5'-

CGGGATCCACTAGTCTCCATGCCAGCATCTT CTTCACTGCCATGCAATGGGTAGG

ATGAGCACACACAGCCGTACACAGCCACAGCA GGACTGGGAGCCCTTTCATAAGCTTCG.

This Transin linker introduced a unique *Spe*I restriction site. The coding sequence of the hAGT protein (SNAP) was amplified using the pSNAP-tag® (T7) vector (NEB, Ipswich, MA, USA) as a template and subcloned in frame into *Bam*HI/*Eco*RI digested pcDNA3 Amp Transin *Strep*-tag providing pcDNA3-T3C/SNAP-strep. The primers used for amplification were: sense 5'-CGGGATCCCTGGAGGTGCTGTTCCAGGGCCC-CATGGACAAAGACTGCGAAATGAAGCG-3' including the coding sequence for the 3C protease recognition site of the human rhinovirus HRV 3C (LEVLFQGP, underlined) and antisense 5'- CGGAATTCACCCAGC CCAGGCTTGCCCAGA.

CHO_{Lec1} cells were used for TS1 expression due to their ability to express only high-mannose glycans, since these cells are lacking *N*-acetyl glucosaminyltransferase I [28]. Transfection of CHO_{Lec1} cells grown in αMEM supplemented with 10% fetal calf serum at 37°C, 5% CO_2 was accomplished with polyethylenimine (PEI) transfection reagent following the manufacturer's

instructions. 24 h after transfection, cells were passaged into 96-well plates in a selection media of αMEM containing varying amounts of hygromycin, ranging from 400 μg/mL to 1000 μg/mL. Expression of recombinant TS1 (120 kDa including SNAP and *Strep* tags) was tested by analyzing cell culture supernatant using Western blots with anti TS1 monoclonal antibody (mAb 7/23) as primary antibody [9]. The presence of SNAP and *Strep* tags was confirmed using anti-SNAP-tag rabbit polyclonal antibody and anti-*Strep*-tag rabbit polyclonal antibody respectively in Western blots analysis. Selected cells were then adapted to Ex-cell® CD CHO media supplemented with 8 mM L-glutamine.

Purification of *Trypanosoma congolense* trans-sialidase

The harvested tissue culture supernatant was supplemented with 10 mM Tris/HCl, pH 7.5, 1 mM EDTA, 1 mM DTT and 0.02% sodium azide (all final concentrations) and centrifuged at 125,000 × g for 1 h. The cleared supernatant was then concentrated 100-fold by ultrafiltration (100 kDa cut off). Buffer was exchanged twice using 250 mL 100 mM Tris-Cl, pH 8.0, 150 mM NaCl, 1 mM EDTA (buffer A) in the same ultrafiltration unit and concentrated to a total volume of 10 mL for 1 L tissue culture supernatant. This was further clarified by centrifugation at 21,000 × g for 30 minutes before applying on a column of 1 mL *Strep*Tactin® beads equilibrated with buffer A. After loading the column was washed with 5 column volumes wash buffer (100 mM Tris-Cl, pH 8.0, 150 mM NaCl, 1 mM EDTA) and TS1 was eluted with 3 column volumes of elution buffer (100 mM Tris-Cl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin) in fractions of 0.5 mL. The affinity purified TS1 was dialyzed (10 kDa cut off) against 10 mM phosphate, pH 7.4. Purification products were analyzed by SDS-PAGE and quantified by BCA assay with bovine serum albumin as standard.

Expression and purification of *Trypanosoma cruzi* trans-sialidase

Recombinant *T. cruzi* TS was produced in *E. coli* M15 (pREP4) according to Agusti et al., 2004 and Neubacher et al., 2005 [29,30]. In brief, cells were grown in 1 L "terrific broth" medium overnight at 18°C. Protein expression was initiated with 0.5 mM IPTG. The cells were dissolved in 40 mL lysis buffer (50 mM phosphate, 300 mM NaCl, pH 8.0 and 0.05% Lubrol), 1 tablet protease-inhibitor (Complete Mini, EDTA free) and 1 spatula tip of lysozyme were before incubation of 30 minutes at 4°C. The cells were disrupted by 5 cycles of sonification on ice. Cells debris was removed by centrifugation at 40,000 × g for 60 minutes at 4°C and the supernatant was filtered using a 0.2 μm pore size filter. 20 mM imidazol was added before application on 0.5 mL Ni-NTA beads. Target proteins were eluted in the same buffer and 250 mM imidazole. The eluted protein was dialyzed against 20 mM Tris, 30 mM NaCl, pH 8.0 and further purified using a Q-Sepharose FF column in the same buffer with a linear gradient up to 1 M NaCl. The activity of the purified protein was tested by a sialidase activity assay using MU-Neu5Ac as donor substrate and lactose as acceptor substrate as described below.

Vibrio cholerae sialidase treatment of fetuin

Asialofetuin (ASF) was prepared from fetuin by *Vibrio cholerae* sialidase (VCS) treatment as described [31]. In brief, fetuin was digested with VCS in 50 mM sodium acetate, 9 mM $CaCl_2$, pH 5.5 overnight at 37°C in a dialysis bag against the same buffer and afterwards against distilled water. Sialylated fetuin and sialidase was separated from ASF by anion exchange chromatography using Q-Sepharose. The proteins were eluted by a linear gradient from 0 to 1 M NaCl in 10 mM Tris, pH 7.4. Collected fractions were assayed for protein at 280 nm and for SA activity with MU-Neu5Ac acid as substrate as described below. The fractions containing ASF but no SA activity were pooled and the buffer was exchanged against 10 mM phosphate, pH 7.4 using VivaSpin6 ultrafiltration units (10 kDa cut off).

Sialidase activity assay

Sialidase activity (+/- lactose) was determined by applying a microtitre plate assay detecting free 4-methylumbelliferone (MU) released from Neu5Ac-MU [6]. In brief, 50 μL sample were incubated with 1 mM Neu5Ac-MU (final concentration) in a black 96-well microtitre plate. To determine *T. cruzi* TS 1 mM lactose was added as acceptor substrate. The plate was centrifuged for 1 minute at 1,000 × g and incubated for 30 minutes at room temperature in the dark. The reactions were stopped with 200 μL 100 mM glycine, pH 10 and the fluorescence intensities were measured at 355 nm excitation and 460 nm emission using a fluorimeter (Ascent Fluoroscan).

Trans-sialidase reactions

The principle of this assay is based on the quantification of sialylated oligosaccharides by high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) as described below. In these reactions either fetuin is used as donor substrate, e.g. lactose as acceptor or 3'-sialyllactose (3'SL) is used as donor substrate with ASF as acceptor. A final volume of 50 μL 10 mM phosphate buffer were used for all TS reactions. Stock solutions of donor and acceptor substrate were mixed in 40 μL buffer (10 mM Tris/HCl, pH 7.5) and the reactions were started with 10 μL TS (50 ng in standard reactions) and incubated at 37°C. The

reactions were stopped with 200 μL ice-cold acetone containing 28 μM glucuronic acid and incubated overnight at -20°C. After centrifugation for 15 minutes at 20,000 × g and 4°C, 225 μL supernatant were removed and both, protein pellets and supernatants were lyophilized.

The dried supernatant of the acetone precipitation was dissolved in 125 μL H_2O for HPAEC-PAD, which was carried out by using a DX600 system (Dionex, Sunnyvale, CA, USA) with an electrochemical detector (ED50), a gradient pump (GP50) and an autosampler (AS50). Carbohydrates were separated by HPAEC on a CarboPAC PA1 (4 × 250 mm) analytical column (Dionex) together with a guard column (4 × 50 mm) using a constant flow rate of 1 mL/min. Sample volumes of 25 μL were injected and the chromatography was performed as follows: 100 mM NaOH for 2 min, followed by 100 mM NaOH/100 mM NaOAc for 22 min. The column was regenerated by washing for 5 min with 100 mM NaOH/500 mM NaOAc, followed by 5 min with 100 mM NaOH. For PAD the typical quadruple waveform was used as described previously [32]. The Dionex software Chromeleon 6.40 SP8 was used for data acquisition and data evaluation.

Calculation of kinetic parameters

V_{max} and K_M were calculated using the curve fit module of SigmaPlot 11 employing the Michaelis-Menten equation $v = v_{max} \times c_s/(c_s + K_M)$.

Siglec-4 binding assay

Murine Siglec-$4_{d1\text{-}3}$-Fc was purified by protein-A affinity chromatography from tissue culture supernatants of stably transfected CHO Lec3.2.8.1 as described before [33]. The protein solution was dialyzed against 10 mM phosphate buffer pH 7.4, sterile filtered and stored at 4°C. Binding assays with Siglec-4 were performed as described previously [33]. In brief, 4 μg/mL fetuin, ASF or TS-treated fetuin were immobilized in microtitre plates and binding of serially diluted Siglec-$4_{d1\text{-}3}$-Fc (8 dilutions starting with 16 μg/mL) was determined using alkaline phosphatase-labeled anti-Fc antibodies. The concentrations sufficient for 50% binding (relative to Siglec-$4_{d1\text{-}3}$-Fc binding to fetuin) were determined from corresponding binding curves. At least three independent titrations were performed.

SDS-PAGE and Western Blot analysis

Samples were separated by SDS-PAGE (MiniProtean III; Bio-Rad, München, Germany) according to Laemmli [34] and stained with PageBlue.

For Western blot analysis, samples were transferred onto polyvinylidene difluoride membranes after SDS-PAGE. The membranes were blocked with 5% BSA in Tris-buffered saline (TBS) buffer containing 0.15% Tween20 (TBS-T) for 1 h. Washing of the membrane was done five times for 5 minutes each using TBS-T. Immunodetection was performed by incubating membranes with a primary antibody diluted in blocking buffer overnight at 4°C. The following antibodies were used: anti-*T. congolense* TS mAb 7/23 (1:1000) and rabbit anti-*Strep*-tag (1:1000). Following four washes with TBS-T of 10 minutes each, the membranes were incubated with a secondary antibody conjugated to horseradish peroxidase for 2 h at room temperature. After washing four times with TBS-T, blots were developed with the enhanced chemiluminescence system using X-ray film.

Matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF-MS)

TS reactions were carried out with 50 ng TS, 100 nmol lactose and 100 μg fetuin as described above. The dried protein pellets after TS reaction were dissolved in 200 μL 50 mM ammonium hydrogen carbonate, pH 7.8 and 1.6 μL 45 mM DTT were added. After 30 minutes incubation at 50°C, 1.6 μL 100 mM IAA were added and further incubated for 30 minutes at 37°C. The tryptic digestion was started with 2 μg trypsin dissolved in 1 μL 50 mM acetic acid and incubated overnight at 37°C and stored at -20°C.

5 μL of the tryptic digest were mixed with 5 μL H_2O and 1 μL 1% trifluoroacetic acid and were directly mixed with matrix and 1 μL was applied to the MALDI-TOF-MS target plate. The remaining 9 μL were desalted using C18-reversed phase pipette tips (ZipTip). Peptides were eluted with 3 × 100 μL 12% and 30% acetonitrile in H_2O and lyophilized. 2,5-Dihydroxybenzoic acid in 0.1% trifluoroacetic acid was used as MALDI matrix. The dried peptides were directly dissolved in 10 μL DHB solution and 1 μL of the mixture were spotted to the target plate for crystallization.

The mass spectrometry was performed using a Voyager DE Pro MALDI-TOF N_2-Laser with a wavelength of 337 nm (Applied Biosystems, Foster City, USA). All spectra were measured in the linear detector mode. Laser intensities and the number of records per spectrum were varied manually.

Voyager software was used for data acquisition and peak detection. For quantification the peak intensity of each peptide was determined relative to the non-sialylated glycopeptide 1-32 (3459.66 Da) as internal standard. This peptide is detected in all spectra of the sialylation variants of the glycopeptides and is not changed by TS reaction (Figure 8). All measurements were repeated for at least three times.

Homology Modeling

A homology model of *T. congolense* TS1 e-1 was calculated using the software Yasara 10.11.8 [35-40]. The

crystal structure of *T. cruzi* TS (UniProt: Q26964; PDB entry: 3B69), previously reported by Buschiazzo et al., 2002 [12] was used as the template structure. A benzoylated *N*-acetylneuraminic acid derivative used as a ligand for *T. cruzi* TS in the template structure was kept in the binding site during the calculation of the homology model.

The following parameters of the Yasara homology modeling module were modified manually from the default settings of the program: Modeling speed: slow, PsiBLASTs: 6, EValue Max: 0.5, Templates total: 1, Templates SameSeq: 1, OligoState: 4, alignments: 10, LoopSamples: 50, TermExtension:10.

Acknowledgements
We like to thank Joel Smith, who initially searched TS genes in the *T. congolense* genome in the WTSI database. We gratefully acknowledge Prof. Patrick Ziegelmüller (University of Hamburg, Germany) for providing a clone of *E. coli* M15 (pREP4) expressing trans-sialidase from *Trypanosoma cruzi*. Neu5Ac-pNP were kindly donated by Prof. Joachim Thiem (University Hamburg, Germany). We are grateful to Evelin Tiralongo and Roland Schauer (University Kiel, Germany) for providing the *T. congolense* genomic DNA, Hans Lange and Hilmar Lemke (University Kiel, Germany) for the cell line producing anti-*T. congolense* TS antibody mAb 7/23. The plasmid vector pDEF was kindly provided by Prof. Paul Crocker (University of Dundee, UK). We would like to thank the German Federal Ministry for Education and Research (BMBF, project BioChangePLUS 031632A), the Tönjes-Vagt Foundation (project XXI) and Deutsche Forschungsgemeinschaft (DFG, project KE 428/8-1) for their financial support of this work.

Authors' contributions
HKB carried out all the HPAEC-PAD based TS assays and its data acquisition, the Siglec-4 binding assay and SDS-PAGE to determine resialylation of asialofetuin, homology modeling, alignment of amino acid sequences, supported MALDI data acquisition and drafted the manuscript. TTG carried out *T. congolense* TS1 cloning, sequencing and expression as well as *T. cruzi* TS expression. MW performed the TS1g enzyme assays. OR participated in development of the HPAEC-PAD based TS assay. PM carried out MALDI-MS analysis and its data acquisition. ED participated in MALDI-MS analysis and its data acquisition. FD established the pDEF-Transin-TS-3C-SNAP-Strep vector construct and carried out *T. congolense* TS1 cloning. SK designed and coordinated the study and supported drafting of the manuscript. All authors read and approved the final manuscript.

Received: 22 April 2011 Accepted: 30 July 2011 Published: 30 July 2011

References

1. Steverding D: **The history of African trypanosomiasis.** *Parasit Vectors* 2008, **1(1)**:3.
2. Schenkman S, Jiang MS, Hart GW, Nussenzweig V: **A novel cell surface trans-sialidase of Trypanosoma cruzi generates a stage-specific epitope required for invasion of mammalian cells.** *Cell* 1991, **65(7)**:1117-1125.
3. Montagna G, Cremona ML, Paris G, Amaya MF, Buschiazzo A, Alzari PM, Frasch AC: **The trans-sialidase from the african trypanosome Trypanosoma brucei.** *Eur J Biochem* 2002, **269(12)**:2941-2950.
4. Garcia ES, Azambuja P: **Development and interactions of Trypanosoma cruzi within the insect vector.** *Parasitol Today* 1991, **7(9)**:240-244.
5. Pontes de Carvalho LC, Tomlinson S, Vandekerckhove F, Bienen EJ, Clarkson AB, Jiang MS, Hart GW, Nussenzweig V: **Characterization of a novel trans-sialidase of Trypanosoma brucei procyclic trypomastigotes and identification of procyclin as the main sialic acid acceptor.** *J Exp Med* 1993, **177(2)**:465-474.
6. Tiralongo E, Schrader S, Lange H, Lemke H, Tiralongo J, Schauer R: **Two trans-sialidase forms with different sialic acid transfer and sialidase activities from Trypanosoma congolense.** *J Biol Chem* 2003, **278(26)**:23301-23310.
7. Engstler M, Schauer R: **Sialidases from African trypanosomes.** *Parasitol Today* 1993, **9(6)**:222-225.
8. Nagamune K, Acosta-Serrano A, Uemura H, Brun R, Kunz-Renggli C, Maeda Y, Ferguson MA, Kinoshita T: **Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors.** *J Exp Med* 2004, **199(10)**:1445-1450.
9. Tiralongo E, Martensen I, Grotzinger J, Tiralongo J, Schauer R: **Trans-sialidase-like sequences from Trypanosoma congolense conserve most of the critical active site residues found in other trans-sialidases.** *Biol Chem* 2003, **384(8)**:1203-1213.
10. Eisenhaber B, Bork P, Yuan Y, Loffler G, Eisenhaber F: **Automated annotation of GPI anchor sites: case study C. elegans.** *Trends Biochem Sci* 2000, **25(7)**:340-341.
11. Cremona ML, Sanchez DO, Frasch AC, Campetella O: **A single tyrosine differentiates active and inactive Trypanosoma cruzi trans-sialidases.** *Gene* 1995, **160(1)**:123-128.
12. Buschiazzo A, Amaya MF, Cremona ML, Frasch AC, Alzari PM: **The crystal structure and mode of action of trans-sialidase, a key enzyme in Trypanosoma cruzi pathogenesis.** *Mol Cell* 2002, **10(4)**:757-768.
13. Green ED, Adelt G, Baenziger JU, Wilson S, Van Halbeek H: **The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy.** *J Biol Chem* 1988, **263(34)**:18253-18268.
14. Frasch AC: **Functional diversity in the trans-sialidase and mucin families in Trypanosoma cruzi.** *Parasitol Today* 2000, **16(7)**:282-286.
15. Buschiazzo A, Campetella O, Frasch AC: **Trypanosoma rangeli sialidase: cloning, expression and similarity to T. cruzi trans-sialidase.** *Glycobiology* 1997, **7(8)**:1167-1173.
16. Engstler M, Schauer R, Brun R: **Distribution of developmentally regulated trans-sialidases in the Kinetoplastida and characterization of a shed trans-sialidase activity from procyclic Trypanosoma congolense.** *Acta Trop* 1995, **59(2)**:117-129.
17. Buschiazzo A, Tavares GA, Campetella O, Spinelli S, Cremona ML, Paris G, Amaya MF, Frasch AC, Alzari PM: **Structural basis of sialyltransferase activity in trypanosomal sialidases.** *Embo J* 2000, **19(1)**:16-24.
18. Amaya MF, Watts AG, Damager I, Wehenkel A, Nguyen T, Buschiazzo A, Paris G, Frasch AC, Withers SG, Alzari PM: **Structural insights into the catalytic mechanism of Trypanosoma cruzi trans-sialidase.** *Structure* 2004, **12(5)**:775-784.
19. Damager I, Buchini S, Amaya MF, Buschiazzo A, Alzari P, Frasch AC, Watts A, Withers SG: **Kinetic and mechanistic analysis of Trypanosoma cruzi trans-sialidase reveals a classical ping-pong mechanism with acid/base catalysis.** *Biochemistry-Us* 2008, **47(11)**:3507-3512.
20. Brown WM, Saunders NR, Mollgard K, Dziegielewska KM: **Fetuin–an old friend revisited.** *Bioessays* 1992, **14(11)**:749-755.
21. Cointe D, Leroy Y, Chirat F: **Determination of the sialylation level and of the ratio alpha-(2– > 3)/alpha-(2– > 6) sialyl linkages of N-glycans by methylation and GC/MS analysis.** *Carbohydr Res* 1998, **311(1-2)**:51-59.
22. Dziegielewska KM, Brown WM, Casey SJ, Christie DL, Foreman RC, Hill RM, Saunders NR: **The complete cDNA and amino acid sequence of bovine fetuin. Its homology with alpha 2HS glycoprotein and relation to other members of the cystatin superfamily.** *J Biol Chem* 1990, **265(8)**:4354-4357.
23. Spiro RG: **Studies on fetuin, a glycoprotein of fetal serum. I. Isolation, chemical composition, and physiochemical properties.** *J Biol Chem* 1960, **235(10)**:2860-2869.
24. Zhang JQ, Biedermann B, Nitschke L, Crocker PR: **The murine inhibitory receptor mSiglec-E is expressed broadly on cells of the innate immune system whereas mSiglec-F is restricted to eosinophils.** *Eur J Immunol* 2004, **34(4)**:1175-1184.
25. Matrisian LM, Glaichenhaus N, Gesnel MC, Breathnach R: **Epidermal growth factor and oncogenes induce transcription of the same cellular mRNA in rat fibroblasts.** *Embo J* 1985, **4(6)**:1435-1440.
26. Walker PA, Leong LE, Ng PW, Tan SH, Waller S, Murphy D, Porter AG: **Efficient and rapid affinity purification of proteins using recombinant fusion proteases.** *Biotechnology* 1994, **12(6)**:601-605.
27. Thakar K, Niedenthal R, Okaz E, Franken S, Jakobs A, Gupta S, Kelm S, Dietz F: **SUMOylation of the hepatoma-derived growth factor negatively influences its binding to chromatin.** *Febs J* 2008, **275(7)**:1411-1426.
28. Kumar R, Stanley P: **Transfection of a human gene that corrects the Lec1 glycosylation defect: evidence for transfer of the structural gene for N-**

acetylglucosaminyltransferase I. *Molecular and cellular biology* 1989, 9(12):5713-5717.
29. Neubacher B, Schmidt D, Ziegelmuller P, Thiem J: **Preparation of sialylated oligosaccharides employing recombinant trans-sialidase from Trypanosoma cruzi.** *Org Biomol Chem* 2005, **3(8)**:1551-1556.
30. Agusti R, Paris G, Ratier L, Frasch AC, de Lederkremer RM: **Lactose derivatives are inhibitors of Trypanosoma cruzi trans-sialidase activity toward conventional substrates in vitro and in vivo.** *Glycobiology* 2004, **14(7)**:659-670.
31. Hirmo S, Kelm S, Schauer R, Nilsson B, Wadstrom T: **Adhesion of Helicobacter pylori strains to alpha-2,3-linked sialic acids.** *Glycoconj J* 1996, **13(6)**:1005-1011.
32. Rocklin R, Clarke A, Weitzhandler M: **Improved long-term reproducibility for pulsed amperometric detection of carbohydrates via a new quadruple-potential waveform.** *Anal Chem* 1998, **70(8)**:1496-1501.
33. Bock N, Kelm S: **Binding and inhibition assays for Siglecs.** *Methods Mol Biol* 2006, **347**:359-375.
34. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, **227(5259)**:680-685.
35. Canutescu AA, Shelenkov AA, Dunbrack RL Jr: **A graph-theory algorithm for rapid protein side-chain prediction.** *Protein Sci* 2003, **12(9)**:2001-2014.
36. Jones DT: **Protein secondary structure prediction based on position-specific scoring matrices.** *J Mol Biol* 1999, **292(2)**:195-202.
37. King RD, Sternberg MJ: **Identification and application of the concepts important for accurate and reliable protein secondary structure prediction.** *Protein Sci* 1996, **5(11)**:2298-2310.
38. Krieger E, Joo K, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K: **Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8.** *Proteins* 2009, **77(Suppl 9)**:114-122.
39. Muckstein U, Hofacker IL, Stadler PF: **Stochastic pairwise alignments.** *Bioinformatics* 2002, **18(Suppl 2)**:S153-160.
40. Qiu J, Elber R: **SSALN: an alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs.** *Proteins* 2006, **62(4)**:881-891.

doi:10.1186/1471-2091-12-39
Cite this article as: Koliwer-Brandl *et al.*: Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*. *BMC Biochemistry* 2011 **12**:39.

3.2 Results

Biochemical diversity in the *Trypanosoma congolense* trans-sialidase family

Thaddeus T. Gbem, Mario Waespy, Bettina Hesse, Frank Dietz, Joel Smith, Gloria D. Chechet, Jonathan A. Nok, Sørge Kelm (2013)

*PLOS Neglected Tropical Diseases, **7 (12)**, p. e2549*

(doi:10.1371/journal.pntd.0002549)

Contribution of Mario Waespy:

- Purification of TconTS enzymes
- TconTS enzyme assays
- HPAEC-PAD data acquisition and evaluation
- Preliminary data regarding the epitope mapping of anti TS mAb 7/23 and TconTS domain swap have been obtained by Bettina Hesse and are part of her Bachelor Thesis planed and supervised by Mario Waespy
- Supported preparation of the manuscript

OPEN ACCESS Freely available online

PLOS | NEGLECTED TROPICAL DISEASES

Biochemical Diversity in the *Trypanosoma congolense* Trans-sialidase Family

Thaddeus T. Gbem[1], Mario Waespy[1], Bettina Hesse[1], Frank Dietz[1], Joel Smith[1], Gloria D. Chechet[2], Jonathan A. Nok[2], Sørge Kelm[1]*

1 Centre for Biomolecular Interactions Bremen, Faculty for Biology and Chemistry, University Bremen, Bremen, Germany, **2** Department of Biochemistry, Ahmadu Bello University, Zaria, Nigeria

Abstract

Trans-sialidases are key enzymes in the life cycle of African trypanosomes in both, mammalian host and insect vector and have been associated with the disease trypanosomiasis, namely sleeping sickness and nagana. Besides the previously reported TconTS1, we have identified three additional active trans-sialidases, TconTS2, TconTS3 and TconTS4, and three trans-sialidase like genes in *Trypanosoma congolense*. At least TconTS1, TconTS2 and TconTS4 are found in the bloodstream of infected animals. We have characterised the enzymatic properties of recombinant proteins expressed in eukaryotic fibroblasts using fetuin as model blood glycoprotein donor substrate. One of the recombinant trans-sialidases, TconTS2, had the highest specific activity reported thus far with very low sialidase activity. The active trans-sialidases share all the amino acids critical for the catalytic reaction with few variations in the predicted binding site for the leaving or acceptor glycan. However, these differences cannot explain the orders of magnitudes between their transfer activities, which must be due to other unidentified structural features of the proteins or substrates selectivity. Interestingly, the phylogenetic relationships between the lectin domains correlate with their specific trans-sialylation activities. This raises the question whether and how the lectin domains regulate the trans-sialidase reaction. The identification and enzymatic characterisation of the trans-sialidase family in *T. congolense* will contribute significantly towards the understanding of the roles of these enzymes in the pathogenesis of Animal African Trypanosomiasis.

Citation: Gbem TT, Waespy M, Hesse B, Dietz F, Smith J, et al. (2013) Biochemical Diversity in the *Trypanosoma congolense* Trans-sialidase Family. PLoS Negl Trop Dis 7(12): e2549. doi:10.1371/journal.pntd.0002549

Editor: Daniella C. Bartholomeu, Universidade Federal de Minas Gerais, Brazil

Received July 16, 2013; **Accepted** October 4, 2013; **Published** December 5, 2013

Funding: The Deutsche Forschungsgemeinschaft (DFG, project KE 428/8-1, www.dfg.de) has provided financial support for this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: skelm@uni-bremen.de

Introduction

Trypanosoma congolense (subgenus: *Nannomonas*) is a major causative agent of the Animal African Trypanosomiasis (AAT) otherwise known as nagana. Other parasites implicated in nagana include *T. vivax* (subgenus: *Duttonella*) and *T. brucei brucei* (subgenus: *Trypanozoon*). These protozoan parasites are transmitted by several species of hematophagous biting flies of the genus *Glossina*. Nagana exhibits a severe negative impact on stock farming, milk and meat production [1]. The impact of the disease is thought to be underestimated as most affected areas are remote, limiting access and hence accurate data acquisition. The need for an alternative arsenal against AAT is heightened as existing drugs are either toxic or rapidly becoming ineffective due to drug resistance [2].

The role of TS in Chaga's disease caused by the South American *T. cruzi* has been extensively studied [3]. On the other hand, studies on the *trans*-sialidases from African trypanosomes responsible for the Human African Trypanosomiasis (HAT) as well as AAT are scanty. In *T. brucei*, TS has been implicated in the cyclical survival of the parasite as evidence supports enhanced survival of parasite in midgut of the insect host [4]. No data exist on *T. congolense* in this regard. Though scanty, the role of blood stream TS and sialidase in anaemia in animals suffering trypanosomiasis caused by *T. congolense* [5] [6] and *T. vivax* [7] has been established.

Multiple copies of TS-like genes exist in *Trypanosoma* genomes. The highest number occurs in *T. cruzi*, but most of the over 1000 genes encod enzymatically inactive proteins [8]. In African trypanosomes, the TS-like gene families are much smaller. For example, 9 members have been identified in *T. brucei* [9] [10] [11]. In *T. congolense*, at least 17 TS-like genes have been identified [12] [13] [14] [6]. Eleven of these, forming the TconTS1 family, are closely related and share over 95% sequence identity [14]. The key element mediating the functions of TS has been ascribed to the N-terminal catalytic domain (CD) harbouring the active site with characteristic conserved amino acids [15] [16] [17] [9] [18], whereas hardly anything is known about possible functions of the lectin-like domain (LD) at the C-terminus of these enzymes.

Here we report that besides TconTS1, three additional members of the *T. congolense* TS gene family transfer sialic acids between glycoconjugates, but have much lower sialidase activities. The identification and biochemical characterisation of *T. congolense* TS genes will enable new studies investigating the role of these genes in nagana disease.

Methods

Unless where stated, all chemicals and reagents used were cell culture and analytical grade. *Vibrio cholerae* sialidase was purchased

Author Summary

Trypanosomiasis is a disease also known as sleeping sickness in humans (Human African Trypanosomiasis) and nagana in animals (Animal African Trypanosomiasis). This disease is caused by protozoan parasites of the genus *Trypanosoma*. Tsetse flies are responsible for the transmission of these parasites. *Trypanosoma congolense* is the main causative agent of nagana in cattle. The clinical signs of the disease have been linked to the presence of an enzyme called trans-sialidase. Interestingly, the enzyme alternates in different forms in the mammalian and the insect vector. Previous knowledge had shown that the parasite requires the enzyme for survival in the fly vector. Our current work has revealed other forms of the enzyme that could be essential for the persistence of the disease in mammalian and vector hosts. These enzymes, though similar in structural architecture, show differences in their activities that could be key in delineating their individual roles in the pathophysiology of the disease.

from Roche Diagnostics (Mannheim, Germany). *Pfu* DNA polymerase, *Hind*III, *Xba*I, *Spe*I, *Eco*RI and *Dpn*I, PageBlue, molecular weight marker (PageRuler), BCA assay protein kit were all purchased from Thermo Scientific (St. Leon-Rot, Germany). VivaSpin 6 and VivaCell250 ultracentrifugation units were from Sartorius (Göttingen, Germany). Anti-SNAP-tag rabbit polyclonal antibody was from GeneScript (Piscataway, USA) while anti-*Strep*-tag rabbit polyclonal antibody, *Strep*-Tactin resin beads and buffers were from IBA (Göttingen, Germany). Hygromycin and Gentamycin were purchased from PAA, (Pasching, Austria). Polyethylenimin transfection reagent, glucuronic acid, *N*-acetyl-neuraminic acid (Neu5Ac), 3'sialyllactose (3'SL) and lactose were purchased from Sigma-Aldrich (Steinheim, Germany). Ex-cell CD CHO media from SAFC, USA, X-ray film, Enhanced Chemiluminescence system, and recProtein-A Sepharose Fast Flow were purchased from GE Healthcare (Uppsala, Sweden). Polyvinylidene difluoride membrane was from Millipore (Schwabach, Germany).

Cloning, sequencing, expression and purification of trans-sialidase genes

The Basic Local Alignment Search Tool (BLAST) was used to search the shot-gun sequences of *T. congolense* at the WSTI (http://www.sanger.ac.uk). Using the BLASTN algorithm, the "*T. congolense* reads" were queried with the partial nucleotide sequences (Genbank Accession numbers TS1: AJ535487 and TS2: AJ535488) previously described [13]. Perfect BLAST hits (smallest sum probability P(N)<10-10) were arranged into contiguous sequences using Contig Express (Invitrogen, Carlsbad, USA). By searching the database with ends of the contiguous sequences, the assembled contigs were expanded until open reading frames (ORF) were obtained. On the basis of the obtained ORFs, primers (Supporting Information, Table S1) were designed to amplify by nested PCR the ORF including flanking regions encoding for TconTS2, TconTS3 and TconTS4 using genomic DNA of *T. congolense* strain STIB249 [13]. The resulting products were cloned into the pBlueScript KS- vector (Stratagene, Santa Clara, Ca, USA) via *Spe*I and *Bam*HI (TconTS2) or via *Eco*RI and *Sma*I (TconTS4) or into the mammalian expression vector pcDNAIII Amp (Invitrogen, Carlsbad, USA) via *Hind*III and *Xba*I (TconTS3) and sequenced (Supporting Information, Table S2).

Cloning and sequencing of *T. brucei* TS genes followed a similar strategy as described for *T. congolense* above except that genes were cloned in pJET1.2/blunt vector (Thermo Scientific) following instructions of the manufacturer (for primers see Supporting Information, Table S1).

For the expression of secreted TconTS proteins in mammalian fibroblasts, corresponding DNA sequences without those encoding the signal peptides and GPI anchors were subcloned into a modified pDEF vector providing a 3C protease recognition site, SNAP and *Strep* tags using *Spe*I and *Bam*HI restriction sites [14]. For this purpose, the *Bam*HI site in TconTS3 as well as the *Spe*I and *Bam*HI sites in TconTS4 were removed by site directed mutagenesis without changing the amino acid sequence encoded (for primers see Supporting Information, Table S1). All sequences and mutations were confirmed by Sanger dideoxy DNA sequencing at the Max Planck Institute for Marine Microbiology, Bremen, Germany.

Recombinant TconTS proteins were purified as described [14]. Briefly, CHO_{Lec1} cells (ATCC CRL-1735) were transfected with polyethylenimine, transfection reagent (Sigma, Steinheim, Germany) and stably expressing cell lines selected with hygromycin. Expression of recombinant protein was tested from cell culture supernatant by SDS-PAGE and Western blots methods using rabbit anti-*Strep* and anti-SNAP antibodies. CHO_{Lec1} cells producing TconTS proteins were subsequently adapted to chemically defined Excel CD CHO media.

Purification of anti-TS1 monoclonal antibody

The 7/23 hybridoma cells [12] were grown for 3 days in RPMI media supplemented with IgG depleted 10% FCS. The tissue culture supernatant was cleared by ultracentrifugation at 105×g for 60 min and anti-TconTS antibody was purified using recProtein-A Sepharose Fast Flow and eluted with 0.1 M glycine/HCl pH 3.0. Antibody containing fractions were neutralised with 1M Tris pH 8.0 and dialysed against 10 mM phosphate buffer. Purified antibodies were used in the detection of TconTS proteins in SDS-PAGE and Western Blot analysis as described [14].

Trans-sialidase and sialidase reactions

Purified recombinant proteins were assayed for sialidase and TS activities using Neu5Ac-MU and fetuin as sialic acid donor substrates and lactose as acceptor substrate as described before [14]. In brief, reactions of 50 µL containing substrates and enzymes were incubated at 37°C for the times indicated. Sialidase activity was determined as free sialic acids released from Neu5Ac-MU, 3'SL or fetuin in the absence and/or presence of an acceptor substrate. TS activity on the other hand was determined as 3'SL produced in the presence of lactose. Both, free Neu5Ac and 3'SL were quantified using high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using the Dionex system, DX600 (Dionex Sunnyvale, CA, USA) [14]. The curve fit module of SigmapPlot 11 was used to calculate v_{max} and K_M employing the Michaelis-Menten equation $v = v_{max} \times C_s/(C_s+K_M)$.

Phylogenetic analysis

For the phylogenetic analysis TconTS1b, TconTS2, TconTS3 and TconTS4 were aligned with TS and sialidase sequences from *T. brucei*, *T. vivax*, *T. cruzi* and *T. rangeli*. As outgroup the sialidase from *Vibrio cholerae* was used (genes listed in Supporting Information Table S2). Full length protein sequences were first aligned using ClustalW in Geneious and then truncated at the N-terminus. CDs started from the FRIP region to the N-terminus of the conserved α-helix (HL) linking the CD to the LD. The LDs were taken immediately after the α-helix linkage to the C-terminus

without the stop codon. DNA sequences encoding either full length proteins, the CDs or the LDs were aligned based on the alignment obtained for the amino acid sequences by T-Coffee algorithm in RevTrans, version 2.0 (http://www.cbs.dtu.dk/services/RevTrans-2.0/web/). DNA sequences used in the phylogenetic calculations are shown in Supporting Information Files S1, S2 and S3. Best parameters (HKY substitution model with 6 gamma rate categories) for phylogenetic constructions were determined using MEGA5 and applied in the phylogenetic calculations using the "MrBayes" plug-in of Geneious.

Results

T. congolense sialidase/*trans*-sialidase genes

Partial coding sequences of TconTS1 and TconTS2 genes had been described [13]. From "reads" of the WTSI *T. congolense* genome sequencing project (http://www.sanger.ac.uk), we assembled the full length sequences coding TconTS1 and TconTS2. Further BLAST hits with smallest sum probabilities ($P(N)<1–10$) were identified and arranged into contiguous sequences leading to further five genes with sequence similarities. Two of the putative gene products shared over 40% sequence identity with TconTS1 and TconTS2 and contained all the conserved amino acids required for transfer reactions [17] [18] [10]. Consistent with the naming of TconTS1 and TconTS2 [13], we refer to them as TconTS3 and TconTS4. The other three genes were distantly related with 20–30% amino acid identity (Table 1) and lack several of the conserved amino acid residues. We assume that these set of genes are likely without sialidase or TS activity and were named TconTS-Like1, TconTS-Like2, and TconTS-Like3.

In order to compare sequence similarities between TconTS genes, we cloned and sequenced full lengths TconTS2, TconTS3 and TconTS4. In an earlier study, we amplified eleven highly similar (about 96% identical amino acids) but clearly different sequences of TconTS1 from *T. congolense* genomic DNA [14]. Sequencing several clones of TconTS2, TconTS3 and TconTS4 provided no evidence for such heterogeneity of these genes. The alignment of these genes is given in Figure 1.

Amaya *et al.* [18] identified amino acids in *T. cruzi* TS involved in the catalytic and substrate binding. Whereas these amino acids are not conserved in the three TconTS-like gene products, they are almost completely conserved in TconTS1, TconTS2, TconTS3 and TconTS4 (Table 2). Most of these are conservative changes, with the exception of position 293 (numbering in the consensus sequence), where a Tyr is replaced by Pro in TconTS2. Furthermore, based on mutagenesis experiments [19] [17] with *T. cruzi* TS, two proline residues corresponding to positions 411 and 465 appear to be required for full TS activity. Whereas at position 465 Pro is conserved across the TconTS, it is not found in the TconTS-like gene products (Table 2). However, at position 411 the Pro is not conserved in TconTS, but replaced by Ala or Ser.

TS orthologues occur in *T. congolense* and *T. brucei* but not in *T. vivax*

To decipher the phylogenetic relationship between TS and TS-like genes of African trypanosomes, we compared the four TconTS and three TconTS-Like sequences together with seven sequences from *T. brucei* and five from *T. vivax* (Supporting Information Table S2) using the alignment of DNA sequences reverse transcribed from the protein alignment (see Supporting Information Files S1, S2 and S3 for DNA sequences used). As shown in Figure 2A, for each TconTS and TconTS-Like gene a corresponding orthologue was identified in *T. brucei*, whereas *T. vivax* gene products cluster separately from TS of the other African trypanosomes. TconTS-Like2 and TconTS-Like3 form a branch together with their *T. brucei* orthologues separate from all South American TS genes. In contrast, TconTS-Like1 and its *T. brucei* orthologue appear to be more closely related with the more distant South American branch than the African genes.

Trypanosomal TS contain an N-terminal CD followed by a C-terminal LD. Besides the phylogenetic analysis with the entire ORFs, analysis were performed using sequences coding for each domain separately. The phylogenetic tree obtained for the CDs resembles that for complete ORFs (not shown). Surprisingly, a different situation was observed for the LDs of TconTS1, TconTS2, TconTS3 and TconTS4 and the *T. brucei* TS genes (Figure 2B). First, within the TconTS genes, the LD of TconTS2 is most closely related to that of TconTS1, whereas the CD of TconTS3 is more closely related to TconTS2. Second, it should be noted that not the same *T. congolense* and *T. brucei* genes group as orthologous pairs, if LDs are compared. Amplification and sequencing ORF of *T. brucei* TS genes confirmed that the combination of the CDs and LDs were as predicted from the contigs in the databases.

Monclonal anti-TS1 antibody cross-reacts with TconTS2, recognising an epitope on the lectin domain

To biochemically characterise TconTS genes, recombinant proteins were made for TconTS2, TconTS3 and TconTS4 as previously described for TconTS1 [14]. Recombinant TconTS proteins were expressed in CHO_{Lec1} [20] and purified by affinity chromatography to obtain pure protein from eukaryotic cells with high mannose-type *N*-glycans. The apparent molecular masses of the recombinant TconTS proteins including the SNAP and *Strep* tags are between 110 and 125 kDA as resolved on SDS-PAGE. The generic *Strep* tag fused to the proteins is recognised by anti-*Strep* polyclonal Ab in all the recombinant TconTS proteins as shown in Figure 3 (upper panel). Surprisingly, the monoclonal anti-TS antibody [12] reacted with both TconTS1 and TconTS2,

Table 1. *Trypanosoma congolense trans*-sialidase sequence similarities expressed as percentage of identical amino acids in pair-wise alignments.

Trans-sialidase	TconTS2	TconTS3	TconTS4	TconTS-Like1	TconTS-Like2	TconTS-Like3
TconTS1	42.2%	43.6%	46.2%	21.1%	26.3%	29.8%
TconTS2	-	48.3%	42.8%	20.8%	26.2%	29.3%
TconTS3		-	48.9%	21.1%	25.1%	29.9%
TconTS4			-	21.3%	27.8%	29.8%

Tcon = *Trypanosoma congolense*

doi:10.1371/journal.pntd.0002549.t001

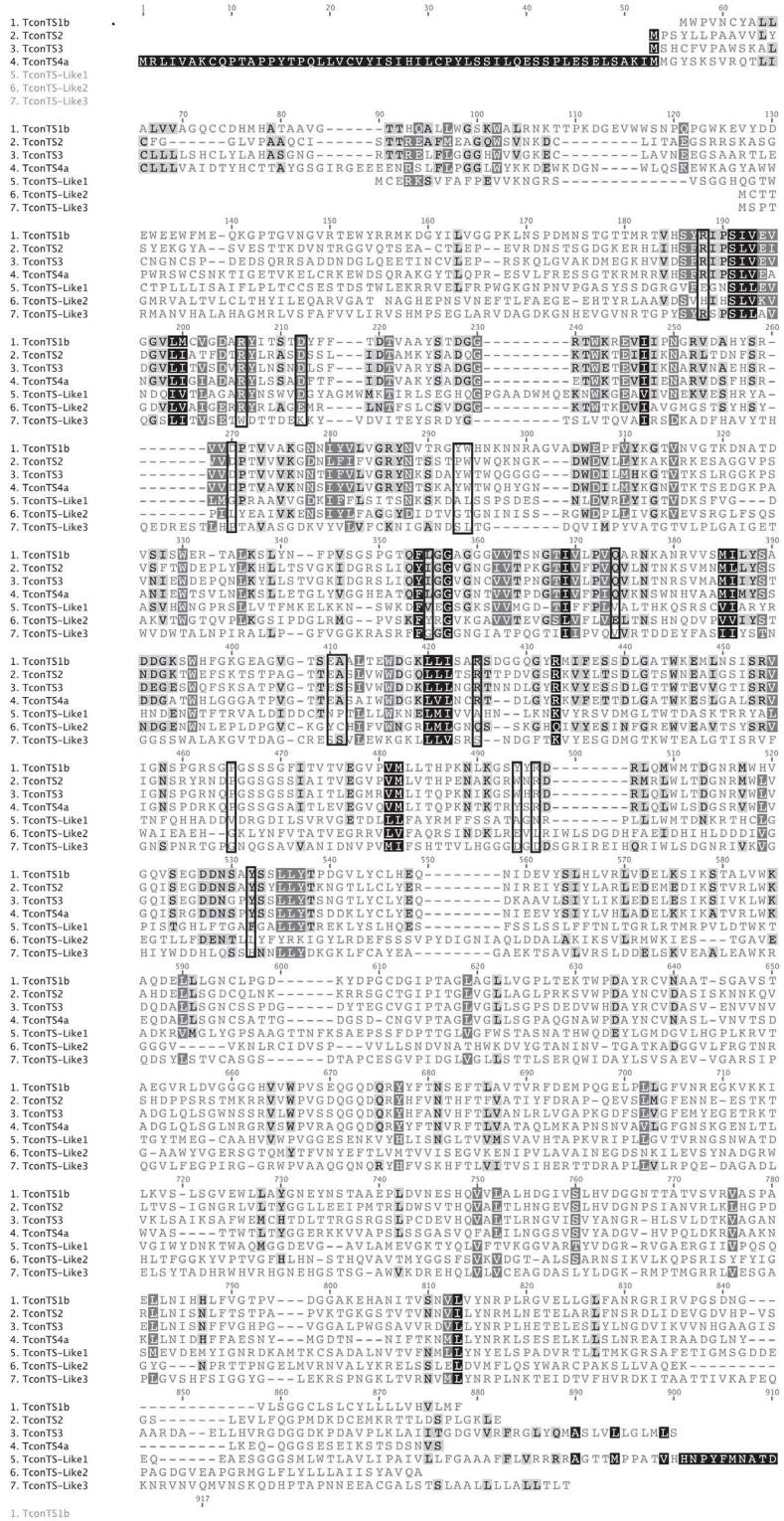

1. TconTS1b
2. TconTS2
3. TconTS3
4. TconTS4a
5. TconTS-Like1
6. TconTS-Like2
7. TconTS-Like3

Figure 1. Primary amino acid sequence alignment of TconTS and TconTS-Like genes. Amino acid sequences of TconTS1b (EMBL:HE583284), TconTS2, TconTS3 and TconTS4 were obtained by sequencing of cloned genes. Those of TconTS-Like 1, TconTS-Like 2 and TconTS-Like 3 were obtained from the WSTI database. CluwstalW alignment plugin of the Geneious software was used employing BLOSUM12 with gap openings and extension penalties of 10 and 0.1, respectively. Increasing darkness of background indicates increasing numbers of identical amino acids at each position. The numbers on top of the sequences indicate the positions in the consensus sequence. Amino acid residues postulated to be critical for catalysis, substrate binding and structure as given in Table 2 are boxed.
doi:10.1371/journal.pntd.0002549.g001

but not TconTS3 and TconTS4 (Figure 3; lower panel). This result points to a similar epitope being present in both TconTS1 and TconTS2. Further experiments provided evidence that the epitope is located in the LDs.

TconTS2, TconTS3 and TconTS4 are trans-sialidases

In order to investigate the enzymatic activities we used the glycoprotein fetuin as donor and lactose as acceptor substrates. Free sialic acid (the product of sialidase activity) and 3'SL (the product of TS activity) could be quantified simultaneously from HPAEC-PAD chromatograms of the reactions. Under standard conditions 25 ng TconTS2 produced about 330 pmol/min 3'SL leading to 200 μM 3'SL in the reaction mix within 30 minutes (Figure 4A). Product formation by TconTS2 was linear for up to 50 ng enzyme under these conditions. The reaction catalysed by TconTS3 was slower than that of TconTS2, since 500 ng of enzyme generated only 4.2 pmol/min 3'SL corresponding to 10 μM 3'SL after 4 h (Figure 4B). 3'SL formation by TconTS3 was almost linear for 4 h. TS activity was also detected for TconTS4. However, the activity was even lower than that of TconTS3 and 500 ng of TconTS4 produced less than 0.1 pmol/min 3'SL (Figure 4C). Therefore, 24 h incubations were routinely used to determine TconTS4 activity.

To determine kinetic parameters (Table 3) reactions were conducted under standard conditions for 30, 120 and 1440 minutes for TconTS2, TconTS3 and TconTS4, respectively. TconTS2 exhibited the lowest K_M for fetuin (299 μM), similar to that reported for TconTS1b (395 μM) [14], whereas those determined for Tcon TS3 (6090 μM) and TconTS4 (949 μM) were higher. The lowest K_M for lactose was found for TconTS2 (602 μM), followed by TconTS3 (1104 μM) and TS4 (1806 μM). Comparing the results with those obtained for two variants of TconTS1 [14], TconTS2 has a K_M value for lactose similar to TconTS1e-1 but about 3-fold less than TconTS1b. For both substrates, TconTS2 showed about 2-, 200- and 2000-fold higher v_{max} values than TconTS1, TconTS3 and TconTS4, respectively.

Sialidase activities

Sialidase activity has been shown for TconTS purified from *T.congolense* axenic culture [12] and for *T. congolense* infected animals [5] [7]. Therefore, we investigated the sialidase activities of TconTS using fetuin as a model glycoprotein. Whereas no release of free Neu5Ac was observed for TconTS1, TconTS2 and TconTS3 under standard conditions of TS assays, TconTS4 clearly showed sialidase activity producing 0.76 pmol/min Neu5Ac up to 48 h (Figure 5A). Indeed, the sialidase activity of TconTS4 is relatively stable and retained a residual sialidase activity of 40% even after incubation at 37oC for 120 days.

We also investigated the effect of lactose on TconTS4 sialidase activity. At lactose concentrations above 3 mM, release of Neu5Ac dropped to undetectable levels (Figure 5B). This result indicates the existence of a competition between lactose and water for the cleaved Neu5Ac from the donor fetuin. This is confirmed by the increased amount of 3'SL produced with increasing lactose concentration.

When we incubated TconTS2 or TconTS3 with fetuin and lactose for extended periods, it was observed that increasing amounts of Neu5Ac were released, similar to TconTS1 [14]. Interestingly, free Neu5Ac appeared only after 3'SL has accumulated as a product of transialylation. For example, in TconTS2 reactions, Neu5Ac was detectable (0.68 μM, 27 pmol) when the 3'SL concentration had reached almost 600 μM, the maximum 3'SL concentration reached. Whereas further incubation did not result in higher 3'SL concentrations, the amount of free Neu5Ac continuously increased. This observation suggests that TconTS2releases free Neu5Ac from 3'SL but not from fetuin. Similar observations were made for TconTS1, but the highest 3'SL concentration reached was about 300 μM. These data indicated that for TconTS1 and TconTS2 at 300 μM and 600 μM 3'SL, respectively, the transfer of Neu5Ac between fetuin and 3'SL has reached an equilibrium. For TconTS3, we could not reach such equilibrium; probably since the maximum 3'SL concentration obtained was 50 μM due to the low specific activity of this enzyme. Nevertheless, small amounts of free Neu5Ac were detected in prolonged TconTS3 reactions leading to 50 μM 3'SL. Also in this case, Sia appears to be released from 3'SL only, since in the absence of lactose no release of free Neu5Ac could be observed.

These observations suggest that TconTS1, TconTS2 and TconTS3 can release free Neu5Ac from glycoconjugates, but the level of product released is too low to be detected under standard TS assay conditions. To address this, higher amounts (500 ng) of TconTS1, TconTS2, TconTS3 and TconTS4 were incubated with 600 μM fetuin-bound Neu5Ac without lactose for 2 h. Under these conditions sialidase activities could also be detected for TconTS1 and TconTS2, but not for TconTS3 (Table 4). However, compared to the corresponding TS activities, the sialidase activities were very low.

Discussion

The enzymatic properties of four TconTS were compared using fetuin as a model for a blood glycoprotein donor substrate. Two of these enzymes, TconTS1 and TconTS2, exhibit about 100- or 1000-fold higher specific TS activities than TconTS3 and TconTS4, respectively. The K_M values for lactose were around 1 mM (0.6 to 1.8 mM) for all four TconTS and did not correlate with their specific activities. The K_M values for the donor fetuin were more different ranging from 0.4 to 6 mM glycoprotein bound sialic acids. The K_M for fetuin also did not correlate with the specific activity, since the highest K_M was determined for TconTS3, one of the enzymes with low activity, and TconTS4 has a similar K_M as TconTS1, but is 1000-fold less active. Since the K_M values are lower than the substrate concentrations used, especially for TconTS4, the specific activities given in Table 3 are lower than those to be expected, if the acceptor substrate lactose would be at saturating concentrations.

Table 2. Amino acids in the catalytic domains of TS and TS-Like genes from *T. congolense* involved in enzymatic activities*.

Consensus	Tcon TS1	Tcon TS2	Tcon TS3	Tcon TS4	Tcon TS-Like1	Tcon TS-Like2	Tcon TS-Like3
catalysis							
212	D150	D135	D142	D207	D110	E85	K86
410	E324	E309	E316	E381	N291	Y257	S262
532	Y438	Y423	Y430	Y493	F404	L375	H382
substrate binding							
188	R126	R111	R118	R183	E86	H61	R62
425	R339	R324	R331	R396	A306	Q272	S277
496	R410	R395	R402	R465	N375	L339	D346
206	R144	R129	R136	R201	R104	R79	W80
270	D188	D173	D180	D245	G160	L122	P132
293	Y211	P196	Y203	Y268	A183	G145	S155
294	W212	W197	W204	W269	L184	T146	L156
374	Q289	Q274	Q281	Q364	V255	E222	V227
494	Y408	W393	W400	Y463	A373	E337	D344
structure							
411	A325	A310	S317	A382	P292	C258	S263
465	P379	P364	P371	P434	V344	G308	G315

*The indicated amino acids have been selected based on structural [18] and mutation [19] [23] studies with *T. cruzi* TS and on the sequence alignment of TconTS1b with *T. cruzi* [14]. Amino acid positions have been numbered based on the consensus of alignment (Figure 1) or starting methionine of each ORF.
doi:10.1371/journal.pntd.0002549.t002

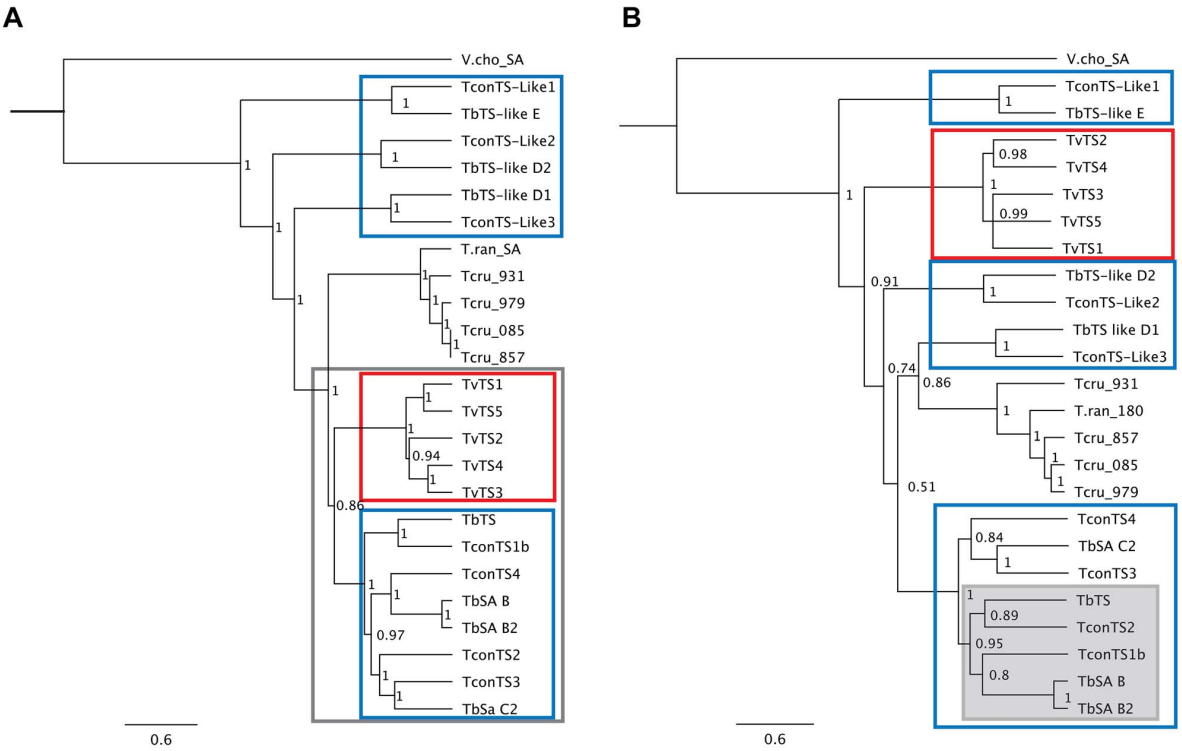


Figure 2. Phylogenetic analysis TS genes. Phylogenetic analyses of DNA sequences were performed as described under Methods using "MrBayes" plug-in of Geneious. Trees are midpoint rooted and nodes supported by posterior probability values and non-parametric bootstraps generated by maximum likelihood analysis in "MrBayes" as described under Methods. TS genes from *T. congolense* and *T. brucei* are marked by blue frames, TS genes from *T. vivax* by red frames. **A**: Phylogenetic tree for full ORFs, the "African TS-branch" is marked by a grey frame; **B**: Phylogenetic tree for LDs, the most active TconTS and their orthologues from *T. brucei* are highlighted by a grey box.
doi:10.1371/journal.pntd.0002549.g002

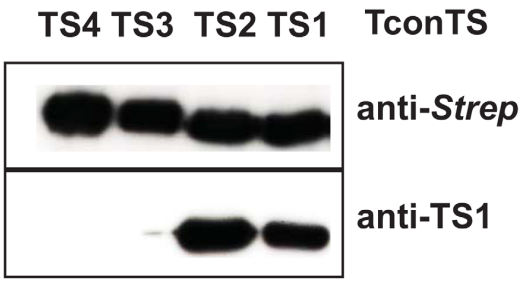


Figure 3. Recognition of TconTS proteins by antibodies. 100 ng of affinity purified TconTS proteins were analysed by Western blot as described under Methods. Blots were probed either with polyclonal rabbit anti-Strep (upper pane) or with monoclonal mouse anti-TS1 antibody, which shows cross reactivity with TconTS2 (lower pane).
doi:10.1371/journal.pntd.0002549.g003

Tiralongo *et al.* [12] purified two TconTS forms with different TS/sialidase activity ratios from procyclic *T. congolense* cultures. Peptide sequences from these preparations have demonstrated that they contained at least TconTS1 [12] [14]. Recombinant TconTS1 variants expressed in eukaryotic cells had lower specific activities for synthetic substrates than described for the purified enzyme [12] suggesting that these TconTS preparations also contained other enzymes and/or factors influencing the TS/sialidase activities [14]. For example, in those preparations, Glutamic Acid-Rich Protein, GARP, a natural substrate for TconTS was co-purified with one of the TS forms [12]. Although it is unclear what role GARP might have played, its presence, as well as other TS enzymes, might have been responsible for the reported higher specific activities of these TS preparations for synthetic substrates. In summary, it appears that TconTS1 and TconTS2 are responsible for most of the TS activity of *T. congolense*. However, it may well be that TconTS3 and TconTS4 are more active on other donor substrates, such as glycoproteins and/or glycolipids from blood components, the natural substrates for trypanosomal TS.

It has been established that procyclic forms of African trypanosomes express TS [21] and emerging evidences point to expression also in the blood stream forms [6] [7]. So far, no information is available on which TS genes are expressed at what stage of the parasite's life cycle. Recently, we have identified mRNAs for TconTS1, TconTS2 and TconTS4 in the blood of infected goats (data not shown). Also the stability and persistence of shed enzymes in the blood stream has to be taken into account. *In vitro* TconTS1 and TconTS3 are the most stable of the four enzymes investigated, retaining full TS activities even after four months at 37°C. Under these conditions TconTS2 lost its activity completely, while TconTS4 retained 40% residual sialidase but no transfer activity (data not shown). It would be interesting to investigate whether this long-term stability correlates with a sustained persistence of enzyme activity in the blood stream.

Lactose was found to suppress the sialidase activity of TconTS4. *In vitro*, lactose appears to be a better acceptor than water (Figure 5B). Therefore, in presence of lactose, the transfer activity of TconTS4 is more efficient than its hydrolytic activity. Anaemia in animals suffering African *Trypanosoma* infections has been attributed to the effects of sialidases [22] [5] [7]. Desialylation of erythrocytes by sialidases exposes underlying galactose residues and their subsequent degradation. The presence of lactose in mammalian blood would lead to lowered efficiency of parasites sialylation and eventual clearance by the immune system. However, desialylation of parasites is equally possible in presence of lactose due to the action of TS. Along this line it is interesting to note that infusion of lactose in the blood of sheep suffering experimental anaemia from *T. congolense* suppressed anaemia (unpublished observation).

All amino acid residues shown to be involved in the catalytic reaction or interaction with the substrate for *T. cruzi* TS are conserved in the TconTS enzymes (Table 2). Only the two residues interacting with the methylumbelliferyl aglycon or the lactose part of 3'SL in the *T. cruzi* TS [18], positions 293 and 494 in consensus sequence (Figure 1), are different in the two most active TconTS1 and TconTS2. This could explain why these enzymes do not use Neu5Ac-MU as substrate (data not shown). Furthermore, these changes could lead to a weaker interaction with the leaving groups and thus facilitate their release during catalysis. In this context it is interesting to note that the most drastic change, Tyr to Pro at position 293, occurs as P196 in TconTS2, the most active enzyme with the highest TS/sialidase ratio (Table 4). Certainly, this modification will reduce the interaction with hydrophobic aglycons or the leaving galactose residue of the donor substrate.

Amaya *et al.* [18] also provided evidence that in *T. cruzi* TS Y119 (position 293 in the consensus sequence) also contributes to hydrogen bonding with O9 of the covalently bound Sia following a conformational change induced by the reaction. Such an interaction would not be possible in TconTS2, but could be compensated by hydrogen bonding with the conserved side chains W197 and Q274 of TconTS2 corresponding to W120 and Q195 in *T. cruzi* TS, two amino acids contributing to the hydrogen bonding network of O9 in the covalently bound Sia [18].

Similarly, although to a lesser degree, the replacement of a Trp at position 494 (corresponding to W312 in *T. cruzi* TS) with a Tyr, as found in TconTS1, is expected to reduce the hydrophobicity of this site leading to a reduced affinity for the leaving group. In *T. cruzi* TS substitution of this Trp (W312) by Ala basically abolished Neu5Ac transfer but only slightly decreased hydrolytic activity for 3'SL [23]. Interestingly, in contrast to the wild type *T. cruzi* TS, this mutant was not able to hydrolyse Neu5Ac-MU, similar to TconTS1, which also does not accept Neu5Ac-MU as a substrate [14].

It appears that TS activity depends on well controlled conformational changes [17] influenced by specific proline residues. This is supported by the potential of the *T. rangeli* sialidase to acquire transfer ability due to a change of Gln to Pro at position 284 [24] and the loss of enzymatic activity in *T. cruzi* TS by the reverse mutation [16]. At the corresponding position 465 Pro is found in all active TconTS (Table 2). The relevance of conformational changes in the enzyme rather than a direct specific interaction of the amino acid were also indicated by mutation of Pro231 to Ala in *T. cruzi* TS [19], corresponding to position 411 in the consensus sequence. Although this mutation led to a significant decrease in *T. cruzi* TS activity, all active TconTS have an Ala or Ser at this position. It would be interesting to see, if higher TS activities can be obtained by introducing a Pro at this position, particularly in TconTS3 or TconTS4, the two enzymes with low TS activities.

Besides these critical amino acids listed in Table 2, other structural features obviously control the ratio of TS versus sialidase activities, since TconTS2 and TconTS4 share identical amino acids at all these positions. Yet, TconTS4 has the highest sialidase to TS ratio amongst the TconTS enzymes, whereas TconTS2 has the lowest ratio (Table 4). Koliwer-Brandl *et al.* [14] observed for TconTS1 that a natural mutation that replaced R144 (206 in

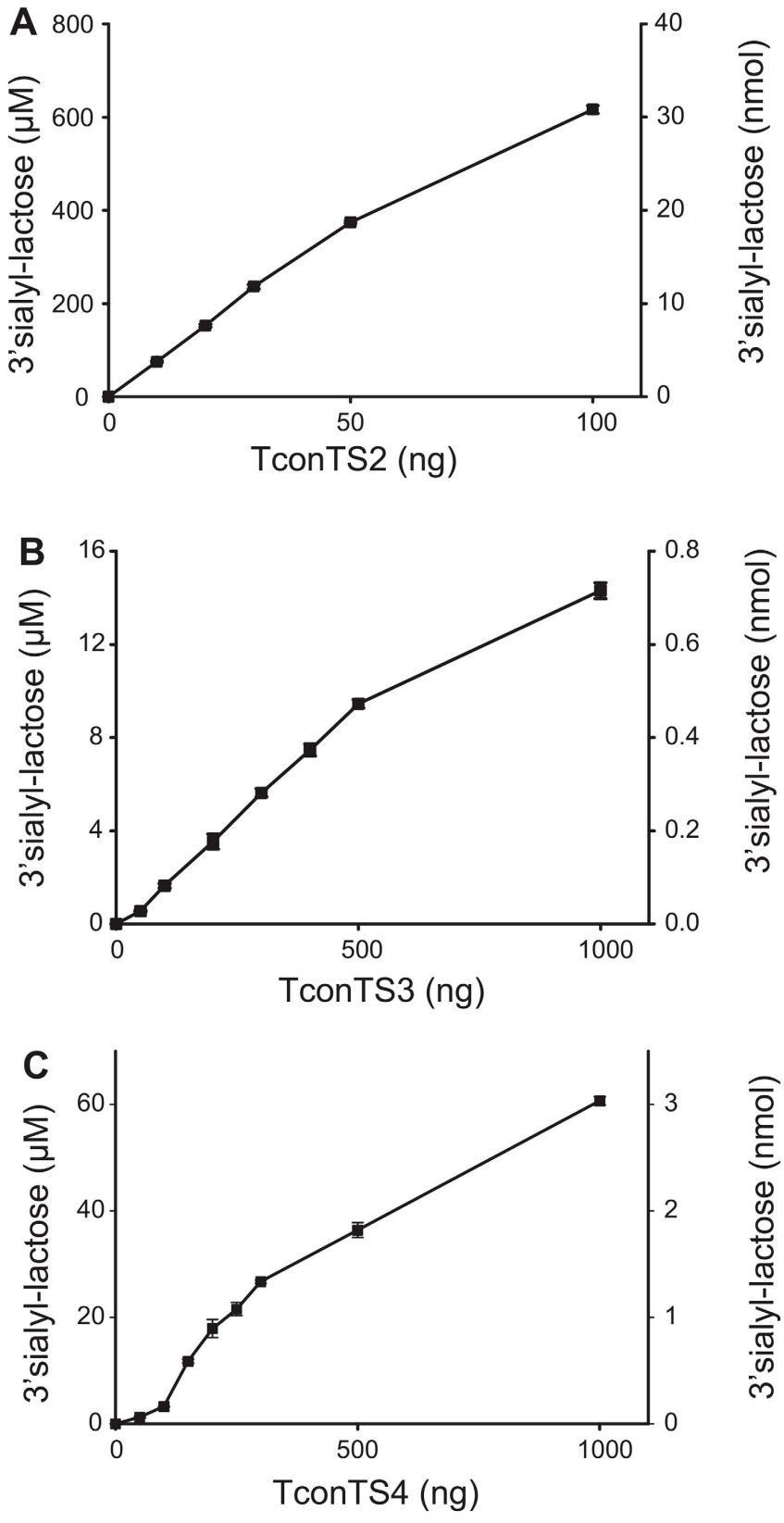

A
800
600
400
200
0
3'sialyl-lactose (μM)
40
30
20
10
0
3'sialyl-lactose (nmol)
0
50
100
TconTS2 (ng)
B
16
12
8
4
0
3'sialyl-lactose (μM)
0.8
0.6
0.4
0.2
0.0
3'sialyl-lactose (nmol)
0
500
1000
TconTS3 (ng)
C
60
40
20
0
3'sialyl-lactose (μM)
3
2
1
0
3'sialyl-lactose (nmol)
0
500
1000
TconTS4 (ng)

Figure 4. Concentration dependency of TS activity. The indicated amounts of recombinant TconTS proteins were incubated with 100 μg fetuin (600 μM bound Sia) and 2 mM lactose for 30 (TconTS2), 240 (TconTS3), or 1440 (TconTS4) minutes. 3'SL produced was determined by HPAEC-PAD as described under Methods. Data points are means ± standard deviations of three independent experiments, each as triplicates. **A**: TconTS2, **B**: TconTS3, **C**: TconTS4.
doi:10.1371/journal.pntd.0002549.g004

consensus) sequence with Cys in the variants TconTS1g (EMBL: HE582290) did not terminate but only reduced TS activity, while increasing relative hydrolytic activity. It can be assumed that a weaker interaction with the hydroxyl group at C4 of sialic acid is responsible for the catalytic properties of TS1g, since the Arg (R53 in *T. cruzi* TS) is in close contact with the bound sialic acid and probably supports the stabilisation of the enzyme-substrate complex [18].

Conservation and/or substitution of amino acids in the active centre of the catalytic domain did not give clear indications of activity differences between the TconTS enzymes. In this context, an interesting aspect is how the different specific activities of TconTS enzymes correlate with those of related gene products from other African trypanosomes. Phylogenetic analyses have allowed the clear assignment of orthologues for *T. brucei*, but not for *T. vivax*, where TS genes clustered exclusively together and away from the TS genes of *T. congolense* and *T. brucei* (Figure 2) [7] [25]. A direct comparison of the TS activities between *T. congolense* and *T. brucei* is difficult, since limited comparable data for enzymatic activities is available. In *T. brucei* TbTS and TbSA C2 have been identified as active TS [9] [10] [11]. This is consistent with the observation that their orthologues (TconTS1 and TconTS2, respectively) are the most active TS in *T. congolense*. RNA*i* based experiments provided evidence that in *T. brucei* TS and sialidase activities are encoded by different genes, TbTS and TbSA C [10], whereas assays with purified recombinant proteins demonstrated both enzyme activities for TbTS and TbSA C2 [11]. Noticeably, the diversity of TS-related genes in *T. vivax* is lower than what was obtained for *T. congolense* and *T. brucei*. Equally, *T. vivax* is distinctively different from *T. congolense* and *T. brucei* in terms of development in the insect host. While the later two develop in the insect midgut and proboscis or salivary glands respectively, *T. vivax* develops exclusively in mouthparts of the tsetse. Moloo and Gray [26] showed that *T. vivax* ingested with blood meal to the midgut is disintegrated. TS-like genes from *T. vivax* share all but two (consensus sequence positions 411 and 494) of the conserved amino acids listed in Table 2 with the active TconTS. Recently Guegan *et al.* [7] reported that TvivTS2 has TS activity. Furthermore, they obtained evidence for the presence of at least TvivTS1, TvivTS3 and TvivTS5 in the bloodstream form and none in epimastigotes and the possible involvement of these proteins in anaemia in infected mice. In summary, it may be possible that *T. vivax* is missing a suitable TS to survive and colonise the fly vector midgut. To this end, it would be interesting to see the survival ability of transgenic *T. vivax* expressing a TS, which is expressed by *T. congolense* in the midgut of tsetse flies.

The presence of multiple highly similar TS genes, as described for TconTS1 [14], suggests that these genes undergo active rearrangements, which could lead to strain specific differences. For example, Coustou *et al.* [6] referred to two highly related TconTS3 genes in the IL3000 strain identified in GeneDB and TrytrypDB databases. However, we could not find evidence for their existence in the STIB294 strain used in this study. Similarly, closely related genes with over 80% sequence identity have been identified for TbSA B and TbSA C in *T. brucei* [11].

TS and sialidase genes of African trypanosomes are organised in two major domains; the CD and the LD. The LDs of TconTS are more varied (40% pairwise identity) when compared with the CDs (58% pairwise identity). Surprisingly, the phylogenetic relationships between the TS are clearly different, if only the LDs are included in the analysis (Figure 2). Furthermore, the LDs of the two most active enzymes TconTS1 and TconTS2 are more closely related than the CDs, where TconTS2 is most closely related to TconTS3. Interestingly, the monoclonal anti-TS1 antibody also binds TconTS2, recognising an epitope in the LD. First preliminary experiments obtained with recombinant proteins, in which the LDs have been swapped between TconTS, provided supporting evidence that the LD influences TS and sialidase activities of the enzymes (data not shown). However, the specific activities of these proteins expressed in bacteria is much lower than those of the proteins expressed in fibroblasts described here,

Table 3. Kinetic parameters of TconTS.

	Donor substrate (fetuin-bound Sia)*		Acceptor (lactose)**	
	v_{max}*** (μmol/(min×mg TS))	K_M*** (μM)	v_{max}*** (μmol/(min×mg TS))	K_M*** (μM)
TS1b****	7.9±0.3	359±45	4.3±0.1	1683±101
TS1e-1****	7.6±0.5	1617±223	2.1±0.1	727±48
TconTS2	17.62±0.13	299.00±7.0	17.85±0.13	602±16
TconTS3	0.17±0.02	6090.00±1267	0.0567±0.0014	1104±79
TconTS4	0.0067±0.0002	949±50	0.0075±0.0002	1806±112

*Approximately 30 nmol Sia per 100 μg fetuin; 2 mM lactose was used as acceptor substrate.
**600 μM fetuin-bound Sia was used as donor substrate.
***K_M and v_{max} were calculated from Michaelis-Menten kinetics (see Supplementary Information, Figure S1) by SigmaPlot. Data points are mean ± standard deviations of three independent experiments, each replicated thrice.
****values from Koliwer-Brandl et al. [14].
doi:10.1371/journal.pntd.0002549.t003

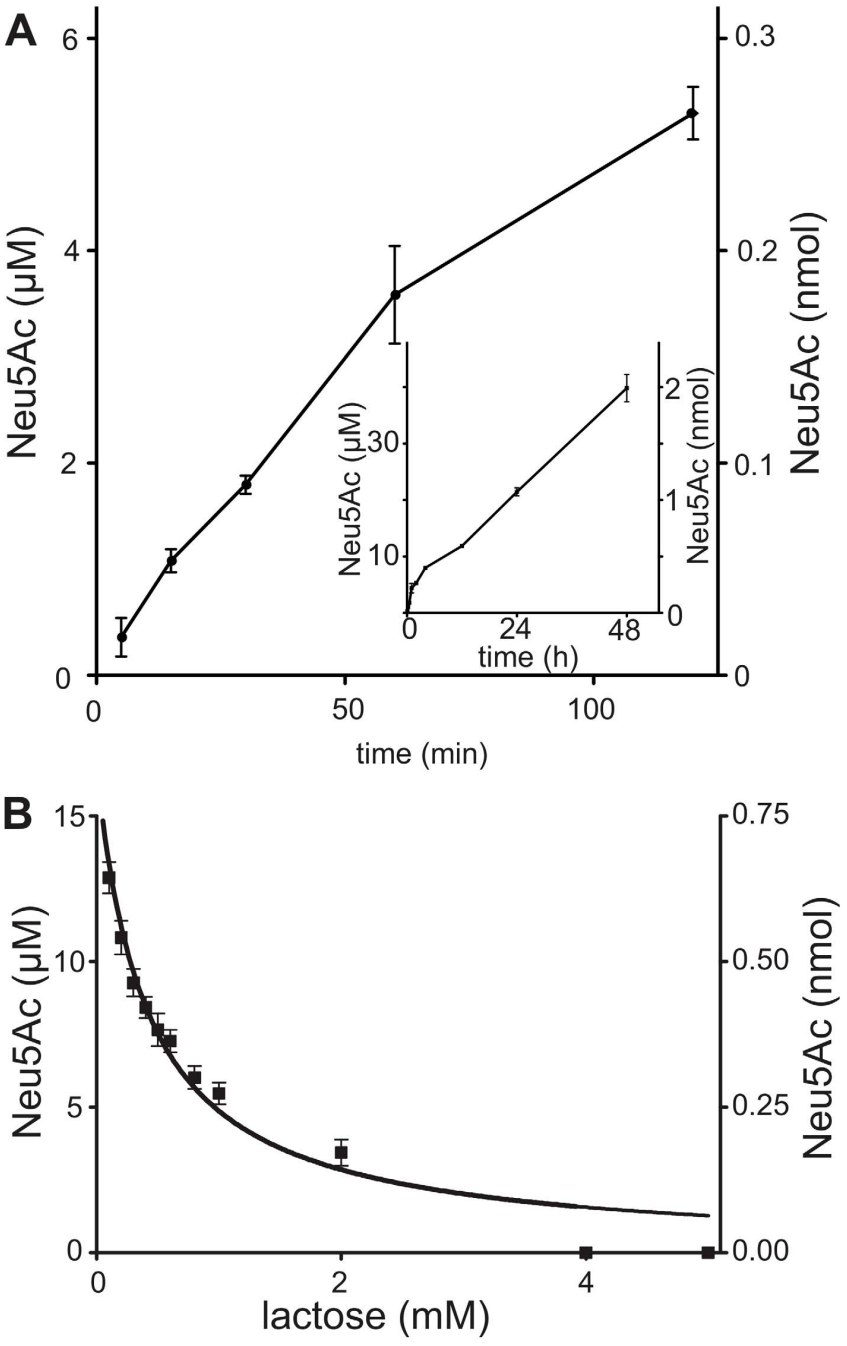


Figure 5. Sialidase activity of TconTS4. A: Recombinant TconTS4 was incubated with 100 µg fetuin (600 µM bound Sia) for the times indicated (see insert for long term reactions) and free Sia determined by HPAEC-PAD as described under Methods. Data points are means ± standard deviations of 3 independent reactions each triplicated. **B**: Recombinant TconTS4 was incubated for 1440 min with 100 µg fetuin (600 µM bound Sia) in the presence of the lactose concentrations indicated and free Sia determined by HPAEC-PAD as described under Methods. Data points are means ± standard deviations of 3 independent reactions each as triplicates.
doi:10.1371/journal.pntd.0002549.g005

suggesting that for conclusive interpretation the domain swapped TconTS have to be expressed in eukaryotic cells and that further studies are necessary to optimise the fusion of the two domains. In summary, these data indicate a more significant role for the LD for the TS activities of TS1 and TS2 and thus possibly in the pathogenesis of African trypanosomiasis.

Table 4. Sialidase activities of TconTS.

	Sialidase* (nmol/(min×mg TS))	Trans-sialidase** (nmol/(min×mg TS))	TS/sialidase
TconTS1	59.7±2.3	4,000±100	67
TconTS2	73.5±4.0	17,850±130	243
TconTS3	n.d.***	34±1	>30
TconTS4	2.7±0.07	9±0.2	3.4

*Sialidase activities were determined by incubating 500 ng of the TconTS indicated for 2 h with 100 µg fetuin (600 µM fetuin-bound Sia). Data points are mean ± standard deviations of three replicates.
**TS activity values see Table 3.
***n.d. not detected.
doi:10.1371/journal.pntd.0002549.t004

Supporting Information

Figure S1 Trans-sialidase reaction velocities depending on substrate concentrations. Product (3′-sialyl-lactose) amounts were determined as described under Methods. v_{max} and K_M for lactose shown in Table 3 were calculated from these data. Data points are mean ± standard deviations of three independent experiments, each replicated thrice. **A) TconTS2 with different donor substrate concentrations.** 50 ng TconTS2 were incubated for 30 minutes with 2 mM lactose and the indicated concentrations of fetuin-bound Sia. **B) TconTS2 with different acceptor substrate concentrations.** 50 ng TconTS2 were incubated for 30 minutes with 600 µM fetuin-bound Sia and the indicated lactose concentrations. **C) TconTS3 with different donor substrate concentrations.** 250 ng TconTS3 were incubated for 120 minutes with 2 mM lactose and the indicated concentrations of fetuin-bound Sia. **D) TconTS3 with different acceptor substrate concentrations.** 500 ng TconTS3 were incubated for 120 minutes with 600 µM fetuin-bound Sia and the indicated lactose concentrations. **E) TconTS4 with different donor substrate concentrations.** 500 ng TconTS4 were incubated for 1440 minutes with 2 mM lactose and the indicated concentrations of fetuin-bound Sia. **F) TconTS4 with different acceptor substrate concentrations.** 500 ng TconTS4 were incubated for 1440 minutes with 600 µM fetuin-bound Sia and the indicated lactose concentrations.
(PDF)

File S1 Nucleotide sequences of the sialidase and trans-sialidase genes used in the phylogenetic comparison of "full length open reading frames" (catalytic plus lectin domains). The gaps inserted for the alignment as described under Methods are indicated by dashes. The file is a text file in FASTA format with the gene names (see Table S2) in the first line for each gene.
(TXT)

File S2 Nucleotide sequences of the sialidase and trans-sialidase genes used in the phylogenetic comparison of "catalytic domains". The gaps inserted for the alignment as described under Methods are indicated by dashes. The file is a text file in FASTA format with the gene names (see Table S2) in the first line for each gene.
(TXT)

File S3 Nucleotide sequences of the sialidase and trans-sialidase genes used in the phylogenetic comparison of "lectin domains". The gaps inserted for the alignment as described under Methods are indicated by dashes. The file is a text file in FASTA format with the gene names (see Table S2) in the first line for each gene.
(TXT)

Table S1 List of primers used for cloning and mutagenesis. Listed are the primers used in this study for cloning, expression plasmids and mutagenesis as described under Methods.
(PDF)

Table S2 Trans-sialidase and sialidase genes used for phylogenetic analysis. Listed are accession numbers and literature references for the genes used in the phylogenetic analysis in this study.
(PDF)

Acknowledgments

We are grateful to Evelin Tiralongo and Roland Schauer (University Kiel, Germany) for providing the *T. congolense* STIB294 genomic DNA, Markus Engstler (University Würzburg, Germany) for providing the *T. brucei* EATRO427 genomic DNA, Hans Lange and Hilmar Lemke (University Kiel, Germany) for the cell line producing anti-TconTS antibody mAb 7/23. The plasmid vector pDEF was kindly provided by Prof. Paul Crocker (University of Dundee, UK). We thank Dr. Hendrik Koliwer-Brandl for fruitful discussions on enzyme purification and characterisation.

Author Contributions

Conceived and designed the experiments: SK JAN. Performed the experiments: TTG MW BH JS GDC SK. Analyzed the data: TTG MW JS SK. Contributed reagents/materials/analysis tools: FD. Wrote the paper: TTG MW JAN SK.

References

1. Steverding D (2008) The history of African trypanosomiasis. Parasites & Vectors 1: 3. doi:10.1186/1756-3305-1-3.
2. Moti Y, Fikru R, Van Den Abbeele J, Büscher P, Van den Bossche P, et al. (2012) Ghibe river basin in Ethiopia: present situation of trypanocidal drug resistance in *Trypanosoma congolense* using tests in mice and PCR-RFLP. Veterinary Parasitology 189: 197–203. doi:10.1016/j.vetpar.2012.04.022.
3. dC-Rubin SSC, Schenkman S (2012) *Trypanosoma cruzi* trans-sialidase as a multifunctional enzyme in Chagas' disease. Cell Microbiol 14: 1522–30. doi:10.1111/j.1462-5822.2012.01831.x.
4. Nagamune K, Acosta-Serrano A, Uemura H, Brun R, Kunz-Renggli C, et al. (2004) Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors. Journal of Experimental Medicine 199: 1445–1450. doi:10.1084/jem.20030635.
5. Nok AJ, Balogun EO (2003) A Bloodstream *Trypanosoma congolense* sialidase could be involved in anemia during experimental trypanosomiasis. Journal of Biochemistry 133: 725–730. doi:10.1093/jb/mvg093.
6. Coustou V, Plazolles N, Guegan F, Baltz T (2012) Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. Cell Microbiol 14: 431–445. doi:10.1111/j.1462-5822.2011.01730.x.

7. Guegan F, Plazolles N, Baltz T, Coustou V (2013) Erythrophagocytosis of desialylated red blood cells is responsible for anaemia during *T. vivax* infection. Cell Microbiol 15: 1285–303. doi:10.1111/cmi.12123.
8. De Pablos LM, Osuna A (2012) Multigene families in *Trypanosoma cruzi* and their role in infectivity. Infection and Immunity 80: 2258–2264. doi:10.1128/IAI.06225-11.
9. Montagna G, Cremona ML, Paris G, Amaya MF, Buschiazzo A, et al. (2002) The trans-sialidase from the african trypanosome *Trypanosoma brucei*. European Journal of Biochemistry 269: 2941–2950. doi:10.1046/j.1432-1033.2002.02968.x.
10. Montagna GN, Donelson JE, Frasch ACC (2006) Procyclic *Trypanosoma brucei* expresses separate sialidase and trans-sialidase enzymes on its surface membrane. J Biol Chem 281: 33949–33958. doi:10.1074/jbc.M604951200.
11. Nakatani F, Morita YS, Ashida H, Nagamune K, Maeda Y, et al. (2011) Identification of a second catalytically active trans-sialidase in *Trypanosoma brucei*. Biochemical and Biophysical Research Communications 415: 421–425. doi:10.1016/j.bbrc.2011.10.085.
12. Tiralongo E, Schrader S, Lange H, Lemke H, Tiralongo J, et al. (2003) Two trans-sialidase forms with different sialic acid transfer and sialidase activities from *Trypanosoma congolense*. Journal of Biological Chemistry 278: 23301–23310. doi:10.1074/jbc.M212909200.
13. Tiralongo E, Martensen I, Grötzinger J, Tiralongo J, Schauer R (2003) Trans-sialidase-like sequences from *Trypanosoma congolense* conserve most of the critical active site residues found in other trans-sialidases. Biological Chemistry 384: 1203–1213.
14. Koliwer-Brandl H, Gbem TT, Waespy M, Reichert O, Mandel P, et al. (2011) Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*. BMC Biochemistry 12: 39. doi:10.1186/1471-2091-12-39.
15. Buschiazzo A, Campetella O, Frasch AC (1997) *Trypanosoma rangeli* sialidase: cloning, expression and similarity to *T. cruzi* trans-sialidase. Glycobiology 7: 1167–1173.
16. Buschiazzo A, Tavares GA, Campetella O, Spinelli S, Cremona ML, et al. (2000) Structural basis of sialyltransferase activity in trypanosomal sialidases. The EMBO Journal 19: 16–24. doi:10.1093/emboj/19.1.16.
17. Buschiazzo A, Amaya MF, Cremona ML, Frasch AC, Alzari PM (2002) The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. Molecular Cell 10: 757–768.
18. Amaya MF, Watts AG, Damager I, Wehenkel A, Nguyen T, et al. (2004) Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. Structure 12: 775–784. doi:10.1016/j.str.2004.02.036.
19. Cremona ML, SInchez DO, Frasch ACC, Campetella O (1995) A single tyrosine differentiates active and inactive *Trypanosoma cruzi* trans-sialidases. Gene 160: 123–128.
20. Kumar R, Stanley P (1989) Transfection of a human gene that corrects the Lecl glycosylation defect: Evidence for transfer of the structural gene for *N*-acetylglucosaminyltransferase I. Molecular and Cellular Biology 9: 5713–5717.
21. Engstler M, Schauer R, Brun R (1995) Distribution of developmentally regulated trans-sialidases in the Kinetoplastida and characterization of a shed trans-sialidase activity from procyclic *Trypanosoma congolense*. Acta Tropica 59: 117–129.
22. Esievo KA, Saror DI, Ilemobade AA, Hallaway MH (1982) Variation in erythrocyte surface and free serum sialic acid concentrations during experimental *Trypanosoma vivax* infection in cattle. Res Vet Sci 32: 1–5.
23. Paris G, Cremona ML, Amaya MF, Buschiazzo A, Giambiagi S, et al. (2001) Probing molecular function of trypanosomal sialidases: single point mutations can change substrate specificity and increase hydrolytic activity. Glycobiology 11: 305–311.
24. Smith LE, Eichinger D (1997) Directed mutagenesis of the *Trypanosoma cruzi* trans-sialidase enzyme identifies two domains involved in its sialyltransferase activity. Glycobiology 7: 445–451.
25. Jackson AP, Allison HC, Barry JD, Field MC, Hertz-Fowler C, et al. (2013) A Cell-surface Phylome for African Trypanosomes. PLoS Negl Trop Dis 7: e2121. doi:10.1371/journal.pntd.0002121.s003.
26. Moloo SKS, Gray MAM (1989) New observations on cyclical development of *Trypanosoma vivax* in *Glossina*. Acta Tropica 46: 167–172.

3.3 Results

Diverse expression and different pH optima of trans-sialidases from *Trypanosoma congolense* is a direct response to changing environments during life cycle

Thaddeus T. Gbem, Mario Waespy, Jonathan A. Nok, Sørge Kelm (2013)

(Manuscript in preparation)

Contribution of Mario Waespy:

- Purification of TconTS enzymes
- TconTS enzyme assays on fetuin-bound Sia
- Contribution to HPAEC-PAD data acquisition and evaluation
- Supported preparation of the manuscript

Diverse expression and different pH optima of trans-sialidases from *Trypanosoma congolense* is a direct response to changing environments during life cycle

Thaddeus T. Gbem[1,3], Mario Waespy[1], Jonathan A. Nok[2], Sørge Kelm[1*]

[1]*Centre for Biomolecular Interactions Bremen, Faculty for Biology and Chemistry, University Bremen, Bremen, Germany*

[2]*Department of Biochemistry, Ahmadu Bello University, Zaria, Nigeria*

[3]*Department of Biological Sciences, Ahmadu Bello University, Zaria, Nigeria*

* corresponding author: skelm@uni-bremen.de

Abstract

Nagana, the animal form of trypanosomiasis, caused by *Trypanosoma* species is a serious livestock production problem in Africa. Trans-sialidases (TS) play a key role in the pathology of nagana. Active TS genes from *T. congolense*, TconTS1, TconTS2, TconTS3 and TconTS4 are expressed in both the tsetse vector and the bloodstream of infected animals. We studied the pH optima and activities of these diverse gene products on blood glycoconjugates and fetuin. TconTS2 was found to be active at pH 9, corresponding to the pH of the tsetse midgut, an unusual pH for sialidases. Though they all transfer sialic acids (Sia) in presence of acceptor molecules, two of them, TconTS1 and TconTS2 exhibit highest specific activities for transfer on serum glycoconjugates. Sialidase activities of TconTS1, TconTS2 and TconTS4 were higher on serum glycoconjugates than on the single substrate, fetuin. TconTS3 formed two products peaks differentially favoured at different serum-Sia concentrations. Serum presents varied targets for TconTS2 than any other TconTS indicated by changes in several peak areas in reactions involving this enzyme. The diversity and different pH optima of TconTS may be a direct response of the parasite to the changing environments during the life cycle.

Author Summary

Trans-sialidases (TS), a multigene family of enzymes have been established as virulent factors in the Animal African Trypanosomiasis (AAT). The disease caused by protozoan parasites of the

genus *Trypanosoma* are transmitted by tsetse flies, exposing the parasites to different environments. It has been established that *T. brucei* require TS to establish infections in the tsetse vector midgut for subsequent transmission to the mammalian host. Both *T. brucei* and *T. congolense* share several common aspects of life cycle. The pH in tsetse midgut is different from that obtained in the mammalian host. In the mammal, *T. congolense* requires sialic acids (Sia) for survival. There are several Sia types and linkages in the blood and serum of mammalian hosts. We studied the pH optima for the active TconTS genes as well as enzyme activities. Our study revealed that different pH optima as well the diverse nature of TS genes are a direct response to changing environments during the life cycle of the parasite as it seeks to survive in the two host systems.

Introduction

Trypanosoma congolense (subgenus: *Nannomonas*) is a major causative agent of the Animal African Trypanosomiasis (AAT) otherwise known as nagana. Nagana has profound negative impact on stock farming, milk and meat production on the African continent [1]. Other *Trypanosoma* species implicated in nagana include but not limited to *T. vivax* (subgenus: *Duttonella*) and *T. brucei brucei* (subgenus: *Trypanozoon*), which are transmitted by bites of *Glossina* spp. Apart from the bites of *Glossina*, another *Trypanosoma* pathogen implicated in nagana, *T. evansi* could also be transmitted mechanically by other bitting flies. Trypanosomes harbour a multigene family expressing unique sets of enzymes, the trans-sialidases (TS). Their role in establishing infection in tsetse midgut has been established for *T. brucei* [2]. While no such information is available for *T. congolense*, TS genes are likely to play a similar role, since *T. congolense* and *T. brucei* exhibit several common aspects of life style, for e.g. establishing infection first in the tsetse midgut. Both *T. brucei*,*T. congolense* [3,4] and *T. evansi* [5] share TS orthologues that are distinctively absent in *T. vivax* [3,6]. The pH in the midgut of tsetse ranges between 9 and 10 [7]. This is rather an unusual pH for known sialidases as sialidases have rather acid pH optima and are inactive at pH between 9 and 10 [8]. Therefore, for any TS/sialidase to be

active in the tsetse midgut, such an enzyme must have evolved to accommodate prevailing pH conditions found therein.

In the mammalian host, TS/sialidase are shed in the serum and blood plasma of infected animals [9,10] and have been identified as virulent factors in trypanosomiasis caused by *T. congolense* [11] and *T. vivax* [6]. A significant level of sialidase activities of *T. congolense* [10] as well as sialidase and TS of *T. vivax* [6] in the blood of infected mice have been reported. Similarly, the mechanism of anaemia induction that directly implicates TS/sialidases has been established. It has been shown that removal of Sia from erythrocytes' major surface sialoglycoproteins, the glycophorins leads to erythrophagocytosis in the course of *T. congolense* [11] and *T. vivax* [6] infections.

Serum contains a wide range of glycoproteins and glycolipids that are heavily sialylated. A comprehensive list of serum components has been compiled [12] and can be assessed at SMDB: http://www.serummetablome.ca/http://www.serummetablome.ca. Sia occur on these glycoconjugates linked as α2,3, α2,6 or α2,8 to a variety of glycan structures and further diversity comes from the different naturally occurring Sia modifications [13]. About 96% of Sia on human erythrocytes are glycoprotein-associated comprising 9 types beside the predominant Neu5Ac [14]. Indeed, the constant interplay between host and pathogens is said to be the driving force behind the rapid evolution of Sia leading to Sia diversification on the one hand, and the evolution of attributes by the parasites to be able to overcome such a rapid Sia modification on the other hand [15]. Pathogens can alter the chemical and physiological composition of serum and blood plasma. For example, the intracellular *Trypanosoma* parasite *T. cruzi* is poorly sialylated when released from cells but becomes densely sialylated upon contact with serum [16]. While a proportion is transferred to the parasites, as described for *T. cruzi* [16,17], a certain proportion is lost becoming free Sia in the serum. The desialylated blood glycoconjugates products are rapidly cleared via galactose specific mechanisms [13], a process which is most likely an important reason for several pathological effects of nagana, such as anaemia and impaired immune response.

Both the procyclic [18,19] and blood stream forms, BSF [3,11] of *T. congolense* express several TS isoforms but the biological relevance of this diversity has not been addressed. This study

investigated pH optima and enzyme activities of single recombinant TconTS and concludes that the diversity of TconTS genes is tied to the complexity of the changing environments during the life cycle of the parasite.

Methods

All chemicals and reagents used were cell culture and analytical grade. Anti-SNAP-tag rabbit polyclonal antibody was from GeneScript (Piscataway, USA) while anti-*Strep* tag rabbit polyclonal antibody, *Strep*-tactin resin beads and buffers were from IBA (Göttingen, Germany). *Pfu* DNA polymerase, *Hind*III, *Xba*I, *Spe*I, *Eco*RI and *Dpn*I, PageBlue, molecular weight marker (PageRuler), BCA assay protein kit were all purchased from Thermo Scientific (St. Leon-Rot, Germany). X-ray film and Enhanced Chemiluminescence system were purchased from GE Healthcare (Uppsala, Sweden). Polyvinylidene difluoride membrane was from Millipore (Schwabach, Germany). VivaSpin 6® and VivaCell250® ultracentrifugation units were from Sartorius (Göttingen, Germany). Hygromycin and gentamycin were purchased from PAA, (Pasching, Austria). Polyethylenimin transfection reagent, glucoronic acid, *N*-acetyl-neuraminic acid (Neu5Ac), 3'sialyllactose (3'SL), lactose, HCl and orcinol were purchased from Sigma-Aldrich (Steinheim, Germany). Excell CD CHO media from SAFC, USA, and 96-well plates were from Sarstedt (Hamburg, Germany).

Recombinant trans-sialidases

Cell lines derived from $CHO_{Lec\ 1}$ cells [20] expressing recombinant TconTS1 through TconTS4, fused with both SNAP and Strep tags and the purification of these proteins have been described [3,21].

Serum preparations and serum Sia quantification

To obtain serum, human blood was collected without anticoagulant and allowed to clot at room temperature for at least 30 minutes. The clotted components were separated from serum by centrifugation at 1500 min^{-1} for 10 minutes at 4°C. To quantify Sia content, serum was serially diluted in 40 µL amounts per well and total sialic acid determined using the resorcinol/Fe^{3+} reagent

as described [22].

Trans-sialidase and sialidase reactions

Enzymes used in control assays were heat inactivated. All reactions were incubated at 37°C in 10 mM phosphate buffer, pH 7.4. Fetuin and human serum served as Sia donor substrates while lactose served as the acceptor substrate. Except for the experiments involving determination of pH optima, all other tests were carried out as previously described [3,21]. For pH optima experiments, pH was adjusted to the required pH of 6, 7, 8, 9 and 10, while reactions in which pH 7.4 was maintained served as control. Quantification of Neu5Ac, 3'SL, lactose and glucoronic acid (internal standard) were carried out using the HPAEC-PAD system as previously described [3,21].

In order to establish kinetic parameters, the ligand binding (single site saturation) curve fit module of SigmapPlot 11 was used to calculate v_{max} and K_M employing the Michaelis-Menten equation $v = v_{max} C_S/(C_S + K_M)$.

Results

TconTS enzymes exhibit different pH optima

The pH in the midgut of the tsetse vector is between 9 and 10 [7]. In the mammal, the parasite encounters a pH of about 7. pH optimum may give a likely indication of which TconTS gene product operates at the optimum in the different host systems. Using fetuin as a model donor blood glycoprotein and lactose as acceptor substrates, we measured enzyme activities at different pH (Figure 1 A-D). Extended incubation periods were required to asses hydrolytic activities for TconTS1 and TconTS2 given their poor sialidase activities compared to the high TS activities [3]. Therefore while TS activities were run for 30 minutes, using 50 ng enzyme for TconTS1 and TconTS2, transfer activity experiments for both TconTS3 and TconTS4 were incubated for 24 h with 250 ng of each enzyme. For sialidase activities, 250 ng of each enzyme (TconTS1, TconTS2, TconTS3 and TconTS4) were incubated with 600 µM fetuin-bound Sia for 24 h without the acceptor lactose.

Transfer activities for TconTS1, TconTS3 and TconTS4 had pH optima at pH 7 with low TS

activities between pH 8 and 10 (Figure 1 A, C and D). TS activity of TconTS2 was highest at pH 9, producing about 370 µM 3'SL as compared to 320 µM 3'SL at pH 7.4 (Figure 1 B). Sia released for TconTS1 and TconTS2 did not vary significantly at different pH points, with both enzymes producing about 100 µM Sia (Figure 1 A and B). Sialidase activities of TconTS3 and TconTS4 in the presence of lactose showed pH optima at pH 6 with progressive reduction in sialidase activity as pH increased (Figure 1 C and D). Interestingly, sialidase activity increased for both TconTS3 and TconTS4 in absence of lactose. However, the increment was more pronounced for TconTS4, attending a pH optima at pH 9 at which 35 µM compared to 15 µM Sia for TconTS3.

Enzyme activities of TconTS on serum glycoconjugates

Sia transfer was determined from 3'SL production in reactions containing serum and lactose as donor and acceptor substrates, respectively. Sialidase activity was monitored by Neu5Ac release from serum either in the absence or presence of lactose. A summary of these results is presented in Table 1. 200 ng TconTS1 produced 42 pmol 3'SL/ min during the first 2 h (Figure 2 A). At longer incubation times, the reaction velocity decreased until a maximum of 6 nmol 3'SL was reached. This did not appear to be due to donor substrate depletion, since this amount corresponds to 120 µM 3'SL, which is only 20 % of the 600 µM serum-bound Sia available in the reaction. Even at prolonged incubations for up to 24 h with up to 1000 ng TconTS1 no significant increase of 3'SL was obtained (Figure 2 B). This suggests that at 120 µM 3'SL, equilibrium has been attained or the residual serum-bound Sia is not available for transfer by TconTS1, for example due to sialidase activity. However, free Sia could only be detected after longer incubations or with high amounts of TconTS1, whereas the maximum Sia released was 1.5 nmol after 24 h with 1000 ng enzyme (Figure 2 B), representing less than 3 % of total serum-bound Sia added. In the absence of lactose, higher sialidase activity was detected and 200 ng TconTS1 released about 8 pmol Sia/ min for up to 4 h, after which the velocity started to decrease, attaining 5 nmol free Sia after 24 h (Figure 2 A). Even 1000 ng TconTS1 released only slightly more Sia (6.5 nmol). Interestingly, this is similar to the 6 nmol 3'SL formed by the same amount of enzyme.

A similar pattern of 3'SL formation was observed for TconTS2. 200 ng TconTS2 produced 75 pmol 3'SL/ min for the first 2 h (Figure 2 C). Reaction velocity decreased at longer incubation times until a maximum of 8 nmol 3'SL was reached. Increasing the amount of enzyme up to 1000 ng TconTS2 did not lead to higher amounts of 3'SL (Figure 2 D) corresponding to 160 µM 3'SL, which represents 27 % of the total serum-bound Sia available in the reaction. In the presence of 2 mM lactose, free Sia could only be detected after longer incubation times and reached a maximum of 0.25 nmol free Sia, less than 1 % total serum-bound Sia available in the reaction (Figure 2 C and 2 D). Higher sialidase activities were observed in the absence of lactose, where 200 ng TconTS2 released 13 pmol free Sia/ min in the first 2 h (Figure 2 C). Reaction velocity decreased at longer incubation times. The maximum amount of free Sia obtained was 5 nmol and did not increase, even when 1000 ng TconTS2 was used for incubation for 24 h.

Sia transfer from serum donor substrate by TconTS3 was almost two orders of magnitude slower relative to TconTS2, since 250 ng TconTS3 produced 0.83 pmol 3'SL/ min. This reaction velocity was linear during the first 8 h (Figure 2 E) before it decreased and a total of 0.65 nmol 3'SL was reached after 24 h. Increasing the enzyme to 1000 ng TconTS3, 1 nmol 3'SL could be obtained (Figure 2 F). This corresponds to 20 µM representing 3 % of the total 600 µM serum-bound Sia available in the reaction. Under these conditions, free Sia was not detected in TconTS3 reactions containing the acceptor lactose, even at longer incubation times and higher enzyme amounts. In absence of lactose however, sialidase activity was detected and 1000 ng enzyme released 0.3 nmol free Sia in 24 h (Figure 2 F).

TconTS4 showed faster Sia transfer from serum donor than TconTS3 with 250 ng enzyme producing 4 pmol 3'SL/ min in 1 h (Figure 2 G). Reaction velocity decreased over extended incubation times leading to 0.8 nmol 3'SL in 24 h. Incubation with 1000 ng TconTS4 for 24 h produced 1.1 nmol (Figure 2 H). In the presence of lactose, 1000 ng TconTS4 released 0.21 nmol free Sia in 24 h and the amount increased to 4.5 nmol in absence of lactose (Figure 2 H).

Kinetic parameters of serum-bound Sia as substrate for sialidase and trans-sialidase activities were determined in the presence of 2 mM lactose as acceptor substrate for Sia transfer activity (Table 1, Figure S1 A-D). Similarly, K_M and v_{max} values of lactose as acceptor substrate for trans-

sialidase was determined in the presence of 600 µM serum-bound Sia (Table 1, Figure S2 A-D). In summary, TconTS1 and TconTS2 have higher specific activities than TconTS3 and TconTS4. TconTS1 and TconTS2 have lower K_M values for the acceptor substrate lactose, whereas TconTS3 and TconTS4 have lower K_M for serum-bound Sia.

Kinetic parameters were also established for sialidase activities on serum-bound Sia by incubating under standard conditions without lactose (Table 2, Figure S3a-c). Since 250 ng TconTS3 did not exhibit sialidase activity within a 24 h incubation, no further reaction was set up for this enzyme. TconTS1 had the lowest K_M (0.16 mM) followed by TconTS2 (0.41 mM). TconTS4 had the highest K_M (1.28 mM) but lowest free Sia release from serum-bound Sia. Free Sia release by TconTS4 is about 50- and 80-fold less than that released by TconTS1 and TconTS2, respectively.

TconTS3 supports an extra product peak in chromatograms while TconTS2 accesses several other Sia types and/ or linkages on donor serum glycoconjugates

It was earlier established that among the TconTS enzymes, TconTS1 and TconTS2 have the highest specific activities on the donor blood glycoprotein fetuin using lactose as acceptor when compared to TconTS3 and TconTS4 [3]. It was speculated on the basis of that finding, the possibility of substrate selectivity being responsible for differences in the specific activities of TconTS enzymes; for example, TconTS3 and TconTS4 exhibiting enhanced specific activities on other donor and acceptor substrates, possibly blood glycoproteins and/or glycolipids.

Utilising HPAEC-PAD analysis for quantification of TconTS reaction products described previously [21], it was possible to detect and differentiate sialyl- as well as asialo-analyte peaks from the resulting chromatograms of the same reaction. When TconTS3 was incubated with human serum as donor substrate, an additional product peak was observed besides the one for 3'SL, Neu5Ac and lactose. This peak eluted after 3'SL, with a retention time, t_R of 8 mins under the conditions used. So far, we could not identify this peak and without suitable standards therefore, quantification could only be done using peak areas. Quantification of the unknown product revealed higher peak areas when compared with the 3'SL peak (Figure 3 A). Like the known 3'SL, the peak areas of the

unknown product increased with increasing lactose concentrations. Interestingly, varying the concentration of serum-bound Sia while keeping constant that of lactose (2 mM) favoured the formation of the unknown product at low serum-Sia concentrations. The unknown peak attained a maximum peak area at 600 µM serum-bound Sia. Above this Sia concentration, the peak area decreased and at serum-Sia concentration of 1200 µM, the unknown peak area became lower than those of 3'SL even though at this point, 3'SL formation has already attained equilibrium (Figure 3 B).

Control serum samples usually contained inherent analyte peaks that were detected by the HPAEC-PAD method used. We observed that when TconTS2 was used, at least two of these inherent peaks labelled a and b (Figure 4 A) had reduced peak areas over consecutive time counts. No changes in peak areas were observed in the control serum samples and those incubated with TconTS1 (Figure 4 B) as well as with TconTS3 and TconTS4 (not shown). Yet there was 3'SL formation in all the TconTS incubated samples but none in the controls. Equally, the peak areas of the internal standard, glucuronic acid remained unchanged over time (Figure 4 A and B), ruling out the possibility of losses. This indicates the ability of TconTS2 to use either different Sia sources or access different types/linkages.

Sialidase activities on fetuin

Sialidase activities of TconTS have been implicated in anaemia in animals suffering *Trypanosoma* infections [6,10,11] due to desialylation of erythrocytes by *Trypanosoma* TS enzymes. This leads to their clearance from the host circulatory system [13]. Serum is a complex substrate with different sialylated glycolipids, and glycoproteins. As a comparative basis, a single substrate, the blood glycoprotein fetuin was used for the following reasons; (i) it is comparatively less complex allowing easy comparison and inferences, (ii) it has been extensively employed to biochemically characterise the Sia transfer activities of TconTS [3,21] and (iii) though found only in foetal serum and may be less relevant in this respect, likely substrates like acute-phase proteins (APPs) e.g. the Alpha-1-acid glycoprotein (AGP) present in serum are highly sialylated and thought to play a role in

modulating immune and inflammatory responses [23].

Gbem *et al.* reported that TconTS enzymes exhibited low sialidase activities with TconTS4 showing the highest activity when fetuin was used as donor in TS reactions [3]. Sialidase activities of TconTS1, TconTS2 and TconTS3 became detectable only when 3'SL accumulated in the reaction. This indicated the release of free Sia from 3'SL but not from the initial donor substrate fetuin. Similarly, in the absence of lactose, except for TconTS4, sialidase activities were observed for TconTS1 and TconTS2 when 500ng enzyme (10-fold more than standard conditions used [21]) were incubated with fetuin for 2 h. TconTS3 did not show sialidase activity even under these conditions. However, the study did not reveal v_{max} and K_M values for sialidase activities of these enzymes. Equally, enzyme amounts higher than 50 ng were not incubated with fetuin for longer times and therefore, no definite statement could be made as to whether or not lower enzyme amounts could also exhibit sialidase activities.

In the present study, sialidase reactions contained approximately 600 µM fetuin-bound Sia without an acceptor substrate. Reaction velocities were measured at indicated time points (Figure 5 A, Table 3). 200 ng TconTS1 released 16,4 µM Sia in 8 h (1.7 pmol/ min). Incubations over extended periods up to 24 h revealed 40 µM Sia, showing reduced reaction velocity. TconTS2 showed a similar pattern but higher Sia release. 200 ng TconTS2 released 20 µM Sia in 8 h (2.2 pmol/ min) with the velocity of free Sia release decreasing with a total of 50 µM free Sia within 24 h. 250 ng TconTS3 released 2.8 µM Sia in 4 h (0.75 pmol/ min). Velocity decreased as prolonged incubations revealed a total of 12 µM Sia after 24 h. 250 ng TconTS4 released 3 µM Sia in 2 h (1.25 pmol/ min). At prolonged incubations up to 24 h, reactions velocities decreased until a total of 22 µM Sia (22 µM). Since fetuin contained 600 µM Sia, the amount of Sia available could not be responsible for the decrease in the velocity of Sia release. However, in typical bacteria sialidases, free Sia has been shown to inhibit sialidase activity [24]. Varied enzyme amounts up to 1000 ng were incubated for 24 h to monitor if Sia release from fetuin depended on the amount of enzyme (Figure 5 B). Under these conditions, Sia release from fetuin by TconTS1 and TconTS2 was concentration dependent. 500 ng enzyme released 120 µM Sia (4.2 pmol/ min) and 1000 ng released 240 µM Sia (8.3 pmol/ min) for both TconTS1 and TconTS2. At below 500 ng enzyme amounts, differences

exist in the amount of Sia released by the two TconTS enzymes. At 200 ng enzyme amounts, TconTS1 released 2 pmol Sia/ min while TconTS2 released 2.4 pmol Sia/ min, indicating faster Sia release by TconTS2. The amount of Sia released by TconTS3 followed a similar pattern with those of TconTS1 and TconTS2 but with lower Sia amounts produced as 500 ng released 42 µM while 1000 ng released 84 µM Sia after 24 h. Increased Sia release was observed with increasing TconTS amounts. 500 ng TconTS4 enzyme released 90 µM while 1000 ng released 130 µM Sia in 24 h.

Kinetics parameters were determined for sialidase activities of TconTS enzymes on fetuin (Tables 1, Figure S1 A-D). To achieve this, we used 200 ng each for TconTS1 and TconTS2, and 250 ng each for TconTS3 and TconTS4 and incubated with 600 µM fetuin-bound Sia for 2 h for both TconTS1 and TconTS2, and 24h for both TconTS3 and TconTS4. TconTS1 and TconTS2 exhibited similar magnitudes of specific activity on fetuin (about 0.1 µmol/(min x mg TS)). This is 13- and 33-fold higher than that exhibited by TconTS4 and TconTS3 respectively on the same substrate. K_M values were also similar for TconTS1 (20 µM) and TconTS2 (30 µM). The K_M values for TconTS3 (137 µM) and TconTS4 (441 µM) were higher than those of TconTS1 and TconTS2.

Discussion

Trypanosoma are digenetic parasites with a life cycle that alternate between two different host systems; the tsetse vector and the mammalian host. Profound differences exist between these two systems, not least the pH. Sialidases have rather acid pH optima and are inactive at pH between 9 and 10 [8]. The pH at which each enzyme showed highest activities was established for TconTS enzymes. Sialidase activities of TconTS3 and TconTS4 in absence of lactose have a pH optimum of 9. This is in contrast to Sia transfer activities for both enzymes with pH optima at pH 7, even though the amount produced is comparatively lower than those produced by TconTS1 and TconTS2 under similar conditions. Sialidase activity on serum glycoconjugates in absence of lactose for all enzymes is higher than when lactose is present (Figure 2 A-D, Table 5). Biologically, it is difficult to envisage a situation warranting a lack of acceptors in host systems hence sialidase

activities in presence of lactose are biologically more relevant. In this light, except for TconTS3 which did not show sialidase activity in presence of lactose (Figure 2 F), the remaining three enzymes are the most likely candidates responsible for the release of free Sia as has been observed [10,11] in the blood of animals suffering *T. congolense* infections.

Considering the pH optimum for TconTS enzymes, the TS activity of TconTS2 can be expected to function optimally in the fly midgut. On the other hand, the sialidase activities of TconTS3 and TconTS4 would function optimally in the fly midgut. However, the magnitude of these activities differs by at least 8-folds in favour of the TconTS2 Sia transfer activity. It would be interesting to see the impact of a TconTS2 knockout or whose activity has been genetically modified to act as a sialidase in the gut of tsetse carrying *Trpanosoma* parasites.

In previous studies [3,21], the enzymatic activities of four TconTS gene products using fetuin as a model blood glycoprotein donor substrate were compared. While the use of fetuin as a substrate provided important insight into the biochemical diversity of these enzymes, a natural medium in which these enzymes operate, such as serum is ideal. Specific activity patterns were similar for both fetuin [3] and serum glycoconjugates in presence of lactose as common acceptor. TconTS2 showed the highest specific activity, followed by TconTS1>TconTS3 >TconTS4 in that order (Table 2, Figure S1 and S2). Except for TconTS4, 3'SL production for the other three TconTS enzymes is at least 10-fold higher in reactions involving fetuin-bound Sia [3,21] than those in which serum glycoconjugates is the donor substrate. Engstler *et al.* observed poor Sia transfer activity of TS from *T. brucei* on equine serum glycoconjugates compared with other donor substrates [25]. Put together, *Trypanosoma* trans-sialidases are poor at transferring Sia in a complex systems like mammalian host fluids. In alternative, it could be that a complex pattern of sialylation and desialylation from the diverse sources of Sia donors and acceptors respectively operates in the complex mammalian host environment. TconTS2 generates analyte peaks, proposed to be due to sialylated glycan structures (Figure 4 A) that were not observed for TconTS1 (Figure 4 B) as well as for TconTS3 and TconTS4 (not shown), yet there is 3'SL formation in TconTS1 (Figure 4 B), TconTS3 and TconTS4 (not shown). This would mean that these other enzymes utilise Sia sources other than the ones used by TconTS2. Interestingly, no free Sia was detected in the TconTS2

reaction medium up to 8 h (Figure 2 C) indicating an efficient Sia transfer to lactose and other acceptors present in serum. Expressing more than one TS could increase the efficiency of transfer onto the parasites body surface in presence of other competing acceptors. The v_{max} obtained for TconTS1 and TconTS2 from assays with varying lactose concentrations are lower than the specific activities determined by varying the donor substrate, since for technical reasons the donor substrate concentration was non-saturating (0.3- and 0.8-fold of the K_M values for the serum-bound Sia for TconTS1 and TconTS2, respectively).

TconTS1 and TconTS2 have the highest specific activities of the TconTS variants on both fetuin [3] and serum glycoconjugates. This however does not preclude the possibility of TconTS3 and TconTS4 being more active on other substrates. TconTS3 showed an additional product peak, which we were unable to identify with methods used (Figure 3 A), in addition to the 3'SL peak. The formation of this additional product, eluting at 8 minutes, 2 minutes after 3'SL under conditions used, was favoured. This collaborates our earlier postulations on activities differing on the basis of substrates [3]. Further, the differential formation of these unidentified product peaks at different serum-Sia concentrations (Figure 3 B) indicates the possibility of Sia transfer to different acceptors under differing Sia concentrations or the use of different Sia types and/or linkages by TconTS3. It is not implausible to envisage such a scenario employed by the parasites at different stages of infection in the mammalian life stage.

Recently, the amino acid composition in the catalytic pocket of active TconTS enzymes; TconTS2, TconTS3 and TconTS4 was reported [3]. When TconTS ORFs were aligned with the reference *T. cruzi* TS [26], two residues at consensus positions 293 and 494 were found substituted in the most active TconTS1 and TconTS2 but conserved in the less active TconTS3 and TconTS4 [3]. Structural and catalysis studies [26] showed the residues in question to be essential in interacting with the methylumbelliferyl aglycon or the lactose-part of 3'SL. However, due to the huge variability in serum composition [12] it is more than likely that other suitable donor and acceptor entities exist in serum that would present better side chains that are readily utilised by TS enzymes. The variability in serum composition is compounded across mammalian species. For example, while Neu5Gc is widely distributed in the animal kingdom, it is absent in the normal tissues of humans

[15,27,28].

Recently, it was suggested that the lectin domain of trans-sialidases play a likely role in the pathogenesis of nagana [3,29,30]. Put together, the diverse pattern of expression, differences in the critical amino acids at the catalytic domain, different pH optima of TconTS gene products as well as structural differences in the TS enzyme structures could contribute in enhancing survival and establishment of infection of the parasites in the different hosts environments.

Acknowledgment

We are grateful to Evelin Tiralongo and Roland Schauer (University of Kiel, Germany) for providing the *T. congolense* STIB249 genomic DNA. The plasmid vector pDEF was kindly provided by Paul Crocker (University of Dundee, UK). We thank Dr. Frank Dietz for modifying the pDEF plasmid and useful discussions on transfection of fibroblasts with plasmids. We are grateful to Nazila Isakovic for help with tissue culture work. We thank Petra Berger for help with tissue culture as well as serum sample preparations. We would like to thank the Deutsche Forschungsgemeinschaft (DFG, project KE 428/8-1) for financial support for this work.

Author contribution

Conceived and designed the experiments: SK, JAN, TTG. Performed the experiments: TTG, MW. Analysis of data: TTG, MW, SK. Wrote the paper: TTG, MW, JAN, SK.

References

1. Steverding D. The history of African trypanosomiasis. Parasit Vectors. BioMed Central Ltd; 2008;1(1):3.

2. Nagamune K, Acosta-Serrano A, Uemura H, Brun R, Kunz Renggli C, Maeda Y, et al. Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors. J Exp Med. Rockefeller Univ Press; 2004 May 17;199(10):1445–50.

3. Gbem TT, Waespy M, Hesse B, Dietz F, Smith J, Chechet GD, et al. Biochemical diversity in the *Trypanosoma congolense* trans-sialidase family. PLoS Negl Trop Dis. 2013;7(12):e2549.

4. Jackson AP, Allison HC, Barry JD, Field MC, Hertz-Fowler C, Berriman M. A Cell-surface phylome for African trypanosomes. Tschudi C, editor. PLoS Negl Trop Dis. 2013 Mar 21;7(3):e2121.

5. Elshafie EI, Sani RA, Hassan L, Sharma R, Bashir A, Abubakar IA. Active infection and morphometric study of *Trypanosoma evansi* among horses in Peninsula Malaysia. Trop Biomed. 2013 Sep;30(3):444–50.

6. Guegan F, Plazolles N, Baltz T, Coustou V. Erythrophagocytosis of desialylated red blood cells is responsible for anaemia during *Trypanosoma vivax* infection. Cellular Microbiology. 2013 Aug;15(8):1285–303.

7. Dyer NA, Rose C, Ejeh NO, Acosta-Serrano A. Flying tryps: survival and maturation of trypanosomes in tsetse flies. Trends Parasitol. Elsevier; 2013 Apr;29(4):188–96.

8. Reuter G, Schauer R, Prioli R, Pereira MEA. Isolation and properties of a sialidase from *Trypanosoma rangeli*. Glycoconj J. Kluwer Academic Publishers; 1987;4(4):339–48.

9. Esievo KA, Saror DI, Ilemobade AA, Hallaway MH. Variation in erythrocyte surface and free serum sialic acid concentrations during experimental *Trypanosoma vivax* infection in cattle. Res Vet Sci. 1982 Jan;32(1):1–5.

10. Nok AJ, Balogun EO. A bloodstream *Trypanosoma congolense* sialidase could be involved in anemia during experimental trypanosomiasis. J Biochem. 2003 Jun;133(6):725–30.

11. Coustou V, Plazolles N, Guegan F, Baltz T. Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. Cellular Microbiology. 2012 Feb 13;14(3):431–45.

12. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. PLoS ONE. Public Library of Science; 2011;6(2):e16957.

13. Kelm S, Schauer R. Sialic acids in molecular and cellular interactions. Int Rev Cytol. 1997;175:137–240.

14. Bulai T, Bratosin D, Pons A, Montreuil J, Zanetta JP. Diversity of the human erythrocyte membrane sialic acids in relation with blood groups. FEBS Lett. 2003 Jan 16;534(1-3):185–9.

15. Angata T, Varki A. Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. Chem Rev. 2002 Feb;102(2):439–69.

16. Schenkman S, Jiang MS, Hart GW, Nussenzweig V. A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian

cells. Cell. 1991 Jun 28;65(7):1117–25.

17. Zingales B, Carniol C, de Lederkremer RM, Colli W. Direct sialic acid transfer from a protein donor to glycolipids of trypomastigote forms of *Trypanosoma cruzi*. Mol Biochem Parasitol. 1987 Nov;26(1-2):135–44.

18. Engstler M, Schauer R, Brun R. Distribution of developmentally regulated trans-sialidases in the Kinetoplastida and characterization of a shed trans-sialidase activity from procyclic *Trypanosoma congolense*. Acta Tropica. 1995 May;59(2):117–29.

19. Tiralongo E, Schrader S, Lange H, Lemke H, Tiralongo J, Schauer R. Two trans-sialidase forms with different sialic acid transfer and sialidase activities from *Trypanosoma congolense*. J Biol Chem. 2003 Jun 27;278(26):23301–10.

20. Kumar R, Stanley P. Transfection of a human gene that corrects the Lec1 glycosylation defect: evidence for transfer of the structural gene for N-acetylglucosaminyltransferase I. Mol Cell Biol. 1989 Dec;9(12):5713–7.

21. Koliwer-Brandl H, Gbem TT, Waespy M, Reichert O, Mandel P, Drebitz E, et al. Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*. BMC Biochem. 2011;12(1):39.

22. Reuter G, Schauer R. Determination of sialic acids. Meth Enzymol. 1994;230:168–99.

23. Lögdberg L, Wester L. Immunocalins: a lipocalin subfamily that modulates immune and inflammatory responses. Biochim Biophys Acta. 2000 Oct 18;1482(1-2):284–97.

24. García Sastre A, Cobaleda C, Cabezas JA, Villar E. On the inhibition mechanism of the sialidase activity from Newcastle disease virus. Biol Chem Hoppe-Seyler. 1991 Oct;372(10):923–7.

25. Engstler M, Reuter G, Schauer R. The developmentally regulated trans-sialidase from *Trypanosoma brucei* sialylates the procyclic acidic repetitive protein. Mol Biochem Parasitol. 1993 Sep;61(1):1–13.

26. Amaya MF, Watts AG, Damager I, Wehenkel A, Nguyen T, Buschiazzo A, et al. Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. Structure. 2004 May;12(5):775–84.

27. Varki A. Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. Am J Phys Anthropol. 2001;Suppl 33(S33):54–69.

28. Sprenger N, Duncan PI. Sialic acid utilization. Adv Nutr. 2012 May;3(3):392S–7S.

29. Ammar Z, Plazolles N, Baltz T, Coustou V. Identification of trans-sialidases as a common mediator of endothelial cell activation by African trypanosomes. Beverley SM, editor. PLoS Pathog. 2013;9(10):e1003710.

30. Waespy M, Gbem TT, Elenschneider L, Jeck A-P, Day CJ, Hartley-Tassell L, et al. Carbohydrate recognition specificity of trans-sialidase lectin domain from *Trypanosoma congolense*. (in press) PLoS Negl Trop Dis. 2015;:1–49. doi:10.1371/journal.pntd.0004120

Figure legends

Figure 1. pH optima for TconTS enzymes. Reactions to determine pH optima were carried out at indicated pH points as described under Methods using 100 µg fetuin (600 µM fetuin-bound Sia) in presence and absence of lactose. **A-B**: 50 ng TconTS1 and TconTS2 were incubated for 30 minutes to measure Sia transfer activities indicated by 3'SL formation, while 250 ng TconTS1 and TconTS2 were incubated with 600 µM fetuin-bound Sia in absence of lactose for 24 h. **C-D**: 250 ng TconTS3 and TconTS4 were incubated in presence and absence of lactose for 24 h. All other experimental conditions are as stated in Methods. Neu5Ac and 3'SL were measured using HPAEC-PAD as described under Methods. Data points are means ± standard deviations of 3 independent experiments, each performed in triplicates.

Figure 2. Concentration and time dependency of TconTS enzymes on human serum glycoconjugates. In concentration dependency experiments, indicated amounts of recombinant TconTS proteins were incubated with 600 µM serum-bound Sia in presence and absence of 2 mM lactose for 24 h. In the time dependency studies, 200 ng each of TconTS1 and TconTS2, and 250 ng each of TconTS3 and TconTS4 were incubated at 37°C for the indicated times. Detection of 3'SL represents Sia transfer activity while the presence of Neu5Ac represents sialidase activity. Both were determined using HPAEC-PAD analysis described under Methods. **A-D**: TconTS1 and TconTS2. **E-H**: TconTS3 and TconTS4. Data points are means ± standard deviations of at least 3 independent experiments, each performed in triplicates.

Figure 3. Additional product peak observed in TconTS3 reactions. An additional product peak was formed by TconTS3 in addition to that of 3'SL when incubated with serum-bound Sia in the presence of lactose. 250 ng recombinant TconTS3 was used against a fixed concentration of one substrate and reactions were incubated for 12 h. **A**: Lactose variation against 600 µM serum-bound Sia. **B**: Serum-bound Sia variation against 2 mM lactose. In all cases, HPAEC-PAD was employed to detect and quantify reaction products as described in Methods. Data points are

means ± standard deviations of at least 3 independent experiments, each performed in triplicates.

Figure 4. Effects of TconTS2 on human serum glycoconjugates. 200 ng TconTS1 and TconTS2 were incubated with 600 µM serum-bound in presence of 2 mM lactose for 24 h. Letters **a**, **b**, **c** and **d** represent unidentified glycoconjugate peaks while GlcAc represent glucuronic acid, used as internal standard. **A**: TconTS2 showing desialylated peaks. **B**: TconTS1 showing 3'SL formation without desialylation of peaks. Colours represent different incubation times: black, blue, pink, brown and green indicate chromatograms of products obtained after 0, 4, 8, 12, 16 and 24 h of incubation, respectively.

Figure 5. Concentration and time dependency of Sia release from fetuin. TconTS enzymes were incubated at 37ºC with 600 µM fetuin-bound Sia in absence of lactose. Formation of Neu5Ac, indicating sialidase activity, was determined by HPAEC-PAD analysis as described under Methods. **A:** 200 ng enzyme was used for both TconTS1 and TconTS2 while 250 ng enzymes were used for TconTS3 and TconTS4 and incubated at 37ºC over indicated times and values for 24 h incubation are given in Table 2. **B:** Varying amounts of TconTS enzymes incubated for 24 h. In all cases, values are means ± standard deviations of at least 3 independent experiments replicated trice.

Supporting Information

Figure S1. Trans-sialidase reaction velocities depending on serum glycoconjugates concentrations. 3'SL amounts were determined as described in Methods. v_{max} and K_M as shown in Table 1 were calculated from these set of data. **A**: 200 ng TconTS1 incubated with 2 mM lactose and varying serum-bound glycoconjugates for 2 h. **B**: 200 ng TconTS2 incubated with 2 mM lactose and varying serum-bound glycoconjugates for 1 h. **C**. 250 ng TconTS3 incubated with 2 mM lactose and varying serum-bound glycoconjugates for 12 h. **D**: 250 ng TconTS4 incubated with 2 mM lactose and varying serum-bound glycoconjugates for 24 h. Data points are means ±

standard deviations of 3 independent experiments, each as triplicates.

Figure S2. Trans-sialidase reaction velocities depending on lactose concentrations. 3'SL amounts were determined employing HPAEC-PAD analysis as described in Methods. v_{max} and K_M as shown in Table 1 were calculated from these set of data using the software SigmaPlot11 as described under Methods. **A**: 200ng TconTS1 was incubated with 600 µM serum glycoconjugates and varying amounts of lactose as indicated for 2 h. **B**: 200 ng TconTS2 was incubated with 600 µM serum glycoconjugates and varying concentrations of lactose as indicated for 1 h. **C**: 250 ng TconTS3 was incubated with 600 µM serum glycoconjugates and varying concentrations of lactose as indicated for 12 h. **D**: 250 ng TconTS4 was incubated with 600 µM serum glycoconjugates and varying concentrations of lactose as indicated for 24 h. Data points are means ± standard deviations of at least 3 independent experiments, each as triplicates.

Figure S3. Sialidase reaction velocities depending on serum glycoconjugates concentrations. Neu5Ac amounts were determined employing HPAEC-PAD analysis as described in Methods. v_{max} and K_M shown in Table 2 were calculated from these set of data. **A**: 200 ng TconTS1 and **B**: 200ng TconTS2, both incubated for 1 and 2 h respectively, with varied serum glycoconjugates concentration without the acceptor lactose. **C**: 250 ng TconTS4 incubated for 24 h with varied concentrations of serum-bound Sia. Data points are means ± standard deviations of 3 independent experiments, each as triplicates.

Figure S4. Sialidase reaction velocities depending on fetuin-bound Sia concentration. Neu5Ac amounts were determined employing HPAEC-PAD analysis as described in Methods. v_{max} and K_M shown in Table 2 are calculated from these set of data. **A**: 200 ng TconTS1 and **B**: 200ng TconTS2, both incubated for 2 h with varied amounts of fetuin-bound Sia. **C**: 250 ng TconTS3 incubated for 12 h with varied amounts of fetuin-bound Sia. **D**: 500ng TconTS4 incubated for 24 h with varied amounts of fetuin-bound Sia. Data points are means ± standard deviations of 3 independent experiments, each as triplicates.

Table 1. Kinetic parameters of TconTS on sialic acid transfer from serum-bound Sia

	Transfer			
TconTS	Acceptor (lactose)*		Donor substrate (serum-bound Sia)**	
	v_{max}*** (nmol/(min x mg TS))	K_M*** (µM)	v_{max}*** (nmol/(min x mg TS))	K_M*** (µM)
TconTS1f	181 ± 1.5	203 ± 16	315 ± 16	669 ± 83
TconTS2	486 ± 9.3	79 ± 7.1	1005 ± 60	750 ± 104
TconTS3	4.2 ± 1.0	1165 ± 91	5.1 ± 0.1	339 ± 25
TconTS4	2.2 ± 0.1	1321 ± 156	3.2 ± 0.1	259 ± 26

* 600 µM serum-bound Sia was used as donor substrate.
**2 mM lactose was used as acceptor substrate.
***K_M and v_{max} were calculated from Michaelis-Menten kinetics (Figure S2 and S3) using SigmaPlot 11. Data points are mean ± standard deviations of three independent experiments, each as triplicates.

Table 2. Kinetic parameters for sialidase activities of TconTS enzymes on fetuin- and serum-bound Sia

TconTS	Fetuin-bound Sia*		Serum-bound Sia*	
	v_{max} ** (µmol/(min x mg TS))	K_M ** (µM)	v_{max} ** (µmol/(min x mg TS))	K_M ** (µM)
TconTS1f	85 ± 1.6	20 ± 1.5	63 ± 0.6	155 ± 7.1
TconTS2	99 ± 1.7	30 ± 1.7	105 ± 3.5	412 ± 40
TconTS3	3.2 ± 0.6	137 ± 7.5	-	-
TconTS4	8.1 ± 0.3	441 ± 16	1.3 ± 0.0	1281 ± 146

* 600µM was used as substrate for both fetuin-bound Sia and serum-bound Sia.
**K_M and v_{max} were calculated from Michaelis-Menten kinetics (Figure S1 and S4) by SigmaPlot11. Data points are mean ± standard deviations of three independent experiments, each replicated thrice
- = no hydrolytic activity on serum

Table 3. Free Sia release from fetuin-bound Sia in absence of lactose

Parameters	TconTS1		TconTS2		TconTS3		TconTS4	
Amount of TconTS used (ng)	200	1000	200	1000	250	1000	250	1000
Sia release (pmol/min., initial velocity)	1.7 ± 0.017		2.2 ± 0.07		0.75 ± 0.004		1.25 ± 0.002	
Total free Sia (nmol) in 24 h*	2.0 ± 0.0.01	12 ± 0.58	2.5 ± 0.03	12 ± 0.62	0.6 ± 0.009	4.2 ± 0.16	1.1 ± 0.00	6.5 ± 0.08

* 30 nmol fetuin-bound Sia (600 µM Sia) was incubated with indicated amounts of various TconTS enzymes

Table 4. 3'SL formation and free Sia release from serum-bound Sia by TconTS enzymes

Parameters	TconTS1		TconTS2		TconTS3		TconTS4	
Amount of TconTS used (ng)	200	1000	200	1000	250	1000	250	1000
3'SL formation (pmol/min., initial velocity)*	42 ± 2.5		75 ± 4.32		0.83 ± 0.01		4.0 ± 0.06	
3'SL produced (nmol) in 24 h*	6.0 ± 0.11	6.0 ± 0.23	8.0 ± 0.04	8.0 ± 0.12	0.65 ± 0.004	1.0 ± 0.004	0.8 ± 0.01	1.1 ± 0.001
Free Sia (nmol) in presence of lactose in 24 h	1.0 ± 0.001	1.5 ± 0.01	0.25 ± 0.01	0.25 ± 0.01	n.d.**	n.d.**	0.1 ± 0.00	0.21 ± 0.00
Free Sia (nmol) in absence of lactose in 24 h***	6.5 ± 0.01	6.5 ± 0.01	5.0 ± 0.09	5.0 ± 0.11	n.d.**	0.3 ± 0.02	0.15 ± 0.001	0.45 ± 0.004

* 600 µM serum-bound Sia was used as donor substrate in presence of 2 mM lactose as acceptor
** not detected, below 0.1 nmol
*** 600 µM serum-bound Sia was used as substrate in absence of acceptor

Table 5. Relationship between sialidase and trans-sialidase activities of TconTS enzymes on serum glycoconjugates

TconTS	Trans-sialidase* activity (nmol/(min x mg TS)	Sialidase activity (nmol/(min x mg TS)		TS/sialidase	
		- lactose**	+ lactose***	- lactose****	+ lactose
TconTS1f	315 ± 16	63 ± 0.6	1.04 ± 0.001	5	303
TconTS2	1005 ± 60	105 ± 3.5	0.17 ± 0.0	10	5912
TconTS3	5.1 ± 0.1	n.d.*****	n.d.*****	>5	>5
TconTS4	3.2 ± 0.1	1.3 ± 0.0	0.31 ± 0.0001	2,5	10

* TS activity values are given in Table 1
** Sialidase activity values in absence of lactose are given in Table 1
*** Sialidase activities in presence of 2 mM lactose were determined by incubating 1000 ng of indicated TconTS enzymes for 24 h with 600 µM serum-bound Sia. Data points are mean ± standard deviations of 3 independent reactions, each in 3 replicates.
**** sialidase activity values in absence of lactose acceptor.
*****n.d. not determined, below 0.1 nmol Sia

Figure 1

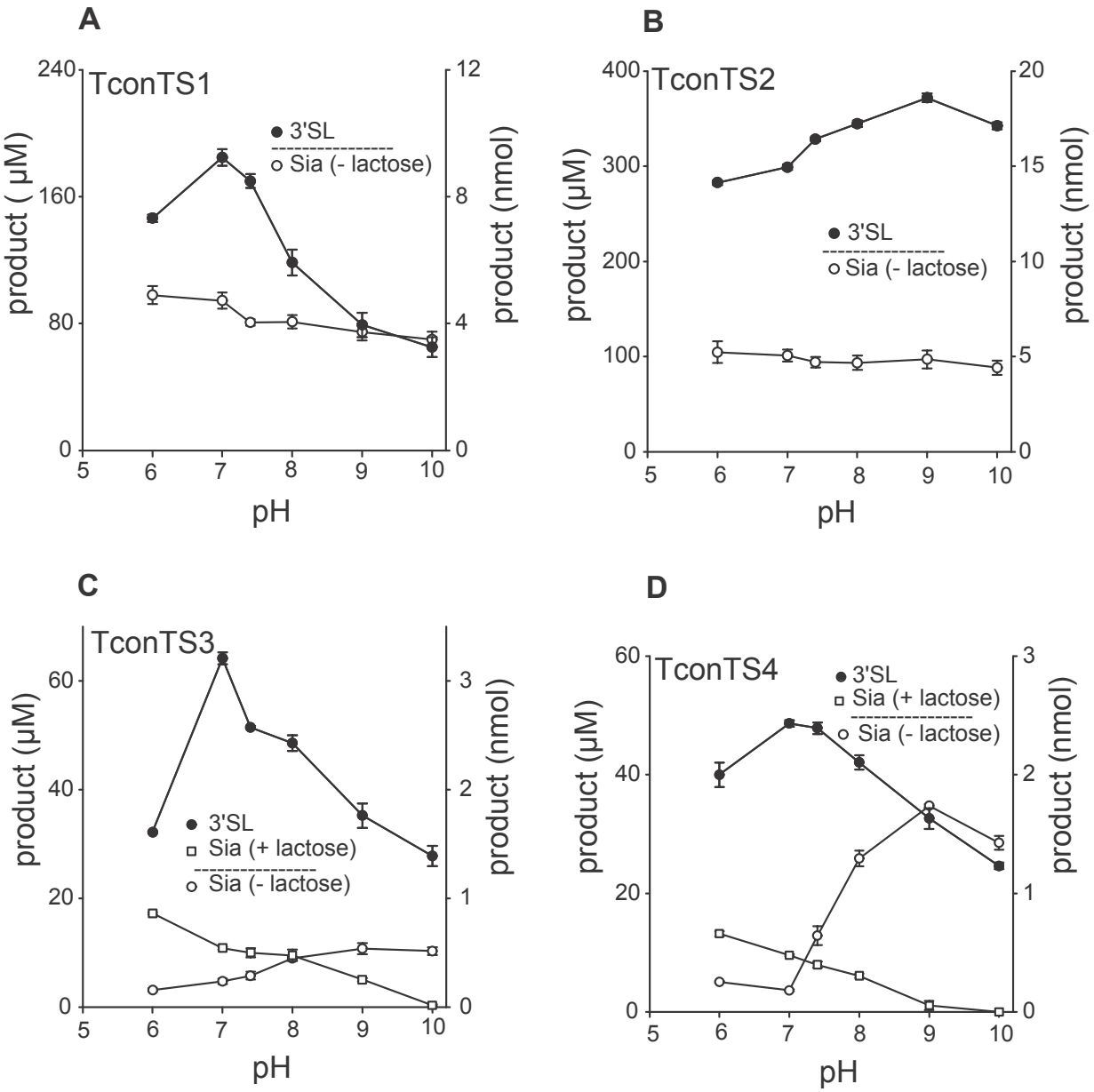

A
TconTS1
3'SL
Sia (- lactose)
product (µM)
product (nmol)
pH
B
TconTS2
3'SL
Sia (- lactose)
product (µM)
product (nmol)
pH
C
TconTS3
3'SL
Sia (+ lactose)
Sia (- lactose)
product (µM)
product (nmol)
pH
D
TconTS4
3'SL
Sia (+ lactose)
Sia (- lactose)
product (µM)
product (nmol)
pH

Figure 2

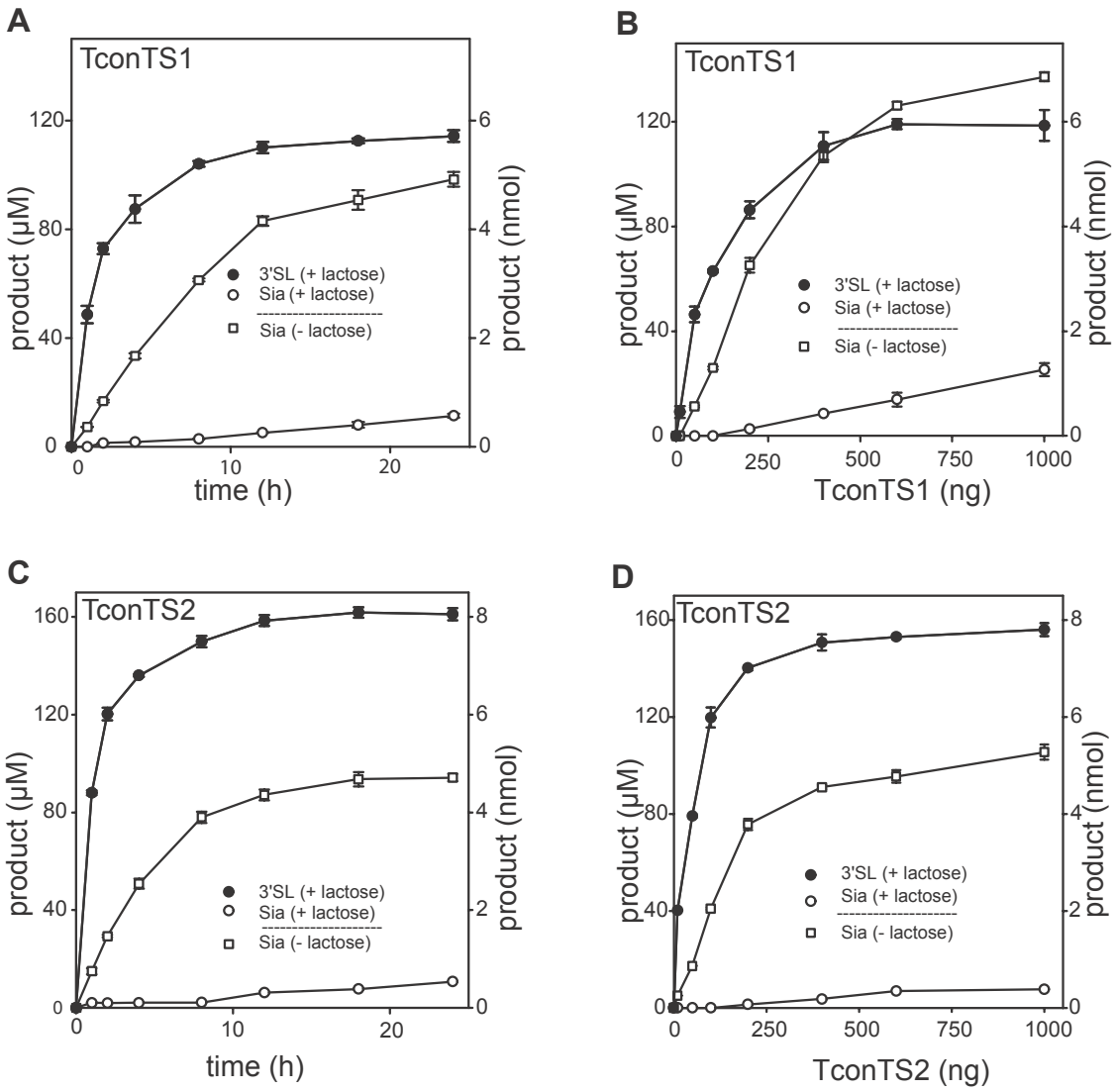

A
TconTS1
product (µM)
product (nmol)
time (h)
3'SL (+ lactose)
Sia (+ lactose)
Sia (- lactose)
B
TconTS1
TconTS1 (ng)
C
TconTS2
time (h)
D
TconTS2
TconTS2 (ng)

Figure 2

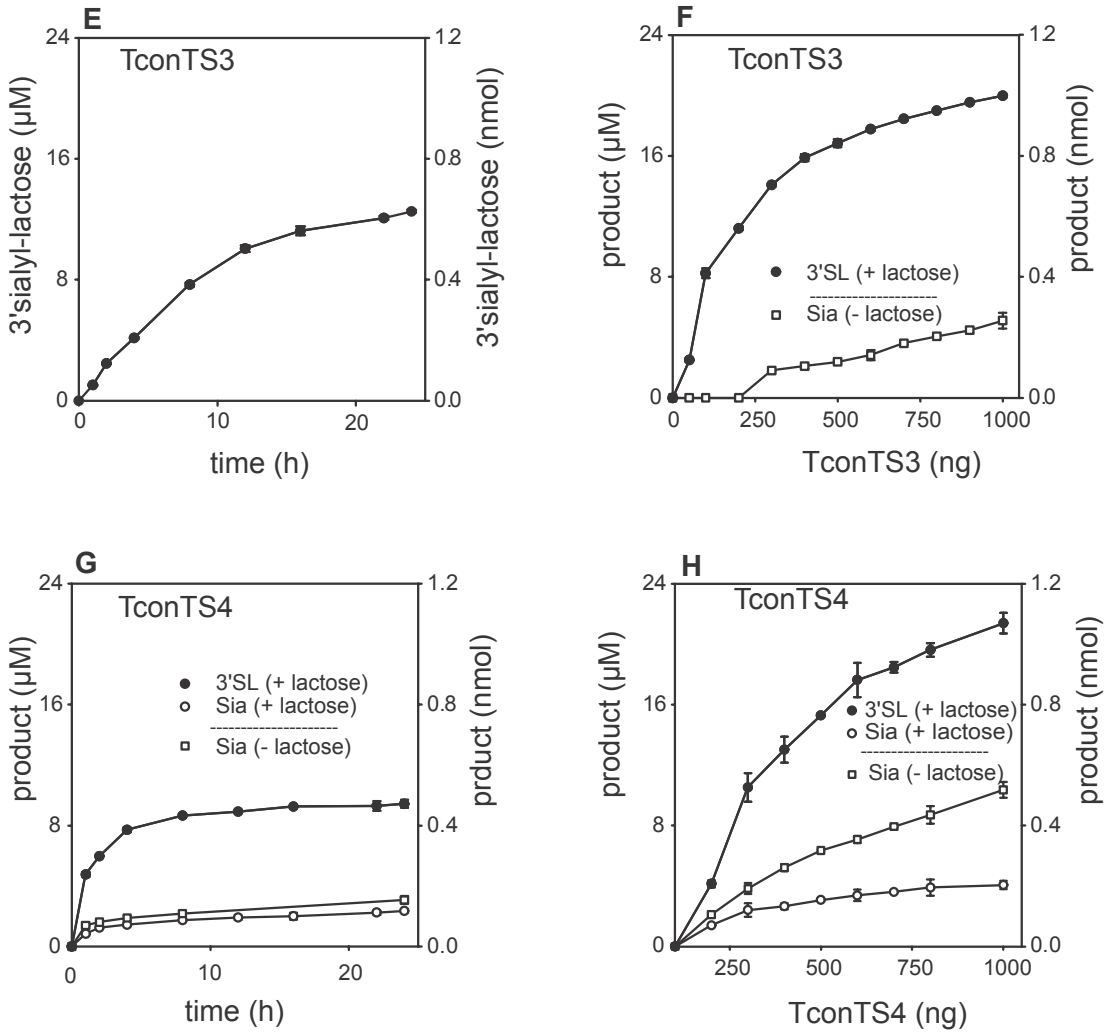

E
TconTS3
3'sialyl-lactose (µM)
3'sialyl-lactose (nmol)
time (h)
F
TconTS3
product (µM)
product (nmol)
3'SL (+ lactose)
Sia (- lactose)
TconTS3 (ng)
G
TconTS4
product (µM)
prduct (nmol)
3'SL (+ lactose)
Sia (+ lactose)
Sia (- lactose)
time (h)
H
TconTS4
product (µM)
product (nmol)
3'SL (+ lactose)
Sia (+ lactose)
Sia (- lactose)
TconTS4 (ng)

Figure 3

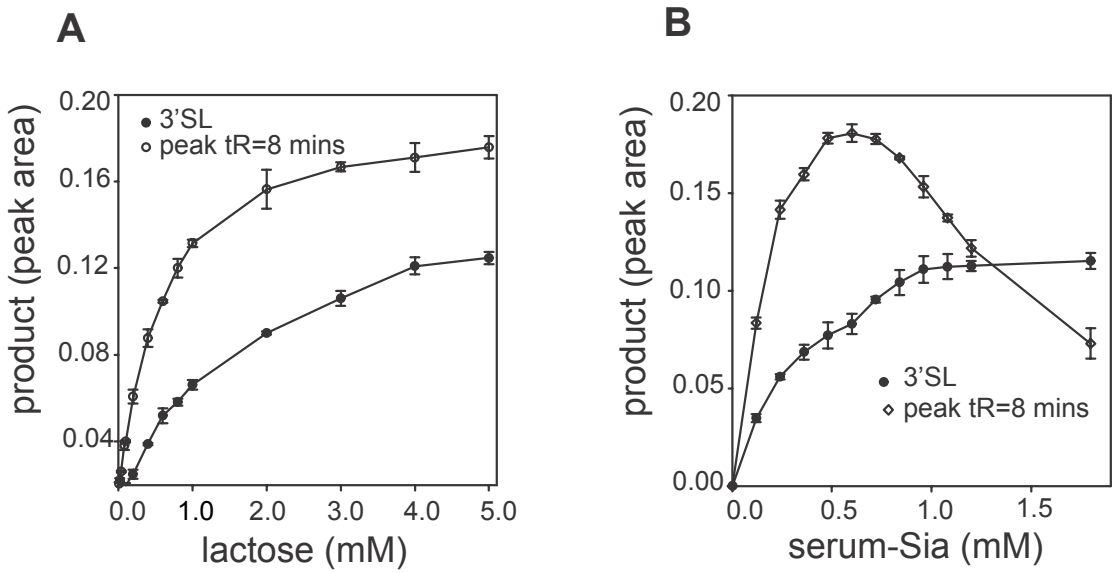

A
product (peak area)
0.20
0.16
0.12
0.08
0.04
3'SL
peak tR=8 mins
0.0
1.0
2.0
3.0
4.0
5.0
lactose (mM)
B
product (peak area)
0.20
0.15
0.10
0.05
0.00
3'SL
peak tR=8 mins
0.0
0.5
1.0
1.5
serum-Sia (mM)

Figure 4

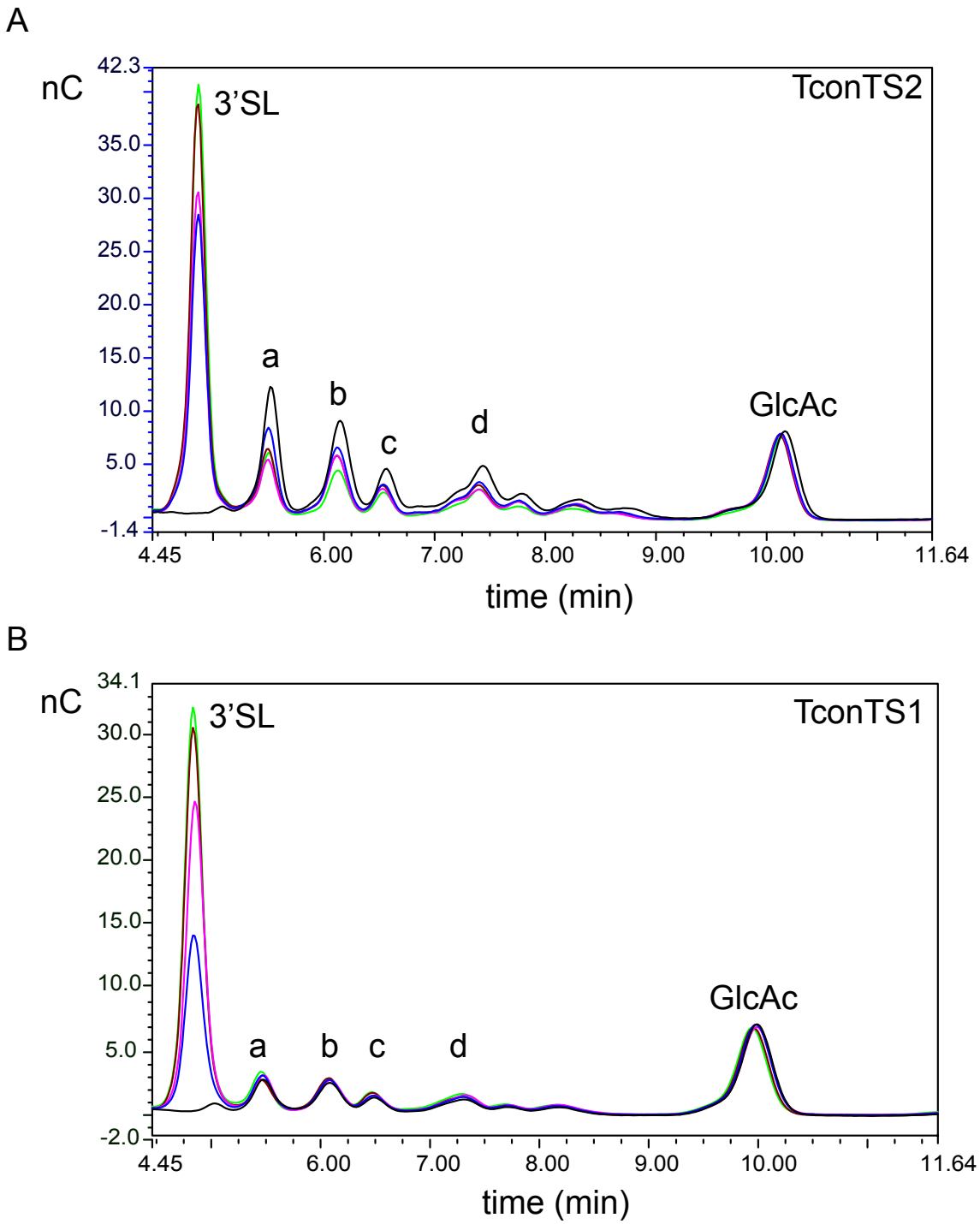

A
nC
42.3
35.0
30.0
25.0
20.0
15.0
10.0
5.0
-1.4
3'SL
TconTS2
a
b
c
d
GlcAc
4.45
6.00
7.00
8.00
9.00
10.00
11.64
time (min)
B
nC
34.1
30.0
25.0
20.0
15.0
10.0
5.0
-2.0
3'SL
TconTS1
a
b
c
d
GlcAc
4.45
6.00
7.00
8.00
9.00
10.00
11.64
time (min)

Figure 5

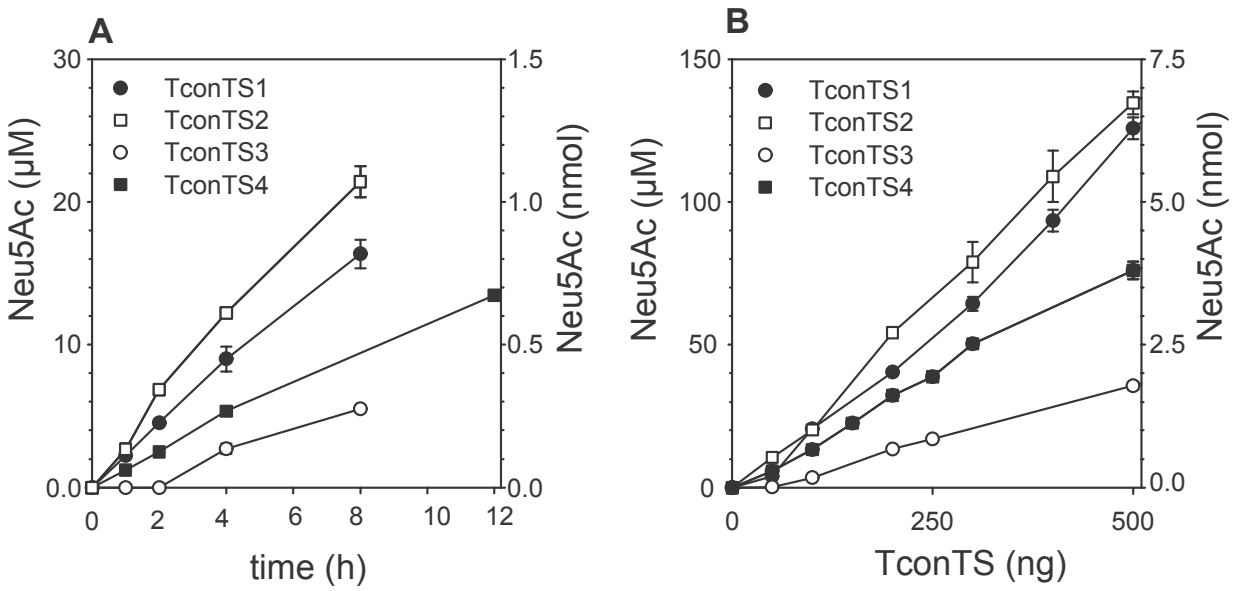

A
B
TconTS1
TconTS2
TconTS3
TconTS4
Neu5Ac (µM)
Neu5Ac (nmol)
time (h)
TconTS (ng)
0.0
10
20
30
0.5
1.0
1.5
0
2
4
6
8
10
12
50
100
150
2.5
5.0
7.5
250
500

Figure S1

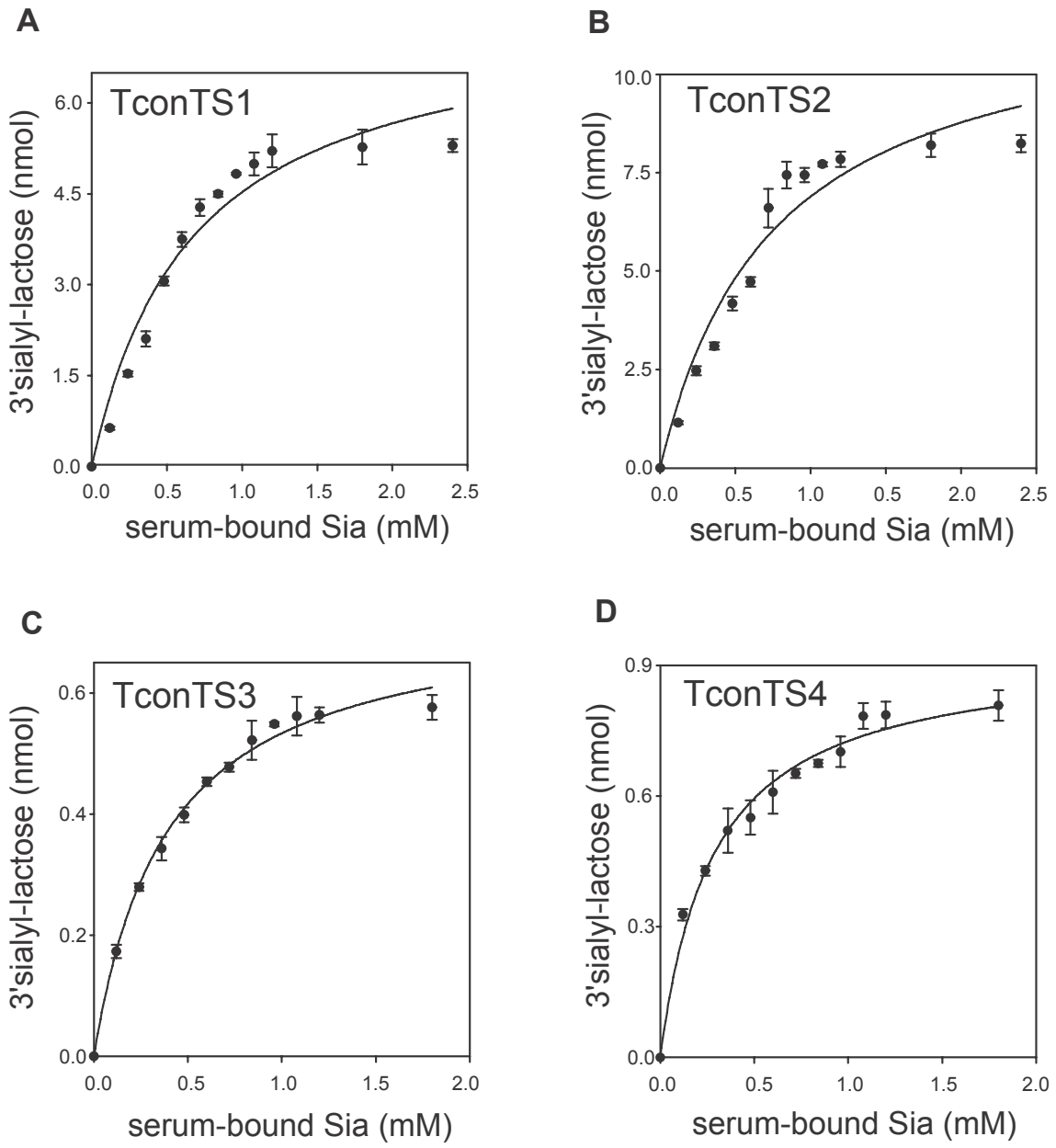

A
TconTS1
3'sialyl-lactose (nmol)
serum-bound Sia (mM)
B
TconTS2
3'sialyl-lactose (nmol)
serum-bound Sia (mM)
C
TconTS3
3'sialyl-lactose (nmol)
serum-bound Sia (mM)
D
TconTS4
3'sialyl-lactose (nmol)
serum-bound Sia (mM)

Figure S2

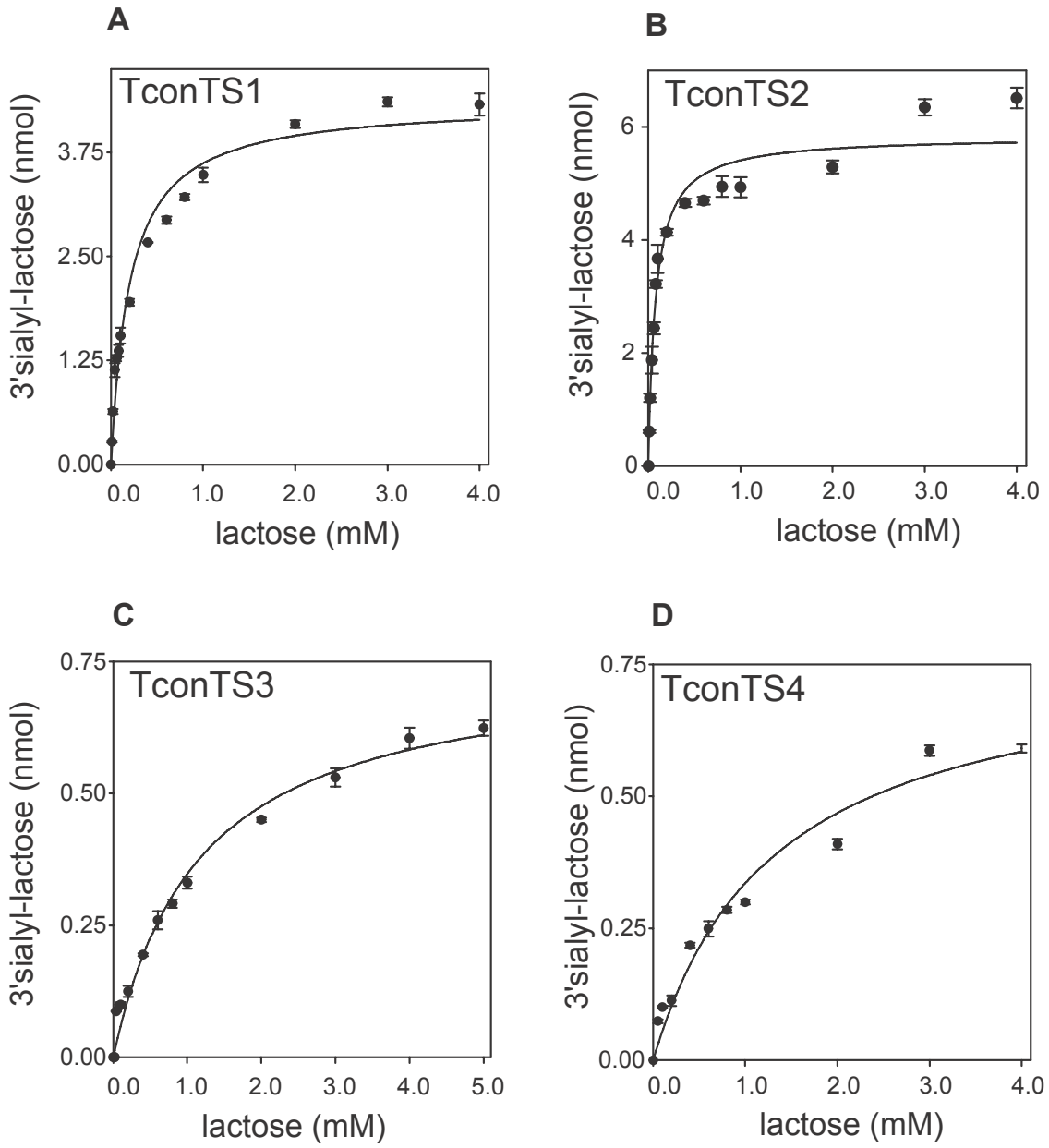

A
TconTS1
3'sialyl-lactose (nmol)
0.00
1.25
2.50
3.75
0.0
1.0
2.0
3.0
4.0
lactose (mM)
B
TconTS2
3'sialyl-lactose (nmol)
0
2
4
6
0.0
1.0
2.0
3.0
4.0
lactose (mM)
C
TconTS3
3'sialyl-lactose (nmol)
0.00
0.25
0.50
0.75
0.0
1.0
2.0
3.0
4.0
5.0
lactose (mM)
D
TconTS4
3'sialyl-lactose (nmol)
0.00
0.25
0.50
0.75
0.0
1.0
2.0
3.0
4.0
lactose (mM)

Figure S3

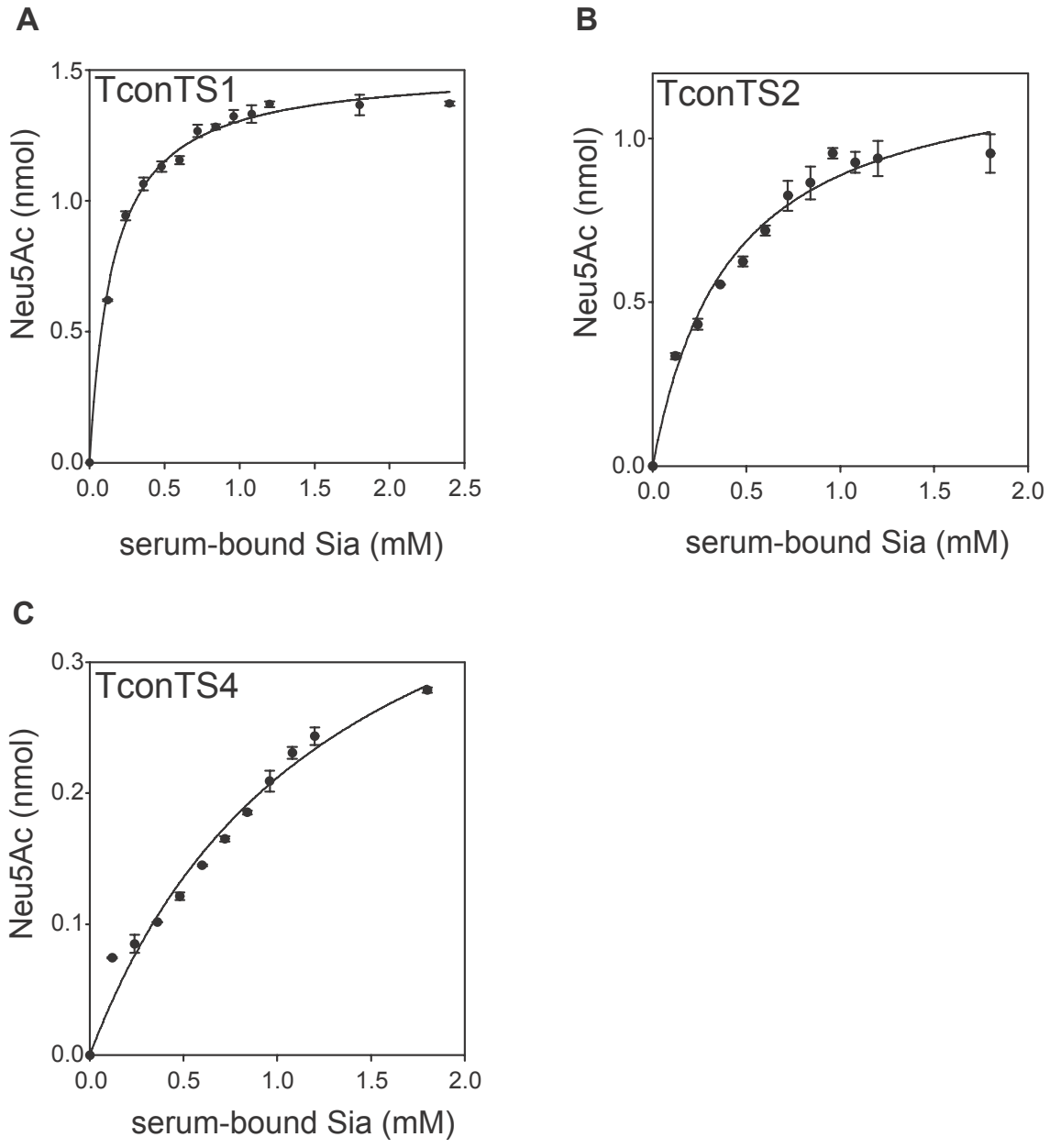
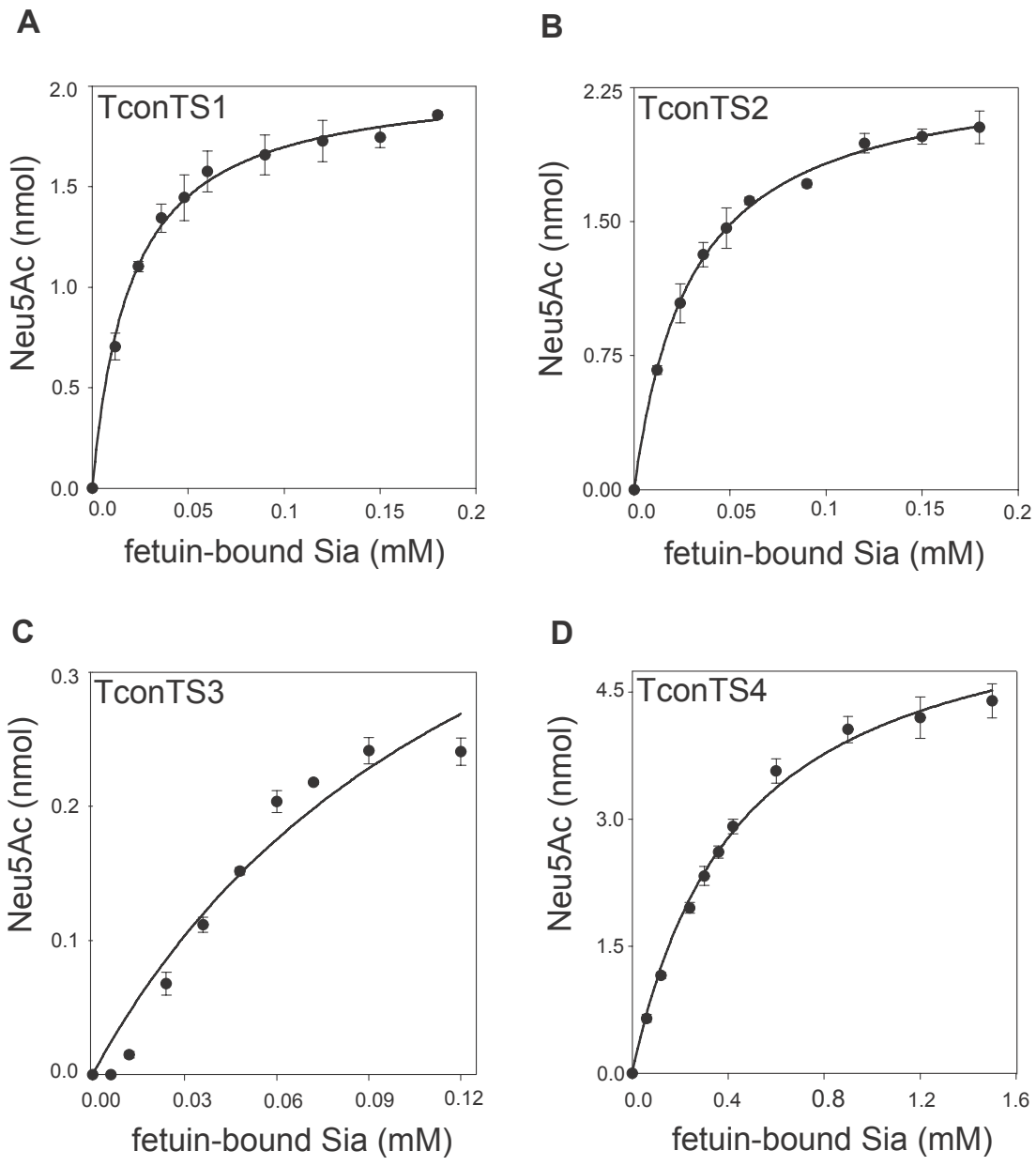

A
TconTS1
Neu5Ac (nmol)
serum-bound Sia (mM)
B
TconTS2
Neu5Ac (nmol)
serum-bound Sia (mM)
C
TconTS4
Neu5Ac (nmol)
serum-bound Sia (mM)

Figure S4


A
TconTS1
Neu5Ac (nmol)
fetuin-bound Sia (mM)
B
TconTS2
Neu5Ac (nmol)
fetuin-bound Sia (mM)
C
TconTS3
Neu5Ac (nmol)
fetuin-bound Sia (mM)
D
TconTS4
Neu5Ac (nmol)
fetuin-bound Sia (mM)

3.4 Results

Carbohydrate recognition specificity of trans-sialidase lectin domain from *Trypanosoma congolense*

Mario Waespy, Thaddeus T. Gbem, Leroy Elenschneider, André-Philippe Jeck, Christopher J. Day, Lauren Hartley-Tassell, Nicolai Bovin, Joe Tiralongo, Thomas Haselhorst, Sørge Kelm (2015)

PLOS Neglected Tropical Diseases (Manuscript in press)
(doi:10.1371/journal.pntd.0004120)

Contribution of Mario Waespy:

- Cloning, expression and purification of recombinant TconTS proteins
- Glycan array and STD NMR data acquisition and evaluation
- Establishment of TconTS-LD binding/inhibition assay, which was optimised by Leroy Elenschneider during his bachelor thesis planed and supervised by Mario Waespy
- Evaluation of TconTS oligomerisation data obtained by André-Philippe Jeck during his bachelor thesis planed and supervised by Mario Waespy
- Homology modelling
- Preparation and draft of the manuscript

Carbohydrate Recognition Specificity of Trans-Sialidase Lectin Domain from *Trypanosoma congolense*

Mario Waespy[1], Thaddeus T. Gbem[1,2], Leroy Elenschneider[1], André-Philippe Jeck[1], Christopher J. Day[3], Lauren Hartley-Tassell[3], Nicolai Bovin[4], Joe Tiralongo[3], Thomas Haselhorst[3], Sørge Kelm[1,2,3#]

[1] Centre for Biomolecular Interactions Bremen, Faculty for Biology and Chemistry, University Bremen, Bremen, Germany

[2] Africa Centre of Excellence for Neglected Tropical Diseases and Forensic Biotechnology, Ahmadu Bello University, Zaria, Nigeria

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

[4] Shemyakin Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

[#] corresponding author: skelm@uni-bremen.de

Abstract

Fourteen different active *Trypanosoma congolense* trans-sialidases (TconTS), 11 variants of TconTS1 besides TconTS2, TconTS3 and TconTS4, have been described. Notably, the specific transfer and sialidase activities of these TconTS differ by orders of magnitude. Surprisingly, phylogenetic analysis of the catalytic domains (CD) grouped each of the highly active TconTS together with the less active enzymes. In contrast, when aligning lectin-like domains (LD), the highly active TconTS grouped together, leading to the hypothesis that the LD of TconTS modulates its enzymatic activity. So far little is known about the function and ligand specificity of these LDs. To explore their carbohydrate-binding potential, glycan array analysis were performed on the LD of TconTS1, TconTS2, TconTS3 and TconTS4. In addition, Saturation Transfer Difference (STD) NMR experiments were done on TconTS2-LD for a more detailed analysis of its lectin activity. Several mannose-containing oligosaccharides, such as mannobiose, mannotriose and higher mannosylated glycans, as well as Gal, GalNAc and LacNAc containing oligosaccharides were confirmed as binding partners of TconTS1-LD and TconTS2-LD. Interestingly, terminal mannose

residues are not acceptor substrates for TconTS activity. This indicates a different, yet unknown biological function for TconTS-LD, including specific interactions with oligomannose-containing glycans on glycoproteins and GPI anchors found on the surface of the parasite, including the TconTS itself. Experimental evidence for such a scenario is presented.

Author Summary

In this study we demonstrated the binding of TconTS lectin domains (TconTS-LD) to high-mannose *N*-glycans and provide evidence for a biological function for this interaction.

TconTS1 and TconTS2 lectin domain bind to galactosyl as well as mannosyl glycans with different affinities as shown by glycan array analysis. Along this line, we have also demonstrated binding of TconTS-LD to high-mannose *N*-glycans on glycoproteins, which is competitively inhibited by the corresponding free high-mannose *N*-glycans, underlining the TconTS-LD substrate specificities. TconTS1 dimerisation is mediated by TconTS-LD binding to its *N*-glycans and enzymatic *N*-deglycosylation of TconTS leads to an increase of monomers. STD NMR results obtained indicate that oligo-mannosylated glycans bind to TconTS-LD to a different site than lactose. In summary, this is good evidence that the lectin domains of TconTS1 and TconTS2 play relevant roles modulating the biological functions of these and possibly other trans-sialidases. Further detailed analysis on TconTS-LD and its role in enzyme activity will lead to a better understanding and possibly to new strategies against Nagana in livestock.

Introduction

The protozoan parasite *Trypanosoma congolense* is the most prevalent cause of animal African Trypanosomiasis (AAT), also called Nagana in cattle and other livestock, causing death to millions of animals resulting in huge economic losses [1-3]. During the parasite's life cycle in the mammalian host and the tsetse fly vector *T. congolense* undergoes different developmental stages utilising various strategies to escape the defence systems of both host and vector. For instance, trypanosomes are unable to synthesise sialic acid (Sia) [4], instead *T. congolense*, like several other trypanosomatids, expresses an unusual glycosyl-transferase called trans-sialidase (TS) that

transfers Sia from host cell glycoconjugates to its own surface structures [5,6].

TS are found in both the African and South American trypanosomes [7-10]. However, their roles in parasite development and pathogenesis appear to be species dependent, as the relevance of TS or sialidase activities has been shown for nagana caused by *T. congolense* [9], but not for sleeping sickness caused by *T. brucei ssp*. The TS from *Trypanosoma cruzi* (TcTS), the causative agent of Chagas' disease in humans [11], is the best characterised [12-18] with the mechanism of sialic acid transfer and catalytic activity being described in detail. It has been suggested that the catalytic domain (CD) of TconTS is located at the N-terminus and folds into a β-propeller structure, similar to that of known bacterial and viral sialidases [19-21]. The CD of trypanosomal TS is presumed to be linked via a well-conserved, relatively long α-helix (22 to 25 amino acids) to a C-terminal domain, whose function has remained unclear. The crystal structure of TcTS [14] revealed that the C-terminal domain folds into a β-barrel topology similar to that of known plant lectins such as GS4 (*Griffonia simplicifolia* lectin 4) [22], GNA (*Galantus nivalis agglutinin*, Snowdrop lectin) [23], LOL (*Lathyrus ochrus* lectin) [24] and WGA (*Wheat germ agglutinin*) [25]. This structural similarity suggests that the C-terminal domain may be a potential carbohydrate-binding site or “lectin-like” domain (LD). In contrast to TcTS, only a few studies have investigated the enzymatic activities of *T. brucei* TS (TbTS) [26-29] and TconTS [6,30,31]. However, given the overall high primary sequence similarity of all TS at the catalytic site it can be assumed that the molecular mechanisms of the African TS are similar to those described for TcTS.

In addition to several other TS-like genes, *T. congolense* possesses eleven TS1 (TconTS1) genes encoding variants with 96.3% overall amino acid identity [30]. Recombinant TconTS1 variants are able to desialylate fetuin in the presence of lactose to generate α2,3-sialyllactose (3‘SL), demonstrating not only their TS activity but also their sialidase activity, reflecting their ability to hydrolyse terminal sialic acids [30]. Three additional *T. congolense* TS family members, TconTS2, TconTS3 and TconTS4, have recently been described, sharing more then 40% amino acid identity [31]. Furthermore, kinetic data show that all TconTS investigated so far have different affinities for glycoprotein and oligosaccharide substrates [30,31].

Previous research on trypanosomal TS has been focussed on investigating the TS CD with respect to substrate specificities, mechanisms of sialic acid transfer and sialidase activities [6,12,14,27,30-

34]. However, apart from sequence and structural data, limited information regarding the LD of the trypanosomal TS is available, with the actual function of the LD in TS remaining unknown. Due to structural similarities with known lectins, it had been proposed [13,35] that TS LD binds carbohydrates and may play a role in mediating cell adhesion. Recently, Ammar et al. suggested that TconTS lectin like domain binds sialic acids and is involved in endothelial cell activation [36]. However, to the best of our knowledge no direct evidence for carbohydrate-binding specificities of TS LD has been described.

Interestingly, a detailed phylogenetic analysis comparing TS domains revealed that the TconTS-CDs grouped the highly active TconTS together with the less active enzymes. In contrast, when aligning TconTS-LDs, the highly active TconTS grouped together [31], indicating a potential role of TconTS-LD in modulating enzyme activities.

Here we report on the biochemical characterisation of four recombinant TconTS-LD (TconTS1-4) employing glycan array, STD NMR and binding/inhibition assays that identified TconTS-LD as a carbohydrate recognition domain (CRD), with several oligosaccharides identified as TconTS-LD binding partners. Interestingly, prominent ligands involved were oligosaccharides with terminal mannose, which are not substrates for trypanosomal TS activity, suggesting previously not reported biological functions. In addition, we provide strong evidence for a second binding site for oligosaccharides with terminal galactose moieties.

Results

Expression of recombinant TS-LD

To characterise the LDs of TconTS the gene sequences encoding the TconTS1-LD, TconTS2-LD, TconTS3-LD and TconTS4-LD were subcloned into a modified pET28a bacterial expression vector as described under Methods. All proteins comprise a N-terminal poly histidine tag (His-tag) directly attached to maltose binding protein (MBP) as well as C-terminal SNAP- and Strep-tags. In these proteins, the tags flanking TconTS-LD can be enzymatically cleaved using the tobacco etch virus (TEV) and human rhinovirus 3C (HRV 3C) proteases (Fig. 1 A).

Fig. 1. Generated TconTS-LD proteins. A: Schematic presentation of recombinant TconTS-LD fusion proteins expressed in bacteria. Fusion tags flanking TconTS-LD are: His: poly histidine tag, MBP: maltose binding protein tag, TEV: *tobacco etch virus* protease cleavage site, 3C: human rhinovirus 3C protease cleavage site, SNAP: SNAP-tag, *Strep*: *Strep-tag*. B: Homology model of TconTS2-LD comprising the α-helix calculated using TconTS2 amino acid sequence and crystal structure of *Trypanosoma cruzi* TS (PDB code: 3b69) as template employing the software Yasara. C: The molecular electrostatic surface of the homology model (B) was calculated using the ESPPME method of Yasara structure. Red colour indicates a positive potential, blue a negative and grey a neutral. A yellow ellipse indicates the groove encompassing the proposed binding site. D**:** SDS-PAGE of purified TconTS-LD proteins. After expression in *E. coli* Rosetta pLacI, 1-2 µg double affinity purified recombinant TconTS-LD, containing and lacking the α-helix, were loaded in each lane of an 10 % SDS polyacrylamide gel as indicated. After electrophorese, the gel was stained using Coomassie Brilliant Blue.

Recombinant protein comprising only His-MBP-SNAP-Strep, but no TconTS-LD was used in control experiments. For TconTS-LD interaction studies, based on homology models (Fig. 1 B and C) two sets of constructs were generated with or without the α-helix connecting CD to the LD (Table 1), to investigate its potential influence on binding activity. Expression conditions were optimised for efficient production of soluble TconTS-LD in amounts ranging from 0.5 – 2 mg/L bacterial culture as described under Methods. After tandem affinity chromatography employing Ni-NTA and Strep-tag consecutively, protein purity was confirmed by gel electrophoresis and Western blot analysis (Fig. 1 D). All TconTS-LD were obtained as pure proteins clearly showing a migration shift due to the presence or absence of the N-terminal α-helix.

Table 1. Bacterial and eukaryotic expressed recombinant TconTS constructs.

		Expression system		Domains					
Nr.	TconTS	*E. coli*	CHO-Lec1	CD	αHel	LD	Length* (AA)	MW (kDa)	Abbreviation
1	1	+	-	-	+	+	258	27.5	TconTS1-αHel-LD
2	1	+	-	-	-	+	236	25.1	TconTS1-LD
3	2	+	-	-	+	+	263	29.3	TconTS2-αHel-LD
4	2	+	-	-	-	+	238	26.5	TconTS2-LD
5	3	+	-	-	+	+	250	27.2	TconTS3-αHel-LD
6	3	+	-	-	-	+	225	24.3	TconTS3-LD
7	4	+	-	-	+	+	261	28.1	TconTS4-αHel-LD
8	4	+	-	-	-	+	236	25.3	TconTS4-LD
9	1	-	+	+	+	+	711	77.5	TconTS1
10	2	-	+	+	+	+	694	77.0	TconTS2
11	3	-	+	+	+	+	682	74.7	TconTS3
12	4	-	+	+	+	+	747	82.7	TconTS4

* Length of the proteins are given in number of amino acids (AA) for the TconTS part excluding the N- and C-terminal tags.

Screening and identification of glycans as potential TconTS-LD ligands

Glycan array analysis was performed to identify potential TconTS-LD oligosaccharides binding partners. Recombinant TconTS-LD containing His and MBP fusion tags (Table 1, Fig. 1 A) were

pre-complexed with anti His mouse polyclonal antibody, anti mouse-IgG-TexasRed conjugated rabbit polyclonal antibody and anti rabbit-IgG-TexasRed conjugated donkey polyclonal antibody. These were then applied to glycan arrays printed onto SuperEpoxy2 glass slides comprising 367 diverse biologically relevant glycan structures (Fig. S2). The major subset of glycans bound by TconTS-LD are summarised in Fig. 2 (full binding data provided in Fig. S2, Table S1). As expected, initial glycan array experiments revealed signals associated with maltose, maltotriose, isomaltotriose, maltotretraose, isomaltotetraose and related glycans due to the binding of MBP (Fig. S1 A). Therefore, 10 mM maltose was added as a competitor during binding and washing steps to inhibit the MBP interaction with maltose and related structures present on the arrays. Under these conditions the majority of maltose related signals disappeared. Only some signals for maltotriose, maltotetraose and other maltodextrins remained. Given that maltotriose has a more than 6-fold higher affinity for MBP (K_d: 0.16 µM) compared to maltose (K_d: 1 µM) [37], 10 mM maltotriose instead of maltose was used during binding and 1 mM in all wash steps. Under these conditions, binding of MBP to all remaining maltose related structures was successfully inhibited (Fig. S1 B). Another option that could have been used to prevent MBP associated binding to our glycan array would have been a proteolytic cleavage using the TEV protease cleavage site of the recombinant TconTS-LD protein (Fig. 1 A). However, the removal of the MBP-tag and subsequent purification of TconTS-LD leads to low yield of pure TconTS-LD, since often the protease digest is not complete. Therefore, we choose to inhibit MBP binding to maltose-related structures on the glycan arrays with maltotriose in the analyses of all eight TconTS-LD constructs. Glycan array analysis of TconTS2-αHel-LD and TconTS2-LD showed clear binding to several different galactobiose and lactose containing oligosaccharides, as well as to some of their *N*-actetylamine derivatives listed in Fig. 2. Also several fucosylated, and two sialylated glycans were bound, although the binding to these structures was less pronounced compared to unsubstituted *N*-acetyllactosamine. Whereas binding to potential TS substrates containing galactose was not unexpected, surprisingly, we also observed binding to α1-6-mannobiose and α1-3,α1-6-mannotriose, which was similar for TconTS2-LD with and without the α-helix. No obvious preference of TconTS2-LD for any of the oligomannose isomers present on the array was identified. The number of glycan structures bound by TconTS1-LD was lower than that observed

for TconTS2-LD, and no binding to any glycan structures was observed for either TconTS3-LD or TconTS4-LD under the conditions used.

Fig. 2. Summary of TconTS-LDs binding to glycans as determined by glycan array analysis. TconTS-LDs binding to the glycan arrays was determined as described under Methods. Black bars indicate glycans bound by the TconTS-LDs. The presence and absence of the α-helix in TconTS-LD constructs is indicated with "+" and "-", respectively. Further binding data (Fig. S2) and all glycans on the arrays (Table S1) are available as Supporting Information.

STD NMR studies provide evidence for a secondary binding site in TconTS-LD

TconTS2-LD showed the highest lectin activity on glycan arrays. Therefore, in further experiments we focused on TconTS2-LD to more fully characterise and confirm the binding of TconTS-LD to both galactose and mannose containing oligosaccharides observed on the glycan array. Several NMR-based methods have been employed to investigate protein carbohydrate interactions on a structural level. For example, line broadening and peak shifts of ^{1}H-NMR signals from amino acid side chains provide information on the type of amino acids involved as well as the occupation of the binding site and thus equilibrium kinetic data, as has been shown for Siglec-1 [38]. Saturation transfer difference (STD) NMR experiments provide important information on the binding epitope of the complexed carbohydrate ligand, since the relative signal intensities of the difference spectra provide direct information on the proximity of the affected protons to the protein [39]. Protein signals are selectively saturated at -1.00 ppm (on-resonance) and subtracted from an off-resonance spectrum (30 ppm) resulting in the final STD NMR spectrum revealing only protons and functional groups of a binding ligand that are in close proximity to the protein surface. Therefore, STD NMR has been widely used to analyse the binding of lectins to their specific carbohydrate ligands. Lactose and α1-3,α1-6-mannotriose were used as ligands for TconTS2-LD as described under Methods. Fig. 3 A shows the ^{1}H NMR (off resonance) and STD NMR spectra of α1-3,α1-6-mannotriose. The relative signal intensities of the STD spectrum (red line) are almost identical to those of the oligosaccharide ^{1}H NMR spectrum (black line). Binding of lactose to TconTS2-LD was also clearly observed (Fig. 3 B). It is important to note that relatively strong STD NMR signals at

3.36 ppm (β-GlcH2) and at 3.92 ppm (GalH4) provide good evidence that both monosaccharide units of lactose are in close contact with the protein.

Taken together, the STD NMR data confirmed binding of both lactose and α1-3,α1-6-mannotriose to TconTS2-LD, which were initially identified by glycan array analysis, and raises the question as to whether both oligosaccharides bind to the same or distinct sites on TconTS2-LD. To try and address this question, an additional STD NMR competition experiment was performed, where an equal quantity of lactose was added to the TconTS2-LD/α1-3,α1-6-mannotriose complex. If lactose was able to bind to the same site as α1-3,α1-6-mannotriose, the two oligosaccharide ligands would compete, which then would lead to a reduction in the STD NMR signals for one or both ligands, depending on their relative affinities for this site [40]. However, no such reduction was observed for either lactose or α1-3,α1-6-mannotriose (Figs. 3 C and D) suggesting that both ligands are likely to bind simultaneously to different binding sites on TconTS2-LD.

Fig. 3. STD NMR experiments of TconTS2-LD. STD NMR experiments with 5.5 μM TconTS2-LD were performed as described under Methods. Off-resonance (black lines) and STD NMR (red lines) spectra are shown. A: In the presence of 1.73 mM for 3α,6α-mannotriose; B: In the presence of 3.45 mM lactose; C: In the presence of 1.73 mM 3α,6α-mannotriose and 1.73 mM lactose. D: STD NMR effects for the signals indicated were determined as ratios between the intensities at the indicated ppm in the off-resonance spectra and corresponding STD NMR spectra using the software TopSpin 3.2. M1, M2, M3 and M4 stand for the NMR signals of 3α,6α-mannotriose, and L1, L2, L3 and L4 for those of lactose, for which the STD NMR effects are shown either for the single ligands (spectra shown in A and B) or for the mixture (spectra shown in C). n.d.: not determined, since in the ligands mixture the STD effect for M3 could not be determined from the spectra.

Binding of TconTS-LD to mannosylated glycoproteins

We established microtitre plate based binding and inhibition assays to further characterise TconTS-LD binding affinity and specificity. Our glycan array and STD NMR experiments revealed TconTS-LD binding to oligo-mannose oligosaccharides. To further investigate how this specificity

mediates interactions of TconTS-LD with glycoproteins, recombinant human Siglec 2 (huS2-Fc, described under Methods) expressed in Chinese hamster ovary Lec1 (CHO-Lec1) cells was used as a model glycoprotein. Due to the lack of *N*-acetylglucosaminyltransferase 1 (GnT1) CHO-Lec1 cells are unable to synthesise complex and hybrid *N*-glycan structures. Therefore, these proteins contain only high-mannose glycans of the type $Man_5GlcNAc_2$-Asn [41]. Purified huS2-Fc immobilised in microtitre plate wells was incubated with different concentrations of TconTS2-LD and binding was detected as described under Methods. TconTS1-catalytic domain (TconTS1-CD) was used as a control for binding specificity. As shown in Fig. 4 A, concentration-dependent binding of TconTS2-LD to immobilised huS2-Fc was clearly observed, reaching a maximum intensity at approximately 2 µg/mL TconTS2LD due to saturation of the binding sites. Wells without immobilised huS2-Fc were used as a control. No detectable binding to immobilised huS2-Fc was observed for TconTS1-CD at 4 µg/mL (Fig. 4 A), confirming the specificity of this assay.

To investigate whether the binding of TconTS2-LD to huS2-Fc was mediated by its high-mannose *N*-glycans, huS2-Fc was treated with Endoglycosidase H ($EndoH_f$), a recombinant glycosidase, which specifically cleaves high-mannose and some hybrid oligosaccharides from *N*-linked glycoproteins [42]. Released *N*-glycans were then isolated and used in a 1:2 serial dilution as potential competitive inhibitors of TconTS2-LD binding. Fig. 4 B shows the concentration-dependent inhibition of TconTS2-LD by the $EndoH_f$ released *N*-glycans. Importantly the undiluted purified *N*-glycans inhibited binding completely, demonstrating that the interaction of TconTS2-LD with huS2-Fc is exclusively mediated by binding to *N*-glycans.

Fig. 4. Binding specificity of TconTS2-LD. A: TconTS2-LD concentration dependent binding to immobilised huS2-Fc (5µg/mL). B: Competitive inhibition of TconTS2-LD binding to huS2-Fc in the presents of serially diluted high-mannose *N*-glycans. Undiluted inhibitor solution was set to 1.0. The maximum increase in relative fluorescence units (RFU) over time was determined as described under Methods. Data points are means ± standard deviation of triplicates.

Oligomerisation of TconTS

The enzymatic activities of TconTSs were previously characterised [30,31] using recombinant

proteins expressed in CHO-Lec1 cells that therefore contain *N*-glycans of the high-mannose-type, similar to the recombinant huS2-Fc used here in binding/inhibition assays. *N*-glycosylation site prediction analysis revealed 8-9 potential sites in TconTSs, and none for the attached SNAP-tag (Fig. S3). In view of the interaction of TconTS-LD with huS2-Fc, we addressed the question of whether TconTS could oligomerise through binding of *N*-glycans. First, the presence of mannosylated glycans on TconTS1 and TconTS2 expressed in CHO-Lec1 cells was confirmed by lectin blot analysis using concanavalin A (ConA) (Fig. 5). To analyse TconTS oligomerisation we used gel permeation chromatography of TconTS expressed in fibroblasts. These proteins contain CD and LD followed by SNAP- and *Strep*-tags, but no His-MBP at the N-terminus. Fig. 6 A shows the chromatograms of recombinant TconTS1 and TconTS2 under identical conditions. A double peak of similar intensities was observed for TconTS1, whereas TconTS2 showed a clear single peak with a small shoulder in front of it. The molecular weight (MW) of the TS1 peak eluting at 13 mL (peak 2) was 293 kDa and for that at 16.9 mL (peak 3) was 119 kDa (Fig. 6 A). This is consistent with peak 3 representing TconTS1 monomers and peak 2 dimers. Furthermore, a molecular mass of 603 kDa calculated for peak 1 eluting at 10 ml is consistent with tetramers of TconTS1. The deviation from the expected monomer (101 kDa without glycosylation) from the calculated MW (119 kDa) can be explained by increased hydrodynamic volumes (Stokes radii) of the oligomeric TconTS1, which is well-known to influence the elution behaviour of a molecule in size-exclusion chromatography [43]. Additionally, glycosylation also influences the protein elution behaviour. For the prominent peak of TconTS2 in Fig. 6 A eluting at 17 mL (peak 2) a MW of about 113 kDa was calculated, which is consistent with the TconTS2 monomer (100 kDa without glycosylation). The small shoulder at 13.3 mL (peak 1) that represents a MW of about 260 kDa is consistent with the dimeric form of TconTS2 (200 kDa without glycosylation). These findings strongly suggest that both TconTS1 and TconTS2 exist as monomers as well as oligomers in solution, however at different ratios. TconTS1 showed an approximate 1:1 ratio of monomer to dimer, whereas TconTS2 mainly migrates as monomer under the conditions used. To address whether oligomerisation is mediated by *N*-linked glycans, TconTS1 was enzymatically deglycosylated using $EndoH_f$, and the resulting oligomeric state assessed by size-exclusion chromatography. As shown in Fig. 5 a clear molecular weight shift as well as a reduction in signal

intensity for ConA binding was observed, strongly indicating the release of mannose containing glycans from TconTS1. Subsequent gel permeation chromatography of the deglycosylated TconTS1 resulted in a changed elution profile (Fig. 6 B, dashed line) compared to the untreated protein (Fig. 6 B, solid line). Calculating the molecular weight, peak 2 of the deglycosylated protein in Fig. 6 B is determined as the dimeric form with 260 kDa (elution volume: 13.3 mL) and peak 3 as the monomer with 109 kDa (elution volume: 17.2 mL). The small differences in MW can again be explained by the reduced glycosylation effect on the Stokes radius after $EndoH_f$ treatment. Comparing the MW of monomer and dimer from the $EndoH_f$ treated sample (peak 2: 260 kDa; 3: 109 kDa, dashed line) to those from the untreated (2: 293 kDa; 3: 119 kDa, solid line), it can be seen that both are decreased due to the loss of high-mannose *N*-glycans released by $EndoH_f$. Interestingly, it was also observed that $EndoH_f$ treatment of TconTS1 reduced the abundance of dimers indicated by the smaller peak 1 (dashed line) in chromatogram B compared to untreated TconTS1 (Fig. 6 B, solid line). In summary, these results provide strong evidence for oligomerisation of TconTS1 by binding to its *N*-glycans. Similar, but less pronounced is the oligomerisation of TconTS2.

Fig. 5. Cleavage of *N*-glycans from TconTS1 and TconTS2. 100 µg TconTS1 and TconTS2 expressed in CHO-Lec1 cells were incubated without (-) or with (+) 4000 units $EndoH_f$ glycosidase under native (-) or denaturing (+) conditions as described under Methods. A: 10 % SDS polyacrylamide gel with subsequent Coomassie Brilliant Blue staining. B: Western blot of deglycosylated TconTS, detected using anti-*Strep*-tag mAb. C: Concanavalin A (ConA) lectin blot using 2 µg/mL biotinylated ConA and an peroxidase conjugated avidin-biotin system (ABC-Kit, VECTASTAIN) for detection. 50 ng TconTS sample were used for ConA and Wester blot analysis and 800 ng for SDS-PAGE.

Fig. 6. Oligomerisation of TconTS. Size exclusion chromatography on Superdex200 column and detection at E_{280nm} was used, MWs of different peaks were determined and assigned to oligomeric (1), dimeric (2) and monomeric (3) TconTS1 as described under Methods. A: Oligomerisation pattern of TconTS1 (solid line) or TconTS2 (dashed line), 300 µg protein was loaded. B: Effect of

TconTS1 deglycosylation on enzyme oligomerisation. 100 µg TconTS1 were deglycosylated using 4000 units EndoH_f glycosidase in phosphate buffer pH 7.4 for 4 hours and directly applied to the column (dashed line). As control 100 µg TconTS1 was treated correspondingly without the addition of enzyme was loaded (solid line).

Discussion

Lectin domains (LD) of the four TconTS1-4 were expressed and characterised with respect to their ability and specificity to bind carbohydrate structures. Structural comparison to other known bacterial *(Salmonella typhimurium* LT2*)* [44] and viral (*Vibrio cholerae* neuraminidase) [45] sialidases, as well as to plant lectins (*Griffonia simplicifolia* lectin 4, GS4; *Lathyrus ochrus* lectin, LOL) [22,46] provided structural evidence for potential carbohydrate-binding of TconTS-LD. Besides typical structural elements seen for several lectins, such as the β-barrel topology, each TconTS-LD also comprises a cluster of histidine, phenylalanine and arginine residues in its potential binding site, a rather shallow indentations (Fig. 1 C). These could presumably be involved in carbohydrate recognition via aromatic side-chain and sugar ring interaction as well as hydrogen bonding, as described for other lectins [47]. Furthermore, it is noticeable that this potential TconTS-LD carbohydrate-binding site is oriented in the same direction as the TconTS-CD catalytic site, similar to *T. cruzi* TS, *T. rangeli* SA and leech IT-sialidase [14,35,48]. This structural organisation of TconTS-CD and LD appears to be stabilised by a relatively extended close contact site between both domains comprising a network of hydrogen bonds and complementary hydrophobic patches (Fig. 7).

Fig. 7. Contact site between TconTS-CD and LD. Homology model of TconTS1 was calculated using the crystal structure of TcTS (PDB: 3b69) as template and the software Yasara. Molecular surface of TconTS1 was calculated using the surface module of Yasara Structure. Illustrated are the parts of TconTS-CD (yellow) and LD (orange), which are in close contact to each other. The α-helix connecting both domains is shown in blue.

We have identified TconTS-LD as a carbohydrate-binding domain. Employing glycan arrays, TconTS2-LD showed binding to a variety of oligosaccharides (Fig. 2), whereas only a few of the glycans presented on the arrays were bound by TconTS1-LD, and none by TconTS3-LD or TconTS4-LD. It remains unclear, why only TconTS1-LD containing the α-helix at its N-terminus showed binding activity in the glycan array experiments. One explanation could be that folding or the accessibility of the binding site is compromised for TconTS1-LD without the α-helix serving as a spacer to the N-terminal MBP tag. It appears unlikely, that the α-helix is directly involved in binding activity, since the binding pattern and signal intensities for TconTS2-LD were not influenced by the presence of the α-helix. Furthermore, our binding assay data (Fig. 3 A) have provided evidence that TconTS1-LD also binds high-mannose glycans, which were not available on the glycan arrays, besides 1N and 4D. In addition, EndoH treatment reduced oligomerisation of TconTS1. Finally, it should also be pointed out that most of the differences in amino acid sequence among the 11 TconTS1 gene variants occur in the LD and are clustered close to the postulated binding site [30]. Therefore, it may well be possible that the other TconTS1-LD variants, besides the TconTS1a-LD used in this study, have different carbohydrate-binding specificities. Interestingly, variants that differ only in this cluster (TconTS1b and TconTS1e) have different kinetic properties for the trans-sialylation reaction, which further supports the hypothesis that TconTS-LD modulates also the enzymatic activities. This may also explain the low enzymatic activity of TconTS3 and TconTS4, since no carbohydrate-binding activity was observed for TconTS3-LD and TconTS4-LD. However, as discussed above for TconTS1-LD, TconTS3-LD and TconTS4-LD may bind to carbohydrate structures not present on the glycan arrays. In this context it should be kept in mind that the amino acid sequence diversity of TconTS-LDs (Table S2) is high enough to allow different binding activities. Furthermore, the structures of spacers used to immobilise the glycans may have prevented binding of TconTS3-LD and TconTS4-LD, since these affect lectin binding to glycan arrays [49,50]. However, given the diverse spacer structures and lengths utilised here, it appears less likely that binding would have escaped detection. Finally, we cannot exclude the possibility that insufficient folding of TconTS-LDs in the bacteria during expression is the reason for undetected lectin activity.

According to a phylogenetic analysis comparing separately TconTS-LDs and TconTS-CDs [31]

TconTS1-LD and TconTS2-LD are more closely related with each other than TconTS3-LD and TconTS4-LD. However, when comparing CDs, TconTS2-CD is more closely related to TconTS3-CD and TconTS4-CD, but least to TconTS1-CD. Interestingly, the highest enzyme activities of TconTS were found with TconTS1 and TconTS2, whereas TconTS3 and TconTS4 were 100- or 1000-fold less active [31]. Together with the carbohydrate-binding activities of TconTS1-LD and TconTS2-LD described in this study, it can be postulated that LDs may directly influence enzyme activities. In agreement with this hypothesis is the observation that disrupting a salt bridge between R642 and E648 in the LD of *T. cruzi* TS enhanced trans-sialidase relative to sialidase activity [35,51]. Along this line it could be postulated that binding of oligosaccharide substrates to TconTS via oligomannose clusters may lead to an improved presentation of the terminal acceptor galactose residues towards the active centre of the CD and therefore lead to an enhanced catalytic efficiency.

Ammar et al. recently demonstrated that TconTS can induce activation of endothelial cells [36] and assumed that the TconTS-LD may be involved in this process. In this context, they prepared a mutant (Y438H) of TconTS1 (TcoTS-A1 according to their nomenclature) to prevent transfer activity. Since this mutant competed with MAL (*Maackia amurensis* lectin) binding to α2,3-linked Sia of the surface of endothelial cells, they concluded that lectin binding to cell surface carbohydrates play the key role in endothelial cell activation. Based on our data regarding the carbohydrate-binding specificities of TconTS-LD, it appears unlikely that the LD mediates the activation of endothelial cells. However, their findings are in agreement with an involvement of the CD.

TconTS2-LD binds structures containing mannose, such as α1-6-mannobiose and α1-3,α1-6-mannotriose (Fig. 2) and several galactose and lactose containing oligosaccharides. Given that the few fucosylated and sialylated glycans recognised by TconTS2-LD all possess core galactose, lactose or *N*-acetyl-lactoseamine units, it can be assumed that fucosylation and/or sialylation at least in some positions does not interfere with binding. For example, when fucose is linked to positions of the glycan, which is more solvent exposed, it will not disturb ligand binding.

Binding of TconTS-LD to galactosyl, lactosyl and potentially also sialyl glycans was not completely unexpected, since they also serve as substrates for TconTS [6,30,31] and lectin-like binding to sialidase substrates is not uncommon. For example, the lectin-like domain of *Vibrio cholerae*

sialidase (VCS) binds *N*-acetylneuraminic acid in a similar manner compared to the catalytic domain but without hydrolytic activity [52].

Engstler et al. 1995 investigated procyclic TconTS substrate specificity using a selection of sialylated donor substrates and galactosylated acceptor substrate oligosaccharides, including the monosaccharide mannose, as substrates for TS [5]. Whereas transfer of Sia to terminal galactose oligosaccharides and even the monosaccharide galactose was observed, mannose did not appear to be a suitable acceptor for sialic acid transfer. Thus, our discovery that TconTS-LD binds to mannosylated oligosaccharides suggests a yet unknown function of the LD, distinct from that of the CD. The overall ligand specificities of TconTS-LD, binding to both mannosylated as well as galactosylated glycans, but not to glucose containing oligosaccharides (Fig. 2, Fig. S2), differs from that of other mannose-specific lectins, such as concanavalin A (ConA, *Canavalia ensiformis*) [53,54], LOL (*Lathyrus ochrus* lectin) [24] or GNA (*Galantus nivalis agglutinin*, Snowdrop lectin) [23]. In contrast to mannose and glucose, which both show equatorial orientation of the C4-OH group, in Gal the orientation of C4-OH is axial, which does not support binding to ConA or LOL, since C4-OH is involved in carbohydrate recognition by these lectins. Similarly, GNA specifically binds Man via an essential hydrogen bond to its C4-OH group. Considering these ligand selection mechanisms, it appears more likely that two structurally independent binding sites provide TconTS2-LD binding to both mannose- and galactose-containing oligosaccharides. This hypothesis is supported by our STD NMR data that clearly indicate that lactose and α1-3,α1-6-mannotriose do not compete for the same site on TconTS2-LD. According to published STD NMR data for lactose [55] and α1-3,α1-6-mannotriose [56], it could be concluded that both moieties of lactose interact with TconTS2-LD, at least partially. For example, this is indicated by the βGlc-H2 and βGal-H4 protons of lactose (Fig. 3 B). All signals observed in the off-resonance spectrum of α1-3,α1-6-mannotriose could also be identified in the STD NMR spectra, showing no clear preference for any proton. This indicates specific binding and that all three mannose units appear to be in close contact with the protein. A similar binding epitope of α1-3,α1-6-mannotriose has recently been described for the antibiotic Pradimicin S [56]. The binding epitope for lactose is less uniform, suggesting that not all protons of the disaccharide are in the same close contact with the protein. The exception constitutes the βGlc-H2-proton, for which a two-fold higher STD signal

intensity was observed compared to several other protons with similar STD effects than most of the protons of α1-3,α1-6-mannotriose (Fig. 3 D). This finding might be explained by selective interaction of the βGlc configuration of lactose with TconTS2-LD. An important result was that the STD NMR effects of both oligosaccharides were independent of the presence of the other ligand, since they were identical, if TconTS2-LD was incubated with an equimolar mixture, to those obtained for the individual compounds. If they did compete for the same site, reduced STD NMR signals would have occurred for either both ligands, if they bind with similar affinities, or at least for that ligand, which binds with much lower affinity, if they bind with different affinities [40]. It can be excluded that the STD NMR signals of lactose or α1-3,α1-6-mannotriose reflect interactions with the MBP tag, since no binding of MBP to these structures have been observed in our glycan array experiments, which is in agreement with previous studies on MBP specificity applying diverse methods [57,58] including STD NMR [59]. In conclusion, our findings suggest that two distinct binding sites exist on TconTS2-LD, similar to the lectins WGA [60] and GNA [23].

Interestingly, in a previous STD NMR study TcTS binding to lactose was only observed in the presence of Neu5Ac [16]. Apparently, TcTS and TconTS2-LD have different carbohydrate-binding activities, since TconTS2-LD clearly binds to lactose and α1-3,α1-6-mannotriose in the absence of Sia. Furthermore, the binding epitope for lactose in complex with TconTS2-LD is distinct from that observed for TcTS [16], most pronounced is this difference for the STD NMR signal of βGlc-H2, which was not observed for TcTS. This further underlines that the binding site for lactose on TconTS2-LD is distinct from the acceptor-binding site of the CD described for TcTS.

The crystal structure of TcTS [14] revealed that the binding pocket of the catalytic domain is located at the same side as carved β-barrel groove of the lectin domain, in which conserved histidine and tyrosine residues were identified, known to be involved in carbohydrate recognition of other lectins [47]. Therefore, we assume this position to be the potential carbohydrate-binding site on TconTS-LDs. Furthermore this hypothesis is in agreement with our data assuming TconTS-LD interacts simultaneously with more then just one monosaccharide, suggesting an extended binding site, also termed sub-site multivalency [61]. However, detailed structural studies, such as X-ray crystallography of TconTS2-LD with these oligosaccharide ligands are required to locate and investigate these binding sites precisely.

For several lectins it has been reported that oligomerisation enhances binding due to interactions of multivalent glycoconjugate ligands to multiple binding sites of oligomeric lectins [47,62,63]. Also for TconTS1-LD and TconTS2-LD multivalent interactions in larger complexes strengthen binding to the target glycoprotein, since in our binding/inhibition assays pre-complexing of TconTS-LD with the anti His-tag mAb and the corresponding secondary antibodies used for detection lead to stronger signals compared to applying every component in consecutive steps (Fig. S4 A and B).

By large, this binding is mediated by the high-mannose *N*-glycans of huS2-Fc, since it could be inhibited with *N*-linked oligosaccharides, enzymatically released from recombinant huS2-Fc, as competitive inhibitors. Interestingly, for the monosaccharide α-methyl-mannopyranoside, which is a well-known inhibitor for ConA [64], only slight inhibition of TconTS-LD binding could be observed at 50 mM (Fig. S4 C). This evidence for binding towards complex oligosaccharide ligands was already reported earlier for several other lectins, including the lectin GS4 (*Griffonia simplicifolia* lectin IV), which shows high affinity to poly- but not to monosaccharides [47,65]. This mechanism, providing poor affinities for lectins to monosaccharides, prevents unspecific interference from competing, structurally similar molecules and enhances ligand selectivity. In addition, recombinant TconTS enzymes expressed in CHO-Lec1 cell lines contain these *N*-linked high-mannose-type oligosaccharides. Eight to nine potential *N*-linked glycosylation sites are found in recombinant TconTS, distributed over CD and LD. As expected, lectin blot analysis using ConA, for detection of mannose oligosaccharides, clearly confirmed the presence of *N*-linked mannosylated glycans on recombinant TconTS expressed in CHO-Lec1 cells (Fig. 5). Recombinant TconTS exhibit much higher enzyme activity, if expressed in these fibroblasts compared to those expressed in bacteria (Fig. S4 D). This is possibly related to the absence of *N*-glycosylation in bacteria, which may have an indirect effect on enzyme activity by influencing proper enzyme folding or even directly by glycan mediated TconTS oligomerisation. In agreement with the latter scenario, binding to *N*-linked mannosyl oligosaccharides of recombinant TconTS expressed in CHO-Lec1 cells leads to oligomerisation. This conclusion is supported by (1) the observation that TconTS1 elutes in about equal amounts as monomer and as dimer in size exclusion chromatography (Fig. 6 A) and (2) that the removal of *N*-linked glycans with $EndoH_f$ glycosidase leads to a clear shift from the dimer towards the monomeric form of TconTS1 (Fig. 6 B). The remaining dimers are likely to be due to

inefficient deglycosylation, as suggested by ConA lectin blot analysis (Fig. 5) or due to other carbohydrate-independent protein-protein interactions. It remains unclear as to why TconTS1 oligomerises to a larger extent than TconTS2. One possible explanation may be differences in the glycosylation pattern of the two recombinant enzymes. That is, potential *N*-glycosylation sites are distributed differently in both enzymes with nine potential sites for TconTS1, eight for TconTS2, only one site being conserved. Finally, it should be mentioned that it is also unclear, which of the predicted *N*-glycosylation sites of recombinant TconTS utilised in this work are actually glycosylated. In addition, it should be noted that the glycosylation pattern of the native TconTS is unknown as well.

Based on the finding that TconTS-LD binds to high-mannose-type *N*-glycans we assumed that glycoconjugates containing high mannosylated structures might be preferred natural acceptor substrates for TconTS. Interestingly, such high-mannose-type glycans have been identified on the surface of *T. congolense* in both, bloodstream (mammalian host) and procyclic (tsetse vector) forms, either linked to amino acids or as part of the GPI anchors [66-69]. The African parasites express two major stage specific, glycosylphosphatidylinositol (GPI)-anchored glycoproteins on their surfaces, the variant surface glycoprotein (VSG) of the bloodstream form and procyclins of the procyclic form. During development of the bloodstream form (BSF) in the mammalian host, the parasites express a surface coat composed of hundreds of immunologically distinct VSG molecules (antigenic variation) to evade host immune response. These VSGs share relatively little primary sequence homology [70] but are structurally related to each other [71]. It has been demonstrated that VSGs from *T. congolense* and *T. brucei* BSF are highly glycosylated, exhibiting glycan structures similar to those of higher eukaryotic *N*-linked oligosaccharides. Interestingly, they are composed of *N*-linked high-mannose-type oligosaccharides (Man_{5-9}) and *N*-acetyllactosamine oligosaccharides, as well as branched poly-*N*-acetyllactosamine oligosaccharides with a Man_{3-4} core $(GalGlcNAc)_3Man_3GlcNAc$, which were also found to be sialylated in *T. congolense* [72-74]. The fact that terminal Sia were found on VSGs indicate that these are substrates for TS present on the cell surface. In this context, it is plausible, that TconTS-LD contributes to the binding of *T. congolense* VSG to TconTS via oligomannose oligosaccharides present on these glycoproteins.

When parasites are taken up by tsetse fly through a blood meal, VSGs are replaced by procyclic

stage specific, membrane bound, major surface proteins known as procyclins or procyclic acidic repetitive proteins (PARP) [75,76] in *T. brucei* and glutamic acid/alanine-rich protein (GARP) in *T. congolense* [66-68,77]. Interestingly, compared to the highly *N*-glycosylated *T. brucei* BSF VSGs, procyclins only contain a single *N*-glycosylation site, substituted with oligomannose oligosaccharide $Man_5GlcNAc_2$, which is unusual and rare, but not unique [78]. The primary sequence of *T. congolense* procyclic GARP does not contain a single potential *N*-glycosylation site, which was also experimentally confirmed, as well as the absence of conventional *O*-linked glycans [77]. However, two large Man and Gal-rich oligosaccharides of the type $Man_{11}Gal_{6-7}$ linked via phosphodiester bonds to two threonine residues were found [77]. Considering these findings, GARP may also be a potential binding partner for TconTS-LD. Indeed, TconTS-mediated sialylation of GARP has been demonstrated and even procyclin was equally efficiently sialylated by the same enzyme, indicating their functional similarities at least as substrates for TS [5]. However, sialylation of procyclin occurs at the glycan moiety of its GPI anchor [26], which has been structurally characterised [79-81]. They share the common core structure of GPI anchor $EtNH_2$-HPO_4-6Man(α1–2)Man(α1–6)Man(α1–4)$GlcNH_2$(α1-6)-PI, but with an additional glycosylation at the Man_3-core, which is unique for African trypanosomes [80,81]. It comprises oligolactosamine oligosaccharides $(Gal\text{-}GlcNAc)_9$ for *T. brucei* procyclin GPI anchors and oligogalactosyl oligosaccharides Gal_{5-7} for *T. congolense*, which both represent substrates for trans-sialylation [77,80]. In this context it is important to note that GARP was co-purified and co-immunoprecipitated with TS-form 1 from procyclic cultures of *T. congolense* [6], indicating a relatively strong interaction between these two surface proteins. It is in complete agreement with the data of this study that this interaction is mediated by TconTS-LD binding to the GPI anchor Man_3-core of GARP, since binding to similar oligomannose oligosaccharides like Man(α1–3)Man and Man(α1–6)Man was observed by glycan array and STD NMR analysis. Homology models of TconTS1-4 revealed, that the distance between the catalytic tyrosine residue in the active centre of TconTS-CD and a conserved phenylalanine residue in the predicted TconTS-LD carbohydrate-binding site, ranges from 40.5 to 42.6 Å. With an average diameter of 7 Å for a single hexose unit, an oligosaccharide of minimum 6 monosaccharide units from the Man_3-core is needed to reach the catalytic centre of TconTS-CD, depending on the glycosidic linkage of the oligosaccharide. Consistent with this, the

oligosaccharide Gal_{5-7} of the GARP GPI-anchor has the appropriate size to reach the catalytic centre when TconTS-LD is bound to the Man_3-core. In this case both TconTS domains could interact simultaneously with different sections of the same oligosaccharide, whereby the binding affinities of each domain contribute to the overall TconTS binding affinity, which is then clearly enhanced, consistent with the observed co-purification of TS-form 1 with GARP from *T. congolense* cultures [6]. This would be somewhat similar to the situation reported for the *Vibrio cholerae* sialidase, where Neu5Ac binds to a lectin domain without hydrolytic activity, leading to an increased affinity of the enzyme for highly sialylated regions [52]. In addition, the apparent organisation of the two native TS-forms isolated by Tiralongo *et al.* [6], who observed TS-form 2 as dimers and TS-form 1 as oligomers is in very good agreement with the oligomannosyl oligosaccharide-mediated interaction of recombinant, high-mannosylated TconTS1 and TconTS2. It should be noted that these are not identical to purified TS-forms 1 and 2 described by Tiralongo *et al.* [6]. Furthermore, *T. brucei* TS has been previously purified by ConA affinity chromatography from procyclic trypanosomes, suggesting that also this TS is highly mannosylated in its native state on the parasite [27].

In summary, we identified TconTS1-LD and TconTS2-LD as a carbohydrate recognition domain (CRD), exhibiting different binding affinities to several oligogalactosyl, oligolactosamine and oligomannosyl glycans, via two independent binding sites. Functionally, the interaction with specific oligomannosyl structures appears to be required to facilitate TconTS oligomerisation, and binding to oligogalactosyl and oligolactosamine oligosaccharides may represent the recognition event associated with TS acceptor substrate binding. Since the LD from *T. congolense*, *T. brucei*, *T. cruzi* and potentially other trypanosomes are structurally related, this may be a general function of TS-LD and may open new avenues for the design of novel inhibitors for therapeutic applications controlling trypanosomiasis in Africa and Latin America.

Methods

Materials

All chemicals and reagents used in this study were analytical grade. Recombinant $EndoH_f$

glycosidase (EndoH$_f$) was from New England Biolabs, UK. *Pfu* and *Taq* DNA polymerase, *HindIII*, *NcoI*, *NotI*, *SalI* Fast Digest restriction enzymes, T4-DNA ligase, isopropyl-β-D-1-thiogalactopyranoside (IPTG), Dithiothreitol (DTT), Coomassie Brilliant Blue (Page Blue), protein molecular weight marker (PageRuler), GeneJET DNA Gel Extraction Kit, BCA Protein Assay Kit, enhanced chemiluminescence system (ECL-Kit), fluorescein diphosphate tetraammonium salt (FDP), Luria Broth (LB) microbial growth medium, anti His-tag mouse polyclonal antibody, anti mouse-IgG-alkaline phosphatase-conjugated donkey polyclonal antibody (serum purified) were from Thermo Scientific, Germany. Biozym LE Agarose was from Biozyme Scientific, Germany. StrepTactin Sepharose, purification buffers and anti *Strep*-tag rabbit polyclonal antibody were from IBA, Germany. Anti mouse-IgG-TexasRed conjugated rabbit polyclonal antibody, anti rabbit-IgG-TexasRed conjugated donkey polyclonal antibody were purchased from Life Technologies. β-D-galactosyl-(1-4)-α-D-glucose (4α-lactose), β-D-galactosyl-(1-4)-α-D-*N*-acetylglucosamine (4α-*N*-acetyllactosamine), 4-α-D-maltose, α-D-glucopyranosyl-(1-4)-α-D-glucose (4α-maltose), α-D-glucopyranosyl-(1-4)-α-D-glucopyranosyl-(1-4)-α-D-glucose (4α-maltotriose), α-methyl-D-mannose, α-D-mannosyl-(1-2)-D-mannose (2α-mannobiose), α-D-mannosyl-(1-3)-D-mannose (3α-mannobiose), α-D-mannosyl-(1-4)-D-mannose (4α-mannobiose), α-D-mannosyl-(1-6)-D-mannose (6α-mannobiose), α-D-mannosyl-(1-3)-[α-D-mannosyl-(1-6)]-D-mannose (3α,6α-mannotriose), polyethylene glycol sorbitan monolaurate (TWEEN 20), Gel Filtration Markers Kit for protein molecular weights between 29,000-700,000 Da were from Sigma-Aldrich, Germany. Concanavalin A (ConA), Sepharose and biotinylated recombinant ConA were purchased from Galab, Germany. VECTASTAIN ABC detection system was from Vector laboratories, UK. Ultrafiltration units Vivacell and Vivaspin6 were from Sartorius, Germany. X-ray film was purchased from GE Healthcare, Sweden. Protino Ni-NTA Agarose and NucleoBond Midi Plasmid DNA Purification Kit were from Macherey-Nagel, Germany. Polyvinylidenedifluoride (PVDF) membranes were from Millipore, Germany. 96-well transparent microtitre plates were from Sarstedt, Germany. High binding 384-well black microtitre plate were purchased from Corning, USA. 6 mL gravity flow columns were from Biorad, Germany.

Cloning and expression of recombinant TconTS-LD

To obtain non-glycosylated TconTS-LD as recombinant proteins in sufficient amounts, a bacterial expression system based on a modified pET28a+ expression vector was established. Modifications made comprised an N-terminal poly-histidine-tag (His-tag) followed by maltose binding protein (MBP) and a tobacco etch virus (TEV) protease cleavage site [82]. MBP was used to enhance expression and solubility of TconTS-LD in *E. coli*. SNAP- Strep- and His-tags were employed for affinity purification, detection and immobilisation of recombinant protein as previously described [30,31] The DNA sequence encoding the His-MBP part was amplified form pETM-41 (EMBL, Germany) using *Pfu* DNA polymerase and the appropriate primers (Table S3). The purified PCR products were ligated into the *NcoI* and *SalI* digested pET28a (Novagene, USA) bacterial expression vector and transformed into chemical competent *E. coli* BL21 (DE3)(BD Bioscience, Clonetech, USA). Sequence identity was confirmed by commercial sequencing at the Max Planck Institute for Marine Biology (MPI) Bremen and results were evaluated using the Geneious Software.

The modified eukaryotic expression pDEF-based vectors coding for TconTS1, TconTS2, TconTS3 and TconTS4 [30,31] were used as template to amplify TconTS1-LD, TconTS2-LD, TconTS3-LD and TconTS4-LD containing C-terminal SNAP and Strep tags. Two sets of sense primers were designed for each TconTS variant (Table S3). The same reverse primer including a *NotI* restriction site (underlined) was used for all TconTS-LD constructs, since all TconTS-LD constructs contain C-terminal Strep-tag. Purified PCR-products were ligated in frame into the *HindIII* and *NotI* digested, modified pET28aMBP vector and transformed into *E. coli* Rosetta (DE3) pLacI (BD Bioscience, Clonetech, USA). Plasmid preparations of pET28aMBP/TconTS-LD were prepared and characterised as described above.

E. coli Rosetta (DE3) pLacI colonies freshly transformed with pET28aMBP/TconTS-LD were inoculated in 20 mL of 50 µg/mL kanamycin containing Luria Broth (LB) medium and incubated overnight at 37°C and 240 rpm shaking. 2 mL of these overnight cultures were transferred into 1 L of 50 µg/mL kanamycin containing LB medium and grown at 37°C and 240 rpm until an optical density at 600 nm of 0.5 was reached. Recombinant protein expression was then induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG), with a final concentration of 0.1 mM and cells were incubated for additional 120 min at 37°C and 240 rpm. Cells were harvested by

centrifugation for 15 min at 1500 x g, 4°C and the pellet was resuspended in 20 mL lysis buffer 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl. Lysis was done by ultrasonication on ice applying 9 pulses of 20 sec each (50 Watts) with 10 sec pauses between pulses. The bacterial lysates were centrifuged for 30 min at 15000 x g, 4°C. Clear supernatants were transferred to 4 mL of equilibrated Ni-NTA beads and incubated on a rotation wheel (6 rpm) at 4°C for 120 min. The suspensions were transferred to 6 mL gravity flow columns in portions, until all beads were settled in the column. Beads were washed with 40 mL wash buffer containing 50 mM NaH_2PO_4, pH 8.0, 150 mM NaCl, 20 mM imidazole. Recombinant TconTS-LD was eluted using 250 mM imidazole in 50 mM NaH_2PO_4, pH 8.0, 150 mM NaCl and directly applied to a new gravity flow column containing 1.6 mL StrepTactin beads equilibrated with wash buffer (100 mM Tris-Cl, pH 8.0, 150 mM NaCl and 1 mM EDTA) and beads were washed with 5 column volumes of wash buffer. Recombinant proteins were eluted with wash buffer containing 2.5 mM desthiobiotin and dialysed against 10 mM phosphate buffer, pH 7.4 using a Vivaspin6 filtration unit with a 100 kDa cut off. Purified TconTS-LD was characterised by SDS-PAGE and Western blot analysis and quantified by BCA assay using bovine serum albumin (BSA) as standard.

T. congolense recombinant TconTS1 and TconTS2 containing catalytic (CD) and lectin domain (LD) expressed by CHO-Lec1 cells were prepared from culture supernatants and characterised using SDS-PAGE, Western Blot and BCA assay analysis as described [30].

Glycan Array

Glycan arrays consisting of 367 diverse glycans with and without the presence of one of three spacers (sp2, sp3 or sp4 [49]) were prepared from two previously described glycan libraries [83,84]. Amine containing glycans with spacer's sp2, sp3 or sp4 were synthesised as previously described [49] and glycans without spacers were amine functionalised as previously published [85]. All glycans were suspended in 1:1 DMF:DMSO at a concentration of 500 mM and were printed onto SuperEpoxy 2 glass slides (ArrayIt, Sunnyvale, CA) using a ArrayIt SpotBot Extreme array spotter in a six pin subarray print per glass slide format. All glycans were printed in replicates of four, including four FITC control spots as well as additional position controls (Fig. S1), per subarray using SMP4 pins and a contact time of 1 second at 60% relative humidity, with pins being

reloaded after every 12 spots.

Prior to performing glycan array experiments, slides were scanned using a ProScanArray Microarray 4-laser scanner (Perkin Elmer, Waltham, MA) using the blue argon 488 laser set to the FITC settings (492 nm excitation and 517 nm emission). Array slides were blocked with 0.1 % BSA in 50 mM phosphate buffered saline (PBS), pH 7.4 for 5 min at 22°C. After washing with PBS, each slide was dried by placing them in an empty 50 mL tube and centrifuging for 5 min at 200 x g (900 rpm). Recombinant TconTS-LD (2 µg) was incubated at a molar ratio of 1:2:3 with anti His-tag mouse polyclonal antibody (10 mg/mL, Cell Signalling Technology), anti mouse-IgG-Alexa555 conjugated rabbit polyclonal antibody (2 mg/mL, Life Technologies) and anti rabbit-IgG-Alexa555 conjugated goat polyclonal antibody (2 mg/mL, Life Technologies) in 50 mM PBS, pH 7.4 containing 0.1 % BSA and 10 mM maltotriose for 15 min on ice protected from light. All subarrays on the slide were isolated using a Gene Frame (1.5 x 1.6 cm, 65 µL, Abgene, Epsom, UK) prior to the addition of the TconTS-LD-antibody mix to the array. A coverslip was applied to the GeneFrame and array slides incubated for 30 min at 22°C in the dark. The GeneFrame and coverslip were subsequently removed and the slide gently washed twice with 50 mM PBS, pH 7.4 containing 0.01 % TWEEN 20 and 10 mM maltose, and once with 50 mM PBS, pH 7.4 containing 10 mM maltose. Slides were dried by centrifugation for 5 min at 200 x g (900 rpm), allowed to air dried for a further 5 min, and the fluorescence associated with the array spots detected using the microarray scanner settings outlined above. Image analysis and spot visualisation was performed using the ProScanArray software, ScanArray Express (Perkin Elmer). The resulting images were visually examined. Fluorescence signals were judged as being positive, if all four replicates for a glycan were clearly detectable (Fig. S1).

STD NMR

TconTS2-LD was several times buffer exchanged to 10 mM deuterated phosphate buffer, pD 7.4 using a Microcon centrifugal ultrafiltration device (cut off 10 kDa). 200 µL of a solution containing 5.5 µM TconTS2-LD was prepared for each experiment. 1024 Scans per STD NMR experiment were acquired as described before [40]. In separate experiments, lactose and 3α,6α-mannotriose were added to the TconTS2-LD solution resulting in 3.45 mM (for lactose) or 1.73 mM (for 3α,6α-

mannotriose or in the mixture of both oligosaccharides) final concentrations. The STD NMR spectra were obtained by subtracting the on- from the off-resonance spectra. As controls, STD NMR spectra for only TconTS2-LD or ligand were recorded under identical conditions used. Data acquisition and evaluation was performed using NMR software TopSpin 3.2 (Bruker Daltonics, Germany).

TS-LD binding/inhibition assay

Microtitre plate based binding and inhibition assays, used for characterising protein binding to sialylated glycoproteins, has been described for siglecs [86]. In this study a modified version of these assays was established to investigate TconTS-LD binding to mannosylated glycoproteins. Recombinant huS2-Fc expressed and purified as previously described [87] was used as binding partner for TconTS-LD, since it contains high-mannose *N*-glycans due to expression in CHO-Lec1 cells. 5 µL of 5 µg/ mL huS2-Fc in 50 mM $NaHCO_3$, pH 9.6, were immobilised on a high binding 384-well microtitre plate (Corning, USA) overnight at 4°C. The plate was washed five times with 20 µL 10 mM Tris-Cl, pH 7.5, 150 mM NaCl containing 0.05% Tween20 (TBS-T) per well. A 1:2 serial dilution of TconTS2-LD ranging from 4 - 0.125 µg/ mL in 10 mM TBS-T was prepared. 0.2 µg/ mL anti His-tag mouse polyclonal antibody and 0.2 µg/ mL anti mouse-IgG alkaline phosphatase(AP)-conjugated donkey polyclonal antibody was added to each dilution step and incubated on ice for 30 min. 5 µL of each sample were transferred in triplicates onto the washed microtitre plate and centrifuged for 1 min at 600 x g. The plate was covered with parafilm and incubated at 4°C for additional 3.5 hours. After washing the plate 4 times with 10 mM TBS-T and twice with 10 mM TBS, 20 µL of freshly prepared fluorescein diphosphate (FDP) substrate solution (50 mM Tris-Cl, pH 8.5, 10 mM $MgCl_2$, 20 µM FDP) was added to each well and the kinetic fluorescence measurement was immediately started employing a Tecan Infinite F200 Pro microtitre plate reader (Tecan, Germany). As controls, wells containing and lacking immobilised huS2-Fc were incubated with both antibodies but in the absence of TconTS2-LD. For comparison, 4 µg/ mL recombinant catalytic domain (CD) of TconTS2 was used instead of TconTS2-LD under the same conditions used.

Inhibiton assays were performed following the same procedure as for the binding assay, but free

oligomannose *N*-glycans were added as potential competitive inhibitors during the incubation with TconTS2-LD. Oligomannose *N*-glycans were released by endoglycosidase H ($EndoH_f$) treatment of 10 µg huS2-Fc in 50 mM sodium citrate, pH 5.5 at 37°C for 4 hours. Proteins were acetone precipitated at -20°C overnight [30]. Following centrifugation the supernatant was transferred into a fresh reaction tube and solvent was removed using SpedVac evaporator for 1 hour at 30°C, 100 mbar. Glycans were resuspended in 10 mM TBS and used in inhibition assay as 1:2 dilution series. Data acquisition was done using the software Magellan 7.2. Binding and inhibition curves were generated using SigmaPlot 11.

Deglycosylation of TconTS

TconTS1 and TconTS2 were enzymatically deglycosylated using recombinant $EndoH_f$ cleaving the chitobiose core (GlcNAc(1-4)-β-GlcNAc) of high-mannose *N*-glycans from glycoconjugates [88]. In brief, 500 µL 10 mM phosphate buffer, pH 7.4 containing 100 µg TconTS and 4000 units $EndoH_f$ were incubated for 4 hours at 37°C. For deglycosylation of TconTS under denaturing conditions, 100 µg TconTS were incubated in 20 µL denaturing buffer containing 40 mM dithiothreitol (DTT) and 0.5% sodium dodecyl sulfate (SDS) for 10 min at 95°C. After the addition of sodium citrate, 50 mM final concentration, pH 5.5 and 4000 units $EndoH_f$, reaction mix was incubated for additional 60 min at 37°C. *N*-deglycosylation efficiency was determined by ConA lectin blot analysis.

Gel permeation chromatography

Oligomerisation of TconTS was analysed employing a fast protein liquid chromatography (FPLC) system (Amersham Pharmacia, USA) using Superdex 200 10/300 GL (GE Healthcare, Sweden) size exclusion column and photometric detection at 280 nm. Chromatographic analysis were done at 4°C. In brief, column was equilibrated with 10 mM phosphate buffer pH 7.4 and calibrated using a gel filtration marker kit for protein molecular weights between 29,000-700,000 Da (Sigma-Aldrich, Germany) according to manufactures instructions. 100 to 300 µg TconTS in 500 µL sample volume were injected and separated at a flow rate of 0.5 mL/min. Absorbance at 280 nm was continuously recorded through an analog writer and subsequently transformed to digital chromatograms using the software SigmaPlot 11. $EndoH_f$ treated samples were analysed in the same manner.

Western Blot and Con A lectin blot analysis

Protein samples were separated employing SDS-PAGE as described previously [89] using a MiniProtean III electrophorese Unit (Bio-Rad, Germany) and stained with Coomassie Brilliant Blue. Western blot analysis were performed as previously described [30], using primary anti S*trep*-tag rabbit antibody (1:1000) and secondary anti rabbit-IgG donkey horseradish peroxidase(HRP)-conjugated antibody (1:40000). Membranes were developed using enhanced chemiluminescence system (ECL-Kit, Thermo Scientific, Germany) and X-ray film (GE Healthcare, Sweden).

ConA lectin blots were performed similar to the procedure for Western blotting. Instead of applying primary antibody, 2 µg/mL solution of biotinylated recombinant ConA in 10 mM phosphate buffer, pH 7.4 was added to the membrane and incubated overnight at 4°C. Avidin-biotin HRP conjugated system (VECTASTAIN ABC-Kit, Vector Labs, UK) was used for detection according to manufactures instructions.

Homology Modelling

Homology models of TconTS-LD containing or lacking the α-helix were calculated employing the molecular modelling software Yasara 13.3.26 [90-95] as previously described [30]. In brief, crystal structure of *Trypanosoma cruzi* TS (PDB: 3b69) [14] was used as a template structure for calculating the models. Yasara *homology modelling* module were modified manually from the default settings of the program as follows: Modelling speed: slow, PsiBLASTs: 6, EValue Max: 0.5, Templates total: 1, Templates SameSeq: 1, OligoState: 4, alignments: 15, LoopSamples: 50, TermExtension:10. The molecular and electrostatic potential surfaces were calculated using the ESPPME (Electrostatic Potential by Particle Mesh Ewald) method of Yasara *Structure* with the following parameters: Force field: AMBER96, Algorithm used to calculate molecular surface: numeric, Radius of water probe: 1.4 Å, Grid solution: 3, Maximum ESP: 300 kJ/mol. Red colour indicates a positive potential, blue a negative and grey a neutral.

Acknowledgments

M.W., T. T. G. and S.K. thank Petra Berger and Nazila Isakovic for expert technical assistance as well as Dr. Frank Dietz and Judith Weber for fruitful discussions.

References

1. Kamuanga M. Socio-economic and cultural factors in the research and control of trypanosomiasis. PAAT Technical and Scientific series. 2003 Sep 17;(4):1–67.

2. WHO. Research priorities for Chagas disease, Human African trypanosomiasis and leishmaniasis. Technical Report Series. 2012 Feb 17;975:1–116.

3. Simarro PP, Diarra A, Ruiz Postigo JA, Franco JR, Jannin JG. The Human African trypanosomiasis control and surveillance programs of the World Health Organization 2000–2009: The Way Forward. Aksoy S, editor. PLoS Negl Trop Dis. 2011 Feb 22;5(2):e1007–7.

4. Schauer R, Kamerling JP. The chemistry and biology of trypanosomal trans-sialidases: virulence factors in Chagas disease and sleeping sickness. Chembiochem. WILEY-VCH Verlag; 2011 Oct 17;12(15):2246–64.

5. Engstler M, Schauer R, Brun R. Distribution of developmentally regulated trans-sialidases in the Kinetoplastida and characterization of a shed trans-sialidase activity from procyclic *Trypanosoma congolense*. Acta Tropica. 1995 May;59(2):117–29.

6. Tiralongo E, Schrader S, Lange H, Lemke H, Tiralongo J, Schauer R. Two trans-sialidase forms with different sialic acid transfer and sialidase activities from *Trypanosoma congolense*. J Biol Chem. 2003 Jun 27;278(26):23301–10.

7. Schenkman S, Eichinger D. *Trypanosoma cruzi* trans-sialidase and cell invasion. Parasitol Today (Regul Ed). 1993 Jun;9(6):218–22.

8. Engstler M, Schauer R. Sialidases from African trypanosomes. Parasitol Today (Regul Ed). 1993 Jun;9(6):222–5.

9. Nok AJ, Balogun EO. A bloodstream *Trypanosoma congolense* sialidase could be involved in anemia during experimental trypanosomiasis. J Biochem. 2003 Jun;133(6):725–30.

10. Coustou V, Plazolles N, Guegan F, Baltz T. Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. Cellular Microbiology. 2012 Feb 13;14(3):431–45.

11. Clayton J. Chagas disease 101. Nature. 2010 Jun 24;465(7301):S4–5.

12. Cremona ML, Sánchez DO, Frasch AC, Campetella O. A single tyrosine differentiates active and inactive *Trypanosoma cruzi* trans-sialidases. Gene. 1995 Jul 4;160(1):123–8.

13. Buschiazzo A, Tavares GA, Campetella O, Spinelli S, Cremona ML, París G, et al. Structural basis of sialyltransferase activity in trypanosomal sialidases. EMBO J. 2000 Jan 4;19(1):16–24.

14. Buschiazzo A, Amaya MF, Cremona ML, Frasch AC, Alzari PM. The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. Mol Cell. 2002 Oct;10(4):757–68.

15. París G, Cremona ML, Amaya MF, Buschiazzo A, Giambiagi S, Frasch AC, et al. Probing molecular function of trypanosomal sialidases: single point mutations can change substrate specificity and increase hydrolytic activity. Glycobiology. 2001 Apr;11(4):305–11.

16. Haselhorst T. NMR spectroscopic and molecular modeling investigations of the trans-sialidase from *Trypanosoma cruzi*. Glycobiology. 2004 May 26;14(10):895–907.

17. Demir O, Roitberg AE. Modulation of catalytic function by differential plasticity of the active site: case study of *Trypanosoma cruzi* trans-sialidase and *Trypanosoma rangeli* sialidase. Biochemistry. 2009 Apr 21;48(15):3398–406.

18. Oliveira IA, Goncalves AS, Neves JL, Itzstein von M, Todeschini AR. Evidence of ternary complex formation in *Trypanosoma cruzi* trans-Sialidase catalysis. Journal of Biological Chemistry. 2014 Jan

3;289(1):423–36.

19. Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. Nature. 1983 May;303(5912):35–40.

20. Taylor G. Sialidases: structures, biological significance and therapeutic potential. Curr Opin Struct Biol. 1996 Dec;6(6):830–7.

21. Taylor NR, Itzstein von M. A structural and energetics analysis of the binding of a series of *N*-acetylneuraminic-acid-based inhibitors to influenza virus sialidase. J Comput Aided Mol Des. 1996 Jun;10(3):233–46.

22. Delbaere LT, Vandonselaar M, Prasad L, Quail JW, Wilson KS, Dauter Z. Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution. Journal of Molecular Biology. 1993 Apr 5;230(3):950–65.

23. Hester G, Kaku H, Goldstein IJ, Wright CS. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. Nat Struct Biol. 1995 Jun;2(6):472–9.

24. Bourne Y, Roussel A, Frey M, Rougé P, Fontecilla-Camps JC, Cambillau C. Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: fine specificity of the monosaccharide-binding site. Proteins. 1990;8(4):365–76.

25. Wright CS. 2.2 Å resolution structure analysis of two refined *N*-acetylneuraminyl-lactose-*wheat germ agglutinin* isolectin complexes. Journal of Molecular Biology. 1990 Oct 20;215(4):635–51.

26. Engstler M, Reuter G, Schauer R. The developmentally regulated trans-sialidase from *Trypanosoma brucei* sialylates the procyclic acidic repetitive protein. Mol Biochem Parasitol. 1993 Sep;61(1):1–13.

27. Pontes de Carvalho LC, Tomlinson S, Vandekerckhove F, Bienen EJ, Clarkson AB, Jiang MS, et al. Characterization of a novel trans-sialidase of *Trypanosoma brucei* procyclic trypomastigotes and identification of procyclin as the main sialic acid acceptor. J Exp Med. 1993 Feb 1;177(2):465–74.

28. Nakatani F, Morita YS, Ashida H, Nagamune K, Maeda Y, Kinoshita T. Identification of a second catalytically active trans-sialidase in *Trypanosoma brucei*. Biochemical and Biophysical Research Communications. Elsevier Inc; 2011 Nov 18;415(2):421–5.

29. Montagna GN, Donelson JE, Frasch ACC. Procyclic *Trypanosoma brucei* expresses separate sialidase and trans-sialidase enzymes on its surface membrane. Journal of Biological Chemistry. 2006 Nov 3;281(45):33949–58.

30. Koliwer-Brandl H, Gbem TT, Waespy M, Reichert O, Mandel P, Drebitz E, et al. Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*. BMC Biochem. 2011;12(1):39.

31. Gbem TT, Waespy M, Hesse B, Dietz F, Smith J, Chechet GD, et al. Biochemical diversity in the *Trypanosoma congolense* trans-sialidase family. PLoS Negl Trop Dis. 2013;7(12):e2549.

32. Schenkman S, Jiang MS, Hart GW, Nussenzweig V. A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. Cell. 1991 Jun 28;65(7):1117–25.

33. Cremona ML, Campetella O, Sánchez DO, Frasch AC. Enzymically inactive members of the trans-sialidase family from *Trypanosoma cruzi* display β-galactose binding activity. Glycobiology. 1999 Jun;9(6):581–7.

34. Tiralongo E, Martensen I, Grötzinger J, Tiralongo J, Schauer R. Trans-sialidase-like sequences from *Trypanosoma congolense* conserve most of the critical active site residues found in other trans-sialidases. Biol Chem. 2003 Aug;384(8):1203–13.

35. Amaya MF, Buschiazzo A, Nguyen T, Alzari PM. The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T.cruzi* trans-sialidase. Journal of

Molecular Biology. 2003 Jan;325(4):773–84.

36. Ammar Z, Plazolles N, Baltz T, Coustou V. Identification of trans-sialidases as a common mediator of endothelial cell activation by African trypanosomes. Beverley SM, editor. PLoS Pathog. 2013;9(10):e1003710.

37. Schwartz M, Kellermann O, Szmelcman S, Hazelbauer GL. Further studies on the binding of maltose to the maltose-binding protein of *Escherichia coli*. Eur J Biochem. 1976 Dec;71(1):167–70.

38. Crocker PR, Vinson M, Kelm S, Drickamer K. Molecular analysis of sialoside binding to sialoadhesin by NMR and site-directed mutagenesis. Biochem J. Portland Press Ltd; 1999 Jul 15;341 (Pt 2)(Pt 2):355–61.

39. Meyer B, Peters T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew Chem Int Ed Engl. 2003 Feb 24;42(8):864–90.

40. Kelm S, Madge P, Islam T, Bennett R, Koliwer-Brandl H, Waespy M, et al. C-4 modified sialosides enhance binding to Siglec-2 (CD22): towards potent Siglec inhibitors for immunoglycotherapy. Angew Chem Int Ed Engl. 2013 Mar 25;52(13):3616–20.

41. Robertson MA, Etchison JR, Robertson JS, Summers DF, Stanley P. Specific changes in the oligosaccharide moieties of VSV grown in different lectin-resistant CHO cells. Cell. 1978 Mar;13(3):515–26.

42. Tarentino AL, Maley F. Purification and properties of an endo-β-*N*-acetylglucosaminidase from *Streptomyces griseus*. J Biol Chem. 1974 Feb 10;249(3):811–7.

43. Batas B, Jones HR, Chaudhuri JB. Studies of the hydrodynamic volume changes that occur during refolding of lysozyme using size-exclusion chromatography. Journal of Chromatography A. 1997 Apr 4;766(1-2):109–19.

44. Crennell SJ, Garman EF, Laver WG, Vimr ER, Taylor GL. Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. Proc Natl Acad Sci USA. 1993 Nov 1;90(21):9852–6.

45. Crennell S, Garman E, Laver G, Vimr E, Taylor G. Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. Structure. 1994 Jun 15;2(6):535–44.

46. Bourne Y, Rougé P, Cambillau C. X-ray structure of a (α-Man(1-3)β-Man(1-4)GlcNAc)-lectin complex at 2.1-Å resolution. The role of water in sugar-lectin interaction. J Biol Chem. 1990 Oct 25;265(30):18161–5.

47. Weis WI, Drickamer K. Structural basis of lectin-carbohydrate recognition. Annu Rev Biochem. 1996;65:441–73.

48. Luo Y, Li SC, Li YT, Luo M. The 1.8 Å structures of leech intramolecular trans-sialidase complexes: evidence of its enzymatic mechanism. Journal of Molecular Biology. 1999 Jan 8;285(1):323–32.

49. Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, Alvarez R, et al. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc Natl Acad Sci USA. 2004 Dec 7;101(49):17033–8.

50. Grant OC, Smith HM, Firsova D, Fadda E, Woods RJ. Presentation, presentation, presentation! Molecular-level insight into linker effects on glycan array screening data. Glycobiology. 2013 Dec 6;24(1):17–25.

51. Smith LE, Eichinger D. Directed mutagenesis of the *Trypanosoma cruzi* trans-sialidase enzyme identifies two domains involved in its sialyltransferase activity. Glycobiology. 1997 Apr;7(3):445–51.

52. Moustafa I. Sialic acid recognition by *Vibrio cholerae* neuraminidase. Journal of Biological Chemistry. 2004 May 28;279(39):40819–26.

53. Kalb AJ, Levitzki A. Metal-binding sites of concanavalin A and their role in the binding of α-methyl-D-glucopyranoside. Biochem J. 1968 Oct;109(4):669–72.

54. Derewenda Z, Yariv J, Helliwell JR, Kalb AJ, Dodson EJ, Papiz MZ, et al. The structure of the saccharide-binding site of concanavalin A. EMBO J. 1989 Aug;8(8):2189–93.

55. Bubb WA. NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. Concepts Magn Reson. Wiley Subscription Services, Inc., A Wiley Company; 2003 Aug 26;19A(1):1–19.

56. Shahzad-ul-Hussan S, Ghirlando R, Dogo-Isonagie CI, Igarashi Y, Balzarini J, Bewley CA. Characterization and carbohydrate specificity of Pradimicin S. J Am Chem Soc. 2012 Aug;134(30):12346–9.

57. Merino G, Shuman HA. Unliganded maltose-binding protein triggers lactose transport in an *Escherichia coli* mutant with an alteration in the maltose transport system. J Bacteriol. 1997 Dec;179(24):7687–94.

58. Oldham ML, Chen S, Chen J. Structural basis for substrate specificity in the *Escherichia coli* maltose transport system. Proc Natl Acad Sci USA. National Acad Sciences; 2013 Nov 5;110(45):18132–7.

59. Albohy A, Richards MR, Cairo CW. Mapping substrate interactions of the human membrane-associated neuraminidase, NEU3, using STD NMR. Glycobiology. Oxford University Press; 2015 Mar;25(3):284–93.

60. Wright CS. Structural comparison of the two distinct sugar binding sites in *wheat germ agglutinin* isolectin II. Journal of Molecular Biology. 1984 Sep 5;178(1):91–104.

61. Rini JM. Lectin structure. Annu Rev Biophys Biomol Struct. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1995;24(1):551–77.

62. Weis WI, Drickamer K, Hendrickson WA. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. Nature. Nature Publishing Group; 1992 Nov 12;360(6400):127–34.

63. Weis WI, Drickamer K. Trimeric structure of a C-type mannose-binding protein. Structure. 1994 Dec 15;2(12):1227–40.

64. Dani M, Manca F, Rialdi G. Calorimetric study of concanavalin A binding to saccharides. Biochim Biophys Acta. 1981 Jan 30;667(1):108–17.

65. Shibata S, Goldstein IJ, Baker DA. Isolation and characterization of a Lewis b-active lectin from *Griffonia simplicifolia* seeds. J Biol Chem. 1982 Aug 25;257(16):9324–9.

66. Beecroft RP, Roditi I, Pearson TW. Identification and characterization of an acidic major surface glycoprotein from procyclic stage *Trypanosoma congolense*. Mol Biochem Parasitol. 1993 Oct;61(2):285–94.

67. Bayne RA, Kilbride EA, Lainson FA, Tetley L, Barry JD. A major surface antigen of procyclic stage *Trypanosoma congolense*. Mol Biochem Parasitol. 1993 Oct;61(2):295–310.

68. Bütikofer P, Vassella E, Boschung M, Renggli CK, Brun R, Pearson TW, et al. Glycosylphosphatidylinositol-anchored surface molecules of *Trypanosoma congolense* insect forms are developmentally regulated in the tsetse fly. Mol Biochem Parasitol. 2002 Jan;119(1):7–16.

69. Utz S, Roditi I, Kunz Renggli C, Almeida IC, Acosta-Serrano A, Bütikofer P. *Trypanosoma congolense* procyclins: unmasking cryptic major surface glycoproteins in procyclic forms. Eukaryotic Cell. 2006 Aug;5(8):1430–40.

70. Carrington M, Miller N, Blum M, Roditi I, Wiley D, Turner M. Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. Journal of Molecular Biology. 1991 Oct 5;221(3):823–35.

71. Blum ML, Down JA, Gurnett AM, Carrington M, Turner MJ, Wiley DC. A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*. Nature. Nature Publishing Group; 1993 Apr 15;362(6421):603–9.

72. Savage A, Geyer R, Stirm S, Reinwald E, Risse HJ. Structural studies on the major oligosaccharides in a variant surface glycoprotein of *Trypanosoma congolense*. Mol Biochem Parasitol. 1984 Apr;11:309–28.

73. Zamze SE, Wooten EW, Ashford DA, Ferguson MA, Dwek RA, Rademacher TW. Characterisation of the asparagine-linked oligosaccharides from *Trypanosoma brucei* type-I variant surface glycoproteins. Eur J Biochem. 1990 Feb 14;187(3):657–63.

74. Zamze SE, Ashford DA, Wooten EW, Rademacher TW, Dwek RA. Structural characterization of the asparagine-linked oligosaccharides from *Trypanosoma brucei* type II and type III variant surface glycoproteins. J Biol Chem. 1991 Oct 25;266(30):20244–61.

75. Mowatt MR, Clayton CE. Developmental regulation of a novel repetitive protein of *Trypanosoma brucei*. Mol Cell Biol. 1987 Aug;7(8):2838–44.

76. Clayton CE, Mowatt MR. The procyclic acidic repetitive proteins of *Trypanosoma brucei*. Purification and post-translational modification. J Biol Chem. 1989 Sep 5;264(25):15088–93.

77. Thomson LM, Lamont DJ, Mehlert A, Barry JD, Ferguson MAJ. Partial structure of glutamic acid and alanine-rich protein, a major surface glycoprotein of the insect stages of *Trypanosoma congolense*. J Biol Chem. 2002 Dec 13;277(50):48899–904.

78. Treumann A, Zitzmann N, Hülsmeier A, Prescott AR, Almond A, Sheehan J, et al. Structural characterisation of two forms of procyclic acidic repetitive protein expressed by procyclic forms of *Trypanosoma brucei*. Journal of Molecular Biology. 1997 Jun 20;269(4):529–47.

79. Ferguson MA, Homans SW, Dwek RA, Rademacher TW. Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. Ullstein Mosby. 1988 Feb 12;239(4841 Pt 1):753–9.

80. McConville MJ, Ferguson MA. The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. Biochem J. 1993 Sep 1;294 (Pt 2):305–24.

81. Mehlert A, Richardson JM, Ferguson MA. Structure of the glycosylphosphatidylinositol membrane anchor glycan of a class-2 variant surface glycoprotein from *Trypanosoma brucei*. Journal of Molecular Biology. 1998 Mar 27;277(2):379–92.

82. Miladi B, Marjou El A, Boeuf G, Bouallagui H, Dufour F, Di Martino P, et al. Oriented immobilization of the *tobacco etch virus* protease for the cleavage of fusion proteins. Journal of Biotechnology. 2012 Apr 15;158(3):97–103.

83. Huflejt ME, Vuskovic M, Vasiliu D, Xu H, Obukhova P, Shilova N, et al. Anti-carbohydrate antibodies of normal sera: findings, surprises and challenges. Mol Immunol. 2009 Sep;46(15):3037–49.

84. Arndt NX, Tiralongo J, Madge PD, Itzstein von M, Day CJ. Differential carbohydrate binding and cell surface glycosylation of human cancer cell lines. J Cell Biochem. 2011 Aug 18;112(9):2230–40.

85. Day CJ, Tiralongo J, Hartnell RD, Logue C-A, Wilson JC, Itzstein von M, et al. Differential carbohydrate recognition by *Campylobacter jejuni* strain 11168: influences of temperature and growth conditions. Heimesaat MM, editor. PLoS ONE. 2009;4(3):e4927.

86. Bock N, Kelm S. Binding and inhibition assays for Siglecs. Methods Mol Biol. 2006;347:359–75.

87. Koliwer-Brandl H, Siegert N, Umnus K, Kelm A, Tolkach A, Kulozik U, et al. Lectin inhibition assays for the analysis of bioactive milk sialoglycoconjugates. International Dairy Journal. Elsevier Ltd; 2011 Jun 1;21(6):413–20.

88. Maley F, Trimble RB, Tarentino AL, Plummer TH. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. Anal Biochem. 1989 Aug 1;180(2):195–204.

89. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970 Aug 15;227(5259):680–5.

90. King RD, Sternberg MJ. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. Protein Sci. Cold Spring Harbor Laboratory Press; 1996 Nov;5(11):2298–310.

91. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology. 1999 Sep 17;292(2):195–202.

92. Mückstein U, Hofacker IL, Stadler PF. Stochastic pairwise alignments. Bioinformatics. 2002;18 Suppl 2:S153–60.

93. Canutescu AA, Shelenkov AA, Dunbrack RL. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci. 2003 Sep;12(9):2001–14.

94. Qiu J, Elber R. SSALN: an alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs. Proteins. Wiley Subscription Services, Inc., A Wiley Company; 2006 Mar 1;62(4):881–91.

95. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins. 2009;77 Suppl 9:114–22.

Supporting Information

Fig. S1. Identification of potential TconTS-LD oligosaccharide ligands. Purified TconTS-LDs (2 µg) were pre-complexed with mouse anti-His, rabbit anti-mouse TexasRed, and donkey anti-rabbit TexasRed and applied to glycan array slides in the presence and absence of 10 mM maltotriose. Following washing and drying array slides were scanned using the ProScanArray Microarray 4-laser scanner, with the images generated analysed using the ScanArray Express software. A: 2 µg His-MBP. B: 2 µg His-MBP in the presence of 10 mM maltotriose during incubation and subsequent washing. C: TconTS1-αHel-LD in the absence of maltotriose. D: TconTS1-LD (10 mM maltoriose). E: TconTS2-αHel-LD (10 mM maltoriose). F: TconTS2-LD (10 mM maltoriose). G: TconTS3-αHel-LD (10 mM maltoriose). H: TconTS3-LD (10 mM maltoriose). I: TconTS4-αHel-LD (10 mM maltoriose). J: TconTS4-LD (10 mM maltoriose). K: Array slide before incubation with MBP in the absence of maltotriose. L: Array slide before incubation with TconTS2-αHel-LD in the absence of maltotriose. Inhibited MBP binding to oligomaltose oligosaccharides are cycled in red. Spots cycled in yellow represent internal print controls that are present on all glycan array slides.

Fig. S2. Results from glycan array analysis. MBP and TconTS-LD binding to several glycans was determined in the presence (+) or absence (-) of 10 mM mannotriose. No binding was detected to non-listed glycans (see Table S1 for all glycans on arrays).

Fig. S3. Amino acid sequence alignment and predicted N-glycosylation sites of TconTS. The primary sequences of TconTS1, TconTS2, TconTS3 and TconTS4 were aligned applying *Geneious Alignment* module of the software Geneious 5.5.5 using the following settings: gap open penalty: 12, Gap extension penalty: 3, Alignment type: Global alignment with free end gaps, Cost Matrix: Blosum62. The predicted *N*-glycosylation sites are illustrated in red triangles and distributed over the catalytic (light blue) and lectin (yellow) domains. No *N*-glycosylation sites have been found in SNAP (orange) and Strep (purple) tags attached to TconTS enzymes.

Fig. S4. Binding specificity of TconTS2-LD and enzymatic activity of bacterial and eukaryotic expressed TconTS1. A, B) Binding activities of TconTS2-LD to high-mannose *N*-glycans of immobilised Siglec-2. TconTS2-LD and the antibodies used for detection were consecutively added (triangles) to the immobilised ligands as well as in form of a precomplexed mixture (squares). C) Competitive inhibition of TconTS2-LD specific binding to high-mannose *N*-glycans of Siglec-2. High-mannose *N*-glycans to be used as competitive inhibitor were obtained by EndoH treatment of Siglec-2 (see Methods section for details). αMan: α-methyl-mannopyranoside, Glc: D-glucose. D) Comparison of the specific enzymatic activities of recombinant TconTS1a expressed in *E.coli* (bacterial) and CHO-Lec1 (eukaryotic) cells. Velocities shown represent the production of 3‘SL (3‘-sialyllactose) or free Neu5Ac (*N*-acetyl-neuraminic acid) under standard conditions: 100 µg fetuin, 2 mM lactose, 37°C, 30 min.

Table S1. Library of all glycan structures present on our glycan array

Table S2. Sequence similarities of TconTS-LDs

Table S3. Primers for amplification of cDNA fragments encoding TconTS-LDs

Figure 1

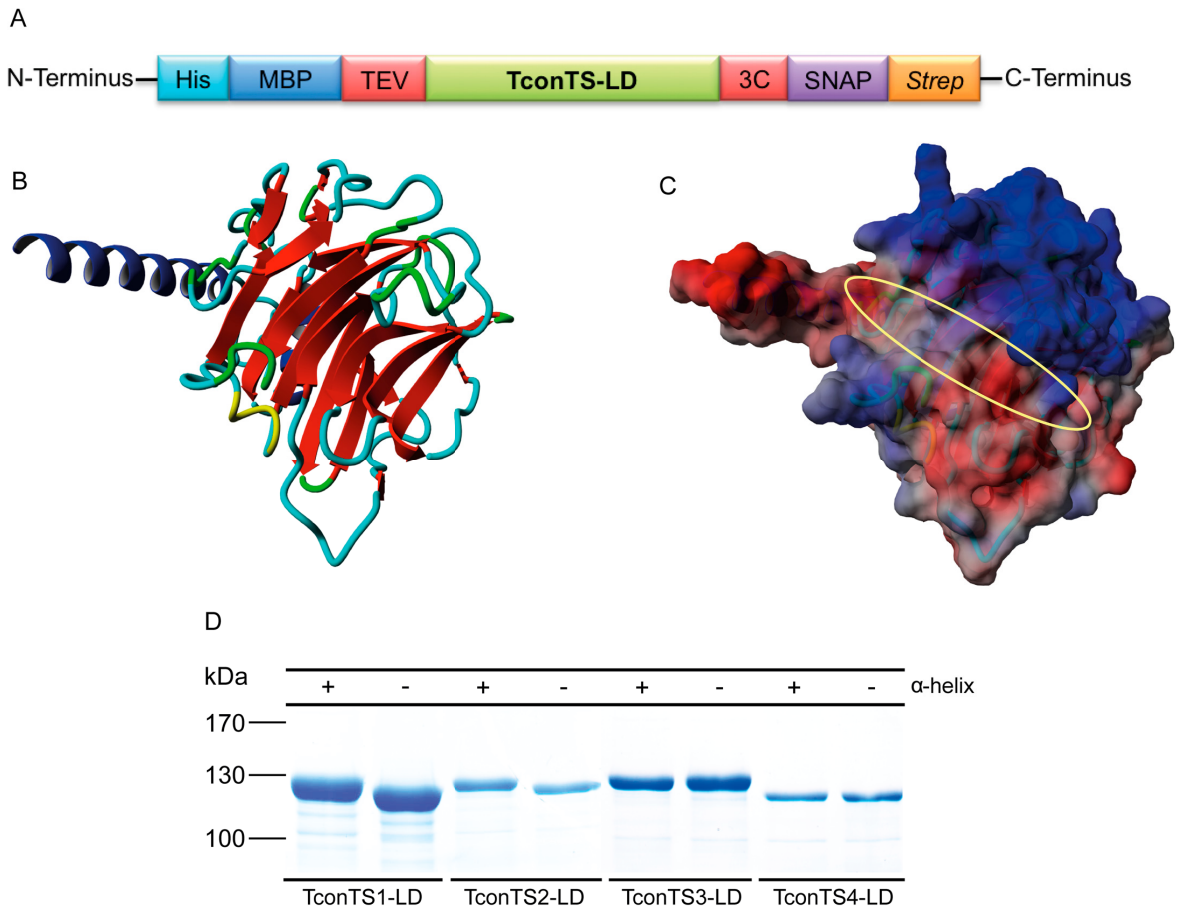
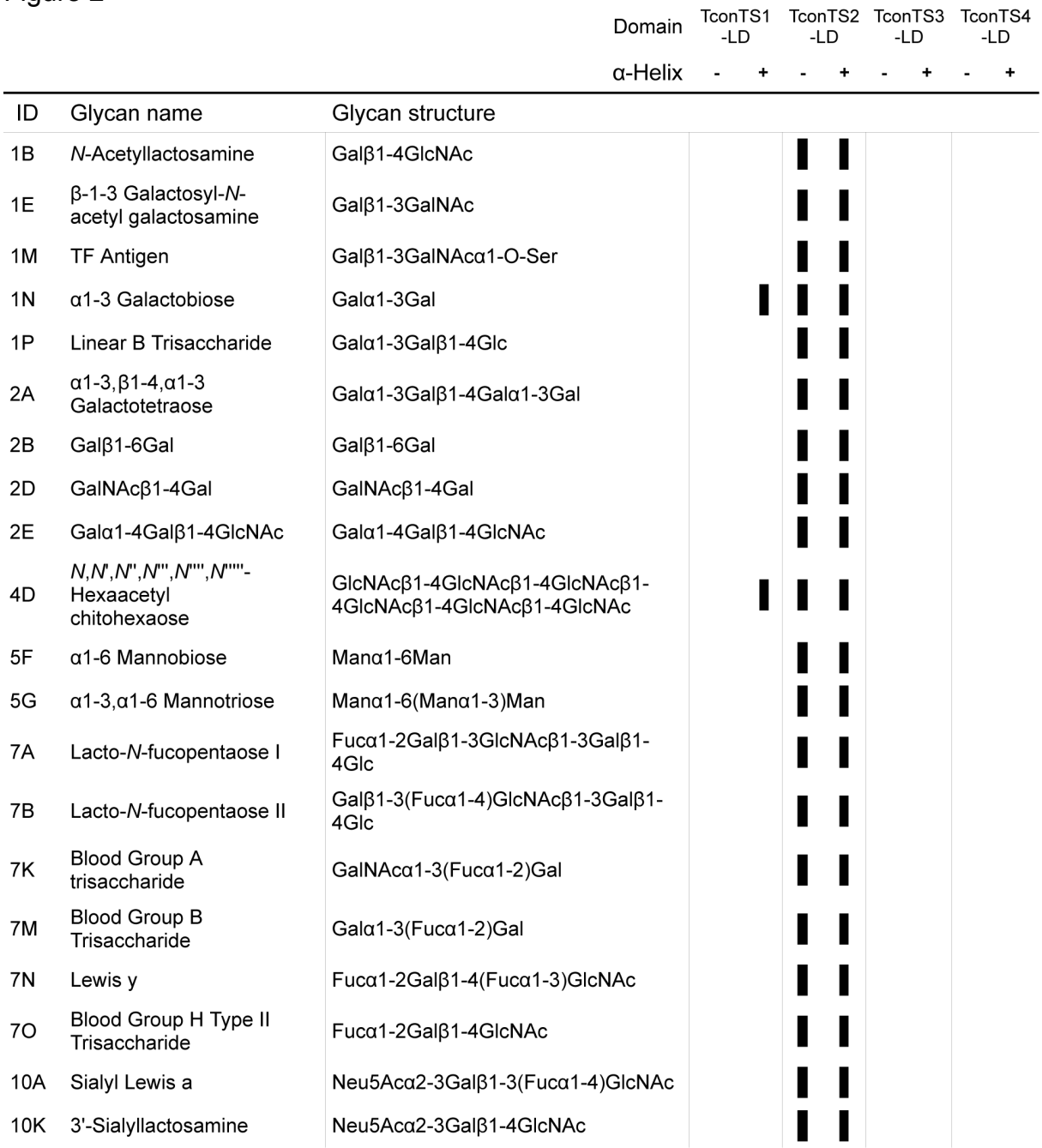
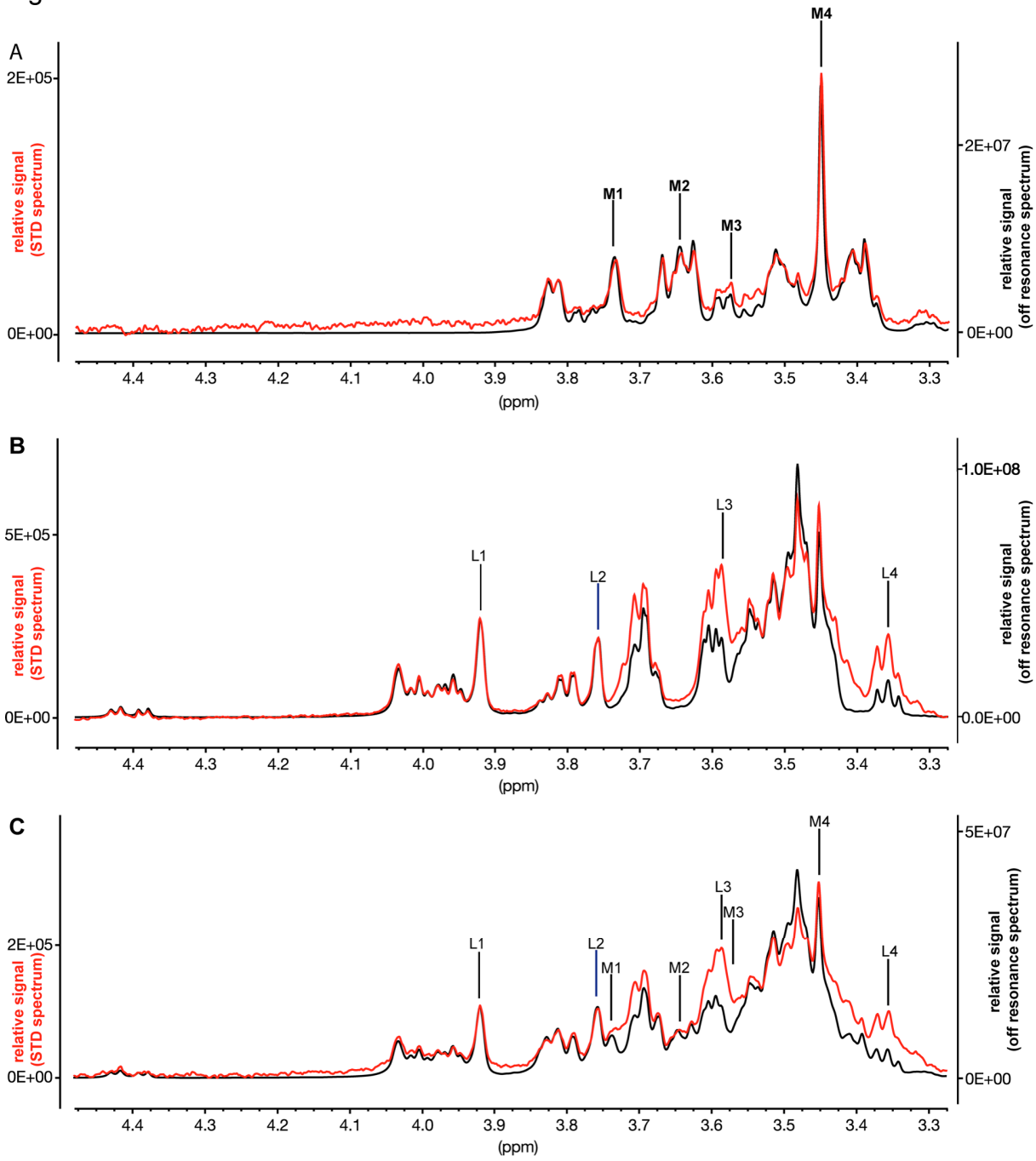

A
N-Terminus
His
MBP
TEV
TconTS-LD
3C
SNAP
Strep
C-Terminus
B
C
D
kDa
+
-
+
-
+
-
+
-
α-helix
170
130
100
TconTS1-LD
TconTS2-LD
TconTS3-LD
TconTS4-LD

Figure 2



			Domain	TconTS1 -LD		TconTS2 -LD		TconTS3 -LD		TconTS4 -LD	
			α-Helix	-	+	-	+	-	+	-	+
ID	Glycan name	Glycan structure									
1B	*N*-Acetyllactosamine	Galβ1-4GlcNAc									
1E	β-1-3 Galactosyl-*N*-acetyl galactosamine	Galβ1-3GalNAc									
1M	TF Antigen	Galβ1-3GalNAcα1-O-Ser									
1N	α1-3 Galactobiose	Galα1-3Gal									
1P	Linear B Trisaccharide	Galα1-3Galβ1-4Glc									
2A	α1-3,β1-4,α1-3 Galactotetraose	Galα1-3Galβ1-4Galα1-3Gal									
2B	Galβ1-6Gal	Galβ1-6Gal									
2D	GalNAcβ1-4Gal	GalNAcβ1-4Gal									
2E	Galα1-4Galβ1-4GlcNAc	Galα1-4Galβ1-4GlcNAc									
4D	*N*,*N*',*N*'',*N*''',*N*'''',*N*'''''-Hexaacetyl chitohexaose	GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAc									
5F	α1-6 Mannobiose	Manα1-6Man									
5G	α1-3,α1-6 Mannotriose	Manα1-6(Manα1-3)Man									
7A	Lacto-*N*-fucopentaose I	Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glc									
7B	Lacto-*N*-fucopentaose II	Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4Glc									
7K	Blood Group A trisaccharide	GalNAcα1-3(Fucα1-2)Gal									
7M	Blood Group B Trisaccharide	Galα1-3(Fucα1-2)Gal									
7N	Lewis y	Fucα1-2Galβ1-4(Fucα1-3)GlcNAc									
7O	Blood Group H Type II Trisaccharide	Fucα1-2Galβ1-4GlcNAc									
10A	Sialyl Lewis a	Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAc									
10K	3'-Sialyllactosamine	Neu5Acα2-3Galβ1-4GlcNAc									

Figure 3



D

	α1-3,α1-6-mannotriose				lactose			
peak	M1	M2	M3	M4	L1	L2	L3	L4
ppm	3.735	3.646	3.574	3.451	3.920	3.759	3.587	3.357
single ligand	0.0063	0.0061	0.0068	0.0074	0.0069	0.0066	0.0124	0.0152
mixture	0.0066	0.0059	n.d.	0.0062	0.0070	0.0066	0.0160	0.0163

Figure 4

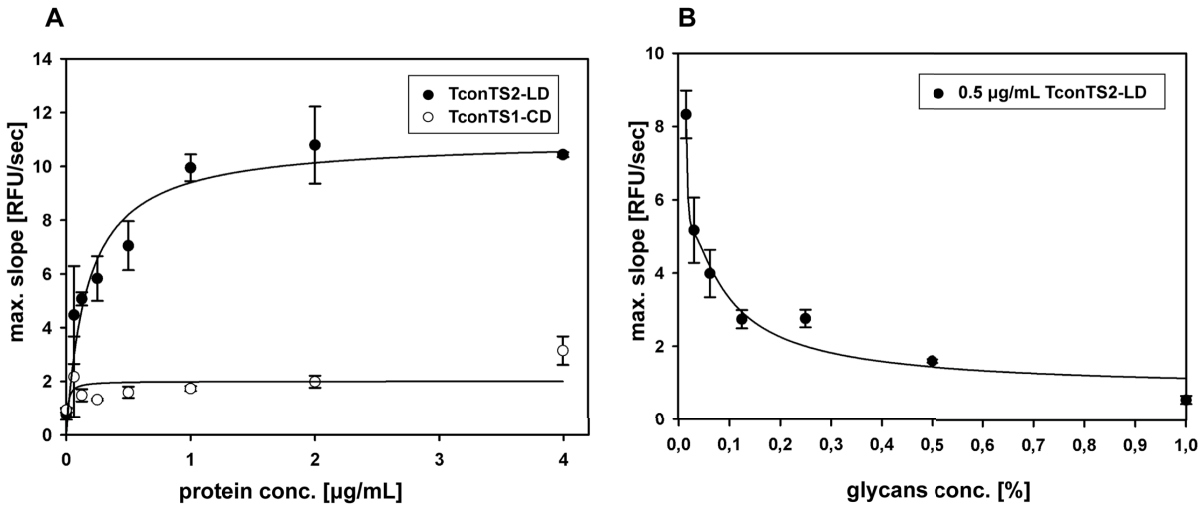

A
max. slope [RFU/sec]
protein conc. [µg/mL]
TconTS2-LD
TconTS1-CD
B
max. slope [RFU/sec]
glycans conc. [%]
0.5 µg/mL TconTS2-LD

Figure 5

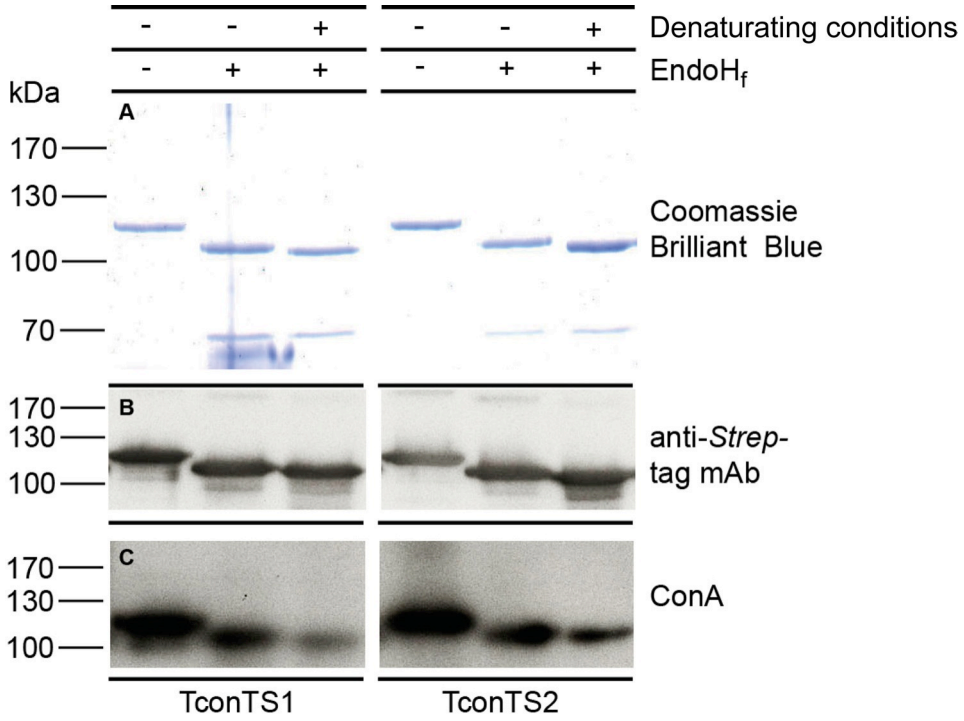

- - + - - +
Denaturating conditions
- + + - + +
EndoHf
kDa
A
170
130
100
70
Coomassie
Brilliant Blue
B
170
130
100
anti-Strep-
tag mAb
C
170
130
100
ConA
TconTS1
TconTS2

Figure 6

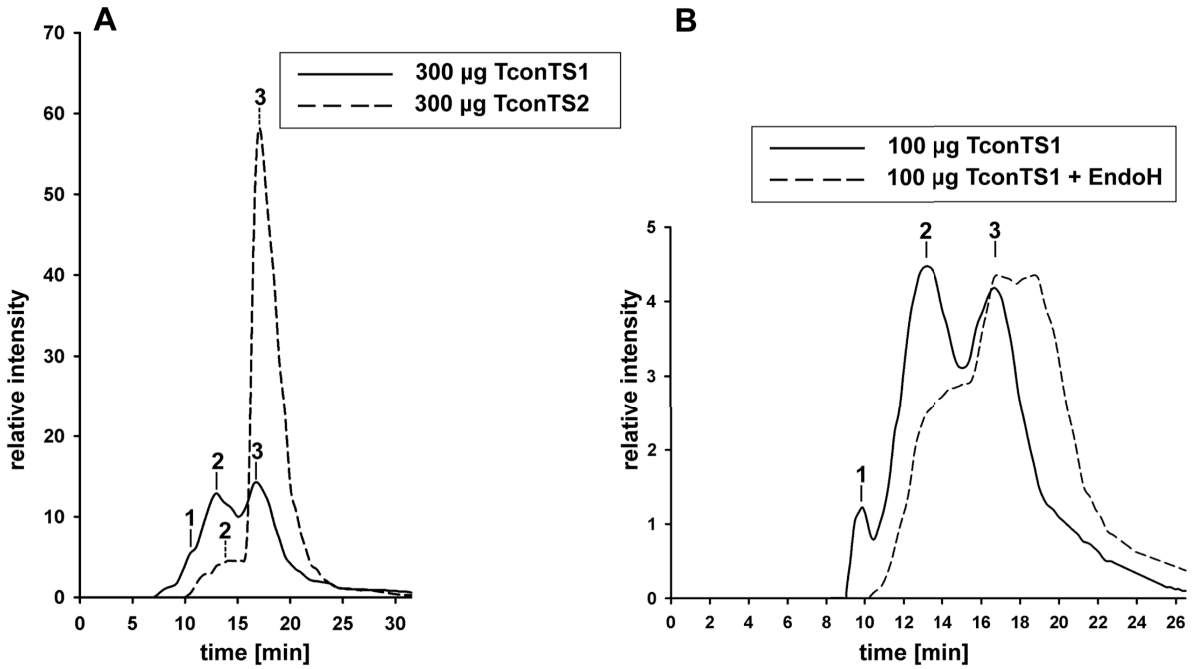

A
300 µg TconTS1
300 µg TconTS2
relative intensity
time [min]
1
2
3
2
3
B
100 µg TconTS1
100 µg TconTS1 + EndoH
relative intensity
time [min]
1
2
3

Figure 7

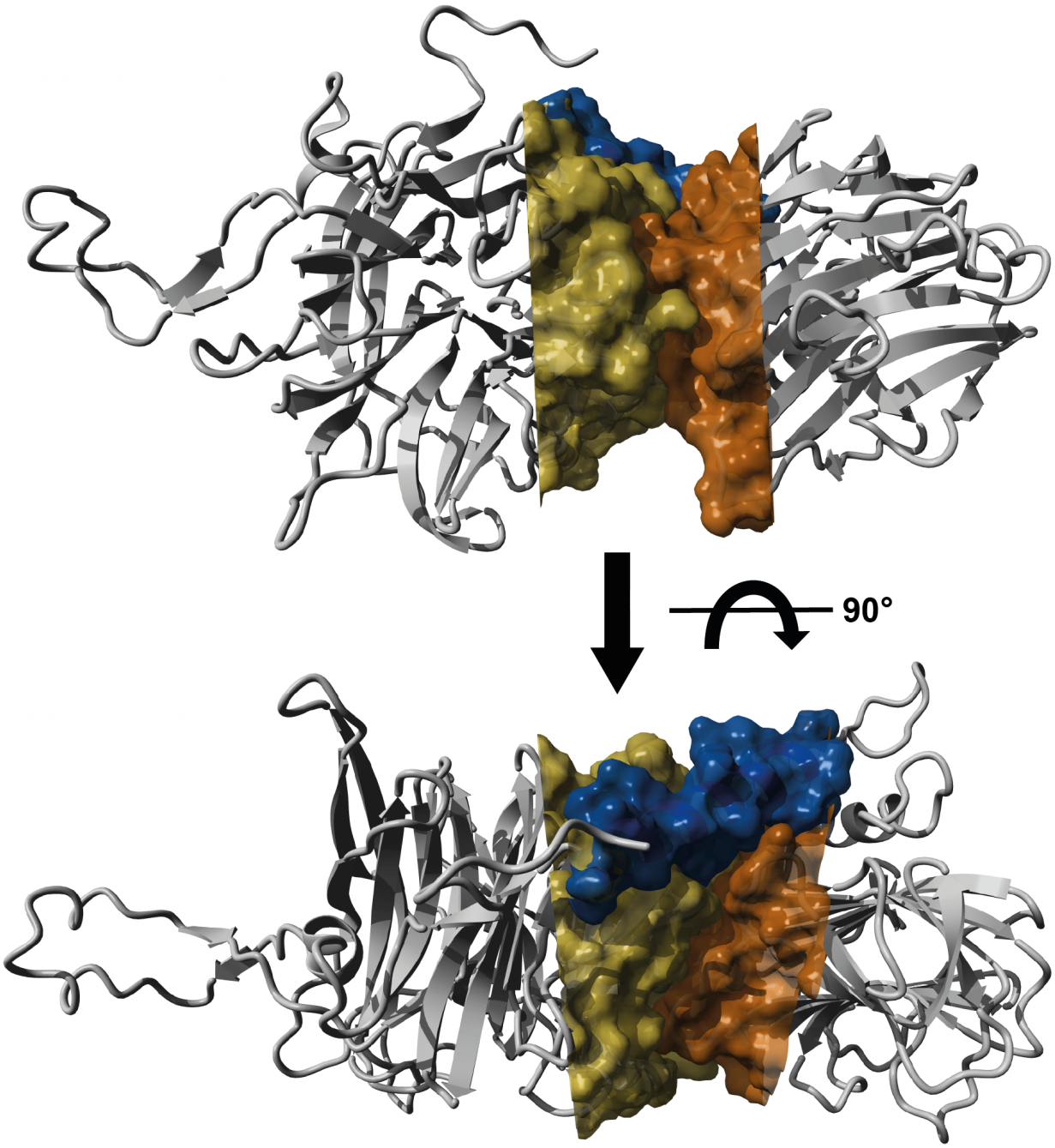
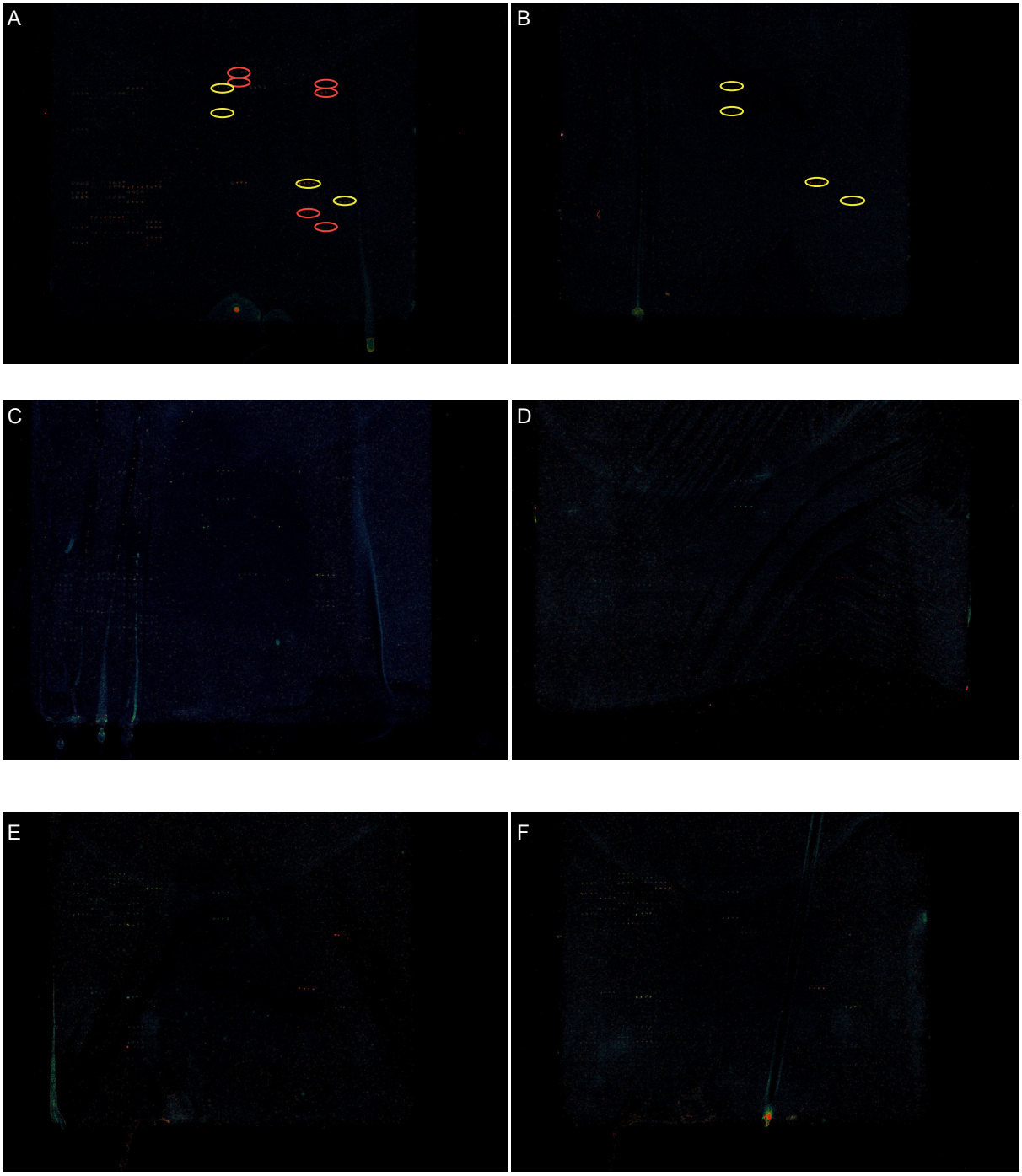


Fig. S1



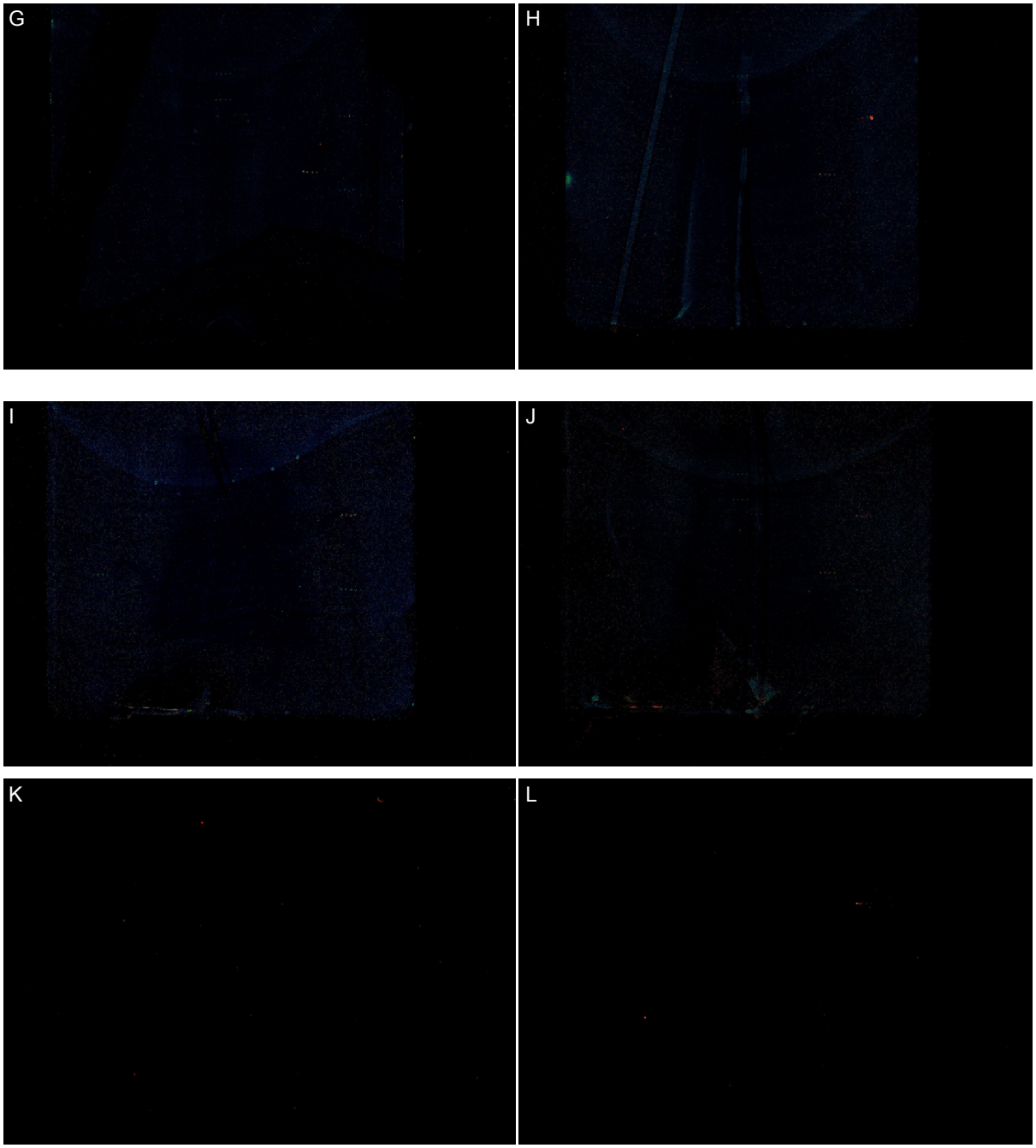
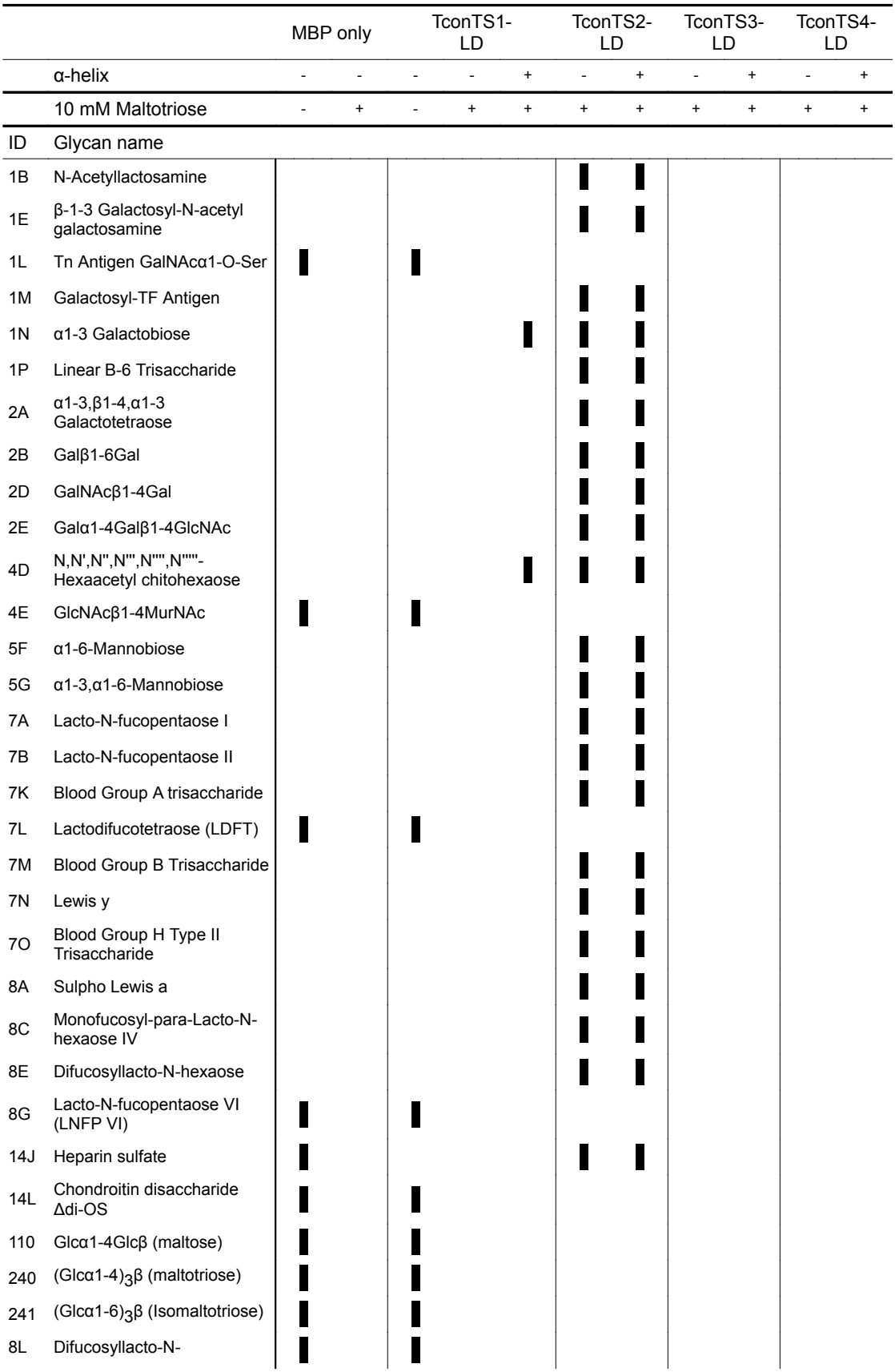

G
H
I
J
K
L

Fig. S2

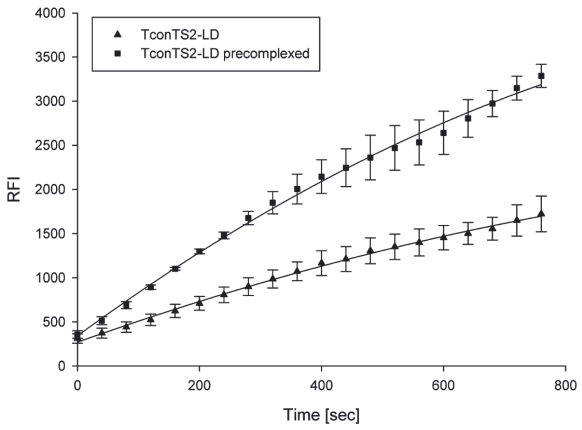

MBP only
TconTS1-LD
TconTS2-LD
TconTS3-LD
TconTS4-LD
α-helix - - - - + - + - + - +
10 mM Maltotriose - + - + + + + + + + +
ID Glycan name
1B N-Acetyllactosamine
1E β-1-3 Galactosyl-N-acetyl galactosamine
1L Tn Antigen GalNAcα1-O-Ser
1M Galactosyl-TF Antigen
1N α1-3 Galactobiose
1P Linear B-6 Trisaccharide
2A α1-3,β1-4,α1-3 Galactotetraose
2B Galβ1-6Gal
2D GalNAcβ1-4Gal
2E Galα1-4Galβ1-4GlcNAc
4D N,N',N'',N''',N'''',N'''''-Hexaacetyl chitohexaose
4E GlcNAcβ1-4MurNAc
5F α1-6-Mannobiose
5G α1-3,α1-6-Mannobiose
7A Lacto-N-fucopentaose I
7B Lacto-N-fucopentaose II
7K Blood Group A trisaccharide
7L Lactodifucotetraose (LDFT)
7M Blood Group B Trisaccharide
7N Lewis y
7O Blood Group H Type II Trisaccharide
8A Sulpho Lewis a
8C Monofucosyl-para-Lacto-N-hexaose IV
8E Difucosyllacto-N-hexaose
8G Lacto-N-fucopentaose VI (LNFP VI)
14J Heparin sulfate
14L Chondroitin disaccharide Δdi-OS
110 Glcα1-4Glcβ (maltose)
240 (Glcα1-4)3β (maltotriose)
241 (Glcα1-6)3β (Isomaltotriose)
8L Difucosyllacto-N-

	neohexaose I										
8M	Difucosyllacto-N-neohexaose II	▮	▮								
8P	Blood Group A Tetrasaccharide	▮	▮								
10A	Sialyl Lewis a	▮	▮			▮	▮				
10C	Sialyllacto-N-tetraose a	▮	▮								
10D	Monosialyl, monofucosyllacto-N neohexose	▮	▮								
10H	Sialyllacto-N-fucopentaose VI (SLNFPVI)	▮	▮								
10K	3'-Sialyllactosamine					▮	▮				
10O	LS-Tetrasaccharide c	▮	▮								
11D	Biantennary 2,6-sialylated-N-glycan-Asn	▮	▮								
12D	Neocarrahexaose-41, 3, 5-tri-O-sulphate (Na+)					▮	▮				
12I	DUA-2S-GlucNS					▮	▮				
12N	DUA-GalNAc-4S	▮	▮								
12O	DUA-GalNAc-6S (Delta Di-6S)					▮	▮				
12P	DUA-GalNAc-4S,6S	▮	▮								
13A	DUA-2S-GalNAc-4S	▮	▮								
13B	DUA-2S-GalNAc-6S (Delta Di-disD)					▮					
13C	DUA-2S-GalNAc-4S-6S	▮	▮		▮	▮	▮				
13F	Hyaluronan fragment (4mer)	▮	▮								
13G	Hyaluronan fragment (8mer)	▮	▮								
13H	Hyaluronan fragment (10mer)	▮	▮								
13I	Hyaluronan fragment (12mer)	▮	▮								
13J	Heparin				▮	▮	▮	▮	▮	▮	▮
13O	HA - 6 10mM	▮	▮								
13P	HA - 8 9.7mM	▮	▮								
14B	HA -12 6.5mM	▮	▮	▮							
14C	HA-14 5.6mM	▮	▮								
14D	HA-16 4.9mM	▮	▮								
390	$(Glc\alpha1\text{-}4)_4\beta$ (Maltotetraose)	▮	▮								
391	$(Glc\alpha1\text{-}6)_4\beta$ (Isomaltotetraose)	▮	▮								

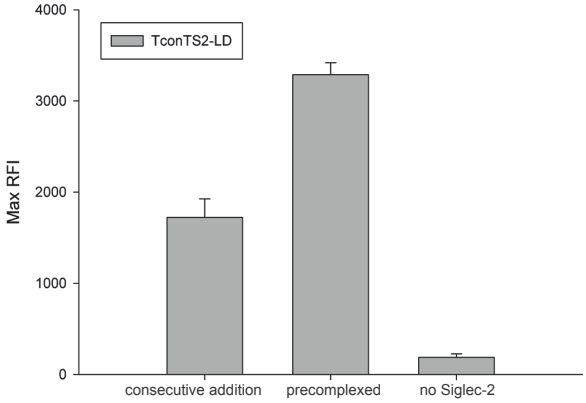
Fig. S3

Fig. S4

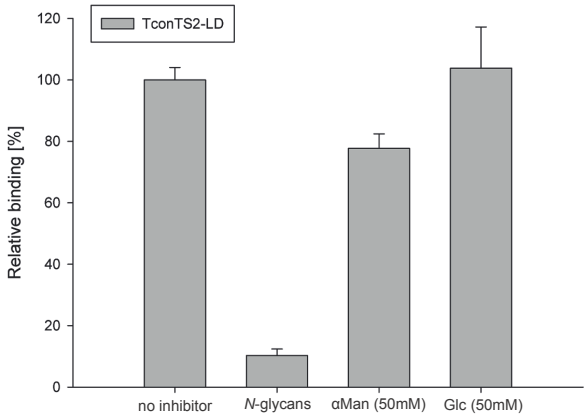
A



B



C



D

TconTS	Specific activity [nmol/(min*mg TconTS)]	
	3'SL	Neu5Ac
TconTS1a (*E. coli*)	5,22 ± 0,03	0,65 ± 0,07
TconTS1a (CHO-Lec1)	2732,33 ± 1,38	88,08 ± 0,74

Table S1: Library of all glycan structures present on our glycan array

Number	Structures	Common name
	Saccharide with additional spacer	
	Monosaccharides	
1	Fucα-sp3	L-α-Fuc
2	Galα-sp3	α-Gal
3	Galβ-sp3	β-Gal
4	GalNAcα-sp0	TnSer
5	GalNAcα-sp3	T_n
6	GalNAcβ-sp3	β-GalNAc
7	Glcα-sp3	α-Glc
9	Glcβ-sp3	β-Glc
10	GlcNAcβ-sp3	β-GlcNAc
14	GlcN(Gc)β-sp4	β-GlcN(Gc)
15	$HOCH_2(HOCH)_4CH_2NH_2$	aminoglucitol
16	Manα-sp3	α-Man
18	Manβ-sp4	β-Man
19	ManNAcβ-sp4	β-ManAc
20	Rhaα-sp3	L-α-Rha
22	GlcNAcβ-sp4	β-GlcNAc
37	3-O-Su-Galβ-sp3	3-O-Su-β-Gal
38	3-O-Su-GalNAcα-sp3	3-O-Su-β-GalNac
43	6-O-Su-GlcNAcβ-sp3	6-O-Su-β-GlcNAc
44	GlcAα-sp3	α-glucoronic acid
45	GlcAβ-sp3	β-glucoronic acid
46	6-H_2PO_3Glcβ-sp4	β-Glc6P
47	6-H_2PO_3Manα-sp3	α-Man6P
48	Neu5Acα-sp3	α-Neu5Ac
49	Neu5Acα-sp9	α-Neu5AcBn
52	Neu5Gcα-sp3	α-Neu5Gc
54	9-NAc-Neu5Acα-sp3	9-Nac-α-Neu5Ac
55	3-O-Su-GlcNAcβ-sp3	3-O-Su-β-GlcNAc
	Disaccharides	
71	Fucα1-2Galβ-sp3	H_{di}
72	Fucα1-3GlcNAcβ-sp3	
73	Fucα1-4GlcNAcβ-sp3	Le
75	Galα1-2Galβ-sp3	
76	Galα1-3Galβ-sp3	B_{di}
77	Galα1-3GalNAcβ-sp3	$T_{\alpha\beta}$
78	Galα1-3GalNAcα-sp3	$T_{\alpha\alpha}$
80	Galα1-3GlcNAcβ-sp3	
81	Galα1-4GlcNAcβ-sp3	α-LacNAc
83	Galα1-6Glcβ-sp4	Melibiose
84	Galβ1-2Galβ-sp3	
85	Galβ1-3GlcNaAcβ-sp3	Le^c
87	Galβ1-3Galβ-sp3	
88	Galβ1-3GalNAcβ-sp3	$T_{\beta\beta}$
89	Galβ1-3GalNAcα-sp3	TF
93	Galβ1-4Glcβ-sp4	Lac
94	Galβ1-4Galβ-sp4	
97	Galβ1-4GlcNAcβ-sp3	LacNAc
100	Galβ1-6Galβ-sp4	
101	GalNAcα1-3GalNAcβ-sp3	Fs-2
102	GalNAcα1-3Galβ-sp3	A_{di}
103	GalNAcα1-3GalNAcα-sp3	core 5
104	GalNAcβ1-3Galβ-sp3	
106	GalNAcβ1-4GlcNAcβ-sp3	LacdiNAc
110	Glcα1-4Glcβ-sp3	maltose
111	Glcβ1-4Glcβ-sp4	cellobiose
112	Glcβ1-6Glcβ-sp4	gentiobiose
113	GlcNAcβ1-3GalNAcα-sp3	core 3
114	GlcNAcβ1-3Manβ-sp4	
115	GlcNAcβ1-4GlcNAcβ-Asn	chitobiose-Asn
117	GlcNAcβ1-4GlcNAcβ-sp4	chitobiose
118	GlcNAcβ1-6GalNAcα-sp3	core 6
119	Manα1-2Manβ-sp4	
120	Manα1-3Manβ-sp4	
121	Manα1-4Manβ-sp4	
122	Manα1-6Manβ-sp4	
123	Manβ1-4GlcNAcβ-sp4	
124	Manα1-2Manα-sp4	
145	Galβ1-3(6-O-Su)GlcNAcβ-sp3	6-O-Su-Le^c
146	Galβ1-4(6-O-Su)Glcβ-sp2	6-O-Su-Lac
147	Galβ1-4(6-O-Su)GlcNAcβ-sp3	6-O-Su-LacNAc
149	GlcNAcβ1-4(6-O-Su)GlcNAcβ-sp2	6-O-Su-chitobiose
150	3-O-Su-Galβ1-3GalNAcα-sp3	3'-O-Su-TF
151	6-O-Su-Galβ1-3GalNAcα-sp3	6'-O-Su-TF
152	3-O-Su-Galβ1-4Glcβ-sp2	SM3
153	6-O-Su-Galβ1-4Glcβ-sp2	6'-O-Su-Lac
155	3-O-Su-Galβ1-3GlcNAcβ-sp3	3'-O-Su-Le^c
157	3-O-Su-Galβ1-4GlcNAcβ-sp3	3'-O-Su-LacNAc
159	4-O-Su-Galβ1-4GlcNAcβ-sp3	4'-O-Su-LacNAc
161	6-O-Su-Galβ1-3GlcNAcβ-sp3	6'-O-Su-Le^c
163	6-O-Su-Galβ1-4GlcNAcβ-sp3	6'-O-Su-LacNAc
164	GlcAβ1-3GlcNAcβ-sp3	
165	GlcAβ1-3Galβ-sp3	
166	GlcAβ1-6Galβ-sp3	
167	GlcNAcβ1-4-[HOOC(CH_3)CH]-3-O-GlcNAcβ-sp4	GlcNAc-Mur
168	GlcNAcβ1--[HOOC(CH_3)CH]-3-O-GlcNAcβ-L-alanyl-D-i-glutaminyl-L-lysine	GMDP-Lys
169	Neu5Acα2-3Galβ-sp3	GM4
170	Neu5Acα2-6Galβ-sp3	
171	Neu5Acα2-3GalNAcα-sp3	3-SiaT_n
172	Neu5Acα2-6GalNAcα-sp3	SiaT_n
174	Neu5Gcα2-6GalNAcα-sp3	NeuGc-T_n
176	3-O-Su-Galβ1-4(6-O-Su)Glcβ-sp2	3',6-di-O-Su-Lac
177	3-O-Su-Galβ1-4(6-O-Su)GlcNAcβ-sp2	3',6-di-O-Su-LacNAc
178	6-O-Su-Galβ1-4(6-O-Su)Glcβ-sp2	6,6'-di-O-Su-Lac
179	6-O-Su-Galβ1-3(6-O-Su)GlcNAcβ-sp2	6,6'-di-O-Su-Le^c
180	6-O-Su-Galβ1-4(6-O-Su)GlcNAcβ-sp2	6,6'-di-O-Su-LacNAc
181	3,4-O-Su_2-Galβ1-4GlcNAcβ-sp3	3',4'-di-O-Su-LacNAc
182	3,6-O-Su_2-Galβ1-4GlcNAcβ-sp2	3',6'-di-O-Su-LacNAc
183	4,6-O-Su_2-Galβ1-4GlcNAcβ-sp2	4',6'-di-O-Su-LacNAc
184	4,6-O-Su_2-Galβ1-4GlcNAcβ-sp3	4',6'-di-O-Su-LacNAc
186	Neu5Acα2-8Neu5Acα2-sp3	$(Sia)_2$
189	3,6-O-Su_2-Galβ1-4(6-O-Su)GlcNAcβ-sp2	3',6,6'-tri-O-Su-LacNAc
192	GalNAcβ1-4(6-O-Su)GlcNAcβ-sp3	6-O-Su-LacdiNAc
193	3-O-Su-GalNAcβ1-4GlcNAcβ-sp3	3'-O-Su-LacdiNAc
194	6-O-Su-GalNAcβ1-4GlcNAcβ-sp3	6'-O-Su-LacdiNAc
195	6-O-Su-GalNAcβ1-4-(3-O-Su)GlcNAcβ-sp3	6'-Su-3-O-Ac-LacdiNAc
196	3-O-Su-GalNAcβ1-4(3-O-Su)-GlcNAcβ-sp3	3,3'-O-Su_2-LacdiNAc

197	3,6-O-Su_2-GalNAcβ1-4GlcNAcβ-sp3	3`,6`-Su_2-LacdiNAc
198	4,6-O-Su_2-GalNAcβ1-4GlcNAcβ-sp3	4`,6`-Su_2-LacdiNAc
199	4,6-O-Su_2-GalNAcβ1-4-(3-O-Ac)GlcNAcβ-sp3	4`,6`-Su_2-3-O-Ac-LacdiNAc
200	4-O-Su-GalNAcβ1-4GlcNAcβ-sp3	4`-O-Su-LacdiNAc
201	3,4-O-Su_2-Galβ1-4GlcNAcβ-sp3	3`,4`-Su_2-LacdiNAc
202	6-O-Su-GalNAcβ1-4(6-O-Su)GlcNAcβ-sp3	6,6`-O-Su-LacdiNAc
203	Galβ1-4(6-O-Su)GlcNAcβ-sp2	6-O-Su-LacNAc
204	4-O-Su-GalNAcβ1-4GlcNAcβ-sp2	4`-O-Su-LacdiNAc
205	Neu5Acα2-6GalNAcβ-sp3	
206	Neu5Gcα2-3Gal-sp3	NeuGcα3Gal
	Trisaccharides	
215	Fucα1-2Galβ1-3GlcNAcβ-sp3	Le^d H (type 1)
216	Fucα1-2Galβ1-4GlcNAcβ-sp3	H (type 2)
217	Fucα1-2Galβ1-3GalNAcα-sp3	H (type 3)
219	Fucα1-2Galβ1-4Glcβ-sp4	H (type 6
220	Galα1-3Galβ1-4Glcβ-sp2	
222	Galα1-3Galβ1-4GlcNAcβ-sp3	Galili (tri)
224	Galα1-4Galβ1-4Glcβ-sp3	P^k, Gb3, $GbOse_3$
225	Galα1-4Galβ1-4GlcNAc-sp2	P_1
226	Fucα1-2(Galα1-3)Galβ-sp3	B_{tri}
228	Galβ1-2Galα1-4GlcNAcβ-sp4	
229	Galβ1-3Galβ1-4GlcNAcβ-sp4	
231	Galβ1-4GlcNAcβ1-3GalNAcα-sp3	
232	Galβ1-4GlcNAcβ1-6GalNAcα-sp3	
233	Galβ1-3(Fucα1-4)GlcNAcβ-sp3	Le^a
234	Fucα1-3(Galβ1-4)GlcNAcβ-sp3	Le^x
235	Fucα1-2(GalNAcα1-3)Galβ-sp3	A_{tri}
238	GalNAcβ1-4Galβ1-4Glcβ-sp3	GA_2, $GgOse_3$
240	(Glcα1-4)$_3$β-sp4	maltotriose
241	(Glcα1-6)$_3$β-sp4	isomaltotriose
246	GlcNAcβ1-2Galβ1-3GalNAcα-sp3	
247	GlcNAcβ1-3Galβ1-3GalNAcα-sp3	
248	GlcNAcβ1-3Galβ1-4Glcβ-sp2	
250	GlcNAcβ1-3Galβ1-4GlcNAcβ-sp3	
251	GlcNAcβ1-4Galβ1-4GlcNAcβ-sp2	
252	GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ-sp4	chitotriose
253	GlcNAcβ1-6Galβ1-4GlcNAcβ-sp2	
254	Galβ1-3(GlcNAcβ1-6)GalNAcα-sp3	core 2
255	GlcNAcβ1-3(GlcNAcβ1-6)GalNAcα-sp3	core 4
258	Manα1-3(Manα1-6)Manβ-sp4	Man_3
262	Galβ1-3GalNAcβ1-3Gal-sp4	$T_{\beta\beta}$-Gal
264	Galβ1-4Galβ1-4GlcNAc-sp3	
287	3-O-Su-Galβ1-3(Fucα1-4)GlcNAcβ-sp3	Su-Le^a
288	Fucα1-3(3-O-Su-Galβ1-4)GlcNAcβ-sp3	Su-Le^x
289	Galα1-3(Neu5Acα2-6)GalNAcα-sp3	6-SiaTF
290	Galβ1-3(Neu5Acα2-6) GalNAcα-sp3	
292	Neu5Acα2-3Galβ1-3GalNAcα-sp3	3`-Sia-TF
293	Neu5Acα2-3Galβ1-4Glcβ-sp3	3`SL
294	Neu5Acα2-3Galβ1-4Glcβ-sp4	3`SL
295	Neu5Acα2-6Galβ1-4Glcβ-sp2	6`SL
298	Neu5Acα2-3Galβ1-4GlcNAcβ-sp3	3`SLN
299	Neu5Acα2-3Galβ1-3GlcNAcβ-sp3	3`-$SiaLe^c$
300	Neu5Acα2-6Galβ1-4GlcNAcβ-sp3	6`SLN
303	Neu5Gcα2-3Galβ1-4GlcNAcβ-sp3	3`SLN (Gc)
304	Neu5Gcα2-6Galβ1-4GlcNAcβ-sp3	6`SLN (Gc)
306	9-NAc-Neu5Acα2-6Galβ1-4GlcNAcβ-sp3	
315	Neu5Acα2-3Galβ1-4-(6-O-Su)GlcNAcβ-sp3	6-Su-3`SLN
317	Neu5Acα2-3Galβ1-3-(6-O-Su)GalNAcβ-sp3	6-Su-3`SiaTF
318	Neu5Acα2-6Galβ1-4-(6-O-Su)GlcNAcβ-sp3	6-Su-6`SLN
319	Neu5Acα2-3-(6-O-Su)Galβ1-4GlcNAcβ-sp3	6`-Su-3`SLN
321	(Neu5Acα2-8)$_3$-sp3	$(Sia)_3$
323	Neu5Acα2-6Galβ1-3GlcNAc-sp3	6`-$SiaLe^c$
324	Neu5Acα2-6Galβ1-3(6-O-Su)GlcNAc-sp3	6Su-6`-$SiaLe^c$
331	Neu5Gcα2-3Galβ1-3GlcNAcβ-sp3	3`$SiaLe^c$ (Gc)
	Tetrasaccharides	
359	Fucα1-2(Galα1-3)Galβ1-3GlcNAcβ-sp3	B (type 1)
360	Fucα1-2(Galα1-3)Galβ1-4GlcNAcβ-sp3	B (type 2)
362	Fucα1-2(Galα1-3)Galβ1-3GalNAcα-sp3	B (type 3)
363	Fucα1-2(Galα1-3)Galβ1-3GalNAcβ-sp3	B (type 4)
364	Fucα1-3(Galα1-3Galβ1-4)GlcNAcβ-sp3	αGalLe^x
366	Fucα1-2(GalNAcα1-3)Galβ1-3GlcNAcβ-sp3	A (type 1)
368	Fucα1-2(GalNAcα1-3)Galβ1-4GlcNAcβ-sp3	A (type 2)
371	Fucα1-2Galβ1-3(Fucα1-4)GlcNAcβ-sp3	Le^b
372	Fucα1-3(Fucα1-2Galβ1-4)GlcNAcβ-sp3	Le^y
373	Galα1-3Galβ1-4GlcNAcβ1-3Galβ-sp3	Galili (tetra)
375	Galα1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-sp3	
376	Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-sp4	LNT
377	Galβ1-3GlcNAcβ1-3Galβ1-3GlcNAcβ-sp2	
378	Galβ1-3GlcNAcα1-3Galβ1-4GlcNAcβ-sp3	
379	Galβ1-3GlcNAcβ1-3Galβ1-4GlcNAcβ-sp3	
380	Galβ1-3GlcNAcα1-6Galβ1-4GlcNAcβ-sp2	
381	Galβ1-3GlcNAcβ1-6Galβ1-4GlcNAcβ-sp2	
382	Galβ1-3GalNAcβ1-4Galβ1-4Glcβ-sp3	Asialo-GM1
383	Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ-sp2	LNnT
385	Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-sp3	i
387	Galβ1-4GlcNAcβ1-6Galβ1-4GlcNAcβ-sp2	
388	Galβ1-3(Galβ1-4GlcNAcβ1-6)GalNAcα-sp3	
389	GalNacβ1-3Galα1-4Galβ1-4Glcβ-sp3	Gb4, P
390	(Glcα1-4)$_4$β-sp4	maltotetraose
391	(Glcα1-6)$_4$β-sp4	isomaltotetraose
392	Fucα1-2(GalNAcα1-6)Galβ1-3GalNAcα-sp3	A (type 3)
395	GlcNAcβ1-3(GlcNAcβ1-6)Galβ1-4GlcNAcβ-sp3	Tk
401	Galβ1-3GlcNAcβ1-3Galβ1-3GlcNAcβ-sp3	Le^c3Le^c
419	3-O-SuGalβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-sp3	
420	4-O-SuGalβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-sp3	
421	Neu5Acα2-3(GalNAcβ1-4)Galβ1-4Glcβ-sp2	GM2
422	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ-sp3	
423	Fucα1-3(Neu5Acα2-3Galβ1-4)GlcNAcβ-sp3	$SiaLe^x$
426	Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAcβ-sp3	$SiaLe^a$
428	Fucα1-3(Neu5Acα2-3Galβ1-4)6-O-Su-GlcNAcβ-sp3	
429	Fucα1-3(Neu5Acα2-3(6-O-Su)Galβ1-4)GlcNAcβ-sp3	
433	Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα-sp3	Sia_2-TF
434	Neu5Acα2-8Neu5Acα2-3Galβ1-4Glcβ-sp4	GD3
	Penta-Nona Saccharides	
479	Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-sp4	LNFP-I
480	Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4GlcNAcβ-sp2	H (type1) penta
481	Galα1-3Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ-sp4	Galili (penta)
483	Fucα1-3(Fucα1-2 (Galα1-3)Galβ1-4)GlcNAcβ-sp3	Ble^y
488	Galβ1-4GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)GalNAcα-sp3	
489	Galβ1-4GlcNAcβ1-3(GlcNAcβ1-6)Galβ1-4GlcNAc-sp2	
490	GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)Galβ1-4GlcNAcβ-sp2	
492	(Glcα1-6)$_5$β-sp4	isomaltopentaose

493	(GlcNAcβ1-4)$_5$β-sp4	chitopentaose
495	Manα1-3(Manα1-3(Manα1-6)Manα1-6)Manβ-sp4	Man5
496	Fucα1-2Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4Glcβ-sp4	Le^b-Lac
497	Fucα1-3(Fucα1-2Galβ1-4)GlcNAcβ1-3Galβ1-4Glcβ-sp4	Le^y-Lac
498	(Galβ1-4GlcNAcβ1-3)$_3$-sp3	(LN)$_3$
499	Galβ1-4GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)Galβ1-4GlcNAc-sp2	I
501	Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ-sp4	Gb5
502	(Glcα1-6)$_6$β-sp4	maltohexaose
503	(GlcNAcβ1-4)$_6$β-sp4	chitohexaose
504	(A-GN-M)$_2$-3,6-M-GN-GNβ-sp4	9-OS
505	(GN-M)$_2$-3,6-M-GN-GNβ-sp4	7-OS
527	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-sp2	3'SLN-LacNAc
528	Fucα1-3(Neu5Acα2-3 Galβ1-4)GlcNAcβ1-3Galβ-sp3	SiaLe^a-3Gal
529	Neu5Acα2-6(Galβ1-3)GlcNAcβ1-3Galβ1-4Glcβ-sp4	LSTb
531	GalNAcβ1-4(Neu5Acα2-8Neu5Acα2-3)Galβ1-4Glc-sp2	GD2
532	Neu5Acα2-8Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc-sp2	GT3
533	(Neu5Acα2-8)2Neu5Acα2-3(GalNAcβ1-4)Galβ1-4Glc-sp2	GT2
534	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-sp3	6'SLN-LacNAc
536	Neu5Acα2-3Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-sp4	LSTa
537	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ-sp4	LSTd
538	Le^{x}1-6'(Le^{c}1-3')Lac-sp4	MFLNH III
539	LacNAc1-6'(Le^{a}1-3')Lac-sp4	MFLNH I
540	Le^{x}1-6'(6'SLN1-3')Lac-sp4	MSMFLNnH
541	Le^{x}1-6'(Le^{a}1-3')Lac-sp4	DFLNH (a)
542	Le^{a}Le^{x}1-6'(Le^{c}1-3')Lac-sp4	MF(1-3)iLNO
543	Le^{x}1-6'(Le^{b}1-3')Lac-sp4	TFLNH
	Higher Oligosaccharides	
625	(GlcAβ1-4GlcNAcβ1-3)$_8$-NH$_2$-ol	hyaluronine acid
627	(Sia2-6A-GN-M)$_2$-3,6-M-GN-GNβ-sp4	11-OS, YDS
	Terminal Galactose	
1A	Galβ1-3Glc*N*Ac	Lacto-N-Biose I
1B	Galβ1-4Glc*N*Ac	N-Acetyllactosamine
1C	Galβ1-4Gal	b-1-4-galactosyl-galactose
1D	Galβ1-6Glc*N*Ac	b-1-6 Galactosyl-N-acetyl glucosamine
1E	Galβ1-3Gal*N*Ac	b-1-3 Galactosyl-N-acetyl galactosamine
1F	Galβ1-3GalNAcβ1-4Galβ1-4Glc	asialo GM1
1G	Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc	Lacto-N-tetraose
1H	Galβ1-4Glc*N*Acβ1-3Galβ1-4Glc	Lacto-N-neotetraose
1I	Galβ1-4Glc*N*Acβ1-6(Galβ1-4Glc*N*Acβ1-3)Galβ1-4Glc	Lacto-N-neohexaose
1J	Galβ1-4Glc*N*Acβ1-6(Galβ1-3Glc*N*Acβ1-3)Galβ1-4Glc	Lacto-N-hexaose
1K	Galα1-4Galβ1-4Glc	Globotriose
1L	Gal*N*Acα1-*O*-Ser	Tn Antigen GalNAcα1-O-Ser
1M	Galb1-3Gal*N*Acα1-*O*-Ser	Galactosyl-Tn Antigen
1N	Galα1-3Gal	a1-3 Galactobiose
1O	Galα1-3Galβ1-4Glc*N*Ac	Linear B-2 Trisaccharide
1P	Galα1-3Galβ1-4Glc	Linear B-6 Trisaccharide
2A	Galα1-3Galβ1-4Galα1-3Gal	a1-3, b1-4,a1-3 Galactotetraose
2B	Galβ1-6Gal	beta1-6galactobiose
2C	Gal*N*Acβ1-3Gal	terminal disaccharide of Globotriose
2D	Gal*N*Acβ1-4Gal	receptor for Paureginosa
2E	Galα1-4Galβ1-4Glc*N*Ac	P1 antigen
2F	GalNAcα1-3Galβ1-4Glc	a-D-N-acetylgalactosamine 1-3Galβ1-4Glc
2G	Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc*N*Acβ1-6(Galβ1-3Glc*N*Acβ1-3)Galβ1-4Glc	iso-Lacto-N-octaose (iLNO)
2H	Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc*N*Acβ1-3Galβ1-4Glc	para-Lacto-N-hexaose (pLNH)
18B	Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glc	Globopentaose
18C	Galβ1-3GalNAcβ1-3Gal	Core type 4/Gb5 triose structure
	Terminal GlcNAc	
4A	Glc*N*Acβ1-4Glc*N*Ac	N,N'-Diacetyl chitobiose
4B	Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Ac	N,N',N''-Triacetyl chitotriose
4C	Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Ac	N,N',N'',N'''-Tetraacetyl chitotetraose
4D	Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Ac	N,N',N'',N''',N'''',N''''-Hexaacetyl chitohexaose
4E	Bacterial cell wall muramyl discaccharide	GlcNAcb1-4MurNAc
4F	Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Acβ1-4Glc*N*Ac	Pentacetyl chitopentaose
	Mannosyl containing glycans	
5A	Glc*N*Acβ1-2Man	b1-2 N-Acetylglucosamine-mannose
5B	Glc*N*Acβ1-2Manα1-6(Glc*N*Acβ1-2Manα1-3)Man	Bianntennary N-linked core pentasaccharide
5C	Manα1-2Man	a1-2-Mannobiose
5D	Manα1-3Man	a1-3-Mannobiose
5E	Manα1-4Man	a1-4-Mannobiose
5F	Manα1-6Man	a1-6-Mannobiose
5G	Manα1-6(Manα1-3)Man	a1-3,a1-6-Mannobiose
5H	Manα1-6(Manα1-3)Manα1-6(Manα1-3)Man	a1-3,a1-3,a1-6-Mannopentaose
	Fucosylated glycans	
7A	Fucα1-2Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc	Lacto-N-fucopentaose I
7B	Galβ1-3(Fucα1-4)Glc*N*Acβ1-3Galβ1-4Glc	Lacto-N-fucopentaose II
7C	Galβ1-4(Fucα1-3)Glc*N*Acβ1-3Galβ1-4Glc	Lacto-N-fucopentaose III
7D	Fucα1-2Galβ1-3(Fucα1-4)Glc*N*Acβ1-3Galβ1-4Glc	Lacto-N-difucohexaose I
7E	Galβ1-3(Fucα1-4)Glc*N*Acβ1-3Galβ1-4(Fucα1-3)Glc	Lacto-N-difucohexaose II
7F	Fucα1-2Gal	H-disaccaride
7G	Fucα1-2Galβ1-4Glc	2'-Fucosyllactose
7H	Galβ1-4(Fucα1-3)Glc	3'-Fucosyllactose
7I	Galβ1-4(Fucα1-3)Glc*N*Ac	Lewisx
7J	Galβ1-3(Fucα1-4)Glc*N*Ac	Lewisa
7K	Gal*N*Acα1-3(Fucα1-2)Gal	Blood Group A trisaccharide
7L	Fucα1-2Galβ1-4(Fucα1-3)Glc	Lactodifucotetraose (LDFT)
7M	Galβ1-3(Fucα1-2)Gal	Blood Group B Trisaccharide
7N	Fucα1-2Galβ1-4(Fucα1-3)Glc*N*Ac	Lewisy
7O	Fucα1-2Galβ1-3Glc*N*Ac	Blood Group H Type II Trisaccharide
7P	Fucα1-2Galβ1-3(Fucα1-4)Glc*N*Ac	Lewisb tetrasaccharide
8A	SO$_3$-3Galβ1-3(Fucα1-4)Glc*N*Ac	Sulpho Lewisa
8B	SO$_3$-3Galβ1-4(Fucα1-3)Glc*N*Ac	Sulpho Lewisx
8C	Galβ1-3Glc*N*Acβ1-3Galβ1-4(Fucα1-3)Glc*N*Acβ1-3Galβ1-4Glc	Monofucosyl-para-Lacto-N-hexaose IV
8D	Galβ1-4(Fucα1-3)Glc*N*Acβ1-6(Galβ1-3Glc*N*Acβ1-3)Galβ1-4Glc	Monofucosyllacto-N-hexaose III
8E	Galβ1-4(Fucα1-3)Glc*N*Acβ1-6(Fucα1-2Galβ1-3Glc*N*Acβ1-3)Galβ1-4Glc	Difucosyllacto-N-hexaose
8F	Galβ1-4(Fucα1-3)Glc*N*Acβ1-6(Fucα1-2Galβ1-3(Fucα1-4)Glc*N*Acβ1-3)Galβ1-4Glc	Trifucosyllacto-N-hexaose
8G	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)Glc	Lacto-N-fucopentaose VI (LNFP VI)
8H	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4Glc	Lacto-N-neodifucohexaose I (LNnDFH I)
8I	Fuca1-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)Glc	Lacto-N-neodifucohexaose II (LNnDFH II)
8J	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3(Fuca1-2)Galb1-4Glc	Trifucosyllacto-N-neoteraose I (TFLNnTI)
8K	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glc	Monofucosyllacto- N-neohexaose I (MFLNnH I)
8L	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-4(Fuca1-3)GlcNAcb1-3)Galb1-4Glc	Difucosyllacto-N-neohexaose I (DFLNnH I)
8M	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glc	Difucosyllacto-N-neohexaose II (DFLNnH II)
8N	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc	Monofucosyl(1-3)-iso-lacto-N-octaose (MFiLNO)
8O	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc	Trifucosyl(1-2,1-2,1-3)-iso-lacto-N-octaose (TFiLNO (1-2,1-2,1-3))
8P	GalNAcb1-3(Fuca1-2)Galb1-4Glc	Blood Group A Tetrasaccharide
9A	Galb1-3(Fuca1-2)Galb1-4(Fuca1-3)Glc	Blood Group B pentasaccharide
18D	Galα1-3(Fucα1-2)Galβ1-4Glc	blood group B antigen tetraose type 5
18E	GalNAcα1-3(Fucα1-2)Galβ1-4(Fucα1-3)Glc	blood group A pentasaccharide
19J	Galβ1-4(Fucα1-3)Glc*N*Acβ1-3Gal	Lewis X tetraose
19L	Fucα1-2Galβ1-4(Fucα1-3)Glc*N*Acβ1-3Gal	lewis Y pentaose
19M	Galβ1-3(Fucα1-4)Glc*N*Acβ1-3Gal	lewis a tetraose
19N	Fucα1-2Galβ1-3(Fucα1-4)Glc*N*Acβ1-3Ga	lewis b pentaose
	Sialylated glycans	

10A	Neu5Acα2-3Galβ1-3(Fucα1-4)Glc*N*Ac	Sialyl Lewisa (S Lea)
10B	Neu5Acα2-3Galβ1-4(Fucα1-3)Glc*N*Ac	Sialyl Lewisx (S Lex)
10C	Neu5Acα2-3Galβ1-3Glc*N*Acβ1-3Galβ1-4Glc	Sialyllacto-N-tetraose a
10D	Galβ1-4(Fucα1-3)Glc*N*Acβ1-6(Neu5Acα2-6Galβ1-4Glc*N*Acβ1-3)Galβ1-4Glc	Monosialyl, monofucosyllacto-N-neohexose
10K	Neu5Acα2-3Galβ1-4Glc*N*Ac	3'-Sialyllactosamine
10L	Neu5Acα2-6Galβ1-4Glc*N*Ac	6'-Sialyllactosamine
10M	Neu5Acα2-3Galβ1-3GlcNAcβ1-3Galβ1-4Glc	LS-Tetrasaccharide a (LSTa)
10N	Galβ1-3(Neu5Acα2-6)Glc*N*Acβ1-3Galβ1-4Glc	LS-Tetrasaccharide b (LSTb)
10O	Neu5Acα2-6Galβ1-4Glc*N*Acβ1-3Galβ1-4Glc	LS-Tetrasaccharide c (LSTc)
10P	Neu5Acα2-3Galβ1-3(Neu5Acα2-6)Glc*N*Acβ1-3Galβ1-4Glc	Disialyllacto-N-tetraose
11A	Neu5Acα2-3Galβ1-4Glc	3'-Sialyllactose
11B	Neu5Acα2-6Galβ1-4Glc	6'-Sialyllactose
11C	(Neu5Acα2-8Neu5Ac)n (n<50)	Colominic acid
18A	Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4Glc	LS-Tetrasaccharide d
19K	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Gal	Sialyl lewis X pentaose
	Glycosaminoglycans - high and low molecular weight	
12A	Neocarratetraose-41, 3-di-O-sulphate (Na+)	$C_{24}H_{36}O_{25}S_2Na_2$ (Mixed anomers. Tetrasaccharide of regular κ - carrageenan)
12B	Neocarratetraose-41-O-sulphate (Na+)	$C_{24}H_{37}O_{22}SNa$ (Mixed anomers. Derived from C1003 by removal of the non-reducing terminal 4-sulphate)
12C	Neocarrahexaose-24,41, 3, 5-tetra-O-sulphate (Na+)	$C_{36}H_{52}O_{46}S_4Na_4$ (Mix. anomers. A hybr. seq. compr. carrageenan disacch. in the order k-i-k, deriv. f. the carrageenan f. *Chondrus crispus*)
12D	Neocarrahexaose-41, 3, 5-tri-O-sulphate (Na+)	$C_{36}H_{53}O_{37}S_3Na_3$ (Mixed anomers. Hexasaccharide of regular κ-carrageenan)
12E	Neocarraoctaose-41, 3, 5, 7-tetra-O-sulphate (Na+)	$C_{48}H_{70}O_{58}S_4Na_4$ (Mixed anomers. Octasaccharide of regular κ-carrageenan)
12F	Neocarradecaose-41, 3, 5, 7, 9-penta-O-sulphate (Na+)	$C_{60}H_{87}O_{71}S_5Na_5$ (Mixed anomers. Decasaccharide of regular κ- carrageenan)
12G	ΔUA-2S-GlcNS-6S	$C_{12}H_{15}NO_{19}S_3Na_4$ (Predominant disaccharide produced from heparin by heparinase I and II)
12H	ΔUA-GlucNS-6S	$C_{12}H_{16}NO_{16}S_2Na_3$ (Produced from heparinase II digestion of heparin and heparin sulphate)
12I	ΔUA-2S-GlucNS	$C_{12}H_{16}NO_{16}S_2Na_3$ (Produced from heparin by digestion with heparinase I and II)
12J	ΔUA-2S-GlcNAc-6S	$C_{14}H_{18}NO_{17}S_2Na_3$ (Minor component produced from heparin by heparinase II)
12K	ΔUA-GlcNAc-6S	$C_{14}H_{19}NO_{14}SNa_2$ (Product of the action of heparinases II and III on heparin and heparan sulphate)
12L	ΔUA-2S-GlcNAc	$C_{14}H_{19}NO_{14}SNa_2$ (Minor product of the action of heparinase II on heparin)
12M	ΔUA-GlcNAc	$C_{14}H_{20}NO_{11}Na$ (Produced from heparin sulphate by digestion With heparinase III)
12N	ΔUA-GalNAc-4S (Delta Di-4S)	$C_{14}H_{19}NO_{14}SNa_2$ (Produced from various chondroitin sulphates By the action of chondroitinases ABC, B and AC-1)
12O	ΔUA-GalNAc-6S (Delta Di-6S)	$C_{14}H_{19}NO_{14}SNa_2$ (Produced from various chondroitin sulphates By the action of chondroitinases ABC, AC-1 and C)
12P	ΔUA-GalNAc-4S,6S (Delta Di-disE)	$C_{14}H_{18}NO_{17}S_2Na_3$ (Produced from various chondroitin sulphates By the action of chondroitinases ABC, B and AC-1)
13A	ΔUA-2S-GalNAc-4S (Delta Di-disB)	$C_{14}H_{18}NO_{17}S_2Na_3$ (Prod. f. var. chondroitin sulphates by action of c. ABC and/or B. Most typ. f. chondroitin sulphate B (dermatan sulphate))
13B	ΔUA-2S-GalNAc-6S (Delta Di-disD)	$C_{14}H_{18}NO_{17}S_2Na_3$ (Produced from various chondroitin sulphates by the action of chondroitinase ABC)
13C	ΔUA-2S-GalNAc-4S-6S (Delta Di-tisS)	$C_{14}H_{17}NO_{20}S_3Na_4$ (Produced as a minor component by the action of chondroitinase ABC on various chondroitin sulphates, particularly B)
13D	ΔUA-2S-GalNAc-6S (Delta Di-UA2S)	$C_{14}H_{18}NO_{14}SNa_2$ (Produced as a minor component from various chondroitin sulphates by the action of chondroitinase ABC)
13E	ΔUA-GlcNAc (Delta Di-HA)	$C_{14}H_{20}NO_{11}Na$ (The only unsaturated disaccharide produced from hyaluronic acid by the action of chondroitinase ABC or AC-1)
13F	(GlcAβ1-3Glc*N*Acβ1-4)n (n=4)	Hyaluronan fragments (4mer)
13G	(GlcAβ1-3Glc*N*Acβ1-4)n (n=8)	Hyaluronan fragment (8mer)
13H	(GlcAβ1-3Glc*N*Acβ1-4)n (n=10)	Hyaluronan fragment (10mer)
13I	(GlcAβ1-3Glc*N*Acβ1-4)n (n=12)	Hyaluronan fragment (12mer)
13J	(GlcA/IdoAα/β1-4Glc*N*Acα1-4)n (n=200)	Heparin
13K	(GlcA/IdoAβ1-3(±4/6S)Gal*N*Acβ1-4)n (n<250)	Chondroitin sulfate
13L	((±2S)GlcA/IdoAα/b1-3(±4S)Gal*N*Acβ1-4)n (n<250)	Dermatan sulfate
13M	(GlcA/IdoAβ1-3(±6S)Gal*N*Acβ1-4)n (n<250)	Chondroitin 6-sulfate
13 N	HA - 4 10mM	
130	HA - 6 10mM	
13P	HA - 8 9.7mM	
14A	HA 10 7.83mM	
14B	HA-12 6.5mM	
14C	HA-14 5.6mM	
14D	HA-16 4.9mM	
14E	HA 30000 da 2.5mg/ml	
14F	HA 107000 da 2.5mg/ml	
14G	HA 190000 da 2.5 mg/ml	
14H	HA 220000 da 2.5 mg/ml	
14I	HA 1600000 da 2.5 mg/ml	
14J	Heparin sulfate 5 mg/ml	
14K	β1-3Glucan	
14L	Chondroitin disaccharide Δdi-OS, sodium salt	$C_{14}H_{20}NNaO_{11}$ (produced from various chondriotin sulfates by the action of chondriotinases ABC, AC-1 and C)
	Ganglioside oligosaccharides	
17A	GalNAcβ1-4Galβ1-4Gl	asialo GM2
17B	Galβ1-3GalNAcβ1-4Galβ1-4Glc	asialo GM1
17C	Galβ1-3GalNAcβ1-4(Neu5Acα2-8Neu5Acα2-8 Neu5Acα2-3)Galβ1-4Glc	GT1c ganglioside sugar
17D	Neu5Acα2-8Neu5Acα2-3Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glc	GT1a ganglioside sugar
17E	Galβ1-3GalNAcβ1-4(Neu5Acα2-8Neu5Acα2-3)Galβ1-4Glc	GD1b ganglioside sugar
17F	Neu5Acα2-3Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glc	GD1a Ganglioside sugar
17G	Neu5Acα2-3Galβ1-3GalNAcβ1-4Galβ1-4Glc	GM1b ganglioside sugar
17H	Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glc	GM1a ganglioside sugar
17I	Fucα1-2Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glc	fucosyl GM1 ganglioside sugar
17J	GalNAcβ1-4(Neu5Acα2-8Neu5Acα2-8Neu5Acα2-3)Galβ1-4Glc	GT2 ganglioside sugar
17K	GalNAcβ1-4(Neu5Acα2-8Neu5Acα2-3)Galβ1-4Glc	GD2 gangliosifde sugar
17L	GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glc	GM2 ganglioside sugar
17M	Neu5Acα2-8Neu5Acβ2-8Neu5Acα2-3Galβ1-4Glc	GT3 ganglioside sugar
17N	Neu5Acα2-8Neu5Acα2-3Galβ1-4Gl	GD3 ganglioside sugar
17O	Neu5Acα2-3Galβ1-4Glc	GM3 ganglioside sugar
	Complex type N-glycans	
19A	Galβ1-4GlcNAcβ1-2Manα1-3(Galβ1-4GlcNAcβ1-2Manα1-6Man)β1-4GlcNAcβ1-4(Fucα1-6)GlcNAc Galβ1-4GlcNAcβ1-2(Galβ1-4GlcNAcβ1-4)Manα1-3(Galβ1-4GlcNAcβ1-2(Galβ1-4GlcNAcβ1-6)Manα1-	Asialo galactosylated, fucosylated biantennary
19B	6Man)β1-4GlcNAcβ1-4GlcNAc	Asialo, galactosylated, tetranatennary, N-linked glycan
19C	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-	Monosialo(2,6), biantennary (A1)
19D	4GlcNAc	Disialo (2,6) biantennary (A2)
19E	Galβ1-4GlcNAcβ1-2Manα1-3(Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-	Asialo, galactoyslated, biantennary (NA2)
19F	4(Fucα1-6)GlcNAc Neu5Acα2-6Galβ1-4GlcNAcβ1-2(Neu5Acα2-6Galβ1-4GlcNAcβ1-4)Manα1-3(Neu5Acα2-6Galβ1-	Disialo, galactosylated, fucosylated, biantennary (A2F)
19G	4GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc	Trisialylated, galactosylated, triantennary (A3)
19H	GlcNAcβ1-2(GlcNAcβ1-4)Manα1-3(GlcNAcβ1-2Manα1-6)GlcNAcβ1-4Manβ1-4GlcNAcβ1-4GlcNAc	Asialo, agalacto, bisected triannary (NGA3B)

Table S2: *Sequence similarities between TconTS-LDs*

	TconTS1-LD	**TconTS2-LD**	**TconTS3-LD**	**TconTS4-LD**
TconTS1-LD	-	42.3% (54.6%)	41.0% (57.7%)	40.7% (51.8%)
TconTS2-LD	42.3% (54.6%)	-	39.9% (54.4%)	37.4% (55.1%)
TconTS3-LD	41.0% (57.7%)	39.9% (54.4%)	-	45.4% (61.7%)
TconTS4-LD	40.7% (51.8%)	37.4% (55.1%)	45.4% (61.7%)	-

The amino acid sequences of TconTS-LD were pairwise aligned using the Blosum62 matrix with a gap opening penalty of 12 and a gap extension penalty of 5. Shown are the percent of positions with identical or similar (giving a positive value in the Blosum62 matrix; in brackets) amino acids.

Table S3. Primers for amplification of cDNA fragments encoding TconTS-LDs

Coding Sequence	Sense primer 5'-3')
His-MBP-TEV	GCCCATGGGCCATCACCATCACCATCAC (*NcoI*)
TconTS1-LD	GC AAG CTT AAC TGC CTC CCG GGC (*HindIII*)
TconTS1-αHel-LD	GC AAG CTT GAC GAG CTG AAA AGC (*HindIII*)
TconTS2-LD	GC AAG CTT TGT CAA CTG AAC AAA AAG CG (*HindIII*)
TconTS2-αHel-LD	GC AAG CTT CTG GAG GAT GAG ATG GAG G (*HindIII*)
TconTS3-LD	GC AAG CTT TGT TCC TCA CCG GAT GGT G (*HindIII*)
TconTS3-αHel-LD	GC AAG CTT CTA GAA GAC GAG CTG GAA AGC(*HindIII*)
TconTS4-LD	GC AAG CTT TGC TCT GCA ACT ACC G (*HindIII*)
TconTS4-αHel-LD	GC AAG CTT CTC GCT GAC GAA CTG AAG (*HindIII*)

Sense primers used in combination with reverse primers (A) GCGTCGACGCCCTGAAAATAAAGATTCTC (*SalI*) for His-MBP-TEV; or (B) GC GCG GCC GCT TAT TTT TCG AAC TGC GG (*NotI*) for the TconTS-LD fragments. The restriction sites used are underlined.

3.5 Results

Trans-sialidase lectin domain from *Trypanosoma congolense* influences enzyme activities

Mario Waespy, Thaddeus T. Gbem, Nilima Dinesh, Shanmugam Solaiyappan, Sørge Kelm

(Manuscript in preparation)

Contribution of Mario Waespy:

- Establishment of TconTS domain-swap strategy and experimental realisation
- Cloning, expression and purification of recombinant TconTS proteins
- TconTS enzyme assays
- HPAEC-PAD data acquisition and evaluation
- Anti TS mAb 7/23 epitope mapping
- Homology modelling and *insilico* data calculations
- Preparation and draft of the manuscript

The influence of *T. congolense* trans-sialidase lectin domain on enzyme activities

Mario Waespy[1], Thaddeus T. Gbem[2], Nilima Dinesh[1], Shanmugam Solaiyappan[1], Sørge Kelm[1, 2 §]

[1] *Centre for Biomolecular Interactions Bremen, Faculty for Biology and Chemistry, University Bremen, Bremen, Germany*

[2] *Africa Centre of Excellence for Neglected Tropical Diseases and Forensic Biotechnology, Ahmadu Bello University, Zaria, Nigeria*

[§] corresponding author: skelm@uni-bremen.de

Abstract

Trans-sialidases (TS) are unusual enzymes that catalyse the transfer of terminal sialic acids from sialoglycoconjuagtes to terminal galactose or *N*-acetylgalactosamine of oligosaccharides without the requirement of CMP-Neu5Ac, the activated Sia used by sialyltransferases. Most work on trypanosomal TS has been done on enzymatic activities of TS from *T. cruzi* the causative agent of Chagas disease in Latin America, *T. brucei*, causing sleeping sickness in humans, and TS from *T. congolense* (TconTS), the causative agent of Nagana in livestock. Recently we demonstrated TconTS-LD binding activities to several carbohydrates, such as oligosaccharides containing mannose or terminal galactose.

To investigate the influence of TconTS-LD on enzyme activities we established a novel strategy swapping LDs and CDs from different TconTS and expressing these recombinant TconTS constructs in bacteria and eukaryotic cells. Domain swapped, chimeric trans-sialidases showed clear sialidase and sialic acid (Sia) transfer activities, when using fetuin and lactose as Sia donor and acceptor substrates, but their specific and relative activities differed drastically from that of wild type TconTS, indicating an influence of the lectin domain on enzyme activities and possibly also substrate specificities. Besides that, we could demonstrate that the presence of 1,4-β-mannotriose during TS reactions, modulates enzyme activities, namely increasing the ratio between trans-sialylation and hydrolysis.

Author's summary

Here we reported a strategy to swap catalytic (CD) and lectin domains (LD) of trans-sialidases (TS) from *Trypanosoma congolense* (TconTS) and provide evidence for a potential influence of LD on

enzymatic activities.

Using the crystal structure of *T. cruzi* TS (TcruTS) homology models for TconTS1, TconTS2, TconTS3 and TconTS4 were calculated. This structure models together with amino acid sequence alignments of TconTS were used to locate a suitable region for the insertion of a unique restriction site utilised to exchange and recombine LDs from different enzymes. Domain swapped recombinant TconTS, comprising CD from a highly active and LD from a less active TconTS were expressed by bacteria *(E. coli* Rosetta pLacI*)* as well as fibroblasts (CHO-Lec1). Enzymes were purified and analysed for their specific catalytic activities using fetuin and lactose as sialic acid (Sia) donor and acceptor substrates. TS reaction products 3'SL and Neu5Ac were quantified employing a HPLC-based method (HPAEC-PAD) specific for carbohydrate analysis. Obtained results clearly demonstrated Sia transfer and sialidase activities for domain swapped enzymes, although activities determined, drastically differed from that observed for wild type TconTS, therefore potentially indicating a modulating function for LD on enzymatic activities. Interestingly, TconTS mutated to comprise the inserted unique restriction site but have not been domain swapped also showed significant changes in enzyme activities compared to wild type TconTS. Additional western blot analysis of these modified TconTS enzymes using an anti TS monoclonal antibody provided evidence that changes in enzymatic activities observed are possibly *N*-glycan dependent.

In summary, the strategy to exchange and recombine CDs and LDs from different TconTS, presented in this study, was successfully used to generate chimeric active enzymes expressed and isolated from bacteria and fibroblasts exhibiting Sia transfer and sialidase activities different from that of wild type TconTS. Furthermore, results also provided evidence that TconTS-LDs potentially play relevant roles in modulating the biological functions of these and possibly other trans-sialidases.

Introduction

Trypanosomes are protozoan parasites causing trypanosomiasis in Latin America (*T. cruzi*), also known as Chagas' disease [1,2], Human African trypanosomiasis (HAT, *T. brucei brucei*) and

Animal African trypanosomiasis (AAT, also called Nagana) in livestock (*T. congolense*) in Sub-Sahara Africa [3-5], bringing death to multiple-thousands humans and millions of cattle every year [6-8]. To evade mammalian host immune response and insect vectors digestive system, parasites stage specifically express an unusual enzyme called trans-sialidase (TS) on their surfaces to transfer terminal sialic acids (Sia) from host glycoconjugates to terminal galactose residues on their own surface glycoproteins [9-12]. Several studies have shown that trypanosomal TS play important roles in the pathology of the disease in mammalian hosts and are necessary for survival of the parasite also in the insect vector [9,13-15]. Structurally, TS contain two domains, a N-terminal catalytic (CD) and a C-terminal lectin-like domain (LD), which are connected via a 23 to 25 amino acid long α-helix, and arranged in close contact to each other [11].

Whereas published studies have concentrated on the enzymatic activities of CDs [16-22], no function of the LDs has been described so far. First evidence for a more pivotal role of the LDs has come from our phylogenetic analysis done separately on CD and LD of TconTS, in which, only the LDs of the most active enzymes TconTS1 and TconTS2 group together, whereas the CDs of these enzymes fall into different branches [21]. For example, the CD of the highly active TconTS2 groups together with the two less active enzymes TconTS3 and TconTS4, but it is less related to the also very active TconTS1 [21]. In our recent work we demonstrated the lectin activity of TconTS-LD to oligomannose and oligogalactose oligosaccharides [23]. Interestingly, oligomannose oligosaccharides are not substrates for the catalytic transfer of Sia [10,12,24]. However, high-mannose *N*-glycans have been found on glycoproteins or as part of their glycosylphosphatidylinositol (GPI)-anchor on the parasite's surface [25-31]. Therefore, these potentially serve as ligand structures for TconTS-LD. Furthermore, also TS were found to be glycosylated, predominantly with *N*-glycans of the high-mannose type [3,4], leading to the suggestion of intermolecular interactions possibly mediated by TS-LD. Our experimental results from size-exclusion chromatography analysis demonstrated the mannose-dependent oligomerisation of recombinant, high-mannosylated TconTS [23]. Based on this fundamental findings we wanted to investigate a possible role of TconTS-LD on catalytic activities of the enzymes.

Here we present a strategy to sufficiently swap CDs and LDs from different TconTS and expressed recombined chimeric enzymes by *E. coli* as well as fibroblasts for subsequent biochemical

characterisation regarding enzymatic activities. Homology models of TconTS revealed that amino acid residues localised at the contact sites between TconTS CD and LD are well conserved in the TS family. In addition, *in silico* data of domain swapped TconTS revealed a similar overall topology with an extended hydrogen bond network at the interface between CD and LD, which was also found in wild type TconTS. Based on this observation we assumed that recombination of CDs and LDs from different TconTS would not interfere with correct folding.

Clear Sia transfer and sialidase activities were observed for purified domain swapped recombinant TconTS expressed by *E. coli* or CHO-Lec1 fibroblasts. Interestingly, specific catalytic activities drastically differed from that of wild type enzymes. In addition, we have demonstrated the relevance of LD for enzymatic activities of TconTS1 and TconTS2, since in the presence of 1,4-β-mannotriose these enzymes have a higher Sia transfer to sialidase ratio. Furthermore, epitope mapping of the anti TS mAb 7/23 revealed that it's binding epitope is glycan dependent and localised in the lectin domains of both, TconTS1 and TconTS2. Interestingly, it has been demonstrated that binding of this antibody to native TS-forms isolated from cultured procyclic trypomastigotes reduced specific enzymatic activity by 75 % [12].

Understanding the role of TconTS-LD will provide advanced knowledge of their biological functions in the parasite during pathogenesis and the establishment of chronic infections. Furthermore, new insights into structure-related TS functions are likely to open new strategies for potent drug development targeting TS and will possibly also provide new opportunities for biotechnological applications employing trypanosomal TS.

Results

The contact sites between TconTS CD and LD

Using the crystal structure of *T. cruzi* TS (TcruTS) [11] as template, homology structure models of TconTS1 through TconTS4 were calculated as described under Methods. Similar to other TS, CD and LD are localised in immediate vicinity, connected by an 23 to 25 amino acid long α-helix connecting both domains (Figure 1). From these homology models of TconTS, we observed that the arrangement of CD and LD is stabilised by the relatively broad contact sites between both

domains [23]. The interface between CD and LD from TcruTS [11] and *T. rangeli* sialidase (TranSA) [32] were previously determined by Buschiazzo *et al.* [11] and Amaya *et al.* [32] to be about 36 to 48 % lager compared to those from other sialidases, such as *Vibrio cholerae* sialidase (VCS) [33] and *Marcobdella decora* intramolecular trans-sialidase (leech IT-sialidase) [34]. Along this line, we calculated areas of the molecular interfaces for several African TS, namely TconTS1, TconTS2, TconTS3, TconTS4 and *T. brucei* TS (TbruTS). Results revealed 33 to 46 % larger areas compared to that of VCS and leech IT-sialidase, consistent with Buschiazzo's and Amaya's observations [11,32] (Table 1).

It can be seen that the overall surface area of the contact site between CD and LD of African TS is around 2833.55 to 3306.47 $Å^2$ and quite similar in size among the TS family (Table 1). Interestingly, amino acid sequence alignments of TS illustrated that regions, which are part of the contact site, are well conserved among TS (Figure 1 D). Further *in silico* analysis regarding potential interactions between amino acids at the contact site of TconTS, revealed a network of hydrogen bonds formed between both domains, providing a comparative rigid overall conformation of the enzyme. Interestingly, calculated structure models revealed two defined locations, in which the majority of inter-domain hydrogen bonds were located (Figure 1). One site is close to the α-helix (Figure 1 B), whereas the second is located at the opposite side of the interface (Figure 1 C). Amino acids from both domains of TconTS1 through TconTS4, forming hydrogen bonds at the contact site, where identified and summarised in Table 2.

In total, 13 hydrogen bonds at the interface between CD and LD for TconTS3, 12 for TconTS1 and 11 for TconTS2 as well as for TconTS4, were identified from homology models, also considering hydrogen bonds between CD/LD and the α-helix. Moreover, the number of these hydrogen bonds in each TS seems to be equally distributed over both locations A and B (Figure 1 B and C). Not surprisingly, amino acid sequence alignments of TS revealed that amino acid residues essential for hydrogen bond formation in A and B are well conserved among the TS family (Table 3). One exception represents TranSA, which showed 9 amino acid deviations from the consensus sequence, 5 in CD and 4 in LD (Table 3). 4 of these were also found in TcruTS, which exhibits 5 changes in total, 3 in CD and 2 in LD. For the African trypanosomes, TconTS and TbruTS, only up to 3 variations in total were identified in comparison to the consensus sequence.

Interestingly, the 3 amino acid residues Trp (10), Lys (11) and Asp (12) of the α-helix, essential for the formation of hydrogen bonds to Glu (6) and Arg (9) of CD and Tyr (15) and Trp (13) of LD in A (Figure 1 B), are conserved through all TS listed in Table 3, indicating a fundamental role of that region for enzyme structure stability. In addition, the Gln (18) of LD, which is also well conserved in the trypanosomal TS family, is located on a loop at the more exposed location B (Figure 1 C). Gln 18 reaches relatively far into the CD, where it forms hydrogen bonds to Gly (4), Lys (7) and Arg (5), suggesting that Gln 18 is one of the key amino acids for structure stability in B, mainly responsible for stabilising the contact between CD and LD (Figure 1).

TconTS domain swap

To investigate the influence of TconTS-LD on enzyme activity, we established a strategy to exchange and recombine CDs and LDs from different TconTS. *In silico* data provided good evidence that a potential domain swap on a molecular level is possible, because amino acid residues at the contact site of TconTS between CD and LD, including those essential for hydrogen bond network formation are well conserved in TconTS1 through TconTS4 (Figure 1 and Table 3). A unique endonuclease restriction site was introduced between the regions coding for TconTS CD and LD, which can subsequently be used to swap domains from different TconTS. At first sight, the α-helix itself might represent a potential target for the restriction site introduction, but amino acid sequence alignments of TconTS revealed that the majority of amino acid residues of the α-helix are well conserved (Figure 1 D and Figure 2 A). Thus the insertion or exchange of 2 - 3 amino acids (corresponding to 6 - 9 nucleotides) in the sequence of the α-helix is expected to interfere with its structure and or interactions (see above).

However, the crystal structure of TcruTS [11] revealed a crystallographically unresolved flexible loop right after the α-helix between the two domains (Figure 2). This structural hairpin loop, which is stabilised by a well conserved disulphide bridge (C477 and C487 in TconTS1), does not exhibit any conserved amino acid residues among TTS (Figure 3 A). Sterically, this loop is more distinct from the enzyme and not in close contact to CD. In summary, these features together encouraged us to utilise that loop segment as a suitable target for the introduction of an endonuclease

restriction site. In addition, this region provides almost the entire LD to be exchanged, without the disruption of any potentially important structure elements such as β-sheets, salt-bridges or α-helices.

A screening was performed to identify endonuclease restriction sites not present in the coding sequences of TconTS1 through TconTS4, from which *Eco*105I (TAC'GTA) was selected. The enzyme generates blunt ends after restriction digest and can therefore also be used to insert any other sequence in frame, for example spacer or other TS-LD for further applications without requiring additional restriction sites. Suitable sense and reverse primer for each TconTS, containing the *Eco*105I restriction site, were designed (Table S1) and mutagenesis experiments were performed as described under Methods, inserting the *Eco*105I site at the appropriate position in the hairpin loop (Figure 2 B). Restriction digest and recombination of TconTS CD and LD were performed to generate all 12 possible domain swapped constructs as plasmid preparations, listed in Table 4 (Nr. 5-16).

Expression and purification of domain swapped TconTS

After introduction of the *Eco*105I restriction site into the modified pET28a MBP plasmid [23], all four mutated recombinant TconTS* (TconTS containing *Eco*105I restriction site) were expressed by *E. coli* Rosetta (DE3) pLacI, as described under Methods. A schematic composition of recombinant domain swapped TconTS expressed by bacteria, is shown in Figure 3 A. After expression, sufficient amounts of mutated TconTS* were obtained and characterised by SDS-PAGE and Western blot analysis (Figure 3 C, ~135 kDa). However, in contrast to the expression of TconTS-LD constructs, mutated TconTS* as well as wild type TconTS were obtained only in relatively low amounts (100 – 300 µg of enzyme per litre of bacterial culture) as soluble protein after tandem affinity chromatography, employing Ni-NTA and *Strep*-tag purification (described under Methods). Even several expression optimisations, including variation of the isopropyl-β-D-1-thiogalacto-pyranoside (IPTG) concentration, as well as induction time and temperature adjustments, only negligibly enhanced yields of about 200 – 500 µg of purified soluble protein per litre of bacterial culture were obtained.

Based on our recent findings that TconTS1 and TconTS2 are highly active, whereas TconTS3 and TconTS4 exhibited drastically lower enzyme activities [20,21], it was obvious to first swap the LDs from a highly and a less active TconTS (e.g. TconTS1 and TconTS3) and analyse the resulting specific activities. Along this line, TconTS3-LD was recombined with TconTS1a-CD to generate the recombinant domain swapped TconTS1a/TS3 (Table 4, construct 6) as described under Methods. Concerning the differences in enzymatic activities of wild type TconTS mentioned above, we assumed that if the LD has any effect on enzyme activity, TconTS3-LD might be able to change the overall enzyme behaviour of TconTS1a/TS3 relative to that of wild type TconTS1a.

The pET28aMBP plasmid encoding TconTS1a/TS3 was transformed into *E. coli* Rosetta pLacI and recombinant protein was expressed and purified using the same conditions described for expression of mutated TconTS* as well as wild type TconTS (described under Methods). After tandem affinity purification a major band close to 130 kDa was observed on SDS-PAGE gel (Figure 3 C).

Besides the bacterial expression of domain swapped TconTS, a eukaryotic expression system was also established using CHO-Lec1 fibroblasts. Furthermore we demonstrated that the enzyme activity of TconTS expressed by bacteria (*E. coli*), is drastically lower compared to that of TconTS expressed by CHO-Lec1 cells [20,21].

For stable transfection in CHO-Lec1 cells, DNA sequences coding for recombinant domain swapped TconTS1a/TconTS3 and TconTS2/TconTS4 as well as sequences for mutated TconTS1a* and TconTS2* were subcloned into pDEF, a modified eukaryotic expression vector [20] (described under Methods). A transin signal peptide for protein secretion was N-terminally fused directly to the TconTS coding sequence for eukaryotic expression (Figure 3 A) instead of His- and MBP-tags for TconTS expressed by bacteria (Figure 3 B). In addition, C-terminal SNAP- and *Strep*-tags are included in both, the bacterial and fibroblast expressed TconTS. Because of the time intensive work starting from transfection to single clone selected, stable CHO-Lec1 cell lines and finally to a sufficient amount of purified recombinant protein for enzyme activity analysis, only TconTS1a*, TconTS1a/TS3, TconTS2* and TconTS2/TS4 were selected for production. Stable transfection, single clone selection, cultivation of TconTS expressing cells and finally purification of recombinant proteins was done as described under Methods. Cells transfected with TconTS1a* and

TconTS1a/TS3 respectively, showed relatively high amounts of secreted protein in the cell culture supernatant (SN) directly after clonal selection. In contrast to that, TconTS2* showed only relatively low amounts of secreted protein in the SN even after an extended cultivation time (2-3 weeks), whereas CHO-Lec1 cells transfected with TconTS2/TS4 did not proliferate at all after transfection. Also repetition of stable transfection, using a dilution series of TconTS2/TS4 plasmid DNA ranging from 5 to 15 µg, revealed the same results. Interestingly, bacterial expression of TconTS2/TS4 followed by SDS-PAGE analysis of the lysate revealed a clear band close to 130 kDa (Figure 3 C), as it was expected based on the calculated molecular weight of TconTS2/TS4. However, reasons for this surprising observations still remain unclear, although corresponding experiments to further investigate these findings are part of our current research.

Binding epitope determination of the monoclonal anti TS mAb 7/23

Besides the anti *Strep*-tag antibody also the anti TS mAb 7/23 [12] was used for TconTS characterisation. This mAb recognises two isolated native TconTS forms, which where later suggested to contain TconTS1 and TconTS2 [20,21]. In addition, this antibody was found to precipitate 75% of the TS activity of a native TS from *T. congolense* cultures [12]. T. Gbem *et al.* showed the specific binding of anti TS mAb 7/23 to recombinant TconTS1 and TconTS2, purified from cell culture supernatants of transfected CHO-Lec1 cells, whereas TconTS3 and TconTS4 were not recognised by the same antibody [21]. Based on these findings we were interested to analyse the corresponding binding epitope of this anti TS mAb 7/23, also to employ the antibody in further applications including characterisation of domain swapped TconTS.

To analyse and possibly locate the binding epitope, different truncated recombinant TconTS1a fragments were generated. A schematic overview of these constructs is provided in Figure 4 A. In total, five TconTS1-CD (Figure 4 A, constructs A-E) and two TconTS1-LD (Figure 4 A, constructs F and G) fragments, varying in length, were cloned and expressed by *E. coli* Rosetta pLacI as described under Methods. Bacterial lysates containing truncated recombinant TconTS fragments were used for epitope mapping, employing Western blot analysis using anti *Strep*-tag and anti TS mAb 7/23.

Results of the corresponding Western blot analysis are illustrated in Figure 4 B-D. It can be seen

that none of the truncated recombinant TconTS1-CD constructs (A-E) were recognised by the anti TS mAb 7/23 (Figure 4 B). Wild type TconTS1a, comprising CD and LD, expressed by bacteria was detected by anti TS mAb 7/23, but with significantly lower signal intensity compared to the control TconTS1a expressed by fibroblasts (Figure 4 B). Interestingly, signal intensities of the truncated constructs as well as wild type TconTS1a expressed by bacteria, were comparable when using the anti *Strep*-tag antibody, indicating that the same amount of each protein was used in both Western blots. In contrast, when analysing lysates of the two recombinant TconTS1a-LD constructs A and B, both proteins were detected by the anti TS mAb 7/23 (Figure 4 C), clearly demonstrating the binding epitope to be localised in the LD. In addition, a similar observation regarding the lower signal intensity for wild type TconTS1a expressed by bacteria compared to the eukaryotic construct (Figure 4 C), was obtained as discussed above. Furthermore, when testing the other recombinant, bacterial wild type TconTS2, TconTS3 and TconTS4 as well as their truncated recombinant LDs (Figure 4 D), none of them were recognised by the anti TS mAb 7/23. Recombinant proteins were generated according to TconTS1a, described under Methods. Surprisingly also TconTS2 constructs (wild type and LD) expressed by bacteria, were not detected by the antibody, which appears in contrast to previous findings, when using TconTS2 expressed by fibroblasts [21].

Since one of the main differences between protein expression in CHO-Lec1 and *E.coli* Rosetta pLac1 represents the missing glycosylation, in case of the latter. Eukaryotic TconTS1a and TconTS2 were treated with peptide-*N*-glycosidase F (pNGaseF) to investigate the influence on antibody binding. Interestingly, it was found that *N*-deglycosylation of TconTS1a and TconTS2 under native conditions drastically reduced the binding affinity of anti TS mAb 7/23 to both enzymes, indicated by the weak bands at 110 kDa (Figure 4 E). Furthermore, pNGaseF treatment under denaturing conditions completely eliminated the binding of anti TS mAb 7/23 to TconTS1a as well as to TconTS2 (Figure 4 E). These observations indicate the *N*-glycan dependency of the anti TS mAb 7/23 binding to both, TconTS1a and TconTS2.

Biochemical characterisation of domain swapped TconTS

Selected recombinant domain swapped TconTS constructs, expressed by bacteria and fibroblasts were purified in sufficient amounts and analysed employing SDS-PAGE and Western blot analysis

as described under Methods. Therefore anti TS mAb 7/23 was used, which can be utilised to differentiate mutated and chimeric TconTS constructs expressed by CHO-Lec1, since the binding epitope of anti TS mAb 7/23 is located in the LD of TconTS1 and TconTS2. For example, TconTS1a* and TconTS2* as well as the corresponding wild type TconTS are detectable using the anti TS mAb 7/23, whereas domain swapped TconTS1a/TS3 and TconTS2/TS4 are not. Figure 3 D illustrates the results of the corresponding Western blot analysis. Anti TS mAb 7/23 and *Strep*-tag antibodies were used separately for detection, although the protein amount was keept constant in both blots. As expected, TconTS1a and TconTS2 wild type (Figure 3 D) were clearly detected using both antibodies. In addition, the domain swapped TconTS1a/TS3 was only detectable using the anti *Strep*-tag antibody but not with the anti TS mAb 7/23 (Figure 3 D). Surprisingly, compared to TconTS1a and TconTS2, binding of 7/23 to TconTS1a* and TconTS2* was drastically reduced and not even detectable for TconTS2* under the conditions used (Figure 3 D). Nevertheless, results clearly demonstrate that the anti TS mAb 7/23 can be used as a specific tool to differentiate domain swapped TconTS enzymes on protein level. It will be of significant importance to further analyse the reasons for the drastic reduction of anti TS mAb 7/23 binding to TconTS1* and TconTS2*.

Determination of the influence of TconTS-LD on enzyme activity employing HPAEC-PAD analysis

TconTS1 and TconTS2 were described as the two most active enzymes with specific activities 100- to 1000 fold higher compared to those observed for TconTS3 and TconTS4, respectively [20,21]. Therefore, we wanted to investigate a potential change in specific catalytic activities when exchanging the LD of a highly active (TconTS1 and TconTS2) with that of a lower active TconTS (TconTS3 and TconTS4). Recombinant domain swapped TconTS1a/TS3 (Table 4) was selected and expressed by *E. coli* Rosetta pLacI and CHO-Lec1 fibroblasts. To determine specific enzymatic activities, fetuin and lactose were used as Sia donor and acceptor substrates in TS reactions using standard conditions described under Methods. Reaction products were isolated and characterised employing high performance anion exchange chromatography with pulsed

amperometric detection (HPAEC-PAD), a method optimised for the analysis of carbohydrates (described under Methods).

In the first instance, we wanted to compare the specific catalytic activities of wild type TconTS expressed by bacteria or fibroblasts. Therefore, TconTS1a was expressed by *E. coli* Rosetta pLacI and subsequently purified as described under Methods. Different amounts of bacterial, recombinant TconTS1a were incubated with fetuin and lactose as Sia donor and acceptor substrates under standard conditions for 30 min at 37°C. Reaction products 3-sialyllactose (3'SL) and Neu5Ac were quantified and data analysis was done as described under Methods (results are illustrated in Figure 5 A - D).

3'SL concentrations determined (up to 8 µM 3'SL), using increasing amounts of enzyme, suggested that the equilibrium is still not reached, even when using 3 µg of bacterial TconTS1a under standard conditions for 30 min of incubation (Figure 5 A). In addition, only relatively low concentrations of free Neu5Ac (70 nM) were detected (Figure 5 B). Along this line, 1 µg of TconTS1a was used in the following TS activity time series under standard conditions, using different reaction times ranging from 0 to 24 h. Results are presented in Figure 5 C and D. It can be seen that after 24 h of incubation using 1 µg enzyme the concentration of 3'SL (approximately 80 µM 3'SL) is not at equilibrium, which possibly would be reached at about 100 µM 3'SL (Figure 5 C). In addition, clear sialidase activity was observed, indicated by the formation of Neu5Ac reaching 3.8 µM after 24 h of incubation (Figure 5 D). However, compared to specific catalytic activities observed for recombinant TconTS1 expressed by CHO-Lec1 cells [20,21], these findings clearly demonstrated approximately 500 to 1700 fold lower specific enzyme activities of recombinant TconTS1a expressed by *E. coli* Rosetta pLacI (Figure 5 E and Table 5). Nevertheless, both recombinant TconTS1a enzymes, expressed by bacteria or fibroblasts, exhibited similar overall enzymatic behaviour with higher Sia transfer to sialidase activity rates (Figure 5 E and Table 5).

In following experiments, 1 µg of bacterial expressed and purified TconTS1a, TconTS3 and domain swapped TconTS1a/TS3, were incubated with fetuin and lactose as Sia donor and acceptor substrates for 24 h under standard conditions (described under Methods).

Wild type TconTS1a expressed by bacteria produced around 80 µM 3'SL and 4.24 µM Neu5Ac

(Figure 6 A-C). In contrast, wild type TconTS3 from *E. coli* generated only 1.2 µM 3'SL and 2.32 µM Neu5Ac, demonstrating a higher sialidase then Sia transfer activity, which is in sharp contrast to TconTS3 expressed by CHO-Lec1, which did not show detectable sialidase activity [21]. Interestingly, the domain swapped TconTS1a/TS3 showed contrary overall enzyme behaviour compared to wild type TconTS1a, indicated by the formation of 2.5 µM 3'SL and 4.5 µM Neu5Ac (Figure 6 A-C). Hence, demonstrating a drastically decrease in Sia transfer activity, by a factor of 32 compared to wild type TconTS1a, but two fold higher to wild type TconTS3 (Figure 6 C). Furthermore, a two fold higher sialidase activity for TconTS1a/TS3 (4.5 µM Neu5Ac) was observed compared to that found for wild type TconTS3 (2.3 µM Neu5Ac), but similar to that of wild type TconTS1a (4.2 µM Neu5Ac). In summary, it can be concluded that domain swapped TconTS1a/TS3 expressed by bacteria, is double as active as TconTS3 or a modified TconTS1a-analogue with a drastically reduced Sia transfer activity.

We also wanted to investigate the enzyme activity of TconTS1a/TS3 expressed by CHO-Lec1 cells, besides the bacterial domain swapped TconTS1a/TS3. For TS reactions, 50 ng recombinant TconTS1a/TS3 expressed and purified from CHO-Lec1 were incubated with fetuin and lactose as Sia donor and acceptor substrates for 30 min under standard conditions (described under Methods). The set up for TS reactions of wild type TconTS1a and mutated TconTS1a* expressed by CHO-Lec1 cells was performed according to that for TconTS1a/TS3. Unfortunately, TconTS2/TS4 could not be prepared in sufficient amounts and purity for reasonable activity analysis (see previous chapters). However, mutated TconTS2* was expressed and purified in adequate concentrations and purity and was used to investigate a possible influence of the two inserted amino acid residues (Tyr-Val) on enzymatic activities. Figure 6 D-F summarise the experimental results from HPAEC-PAD analysis of the reaction products 3'SL and Neu5Ac.

82 and 54 µM 3'SL (Figure 6 D and F) as well as 2.64 and 1.7 µM Neu5Ac (Figure 6 E and F) were produced by wild type TconTS1a and TconTS2, respectively. Surprisingly, mutated TconTS1a* showed a drastically reduced production of about 31 µM 3'SL compared to wild type TconTS1a (82 µM 3'SL). On the other hand, 1.48 µM Neu5Ac were generated by mutated TconTS1a*, which is similar to the 1.7 µM Neu5Ac produced by wild type TconTS2 (Figure 6 E and F). Surprisingly, in addition to that, domain swapped TconTS1a/TS3 generated only a slightly lower 3'SL concentration

of about 72 µM compared to wild type TconTS1a (82 µM 3'SL), but a clearly reduced Neu5Ac concentration of 0.47 µM (Figure 6 D-F). Thus demonstrating a reduction in Sia transfer activity by only 12 % but a reduction in sialidase activity by even 82 %, compared to wild type TconTS1a (Figure 6 D-F).

In addition, similar 3'SL concentrations were found for wild type TconTS2 (54 µM 3SL) and mutated TconTS2* (55 µM 3'SL), demonstrating similar Sia transfer activities (Figure 6 D and F). However, drastic differences in Neu5Ac production were observed for wild type TconTS2 and mutated TconTS2*, whereas 1.7 µM Neu5Ac were generated by the former and 6.7 µM by the latter (Figure 6 E and F). In summary, both CHO-Lec1 expressed, mutated enzymes TconTS1a* and TconTS2* exhibited clear Sia transfer and sialidase activities, but significantly different from that of wild type TconTS1a and TconTS2, respectively under conditions used. Interestingly, domain swapped TconTS1a/TS3 showed a more similar Sia transfer (72 µM 3'SL) then sialidase activity (0.47 µM Neu5Ac) compared to wild type TconTS1a (82 µM 3'SL and 2.64 µM Neu5Ac).

In summary, taking all these enzyme activity data together, it can be concluded that generally TconTS enzymes expressed by CHO-Lec1 cells exhibit orders of magnitude higher overall enzymatic activities compared to those expressed by *E. coli*. Furthermore, sialidase activity is more pronounced for TconTS expressed by *E. coli* compared to enzymes expressed by CHO-Lec1, which exhibited clearly higher Sia transfer activities. Regarding the domain swapped TconTS1a/TS3, it can be suggested that the LD of TconTS3 in the chimeric enzyme TconTS1a/TS3 suppresses its sialidase activity when expressed by CHO-Lec1 cells. In contrast to that, the exact contrary effect was observed for the bacterial domain swapped TconTS1a/TS3. It was demonstrated, that the LD of TconTS3 maintain sialidase activity but suppress Sia transfer activity of recombinant TconTS1a/TS3, when expressed by *E. coli*.

In addition to the biochemical characterisation of domain swapped TconTS, also the influence of 1,4-β-mannotriose on the enzymatic activities of wild type TconTS1a and TconTS2 was investigated. Our previous findings have demonstrated that TconTS-LD is a carbohydrate binding domain with a specific affinity for oligomannose oligosaccharides, besides serval other oligogalactosyl oligosaccharides [23]. Hence it was hypothesised that due to its carbohydrate binding ability, the LD of TconTS may additionally bind the same substrate as its CD and therefore

potentially directly effecting the overall enzyme activity. To investigate this hypothesis, CHO-Lec1 expressed wild type TconTS1a and TconTS2 were incubated with fetuin and lactose as Sia donor and acceptor substrates under standard conditions for 30 min, in the presence and absence of 5 mM 1,4-β-mannotriose (described under Methods). Reaction products 3'SL and Neu5Ac were analysed and quantified as described under Methods. Indeed, a significant effect on enzymatic activity for both, TconTS1a and TconTS2, in the presents of 1,4-β-mannotriose could be observed (Figure 7 A-C). In the absence as well as in the presence of 1,4-β-mannotriose, no significant change in 3'SL production was observed neither for TconTS1a (66 µM and 65 µM) nor TconTS2 (60 µM and 61 µM). Interestingly, both enzymes showed a clear decreased Neu5Ac production in the presence of 1,4-β-mannotriose (TconTS1a: 0.45 µM, TconTS2: 0.32 µM) compared to the corresponding samples without the addition of 1,4-β-mannotriose (TconTS1a: 2.74 µM, TconTS2: 1.7 µM) after 30 min of incubation.

Discussion

Based on our findings, namely identifying TconTS-LD as a carbohydrate-binding domain with specific affinities to oligomannosyl oligosaccharides [23], we have hypothesised that the LD modulates TconTS catalytic activities, due to additional binding of TconTS-LD to the same substrate, influencing the overall binding affinities or catalytic turnover. To investigate this hypothesis, we established a novel strategy to simply and efficiently swap CDs and LDs from different TconTS. Chimeric, domain-swapped recombined TconTS were expressed by *E.coli* and CHO-Lec1 cells and subsequently purified to characterise enzymatic activities.

So far, only very minor attention has been payed to the lectin domain of TS. Smith and Eichinger described the expression and characterisation of TcruTS/TranSA hybrid proteins, exhibiting different Sia transfer and sialidase activities [35]. They found that the C-terminal Fn3 domain (fibronectin type III), named according to its structural relation to fibronectin type III, is not only required for expression of enzymatically active TcruTS and TranSA [1,36,37], but also influences the overall Sia transfer and sialidase activities [35]. Amino acid sequence alignments of a well characterised sialidase from *M. viridifaciens* [38] with TcruTS revealed that R572 and E578, which are known to be essential for galactose binding of *M. viridifaciens* sialidase [38], are well

conserved in TcruTS and TranSA [35]. Point mutation of one of these residues resulted in reduced sialidase activities for both enzymes and enhanced Sia transfer activity in TcruTS [35]. As a consequence of these findings Smith and Eichinger predicted both amino acid residues (R572 and E578 in TcruTS) to be involved in galactose binding of the acceptor and/or donor substrates, which would necessarily require an overall protein folding that brings the catalytic domain and the region containing the Arg and Glu of Fn3 domain (at least R and E) close together [35]. However, the resolved crystal structures of TcruTS [11] and TranSA [32] demonstrated that these two amino acid residues are located far away from the actual Sia binding pocket of the catalytic domain, and therefore, they are unlikely to be directly involved in substrate binding as proposed by Smith and Eichinger [35]. Nevertheless, even if these residues do not interact with the galactose moiety of the acceptor/donor substrate, R611 and E617 of TranSA (correspond to R572 and E578 in TcruTS) were found to form a intramolecular salt bridge at the C-terminus of the lectin domain [32] apparently indirectly influencing enzymatic activities, since its disruption was found to change Sia transfer and sialidase activities, respectively [35]. Interestingly, structural and amino acid sequence alignments of TcruTS (EMBL: AAA66352.1), TbruTS (EMBL: AAG32055.1), TranSA (EMBL: AAC95493.1) and TconTS revealed that this salt bridge is conserved among all TS used (Figure S1 A), indicating a possibly essential role in trypanosomal trans-sialidase activities.

Furthermore, our results for TconTS enzymatic activities using mixtures of two different TconTS enzymes revealed an altered overall enzymatic behaviour compared to that determined for each TconTS individually (T.T. Gbem, personal communication). Interestingly, none of the four TconTS1 through TconTS4 enzymes was able to efficiently release Sia from methylumbellyferone-Neu5Ac (MU-Neu5Ac) neither in the presence nor in the absence of a suitable acceptor substrate [21]. In contrast, if any other TconTS was added to either TconTS1 or TconTS2, cleavage of MU-Neu5Ac was clearly observed (T.T. Gbem, personal communication). This supports our hypothesis of a potential cooperative interaction of TconTSs, possibly mediated by the lectin domain, leading to modulation of substrate specificities.

Furthermore, it should be noted that besides the active TconTS described [20,21] three additional genes were identified, sharing only 20-30 % amino acid sequence identity [21]. They lack most of the conserved amino acid residues known to be involved in enzymatic activity and therefore, it has

been assumed that these genes coding for catalytically inactive TconTS, thus they were termed TconTS-like proteins [21]. Their function is not known but it has been proposed that they serve as cell adhesion proteins or may just be involved in cooperative interactions with active TconTS to modulate their enzymatic activities [21].

Structural insights into the contact site between CD and LD of TTS

Buschiazzo *et al.* first observed that the interface between TcruTS CD and LD is significantly larger (2550 $Å^2$) compared to other bacterial and viral sialidases (1300 – 1600 $Å^2$) [11]. Furthermore, Amaya *et al.* found that also the molecular surface of the contact site in TranSA is more extended (about 2600 Å) [32]. Along this line, we calculated the area of the molecular interfaces of TconTS1, TconTS2, TconTS3 and TconTS4 and observed area sizes analogous to those obtained for TcruTS and TranSA (Table 1). Slight difference in area size between TconTS and TcruTS as well as TranSA might be due to different settings used in *in silico* calculations, but it should be noted that the relative areas are very similar. In conclusion, in both publications [11,32] the authors predicted a relatively rigid overall TS core structure precluding a direct involvement of the LD in enzymatic catalysis, which is in agreement with our hypothesis that the LD is indirectly involved in Sia transfer activity instead, potentially by modulating the affinities of the enzymes for several donor/acceptor substrates [23]. In conclusion, the extended interfaces between CD and LD seems to be a typical feature for trypanosomal TS.

Therefore, we investigated the structural architecture at the contact site between TconTS CD and LD in more detail. Interestingly, our TconTS homology models revealed that the majority of amino acids localised at the contact site between CD and LD are well conserved among TconTS family members (Figure 1 D, Table 3), indicating a high pressure to maintain these during evolution of these TconTS genes [21]. When calculating the hydrogen bond network formed between residues at the interface between CD and LD, eleven to thirteen potential hydrogen bonds were observed for TconTS1 through TconTS4, which is similar to the number found for TranSA and in clear contrast to other sialidases, where only about half of this have been found [32]. Mainly two separate locations, A and B (Figure 1), comprise the majority of hydrogen bonds formed, evenly

distributed over both locations. Area A is located at the region where CD, LD and the α-helix are in close contact (Figure 1 A, B), whereas area B is located more closely to the Sia binding pocket of CD (Figure 1 A, C), opposite of A. We assume that the contacts at area B keep the catalytic grove of CD close to the hypothesised binding site of LD, thus providing a well defined orientation of both domains relative to each other and preventing even minor rotations around theoretical axis along the α-helix, which would induce a more drastic effect on more distant locations. Interactions in area A are predicted to stabilise the protein core conformation in a similar manner as described for B. Strikingly, amino acid residues involved in hydrogen bond network formation at both sites, A and B, are particularly well conserved (Table 3). For example Gln (18 in Table 3) in the LD, which is part of area B reaches deep into the CD and forms hydrogen bonds to a conserved Gly (4), to Lys (7), conserved in all TconTS except TconTS2, where it is replaced by Glu, which also forms a hydrogen bond, and finally to Arg (5), which is well conserved among African TS including TranSA (Figure 1 A and C, Table 2 and 3). In area A a similar situation was observed, where five essential amino acid residues, Arg (9), Trp (10), Asp (12), Leu (14) and Tyr (15), are mainly responsible for the distinct hydrogen bond network in that region (Figure 1 A and B, Table 2 and 3). Since all these amino acid residues are highly conserved in the TS family despite their different catalytic activities (e.g. of TranSA), it appears unlikely that this has a direct influence on the catalytic Sia transfer. Nevertheless, the proposed simultaneous binding of both domains to the same substrate requires such a stabilised conformation of the two domains as provided by the conserved hydrogen bonding networks, which thus keep the substrate in an appropriate orientation for a Sia transfer.

A novel strategy to efficiently swap catalytic and lectin domains of TS

Based on the fact that the majority of amino acid residues localised at the contact sites between CD and LD are well conserved in the TconTS family, we rationed that it might be possible to swap the domains of different TconTS to investigate the influence of LD on enzymatic activities. To achieve this we introduced a unique restriction site into the sequence coding for a hairpin loop, which is located right after the α-helix, connecting CD and LD. This appeared to be a good location to join CD and LD from different TS for several structural reasons: (I) This loop seems to be highly

flexible, since it is not resolved in the crystal structure of TcruTS or TranSA [11,39]. (II) The hairpin loop segment is not well conserved showing a high amino acid sequence diversity between all TconTS (Figure 2 A). (III) The loop is sterically distant from the catalytic pocket of CD and the predicted carbohydrate binding-site of LD. To further investigate the theoretical structure stability of the chimeric TconTS a structure homology model was generated *in silico,* using as example the domain swapped TconTS1a/TS3 comprising CD of TconTS1a and LD of TconTS3. Indeed, a homology model of TconTS1a/TS3 (data not shown) was obtained, in which CD and LD exhibited similar structure topology to CD of wild type TconTS1a and LD of wild type TconTS3, respectively. Interestingly, when investigating the interface between CD and LD of TconTS1a/TS3, 11 hydrogen bonds were observed formed by amino acid residues, which were predicted to be essential for conformation stability of the wild type TconTS as discussed above. These observations underline the possibility of such a rearrangement of TconTS genes, which would provide an explanation for the separate evolution of TconTS domains [21].

These considerations encouraged us to employ this strategy for the recombination of CDs and LDs from different active TconTS. Expression of recombinant domain swapped TconTS1a/TS3 by bacteria (*E. coli* Rosetta DE3 pLacI) and fibroblasts (CHO-Lec1) revealed sufficient amounts of soluble and most notably active enzyme, whereas catalytic activities observed for both recombinant proteins differed by orders of magnitude. TconTS1a/TS3 expressed by CHO-Lec1 cells showed 500 to 1800 fold higher Sia transfer and sialidase activities compared to that expressed by *E. coli* (Figure 6 A-F). One explanation for this effect might be the missing glycosylation of bacterial expressed proteins. In addition, a high amount of improperly folded TconTS1a/TS3 in preparations of bacterial enzymes cannot be excluded. Furthermore, successful bacterial expression of active CD and LD separately, was already performed and demonstrated the effective folding of each domain independent of the other [23].

Nevertheless, it can be concluded that in general employing the strategy established in this study has worked well to exchange LDs from different TconTS enzymes.

Also the domain swap of CDs and LDs from different TS family members, such as TcruTS, TbruTS, TranSA or *Trypanosoma vivax* TS (TvivTS) to further investigate the role of LD in enzymatic activity, might be of reasonable interest. Smith and Eichinger expressed several TcruTS/TranSA

hybrid proteins, varying in length, to identify the minimum amino acid sequence, which is required for Sia transfer activity [35]. Results revealed that Sia transfer activity was increased, ether when mutating the C-terminal located R642 or E648, which were suggested to be directly involved in substrate binding [35]. In addition, the relevance of these findings were supported by Amaya *et al.*, who demonstrated that both residues form a salt bridge in the LD of TcruTS [40]. Homology models and amino acid sequence alignments of TS calculated in this study showed that this crucial salt bridge is well conserved in the TS family (Figure S1 A) and furthermore, suggested that it remains intact when swapping TS domains, using the strategy presented here.

A shared oligosaccharide dependent epitope for the anti TS mAb 7/23 localised in the LD of TconTS1 and TconTS2

Tiralongo *et al.* generated an anti TS mAb 7/23 [12], which was later found to specifically bind both, recombinant TconTS1 and TconTS2, but not TconTS3 and TconTS4, [20,21]. Furthermore, also recombinant TcruTS and TbruTS were not detectable on western blots using the anti TS mAb 7/23 [12], indicating its specificity for TS from *T. congolense*. Here we partially identified the antibody binding epitope, using recombinant truncated TconTS1 proteins expressed by bacteria and varying in length (Figure 4 A). Interestingly, only expression products containing TconTS1-LD (Figure 4 C) were detectable in Western blot analysis, but none of the truncated TconTS comprising only parts of CD. These data suggested that the binding epitope is located somewhere in the LD of TconTS1. Surprisingly, compared to TconTS1 expressed by CHO-Lec1 cells, drastically reduced binding to bacterial protein was observed, even when applying high amounts of antigen in Western blots. This effect was not likely to be due to improper folding in the bacteria, since the anti TS mAb 7/23 bound well to TconTS1 denatured under reducing conditions, which consequently indicated a linear epitope not requiring appropriate folding of the protein for antibody recognition.

In a previous study we have demonstrated that oligomerisation of TconTS1a is mediated by binding to *N*-linked oligomannose glycans, since pNGaseF glycosidase treated TconTS1a reduced the degree of oligomerisation [23]. Surprisingly, these experiments revealed new insight in the binding epitope of the anti TS mAb 7/23. When employing the anti TS mAb 7/23 to detect

TconTS1a after deglycosylation with pNGaseF, the corresponding signal intensities drastically decreased (Figure 4 E). These surprising findings clearly demonstrated that the presence of a *N*-glycan is required for efficient binding of anti TS mAb 7/23 to its epitope. However, the fact that TconTS1a expressed by bacteria, which is not glycosylated, is still detectable with the anti TS mAb 7/23 indicates that the binding epitope comprises an amino acid stretch containing a *N*-glycosylation site, which is present in both, TconTS1a and TconTS2. The observation, that we could not detect binding of anti TS mAb 7/23 to bacterial TconTS2 might be explained with minor differences in the amino acid sequence in the epitopes of TconTS1a and TconTS2.

In summary, it can be concluded that the anti TS mAb 7/23 recognises an epitope, which is *N*-glycan dependent and localised in the lectin domain of TconTS1 and TconTS2. Thus, it can be expected to be a more exposed region in TconTS1-LD and TconTS2-LD, which (I) comprise an *N*-glycosylation site, which (II) is not present in TconTS3, TconTS4, TcruTS, TbruTS or TranSA. Indeed, when screening the TS amino acid sequence alignments, we found a short stretch complying with both these criteria (Figure S1 A). Furthermore, in structure homology models of TconTS1a and TconTS2 this regions are relatively exposed to solvent (Figure S1 B), thus suggesting that this region is at least part of the epitope and that the Asn residue in that sequence (N657 in TconTS1, N640 in TconTS2) is glycosylated. However, in order to confirm the proposed binding epitope of anti TS mAb 7/23, it will be necessary to change this region by mutagenesis experiments, including in particular mutants eliminating the glycosylation motif of N657 and N640 in TconTS1 and TconTS2, respectively.

Interestingly, Tiralongo *et al.* demonstrated the ability of anti TS mAb 7/23 to immunoprecipitate 75 % of TS activity present in a native TS-containing sample from *T. congolense* culture supernatants [12]. Thus, the majority of the TS activity in this TS-containing sample is due to TconTS1 and possibly TconTS2, but the presence of other TconTS is well possible [21]. Since the anti TS mAb 7/23 has only week inhibitory activity (20%), it will be interesting to investigate its influence on recombinant TconTS1 and TconTS2 expressed by CHO-Lec1 cells. Along this line, it will be also interesting to determine enzymatic activities of mutants where the *N*-glycosylation sites at N657 (TconTS1) or N640 (TconTS2) is changed, to investigate the influence of that *N*-glycosylation site on catalytic activities.

Specific enzymatic activities of domain swapped TconTS demonstrated the complexity of catalytic mechanisms

The specific enzymatic activities of TconTS1a and TconTS3 expressed by *E. coli* was orders of magnitudes lower (500 to 1800 fold) compared to previously reported activities of these enzymes expressed by CHO-Lec1 cells [20,21]. Nevertheless, for bacterial TconTS1 the ratio of Sia transfer to sialidase activities were similar to those reported for the eukaryotic protein [20]. Also the ratio for specific Sia transfer activities, TconTS1a being about 100-fold more active than TconTS3, was similar to that for enzymes from CHO-Lec1 cells [21]. However, the opposite ratio for Sia transfer to sialidase activity was obtained for TconTS3 expressed by *E. coli*, which surprisingly had a higher sialidase than TS activity (Table 5). The drastically lower activities of bacterial TconTS compared to that expressed by CHO-Lec1 cells is most likely due to either inefficient protein folding or lack of glycosylation, which might influence enzymatic activities, but to our knowledge has never been investigated in detail. The lack of glycosylation may have a bigger effect on TconTS compared to TcruTS, since only two *N*-glycosylation sites were predicted for TcruTS compared to 8-9 potential sites for TconTS1 through TconTS4, in comparison to TbruTS with 4 and TranSA with 5 predicted *N*-glycosylation sites [23]. However, it should be noted that only very little is known about glycosylation of native TS so far. TbruTS from procyclic trypanosomes has been found to be *N*-glycosylated with high-mannose type oligosaccharides [3], leading to the assumption of a similar situation for TconTS. Both *T. brucei* and *T. congolense* bloodstream (mammalian host) and procyclic (tsetse vector) forms contain high-mannose type *N*-glycans on the parasites surface, linked to amino acid residues of the polypeptide or as part of the GPI (glycosylphosphatidylinositol) anchor [28-31,41-43].

In a next step, enzymatic activities were determined for domain swapped TconTS1a/TS3 expressed by *E. coli*. According to the assumption that the LD of TconTS indirectly influences enzymatic activities, it was expected that LD of the less active TconTS3 may decrease the enzymatic activities, if attached to the CD of a highly active TconTS. Indeed, an over 30-fold lower 3'SL production (2.5 µM 3'SL) was observed for chimeric TconTS1a/TS3 compared to wild type

TconTS1a (80 µM 3'SL), but it was still two-fold higher compared to wild type TconTS3 (1.2 µM 3'SL) (Figure 6 C). Surprisingly, the situation was different for the sialidase assays, where similar specific sialidase activities (Figure 6 C) were observed for TconTS1a/TS3 (4.5 µM Neu5Ac) and wild type TconTS1a (4.2 µM Neu5Ac), whereas only the half (2.3 µM Neu5Ac) was generated by wild type TconTS3. Summarising these findings, it can be concluded that the exchange of TconTS1a-LD with that from TconTS3 lead to an active chimeric TconTS1a/TS3 expressed by bacteria, which can be described either as a modified wild type TconTS1a variant with a drastically lower Sia transfer, but equal sialidase activity, or as a twice as active wild type TconTS3. The impact of the LD exchange in TconTS1a/TS3 on enzymatic activities becomes even more apparent, when comparing the ratios of Sia transfer/sialidase activities of wild type TconTS1a and domain swapped TconTS1a/TS3 (Table 5). A ratio of 0.55 for TconTS1a/TS3 illustrated that the ratio is clearly shifted from primarily Sia transfer in TconTS1a (ratio 109) to primarily sialidase activity obtained for the bacterial chimeric enzyme. Thus, it can be concluded that, if LD of TconTS3 is attached to CD of TconTS1a the Sia transfer/sialidase ratio is decreased by the factor of 200, or in other words TconTS3-LD appears to suppress the Sia transfer activity in TconTS1a/TS3, while retaining its sialidase activity.

In the next step enzymatic activities of the domain swapped TconTS1a/TS3 expressed by CHO-Lec1 cells were determined. Interestingly, a completely different situation regarding the overall catalytic activities were observed. For example, similar Sia transfer activities for TconTS1a/TS3 (2409 nmol/(min x mg TS)) and TconTS1a (2732 nmol/(min x mg TS)) were determined, whereas sialidase activity of domain swapped TconTS1a/TS3 was reduced by 82% relative to that of wild type TconTS1a (Table 5). In summary, the domain swap of TconTS1a-CD and TconTS3-LD in CHO-Lec1 cells significantly reduced the sialidase relative to the Sia transfer activity for TconTS1a/TconTS3 compared to wild type TconTS1a. Regarding the Sia transfer/sialidase activity ratio for wild type TconTS1a (31) and domain swapped TconTS1a/TS3 (153) expressed by CHO-Lec1 cells, it can be concluded that TconTS3-LD significantly suppressed sialidase activity in TconTS1a/TS3 by the factor of 5, relative to that of wild type TconTS1a. Interestingly, these findings are exactly contrary to those observed for bacterial expressed enzymes discussed above. Finally, it can be concluded that TconTS3-LD in TconTS1a/TconTS3 enhanced sialidase activity of

the chimeric enzyme when expressed by *E. coli* and suppressed sialidase activity when expressed by CHO-Lec1 cells. Actually, this completely matches the situation observed for TconTS3 wild type expressed by bacteria, predominantly a sialidase (Table 5) compared to TconTS3 produced in fibroblasts, for which basically no sialidase activity could be detected [21]. It is tempting to assign this to the lack of glycosylation, predominantly in the LD of TconTS3, since no such effect was observed for TconTS1a. However, the associated mechanisms still remain unclear. It should be also noted that a contribution of misfolding of the bacterial and/or eukaryotic expressed chimeric TconTS1a/TS3 cannot be completely excluded, and may also contribute to the different activities determined.

In addition to these findings, an interesting effect was observed for the enzymatic properties of TconTS1a* mutant expressed by CHO-Lec1 cells, whose only modification is the addition of two amino acids (Tyr-Val) in the hair-pin loop after the domain connecting α-helix. In contrast to our expectations, both, Sia transfer and sialidase activities were reduced by 62 % and 44 % respectively, compared to those of wild type TconTS1a (Table 5). Obviously, this is not a general consequence of loop extension, since no such effect was observed for TconTS2*. In contrast, in that case the Sia transfer of TconTS2* was not changed, but its sialidase activity was fourfold higher. A possible reason for this unexpected loss of general enzymatic activities of TconTS1a* might be an effect on the glycosylation of N657, which is in close sterical vicinity to the crucial salt-bridge between Arg and Glu in TconTS-LD mentioned above and, possibly more important, close to the hairpin-loop in which the two additional amino acids were introduced (Figure S1 A and B, Figure 2).

In this context it is important to note that N657 is the predicted *N*-glycosylation site in TconTS1a possibly contributing to the epitope of anti TS mAb 7/23. Western blot analysis of wild type TconTS1a, mutated TconTS1a* and chimeric TconTS1a/TS3 expressed by CHO-Lec1 cells revealed that anti TS mAb 7/23 binds to both, TconTS1a as well as TconTS1a*, but not to TconTS1a/TS3 as expected, since the two former comprise the predicted epitope located in their LD (Figure 3 D). However, the signal intensity for TconTS1a* was clearly decreased compared to that of wild type TconTS1a (Figure 3 D). A reasonable explanation for this surprising observation is a less effective or missing *N*-glycosylation of N657. Simple concentration differences can be

excluded, since similar signal intensities for wild type TconTS1a and mutated TconTS1a* were observed, when using anti *Strep*-tag mAb (Figure 3 D). Similar results were obtained for anti TS mAb 7/23 binding to wild type TconTS2 and mutated TconTS2*, where the effect was even more pronounced, as no signal was detected in case of mutated TconTS2*, when using anti TS mAb 7/23 in Western blots (Figure 3 D). It is in complete agreement with our results with enzymatically deglycosylated proteins discussed above, that the decrease or complete missing of mAb 7/23 binding is most likely due to a lack of *N*-glycosylation in the predicted binding epitope.

In summary it can be concluded that introduction of the two amino acids Tyr-Val may change the conformation of the hairpin-loop resulting in a sterically more restricted conformation of that region, which potentially prevents efficient *N*-glycosylation of N657. In conclusion, we hypothesise that differences in enzymatic activities of mutated TconTS1a* and wild type TconTS1a are due to changes in the *N*-glycosylation pattern of the enzymes.

I this context, it should be noted here that *N*-glycan-dependent oligomerisation of TconTS1 has been demonstrated [23], which might also have an impact on enzymatic activities due to TconTS cluster formation and consequently enhanced substrate binding as discussed for MU-Neu5Ac using mixtures of different recombinant TconTS. However, the higher Sia transfer activity of domain swapped TconTS1a/TS3 expressed by CHO-Lec1 compared to mutated TconTS1a* might be due to the *N*-glycosylation site N667 in TconTS3-LD. Interestingly, from amino acid sequence alignments of TconTSs (Figure S1 A) it can be seen that the predicted *N*-glycosylation site N667 in TconTS3-LD is just 16 amino acids further C-terminally located than N657 of TconTS1a (Figure S1 A). Therefore, we propose that in TconTS1a/TS3 N667 possibly adopts a similar function as N657 in TconTS1a, but is slightly less efficient, which would explain its maintained Sia transfer activity (Table 5). Furthermore, we assume that a possible *N*-glycosylation of N667 will not be effected by the extended hair-pin loop of TconTS1a/TS3, since it is more distantly located from the insertion.

Apparently, for TconTS2 the situation is different in some details, since the introduction of Tyr-Val as in TconTS2* enhanced sialidase activity, whereas Sia transfer is not effected (Figure 6). Similar to TconTS1a*, this might be explained by modulation of *N*-glycosylation, in this case of N658 in TconTS2 (Figure S1 A). This *N*-glycosylation site corresponds to N667 in TconTS3 and is only conserved in this two TconTS enzymes (Figure S1 A). Similar to the effect of the maintained Sia

transfer activity of chimeric TconTS1a/TS3 compared to wild type TconTS1a, we hypothesise that N658 in TconTS2* adopts the function of N640 (corresponding to N657 in TconTS1a), which was proposed not to be glycosylated in TconTS2*, a situation which lead to a reduction in overall enzymatic activity of TconTS1a*. Additionally, considering Sia transfer/sialidase activity ratios of mutated TconTS1a* and TconTS2* compared to the corresponding wild type TconTS1a and TconTS2 it can be concluded that extending the loop by Tyr-Val shifted the transfer/sialidase ratio in favour of sialidase activity (Table 5).

The results discussed so far suggest that *N*-glycans play important roles in the modulation of TconTS activities by the LD. Since at least the LD of TconTS2 and TconTS1a interact with high-mannose *N*-glycans, we investigated whether 1,4-β-mannotriose, the binding determinant of these LD, can influence the activities of TconTS1a and TconTS2 produced in CHO-Lec1 cells. Our results clearly demonstrated that in the presence of 5 mM 1,4-β-mannotriose sialidase activities of both, TconTS1a and TconTS2, was reduced by the factor of 6. Obviously, a concentration dependency of 1,4-β-mannotriose on enzymatic activities should be determined, to evaluate this inhibitory effect further. Interestingly, Sia transfer activities for both enzymes remained unchanged by the presence of 5 mM 1,4-β-mannotriose (Figure 7 A-C). The effect of 1,4-β-mannotriose on enzyme activities becomes even more apparent, when considering Sia transfer/sialidase activity ratios (Table 6). Ratios for TconTS1a (24) and TconTS2 (34) were shifted in favour to the Sia transfer activity site by the factor of 6 (144 for TconTS1a and 193 for TconTS2), when 1,4-β-mannotriose was present in enzyme reactions (Table 6). It should be noted that these ratios are very close to that of domain swapped TconTS1a/TS3. Therefore, we postulate that TconTS-LD activities are influenced by *N*-glycosylation and the presence of 1,4-β-mannotriose, and that this effect is also responsible for the decrease of sialidase relative to Sia transfer activity observed for chimeric TconTS1a/TS3, if produced in CHO-Lec1 cells, as well as for TconTS3 wild type, which has a much lower sialidase activity, if expressed by fibroblasts compared to the bacterial protein.

Summarising these results, the influence of LD and probably *N*-glycosylation on TconTS enzymatic activities was demonstrated in this study, indicating a more complexity of catalytic mechanisms as supposed so far. However, the strategy to swap CD and LD of TconTS presented here, can be further optimised, for example by mutating appropriate amino acids present in the hairpin-loop

instead of adding two additional residues. This approach might prevent sterical demanding structure of that region, which appears to result in impaired *N*-glycosylation.

Nevertheless, it would be also necessary to generate other domain swapped TconTS constructs for a better understanding of the interplay between CD and LD of trypanosomal trans-sialidases. Also the proposed glycosylation dependency of enzymatic activities has to be further investigated employing completely degylcosylated TconTS expressed by CHO-Lec1. In conclusion, these new insights into enzymatic properties of TS have opened new perspectives in understanding the mechanisms of these interesting enzyme reactions and finally may support the development of new strategies using TS as potential target in efforts towards the control of African trypanosomiasis.

Methods

Materials

Unless stated otherwise, all chemicals and reagents used in this study were cell culture and analytical grade. Recombinant pNGaseF endoglycosidase was from New England Biolabs, United Kingdom. *Pfu* and *Taq* DNA polymerase, *Eco*105I*, Hind*III, *Bam*HI, *Sal*I and *Spe*I Fast Digest restriction enzymes, T4-DNA ligase, isopropyl-β-D-1-thiogalactopyranoside (IPTG), Dithiothreitol (DTT), Coomassie Brilliant Blue (Page Blue), protein molecular weight marker (PageRuler), GeneJET DNA Gel Extraction Kit, BCA Protein Assay Kit, enhanced chemiluminescence system (ECL-Kit), Luria Broth (LB) microbial growth medium, were from Thermo Scientific, Germany. Biozym LE Agarose was from Biozyme Scientific, Germany. StrepTactin Sepharose, purification buffers and anti *Strep*-tag rabbit polyclonal antibody were from IBA, Germany. β-D-galactosyl-(1-4)-α-D-glucose (4α-lactose), *N*-acetyl-neuraminic acid (Neu5Ac), 3'sialyllactose (3'SL), β-D-glucopyranuronic acid (glucuronic acid), lyophilised fetuin from fetal calf serum, polyethylene glycol sorbitan monolaurate (TWEEN 20), Ex-cell CD CHO media, PEI (Polyethylenimin) transfection reagent were from Sigma-Aldrich, Germany. Hygromycin and gentamycin were purchased from PAA, Austria. 1-4β-D-mannotriose was from Megazyme, Ireland. Ultrafiltration units Vivacell and Vivaspin6 were from Sartorius, Germany. X-ray film were purchased from GE Healthcare, Sweden.

Protino Ni-NTA Agarose and NucleoBond Midi Plasmid DNA Purification Kit were from Macherey-Nagel, Germany. Polyvinylidenedifluoride (PVDF) membranes were from Millipore, Germany. 96-well transparent microtitre plate were from Sarstedt, Germany. 6 mL gravity flow columns were from Biorad, Germany.

Cloning and expression of recombinant, domain swapped TconTS

Cloning of TconTS into modified pET28aMBP bacterial expression vector

DNA sequences encoding for TconTS1 (including truncated CD or LD containing variants, Figure 4 A-G), TconTS2, TconTS3 and TconTS4 as well as their LDs were amplified from modified pDEF vector [20,21], using the corresponding set of sense and reverse primer listed in Table S1. The resulting PCR products were subcloned into modified pET28aMBP vector [23] via *Hind*III and *Bam*HI (TconTS1, TconTS2 and TconTS3) or *Sal*I and *Bam*HI (TconTS4) following instructions of the manufacturers.

Introduction of Eco105I restriction site into TconTS

To insert the *Eco*105I endonuclease restriction site into the appropriate location at the hair-pin loop following the domain-connecting α-helix of TconTS, sense and reverse primer were designed annealing at this target location and comprising the *Eco*105I site (Table S1). DNA sequences coding for TconTS-CD-α-helix with the *Eco*105I site attached at the 5'-end were amplified using *Hind*III (TconTS1, TconTS2 and TconTS3) or *Sal*I (TconTS4) sense primer in combination with the corresponding *Eco*105I reverse primer (Table S1). In addition, DNA sequences encoding the corresponding TconTS-LD sequence with the *Eco*105I site attached to the 3'-end were amplified using *Bam*HI (TconTS1 through TconTS4) reverse and the appropriate *Eco*105I sense primer (Table S1). Both PCR products were digested using *Eco*105I Fast Digest restriction enzyme (Thermo Scientific, Germany), purified using GeneJET DNA Gel Extraction Kit (Thermo Scientific, Germany) and blunt-end ligated using T4-DNA ligase, following instructions of the manufacturers. Religated TconTS DNA sequences comprising the *Eco*105I site were cloned into modified pET28aMBP expression vector using *Bam*HI (TconTS1, TconTS2 and TconTS3) or *Sal*I (TconTS4) and *Hind*III restriction enzymes according to manufacturers instructions. All sequences and insertions were confirmed by DNA sequencing at the Max Planck Institute for Marine Microbiology,

Bremen, Germany.

Recombination of CDs and LDs from different TconTS

To generate domain swapped TconTS constructs, corresponding pET28aMBP plasmids encoding mutated TconTS enzymes were digested with *Eco*105I and *Hind*III Fast Digest restriction enzymes (Thermo Scientific, Germany) to isolate the LD, which was subsequently cloned into a *Eco*105I and *Hind*III digested pET28aMBP plasmid coding for the CD of a TconTS variant different from that of the isolated LD, following the manufacturers instructions.

For expression of secreted TconTS constructs in mammalian fibroblasts, DNA sequences coding for mutated TconTS, only comprising the *Eco*105I endonuclease restriction site, and domain swapped TconTS constructs were subcloned into the modified pDEF expression vector using *Spe*I and *Hind*III restriction sites as described previously [20].

Purification of recombinant TconTS expressed by E. coli Rosetta (DE3) pLacI or CHO-Lec1 fibroblasts

Recombinant TconTS constructs were expressed by CHO-Lec1 fibroblasts or *E. coli* Rosetta (DE3) pLacI and subsequently purified as described previously [20,23].

In brief, for expression of TconTS constructs by *E. coli* Rosetta (DE3) pLacI, colonies freshly transformed with pET28aMBP plasmid, encoding the TconTS construct, were used for an overnight culture in Luria Broth (LB) medium containing 50 µg/ mL kanamycin, incubated at 37°C and 240 rpm shaking. For large scale, 1 L of LB medium containing 50 µg/ mL kanamycin was inoculated with 2 mL of the overnight culture and grown at 37°C and 240 rpm until an optical density of 0.5 at 600 nm was reached. Induction was done using isopropyl-β-D-1-thiogalactopyranoside (IPTG, 0.1 mM final concentration) for 30 min at 37°C followed by incubation at 4°C and 240 rpm for additional 14 h. *E. coli* Rosetta (DE3) pLacI cultures expressing TconTS-LDs were induced with IPTG (0.1 mM final concentration) for 2 h at 37°C and 240 rpm. Purification of recombinant TconTS constructs was done as described previously employing double affinity chromatography using Ni-NTA and *Strep*-tag chromatography [23]. Purified proteins were characterised by SDS-PAGE and Western blot analysis and quantified using BCA assay according to instructions of the manufacturers.

Recombinant TconTS constructs expressed by CHO-Lec1 fibroblasts were purified employing *Strep*-tag chromatography and characterised by SDS-PAGE and Western blot analysis as described previously [20] and subsequently quantified by BCA assay.

Trans-sialidase reactions of recombinant TconTS constructs expressed by *E. coli* and fibroblasts

Purified recombinant TconTS enzymes were assayed for Sia transfer and sialidase activities using fetuin and lactose as Sia donor and acceptor substrates as described previously [20]. In general, TconTS reactions in 50 µL reaction volume containing 10 mM phosphate buffer pH 7.4, the appropriate amount of recombinant TconTS enzyme (50 ng of TconTS expressed by CHO-Lec1 or 1 µg of enzyme expressed by *E. coli*) as well as fetuin (100 µg corresponding to 600 µM fetuin bound Sia) and lactose (2 mM final concentration) as Sia donor and acceptor substrates were incubated at 37°C for the times indicated. To determine the influence of 1,4-β-mannotriose on enzyme activities of recombinant TconTS1a and TconTS2 expressed by CHO-Lec1 fibroblasts, 0.25 µmol (5 mM final concentration) of the trisacchride were additionally added to the reaction mix described above. Reaction products 3'SL and Neu5Ac were quantified from chromatograms obtained, employing **h**igh **p**erformance **a**nion **e**xchange **c**hromatography with **p**ulsed **a**mperometric **d**etection (HPAEC-PAD) utilising a Dionex DX600 system in combination with a CarboPac PA100 column (Dionex/ Thermo Scientific, Germany) as described previously [20].

Deglycosylation of TconTS using pNGaseF endogylcosidase

Purified TconTS expressed by CHO-Lec1 fibroblasts were deglycosylated using pNGaseF endoglycosidase under native and denaturing conditions according to instructions of the manufacturers (New England Biolabs, United Kingdom). In brief, for degylcosylation under denaturing conditions, 10 µg of TconTS enzyme in 20 µL of denaturing buffer (0.5 % SDS, 40 mM DTT) were incubated at 95°C for 10 min. After incubation, NP40 and Na_2P_4O were added to final concentrations of 1 % (v/v) and 40 mM, pH 7.5, respectively. Finally, 2 units of pNGaseF endoglycosidase were added and the reaction mix was incubated at 37°C for 4 h. Deglycosylation

reactions under native conditions were set up according to that under denaturing conditions but without incubation in denaturing buffer at 95°C.

SDS-PAGE and Western Blot analysis

Protein samples were separated employing SDS-PAGE as described previously [44] using a MiniProtean III electrophorese Unit (Bio-Rad, Germany) and stained with Coomassie Brilliant Blue (Thermo Scientific, Germany).

Western blot analysis were performed using primary rabbit anti *Strep*-tag and mouse anti TS mAb 7/23 for detection of recombinant TconTS as previously described [20,23]. After blotting, membranes were developed using enhanced chemiluminescence system (ECL-Kit, Thermo Scientific, Germany) and X-ray film (GE Healthcare, Sweden) according to instructions of the manufacturers.

Homology Modelling and *in silico* calculations

Homology models of TconTS-LD containing and lacking the α-helix were calculated employing the molecular modelling software Yasara 13.3.26 [45-50] as previously described [20]. In brief, crystal structure of *T. cruzi* trans-sialidase [11] was used as a template structure (PDB: 3b69) for calculating the models. Yasara *homology modeling* module were modified manually from the default settings of the program: Modeling speed: slow, PsiBLASTs: 6, EValue Max: 0.5, Templates total: 1, Templates SameSeq: 1, OligoState: 4, alignments: 15, LoopSamples: 50, TermExtension:10. The molecular surface was calculated using the *surface module* of Yasara *Structure* with the following parameters: Force field: AMBER96, Algorithm used to calculate molecular surface: numeric, Radius of water probe: 1.4 Å, Grid solution: 3.

Amino acid sequences alignments of TTS were performed employing the *Geneious Alignment* module of the software Geneious 5.5.5, using Blosum62 Cost Matrix [51] with gap openings and extension 10 and 0.1 respectively. Adaptations and modifications were made using the same software. Increasing darkness of sheds indicates increasing number of identical amino acid residues at each position (black: 100 %; dark grey: 80 to 100 %; light grey 60 to 80 %; white: less

then 60 % similarity). Numbers on top of each sequence indicate the corresponding residue number for the appropriate TTS sequence.

Author contributions

MW cloned, expressed and purified TconTS constructs, performed HPAEC-PAD analysis of reaction products for all TconTS enzyme reactions, calculated TconTS homology models, performed *in silico* calculations and epitope mapping of anti TS mAb 7/23, evaluated experimental data and drafted the manuscript. TTG conducted part of experiments on TconTS enzyme activities. ND significantly contributed to TconTS domain swap experiments and HPAEC-PAD data analysis. SS supported cloning and cell culture work on domain swapped TconTS. SK designed and coordinated the study and supported drafting the manuscript.

References

1. Schenkman S, Chaves LB, Pontes de Carvalho LC, Eichinger D. A proteolytic fragment of *Trypanosoma cruzi* trans-sialidase lacking the carboxyl-terminal domain is active, monomeric, and generates antibodies that inhibit enzymatic activity. J Biol Chem. 1994;269: 7970–7975.

2. Clayton J. Chagas disease 101. Nature. 2010;465: S4–5. doi:10.1038/nature09220

3. Pontes de Carvalho LC, Tomlinson S, Vandekerckhove F, Bienen EJ, Clarkson AB, Jiang MS, et al. Characterization of a novel trans-sialidase of *Trypanosoma brucei* procyclic trypomastigotes and identification of procyclin as the main sialic acid acceptor. J Exp Med. 1993;177: 465–474.

4. Engstler M, Schauer R. Sialidases from African trypanosomes. Parasitol Today (Regul Ed). 1993;9: 222–225.

5. Truc P, Büscher P, Cuny G, Gonzatti MI, Jannin J, Joshi P, et al. Atypical human infections by animal trypanosomes. PLoS Negl Trop Dis. 2013;7: e2256. doi:10.1371/journal.pntd.0002256.t001

6. WHO. Control of Chagas disease. World Health Organization technical report series. 2002;: 1–109.

7. World Health Organization. Control and surveillance of human African trypanosomiasis. World Health Organ Tech Rep Ser. 2013;: 1–237.

8. Kamuanga M. Socio-economic and cultural factors in the research and control of trypanosomiasis. PAAT Technical and Scientific series. 2003;: 1–67.

9. Schenkman S, Jiang MS, Hart GW, Nussenzweig V. A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. Cell. 1991;65: 1117–1125.

10. Engstler M, Schauer R, Brun R. Distribution of developmentally regulated trans-sialidases in the Kinetoplastida and characterization of a shed trans-sialidase activity from procyclic *Trypanosoma congolense*. Acta Tropica. 1995;59: 117–129.

11. Buschiazzo A, Amaya MF, Cremona ML, Frasch AC, Alzari PM. The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. Mol Cell. 2002;10: 757–768.

12. Tiralongo E, Schrader S, Lange H, Lemke H, Tiralongo J, Schauer R. Two trans-sialidase forms with different sialic acid transfer and sialidase activities from *Trypanosoma congolense*. J Biol Chem. 2003;278: 23301–23310. doi:10.1074/jbc.M212909200

13. Montagna G, Cremona ML, Paris G, Amaya MF, Buschiazzo A, Alzari PM, et al. The trans-sialidase from the african trypanosome *Trypanosoma brucei*. Eur J Biochem. 2002;269: 2941–2950.

14. Nok AJ, Balogun EO. A bloodstream *Trypanosoma congolense* sialidase could be involved in anemia during experimental trypanosomiasis. J Biochem. 2003;133: 725–730.

15. Coustou V, Plazolles N, Guegan F, Baltz T. Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. Cellular Microbiology. 2012;14: 431–445. doi:10.1111/j.1462-5822.2011.01730.x

16. Campetella OE, Uttaro AD, Parodi AJ, Frasch AC. A recombinant *Trypanosoma cruzi* trans-

sialidase lacking the amino acid repeats retains the enzymatic activity. Mol Biochem Parasitol. 1994;64: 337–340.

17. Cremona ML, Sánchez DO, Frasch AC, Campetella O. A single tyrosine differentiates active and inactive *Trypanosoma cruzi* trans-sialidases. Gene. 1995;160: 123–128.

18. París G, Cremona ML, Amaya MF, Buschiazzo A, Giambiagi S, Frasch AC, et al. Probing molecular function of trypanosomal sialidases: single point mutations can change substrate specificity and increase hydrolytic activity. Glycobiology. 2001;11: 305–311.

19. Haselhorst T. NMR spectroscopic and molecular modeling investigations of the trans-sialidase from *Trypanosoma cruzi*. Glycobiology. 2004;14: 895–907. doi:10.1093/glycob/cwh108

20. Koliwer-Brandl H, Gbem TT, Waespy M, Reichert O, Mandel P, Drebitz E, et al. Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*. BMC Biochem. 2011;12: 39. doi:10.1093/bioinformatics/18.suppl_2.S153

21. Gbem TT, Waespy M, Hesse B, Dietz F, Smith J, Chechet GD, et al. Biochemical diversity in the *Trypanosoma congolense* trans-sialidase family. PLoS Negl Trop Dis. 2013;7: e2549. doi:10.1371/journal.pntd.0002549

22. Oliveira IA, Goncalves AS, Neves JL, Itzstein von M, Todeschini AR. Evidence of ternary complex formation in *Trypanosoma cruzi* trans-sialidase catalysis. Journal of Biological Chemistry. 2014;289: 423–436. doi:10.1074/jbc.M112.399303

23. Waespy M, Gbem TT, Elenschneider L, Jeck A-P, Day CJ, Hartley-Tassell L, et al. Carbohydrate recognition specificity of trans-sialidase lectin domain from *Trypanosoma congolense*. (in press) PLoS Negl Trop Dis. 2015;: 1–49. doi:10.1371/journal.pntd.0004120

24. Engstler M, Reuter G, Schauer R. The developmentally regulated trans-sialidase from *Trypanosoma brucei* sialylates the procyclic acidic repetitive protein. Mol Biochem Parasitol. 1993;61: 1–13.

25. Savage A, Geyer R, Stirm S, Reinwald E, Risse HJ. Structural studies on the major oligosaccharides in a variant surface glycoprotein of *Trypanosoma congolense*. Mol Biochem Parasitol. 1984;11: 309–328.

26. Zamze SE, Wooten EW, Ashford DA, Ferguson MA, Dwek RA, Rademacher TW. Characterisation of the asparagine-linked oligosaccharides from *Trypanosoma brucei* type-I variant surface glycoproteins. Eur J Biochem. 1990;187: 657–663.

27. Zamze SE, Ashford DA, Wooten EW, Rademacher TW, Dwek RA. Structural characterization of the asparagine-linked oligosaccharides from *Trypanosoma brucei* type II and type III variant surface glycoproteins. J Biol Chem. 1991;266: 20244–20261.

28. Bayne RA, Kilbride EA, Lainson FA, Tetley L, Barry JD. A major surface antigen of procyclic stage *Trypanosoma congolense*. Mol Biochem Parasitol. 1993;61: 295–310.

29. Beecroft RP, Roditi I, Pearson TW. Identification and characterization of an acidic major surface glycoprotein from procyclic stage *Trypanosoma congolense*. Mol Biochem Parasitol. 1993;61: 285–294.

30. Bütikofer P, Vassella E, Boschung M, Renggli CK, Brun R, Pearson TW, et al. Glycosylphosphatidylinositol-anchored surface molecules of *Trypanosoma congolense* insect forms are developmentally regulated in the tsetse fly. Mol Biochem Parasitol. 2002;119: 7–16.

31. Thomson LM, Lamont DJ, Mehlert A, Barry JD, Ferguson MAJ. Partial structure of glutamic acid and alanine-rich protein, a major surface glycoprotein of the insect stages of *Trypanosoma congolense*. J Biol Chem. 2002;277: 48899–48904. doi:10.1074/jbc.M208942200

32. Amaya MF, Buschiazzo A, Nguyen T, Alzari PM. The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T.cruzi* trans-sialidase. Journal of Molecular Biology. 2003;325: 773–784. doi:10.1016/S0022-2836(02)01306-2

33. Crennell S, Garman E, Laver G, Vimr E, Taylor G. Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. Structure. 1994;2: 535–544.

34. Luo Y, Li SC, Li YT, Luo M. The 1.8 Å structures of leech intramolecular trans-sialidase complexes: evidence of its enzymatic mechanism. Journal of Molecular Biology. 1999;285: 323–332. doi:10.1006/jmbi.1998.2345

35. Smith LE, Eichinger D. Directed mutagenesis of the *Trypanosoma cruzi* trans-sialidase enzyme identifies two domains involved in its sialyltransferase activity. Glycobiology. 1997;7: 445–451.

36. Pereira ME, Mejia JS, Ortega-Barria E, Matzilevich D, Prioli RP. The *Trypanosoma cruzi* neuraminidase contains sequences similar to bacterial neuraminidases, YWTD repeats of the low density lipoprotein receptor, and type III modules of fibronectin. J Exp Med. 1991;174: 179–191.

37. Smith LE, Uemura H, Eichinger D. Isolation and expression of an open reading frame encoding sialidase from *Trypanosoma rangeli*. Mol Biochem Parasitol. 1996;79: 21–33.

38. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure. 1995;3: 1197–1205.

39. Buschiazzo A, Tavares GA, Campetella O, Spinelli S, Cremona ML, París G, et al. Structural basis of sialyltransferase activity in trypanosomal sialidases. EMBO J. 2000;19: 16–24. doi:10.1093/emboj/19.1.16

40. Amaya MF, Watts AG, Damager I, Wehenkel A, Nguyen T, Buschiazzo A, et al. Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. Structure. 2004;12: 775–784. doi:10.1016/j.str.2004.02.036

41. Ferguson MA, Homans SW, Dwek RA, Rademacher TW. Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. Ullstein Mosby. 1988;239: 753–759.

42. Treumann A, Zitzmann N, Hülsmeier A, Prescott AR, Almond A, Sheehan J, et al. Structural characterisation of two forms of procyclic acidic repetitive protein expressed by procyclic forms of *Trypanosoma brucei*. Journal of Molecular Biology. 1997;269: 529–547. doi:10.1006/jmbi.1997.1066

43. Utz S, Roditi I, Kunz Renggli C, Almeida IC, Acosta-Serrano A, Bütikofer P. *Trypanosoma congolense* procyclins: unmasking cryptic major surface glycoproteins in procyclic forms. Eukaryotic Cell. 2006;5: 1430–1440. doi:10.1128/EC.00067-06

44. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227: 680–685.

45. King RD, Sternberg MJ. Identification and application of the concepts important for accurate

and reliable protein secondary structure prediction. Protein Sci. Cold Spring Harbor Laboratory Press; 1996;5: 2298–2310. doi:10.1002/pro.5560051116

46. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology. 1999;292: 195–202. doi:10.1006/jmbi.1999.3091

47. Mückstein U, Hofacker IL, Stadler PF. Stochastic pairwise alignments. Bioinformatics. 2002;18 Suppl 2: S153–60.

48. Canutescu AA, Shelenkov AA, Dunbrack RL. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci. 2003;12: 2001–2014. doi:10.1110/ps.03154503

49. Qiu J, Elber R. SSALN: an alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs. Proteins. Wiley Subscription Services, Inc., A Wiley Company; 2006;62: 881–891. doi:10.1002/prot.20854

50. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins. 2009;77 Suppl 9: 114–122. doi:10.1002/prot.22570

51. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. Oxford University Press; 2012;28: 1647–1649. doi:10.1093/bioinformatics/bts199

Figure legends

Figure 1: Hydrogen bond network at the interface between CD and LD of TconTS. **A:** The homology model of TconTS1a, which was calculated using the crystal structure of TcruTS (PDB code: 3B69) as template structure employing the software Yasara (described under Methods). Locations A and B, comprising the majority of conserved hydrogen bonds formed by amino acid residues at the interface between catalytic domain (CD in grey), α-helix (blue) and lectin domain (LD in green) are marked with yellow (A) and red (B) cycles. **B:** Enlarged view of location A (yellow). **C:** Enlarged view of location B (red). **D:** Amino acid sequence alignment of TconTS1a through TconTS4, *T. cruzi* TS (TcruTS EMBL: AAA66352.1), *T. rangeli* sialidase (TranSA EMBL: AAC95493.1) and *T, brucei* TS (TbruTS EMBL: AAG32055.1) with assigned structural elements (light blue, dark blue and yellow for CD, α-helix and LD, respectively). Alignments were calculated employing the software Geneious (described under Methods). Sequence segments being part of the interface between CD, α-helix and LD of TS are marked with green squares. Increasing darkness of background for each residue indicates increasing number of identical amino acid residues at the corresponding position (described under Methods).

Figure 2: Insertion of Eco105I endonuclease restriction site into TconTS. **A:** Amino acid sequence alignment of TconTS1a through TconTS4 with assigned structural elements (light blue, dark blue and yellow for CD, α-helix and LD, respectively). Cysteine residues forming a well conserved disulphide bridge in TS are marked with light blue frames, whereas the resulting hair-pin loop flanked by these cysteine residues is marked with a red frame. **B:** Amino acid sequence alignment of the hair-pin loop region with the Eco105I restriction site inserted (red annotation above each TconTS sequence). Alignments were calculated employing the software Geneious (described under Methods). Numbers on top of each sequence indicate the corresponding residue number. TconTS comprising the inserted Eco105I restriction site are labelled (*). Increasing darkness of background for each residue indicates increasing number of identical amino acid residues at the corresponding position. **C:** Homology model of TconTS1a showing a region of the lectin domain (grey), including the well conserved disulphide bridge (C493 and C505) and the hair-

pin loop with the Eco105I insertion (red label). TconTS1a homology model was calculated using the software Yasara (described under Methods).

Figure 3: Recombinant TconTS constructs generated for expression by CHO-Lec1 fibroblasts and *E. coli*. A-B: Schematic presentation of recombinant TconTS constructs for expression by CHO-Lec1 fibroblasts (A) and *E.coli* (B). Fusion tags flanking TconTS are: Transin: translocation signal peptide (only in A), His: poly histidine tag (only in B), MBP: maltose binding protein tag (only in B), TEV: *tobacco etch virus* protease cleavage site (only in B), 3C: human rhinovirus 3C protease cleavage site, SNAP: SNAP-tag, *Strep*: Strep-tag. **C:** SDS-PAGE of purified TconTS constructs expressed by *E. coli* Rosetta pLacI. 1-2 µg of protein was loaded in each lane of a 10 % SDS polyacrylamide gel as indicated. The gel was stained with Coomassie Brilliant Blue after electrophorese. Lane 1-4 contain TconTS constructs with *Eco*105I insertion as indicated (*). Lane 5 and 6 comprise domain swapped proteins as labelled (CD/LD). **D:** Western blot analysis of TconTS constructs expressed by CHO-Lec1 fibroblasts. 100 ng of each TconTS construct was used and detection was done employing anti *Strep*-tag mAb and anti TS mAb 7/23 as indicated (described under Methods).

Figure 4: Epitope mapping of anti TS mAb 7/23 binding epitope. **A:** Amino acid sequence alignment of wild type TconTS1a and truncated TconTS1a constructs (A-G, CD: catalytic domain, LD: lectin domain) used for anti TS mAb 7/23 epitope mapping. Alignments were calculated employing the software Geneious (described under Methods). Numbers on top of each sequence indicate the length of the corresponding protein. Structural elements, such as catalytic domain (CD), α-helix and lectin domain (LD) are labelled. **B-D:** Western blot analysis of TconTS1a constructs (A-G) using anti TS mAb 7/23 and anti *Strep*-tag mAb as indicated (described under Methods). All TconTS1a constructs were expressed by *E. coli* Rosetta pLacI and bacterial lysates were used for SDS polyacrylamide gel electrophorese. 100 ng TconTS1a expressed and purified from CHO-Lec1 cells was used as a control in Western blot experiments. **E:** pNGaseF treatment of purified recombinant TconTS1a and TconTS2 expressed by CHO-Lec1 cells. Both proteins were

deglycosylated using pNGaseF under native and denaturing conditions (described under Methods) as indicated. 1 µg of each TconTS sample was used in SDS-PAGE analysis with subsequent Coomassie Brilliant Blue staining (upper line) and 100 ng in western blots (lower line) using anti TS mAb 7/23 and anti *Strep*-tag for detection (described under Methods).

Figure 5: Enzymatic activities of recombinant TconTS1a expressed by *E. coli*. A-B: Production of 3'SL (A) and Neu5Ac (B) using different amounts of purified, bacterial expressed, recombinant TconTS1a ranging from 0-3 µg of enzyme. TS reactions were set up as described under Methods using standard conditions employing fetuin and lactose as Sia donor and acceptor substrates for 30 min at 37°C. Reaction products 3'SL and Neu5Ac were quantified using HPAEC-PAD analysis (described under Methods). **C-D:** Time dependency of TconTS1a activity using 1 µg of purified bacterial expressed enzyme. Reactions were set up according to TconTS1a concentration series, using different incubation times ranging from 0-24 h. Data points are means ± standard deviation of triplicates.

Figure 6: Enzymatic activities of mutated and domain swapped recombinant TconTS enzymes. **A-C:** Sia transfer (A and C) and sialidase activities (B and C) of purified, bacterial expressed, recombinant TconTS1a, TconTS3 and domain swapped TconTS1a/TS3. 1 µg of enzyme was incubated with fetuin and lactose as Sia donor and acceptor substrates, respectively, under standard conditions (described under Methods) for 24 h at 37°C. **D-F:** Sia transfer and sialidase activities of recombinant TconTS1a, mutated TconTS1a*, domain swapped TconTS1a/TS3, TconTS2 and mutated TconTS2* expressed and purified from CHO-Lec1 fibroblasts. TS reactions were set up according to bacterial expressed TconTS enzymes, whereas 50 ng of CHO-Lec1 expressed recombinant TconTS were used and incubated under standard conditions for 30 min at 37°C. Reaction products 3'SL and Neu5Ac were quantified using HPAEC-PAD analysis (described under Methods). Data points are means ± standard deviation of triplicates.

Figure 7: Enzymatic activities of recombinant TconTS1a and TconTS2 in the presence or absence of 1,4-β-mannotriose. A-C: Sia transfer (A and C) and sialidase activities (B and C) of recombinant TconTS1a and TconTS2 expressed and purified from CHO-Lec1 fibroblasts. 50 ng of enzyme were incubated under standard conditions using fetuin and lactose as Sia donor and acceptor substrates for 30 min at 37°C (described under Methods), in the presence or absence of 5 mM 1,4-β-mannotriose as indicated. Reaction products 3'SL and Neu5Ac were quantified using HPAEC-PAD analysis (described under Methods). Data points are means ± standard deviation of triplicates.

Supporting Information

Figure S1. Predicted binding epitope of the anti TS mAb 7/23. A: Amino acid sequence alignment of TconTS1a through TconTS4, *T. cruzi* TS (TcruTS EMBL: AAA66352.1), *T. rangeli* sialidase (TranSA EMBL: AAC95493.1) and *T, brucei* TS (TbruTS EMBL: AAG32055.1) was calculated using the software Geneious 5.5.5 as described under Methods. Structural elements, such as predicted *N*-glycosylation sites, α-helix, salt-bridge (between arginine and glutamine) and the predicted anti TS mAb 7/23 binding epitope are marked with red, purple, green and blue labels, respectively. Numbers on top of each sequence indicate the corresponding residue number. Increasing darkness of background for each residue indicates increasing number of identical amino acid residues at the corresponding position (described under Methods). **B:** Homology model of TconTS1a showing a detailed view of the lectin domain (LD). Structural elements, such as the C-terminal α-helix and the predicted *N*-glycosylation site are labelled in purple and red, respectively. The salt-bridge between R704 and E 710, which is well conserved among TS, is illustrated in stick style. **C:** Structural alignment of TconTS2 (homology model), TcruTS (PDB: 3B69) and TranSA (PDB: 1N1Y) calculated using the software Yasara (described under Methods). The tryptophan residue, which is part of the Sia donor substrate binding-site is shown in stick style and labelled in yellow (TconTS2), green (TcruTS) and red (TranSA), respectively. Bound Sia is illustrated in stick style (light blue).

Table S1. List of primers used for cloning and *Eco*105I restriction site insertion.

Table 1: Calculated molecular surface of the contact site between CD and LD of TTSs.

	Area Å^2	
Enzyme (PDB)	CD/LD	total surf.
TconTS1	1526.68/1500.01	3026.69
TconTS2	1440.96/1392.59	2833.55
TconTS3	1509.82/1331.59	2841.41
TconTS4	1674.97/1631.50	3306.47
TcruTS (3B69)	1483.04/1509.52	2992.56
TbruTS*	1677.44/1604.22	3281.66
TranSA (1N1S)	1470.99/1563.01	3034.00
VCS (1W0O)	960.88/935.29	1896.17
LeechMD-SA (1SLI)	868.53/918.46	1786.99

* Homology model for TbruTS was calculated using the amino acid sequence: EMBL AAG32055.1

Table 2: Hydrogen bonds formed by amino acid residues at the contact site between CD and LD of TconTSs.

TconTS1			TconTS2			TconTS3			TconTS4		
Atom A	Atom B	Distance [Å]	Atom A	Atom B	Distance [Å]	Atom A	Atom B	Distance [Å]	Atom A	Atom B	Distance [Å]
V369-O	Q566-N	2.269				V361-O	Q559-N	1.955			
			G349-O	N687-$N^{\delta 2}$	2.400				G419-O	N751-$N^{\delta 2}$	2.183
									R423-$N^{\eta 1}$	N725-$O^{\delta 1}$	2.164
			V354-O	Q552-N	2.294				V424-O	Q621-N	1.873
G371-O	Q564-$N^{\varepsilon 2}$	2.190	G356-O	Q550-$N^{\varepsilon 2}$	1.974	G363-N	Q557-O	1.921	G426-N	D620-$O^{\delta 1}$	1.869
						G363-O	Q557-$N^{\varepsilon 2}$	1.944	G426-O	Q619-$N^{\varepsilon 2}$	2.019
R376-$N^{\eta 2}$	G563-O	2.261	R361-$N^{\eta 1}$	Q550-O	1.923						
R376-$N^{\eta 2}$	S560-O	2.212	R361-N^{ε}	Q550-O	2.294						
			E376-O	K468-N^{ζ}	2.336	E383-O	K475-N^{ζ}	2.386			
K402-O	Q564-$N^{\varepsilon 2}$	1.972	E387-O	Q550-$N^{\varepsilon 2}$	1.919	K394-O	Q557-$N^{\varepsilon 2}$	2.119	K457-N^{ζ}	D620-$O^{\delta 2}$	1.897
						K394-N^{ζ}	Q557-$O^{\varepsilon 1}$	1.983			
						K394-N^{ζ}	D558-$O^{\delta 2}$	1.798			
N421-$O^{\delta 2}$	N703-$N^{\delta 2}$	2.396				N413-$O^{\delta 1}$	G503-N	1.907	S476-O^{γ}	L566-O	1.809
R422-N^{ε}	L511-O	1.836				R414-N^{ε}	L504-O	1.864	R477-$N^{\eta 1}$	D541-$O^{\delta 2}$	1.888
R422-$N^{\eta 1}$	D486-$O^{\delta 1}$	1.902				R414-$N^{\eta 1}$	D478-$O^{\delta 1}$	1.930	R477-$N^{\eta 2}$	D541-$O^{\delta 1}$	1.927
R422-$N^{\eta 2}$	D486-$O^{\delta 2}$	2.016				R414-$N^{\eta 2}$	D478-$O^{\delta 2}$	1.853			
			R399-$N^{\eta 1}$	D551-$O^{\delta 2}$	2.129						
			R399-$N^{\eta 2}$	D551-$O^{\delta 2}$	2.160						
W482-$N^{\eta 1}$	Y528-O	2.047	W467-$N^{\varepsilon 1}$	Y513-O	1.981	W474-$N^{\varepsilon 1}$	Y521-O	1.778	W537-$N^{\varepsilon 1}$	Y583-O	1.886
Q485-$O^{\varepsilon 1}$	S726-O^{γ}	2.073									
D486-$O^{\delta 1}$	Y528-OH	1.811	D471-$O^{\delta 1}$	Y513-OH	1.818	D478-$O^{\delta 1}$	T501-$O^{\gamma 1}$	1.870	D541-$O^{\delta 1}$	Y583-OH	1.898

Table 3: Conserved amino acid residues involved in hydrogen bond formation between CD and LD at the contact site of TSs.

	Nr.	Consensus	TconTS1	TconTS2	TconTS3	TconTS4	TcruTS*	TranSA**	TbruTS***
CD	1	Gly	N364	G349	G356	G419	G269	G291	G376
	2	Arg	R368	R353	R360	R423	R273	H295	R380
	3	Val	V369	V354	V361	V424	V274	V296	V381
	4	Gly	G371	G356	G363	G426	G276	T298	G383
	5	Arg	R376	R361	R368	R431	S281	S303	R388
	6	Glu	E391	E376	E383	E446	E296	E318	G410
	7	Lys	K402	E387	K394	K457	L307	L329	K421
	8	Asn	N421	N406	N413	S476	Q326	Q348	N440
	9	Arg	R422	R407	R414	R477	R327	R349	R441
αHel	10	Trp	W482	W467	W474	W537	W386	W408	W501
	11	Lys	K483	K468	K475	K538	K387	K409	K502
	12	Asp	D486	D471	D478	D541	D390	D412	D505
LD	13	Thr	T508	I493	T501	T563	T417	T439	T528
	14	Leu	L511	L496	L504	L566	L420	L442	L531
	15	Tyr	Y528	Y513	Y521	Y583	Y437	Y459	Y548
	16	Ser	S560	G546	S553	R615	S468	A490	S580
	17	Gly	G563	G549	G556	G618	G471	G493	G583
	18	Gln	Q564	Q550	Q557	Q619	Q472	Q494	Q584
	19	Asp	D565	D551	D558	D620	N473	T495	D585
	20	Gln	Q566	Q552	Q559	Q621	Q474	R496	Q586
	21	Asn	N674	N658	N667	N725	D581	D603	H694
	22	Asn	N703	N687	N696	N751	N610	N632	N723

* TcruTS sequence: EMBL AAA66352.1; ** TranSA sequence: EMBL AAC95493.1; *** TbruTS sequence: EMBL AAG32055.1. It should be noted that different TbruTS have been identified, which are similar to certain TconTS enzymes, respectively [21].

Table 4: Possible domain swapped and mutated TconTS constructs.

Construct Nr.	Catalytic domain	Lectin domain	Swap construct
1	TconTS1a	TconTS1a	TconTS1a*
2	TconTS2	TconTS2	TconTS2*
3	TconTS3	TconTS3	TconTS3*
4	TconTS4	TconTS4	TconTS4*
5	TconTS1a	TconTS2	TconTS1a/TS2
6	TconTS1a	TconTS3	TconTS1a/TS3
7	TconTS1a	TconTS4	TconTS1a/TS4
8	TconTS2	TconTS1a	TconTS2/TS1a
9	TconTS2	TconTS3	TconTS2/TS3
10	TconTS2	TconTS4	TconTS2/TS4
11	TconTS3	TconTS1a	TconTS3/TS1a
12	TconTS3	TconTS2	TconTS3/TS2
13	TconTS3	TconTS4	TconTS3/TS4
14	TconTS4	TconTS1a	TconTS4/TS1a
15	TconTS4	TconTS2	TconTS4/TS2
16	TconTS4	TconTS3	TconTS4/TS3

* Mutated TconTS containing the inserted *Eco*105I restriction site

Table 5: Specific catalytic activities of different TconTS enzymes expressed by *E. coli* and CHO-Lec1 cells.

CHO-Lec1

Enzyme	Trans-sialidase activity [nmol **3'SL**/(min x mg TS)]	Sialidase activity [nmol **Neu5Ac**/(min x mg TS)]	TS/sialidase
TconTS1a	2732 ± 28	88 ± 1	31
TconTS1a*	1029 ± 41	49 ± 7	21
TconTS1a/TS3	2409 ± 109	16 ± 2	153
TconTS2	1820 ± 204	56 ± 7	32
TconTS2*	1852 ± 107	223 ± 26	8

E. coli

Enzyme	Trans-sialidase activity [nmol **3'SL**/(min x mg TS)]	Sialidase activity [nmol **Neu5Ac**/(min x mg TS)]	TS/sialidase
TconTS1a	5.22 ± 0.03	0.05 ± 0.01	109
TconTS1a/TS3	0.086 ± 0.004	0.157 ± 0.014	0.55
TconTS3	0.042 ± 0.008	0.081 ± 0.011	0.51

Quantifications of 3'SL and Neu5Ac were done employing HPAEC-PAD analysis as described under Methods. * Mutated TconTS containing the inserted *Eco*105I endonuclease restriction site. Data points are means of triplicates ± standard deviation.

Table 6: Specific catalytic activities of TconTS1a and TconTS2 expressed by CHO-Lec1 cells under standard conditions for 30 min in the presence and absence of 5 mM 1-4β-mannotriose.

	Trans-sialidase activity [nmol **3'SL**/(min x mg TS)]		**Sialidase activity** [nmol **Neu5Ac**/(min x mg TS)]		**TS/sialidase**	
1-4β-mannotriose	-	+	-	+	-	+
TconTS1a	2213 ± 133	2163 ± 190	91 ± 1	15 ± 1	24	144
TconTS2	1983 ± 26	2038 ± 69	58 ± 7	11 ± 2	34	193

Quantifications of 3'SL and Neu5Ac were done employing HPAEC-PAD analysis as described under Methods. Data points are means of triplicates ± standard deviation.

Figure 1

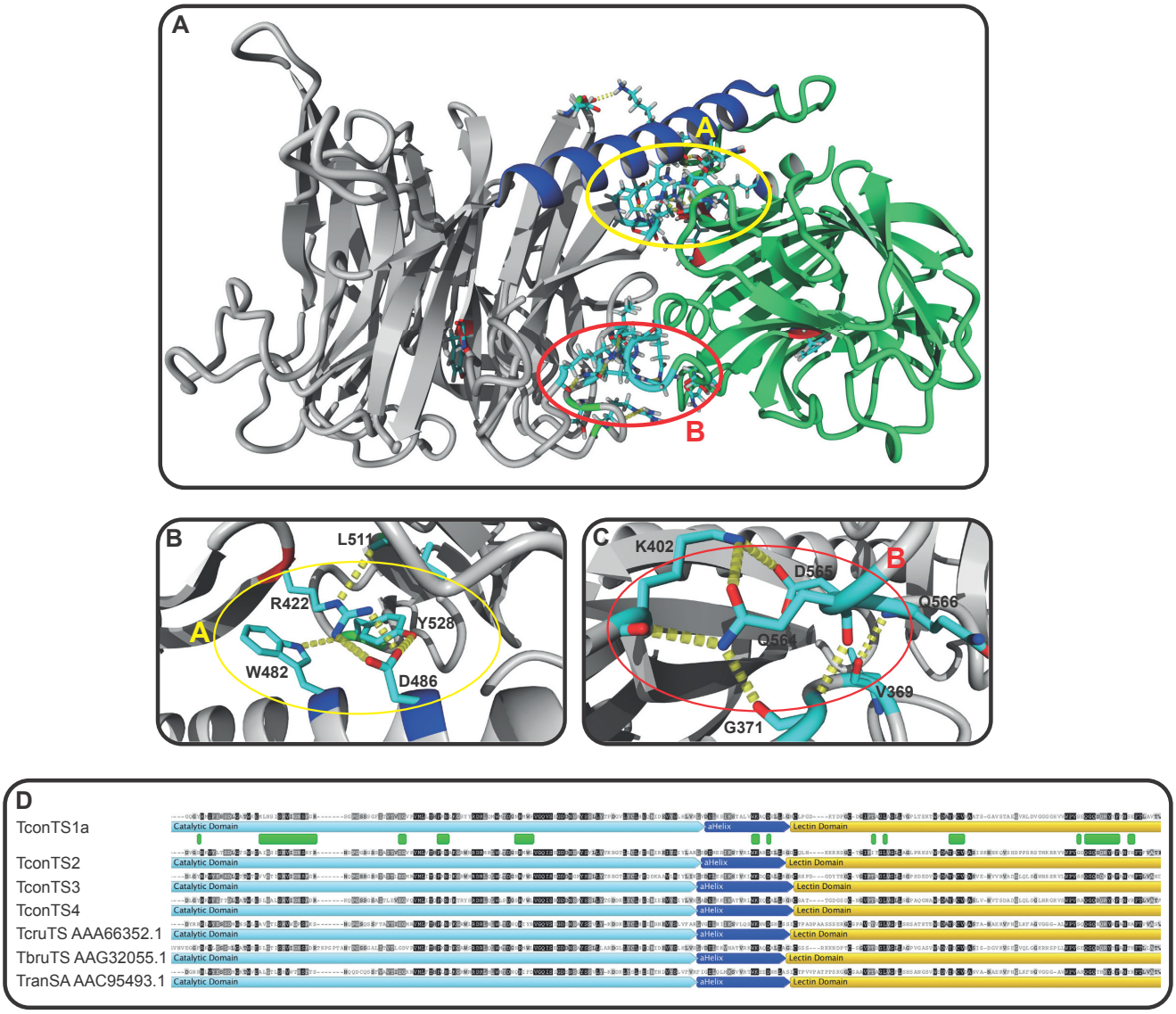

A
A
B
B
L511
R422
A
Y528
W482
D486
C
K402
D565
B
Q566
Q564
V369
G371
D
TconTS1a
TconTS2
TconTS3
TconTS4
TcruTS AAA66352.1
TbruTS AAG32055.1
TranSA AAC95493.1
Catalytic Domain
aHelix
Lectin Domain

Figure 2

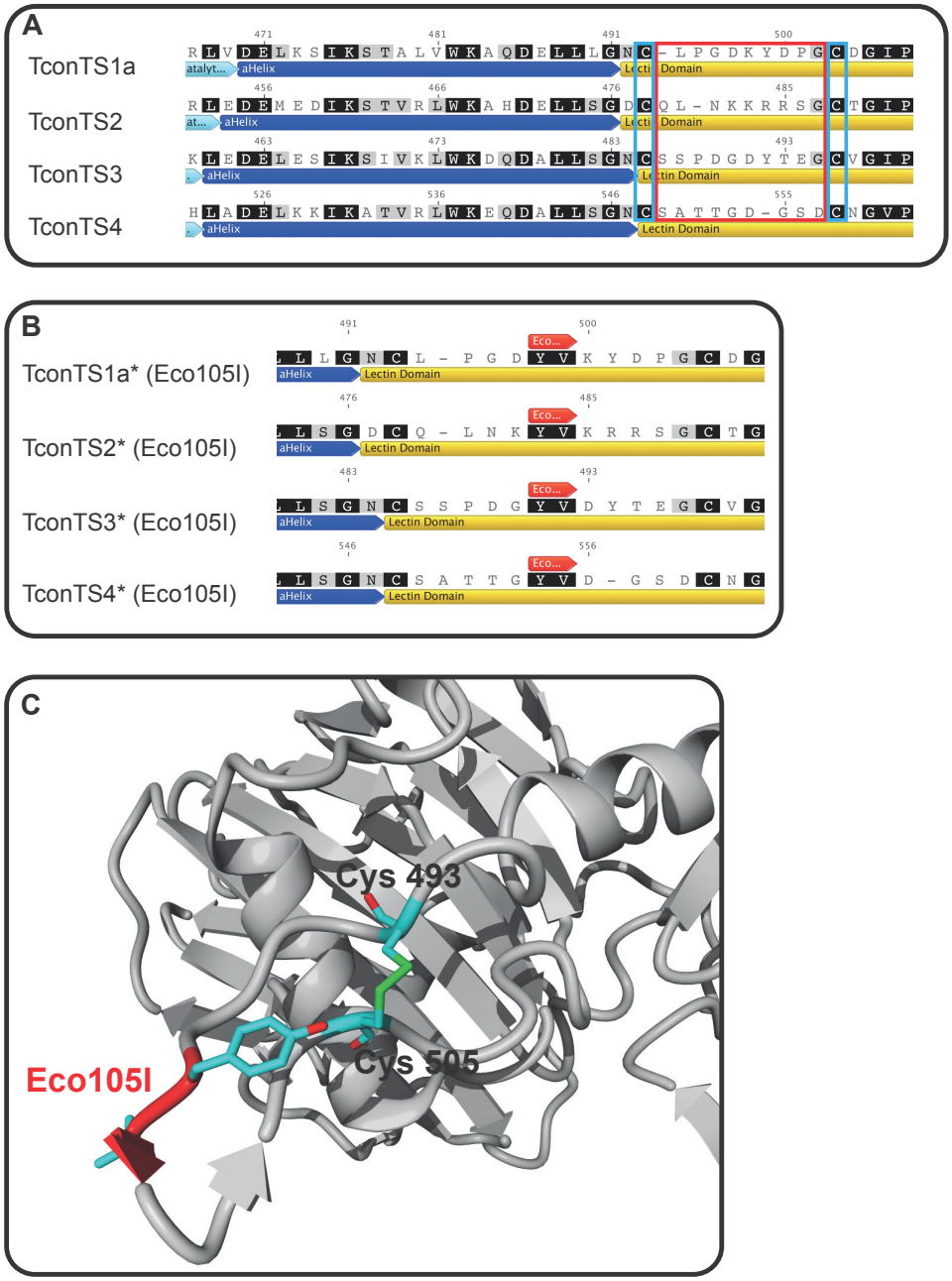

A
TconTS1a
TconTS2
TconTS3
TconTS4
aHelix
Lectin Domain
B
TconTS1a* (Eco105I)
TconTS2* (Eco105I)
TconTS3* (Eco105I)
TconTS4* (Eco105I)
C
Cys 493
Cys 505
Eco105I

Figure 3

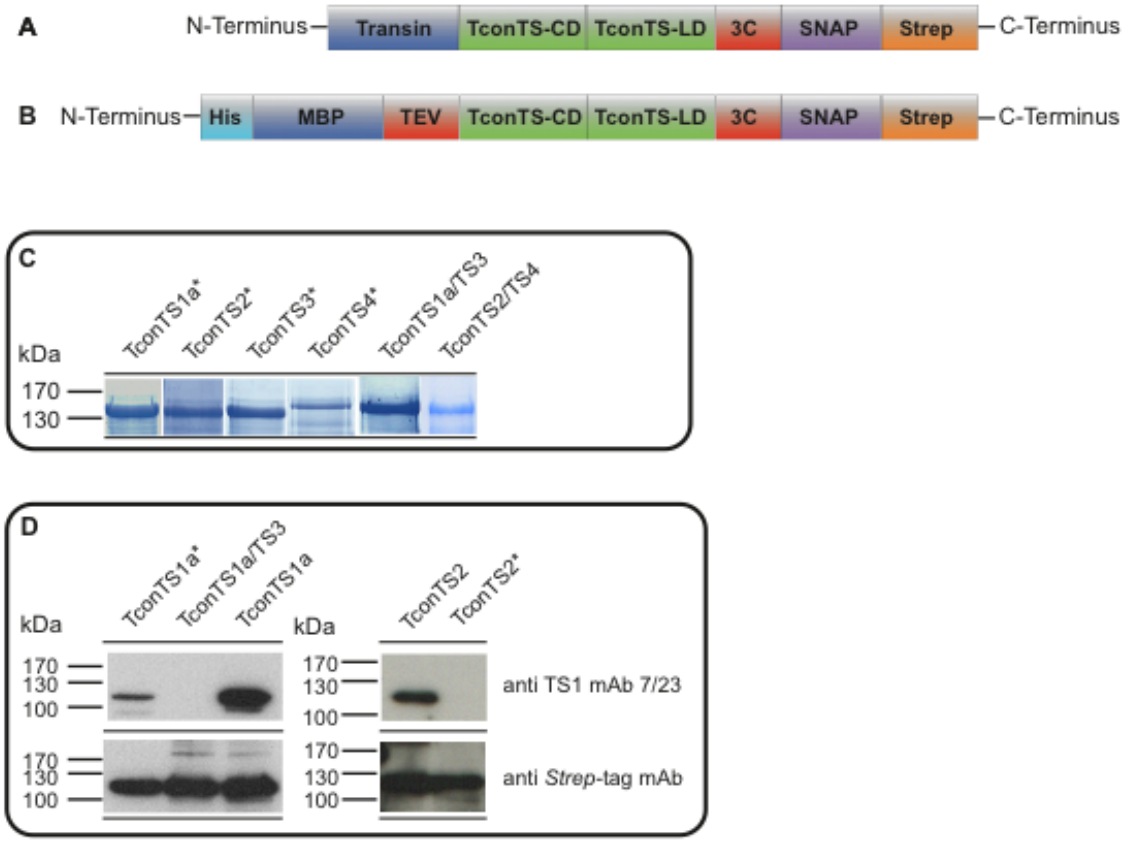

A
N-Terminus
Transin
TconTS-CD
TconTS-LD
3C
SNAP
Strep
C-Terminus
B
N-Terminus
His
MBP
TEV
TconTS-CD
TconTS-LD
3C
SNAP
Strep
C-Terminus
C
kDa
TconTS1a*
TconTS2*
TconTS3*
TconTS4*
TconTS1a/TS3
TconTS2/TS4
170
130
D
kDa
TconTS1a*
TconTS1a/TS3
TconTS1a
kDa
TconTS2
TconTS2*
170
130
100
anti TS1 mAb 7/23
anti Strep-tag mAb

Figure 4

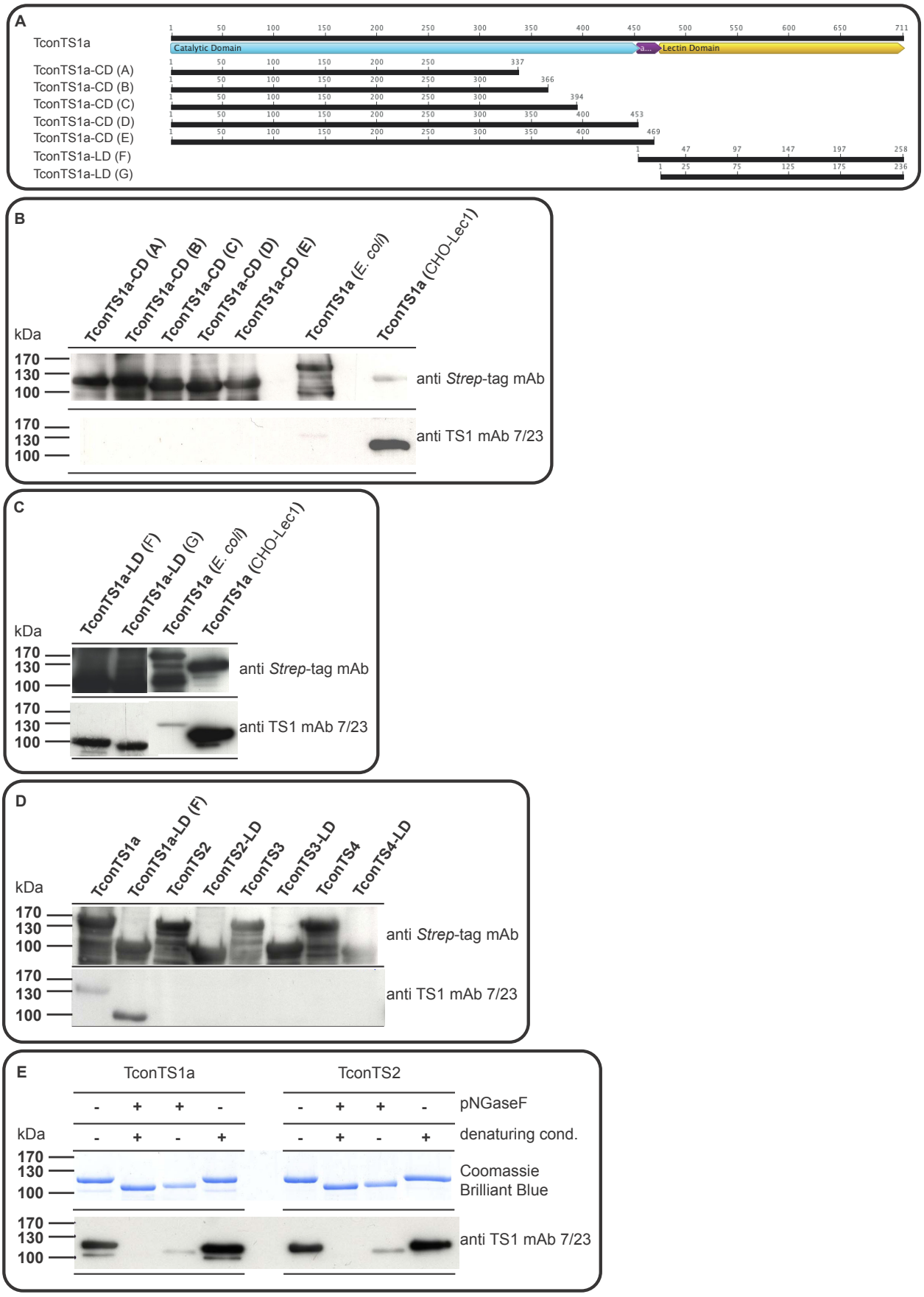

A
TconTS1a
Catalytic Domain
Lectin Domain
TconTS1a-CD (A)
TconTS1a-CD (B)
TconTS1a-CD (C)
TconTS1a-CD (D)
TconTS1a-CD (E)
TconTS1a-LD (F)
TconTS1a-LD (G)
B
TconTS1a-CD (A)
TconTS1a-CD (B)
TconTS1a-CD (C)
TconTS1a-CD (D)
TconTS1a-CD (E)
TconTS1a (E. coli)
TconTS1a (CHO-Lec1)
kDa
170
130
100
anti Strep-tag mAb
anti TS1 mAb 7/23
C
TconTS1a-LD (F)
TconTS1a-LD (G)
TconTS1a (E. coli)
TconTS1a (CHO-Lec1)
kDa
anti Strep-tag mAb
anti TS1 mAb 7/23
D
TconTS1a
TconTS1a-LD (F)
TconTS2
TconTS2-LD
TconTS3
TconTS3-LD
TconTS4
TconTS4-LD
kDa
anti Strep-tag mAb
anti TS1 mAb 7/23
E
TconTS1a
TconTS2
pNGaseF
denaturing cond.
kDa
Coomassie Brilliant Blue
anti TS1 mAb 7/23

Figure 5

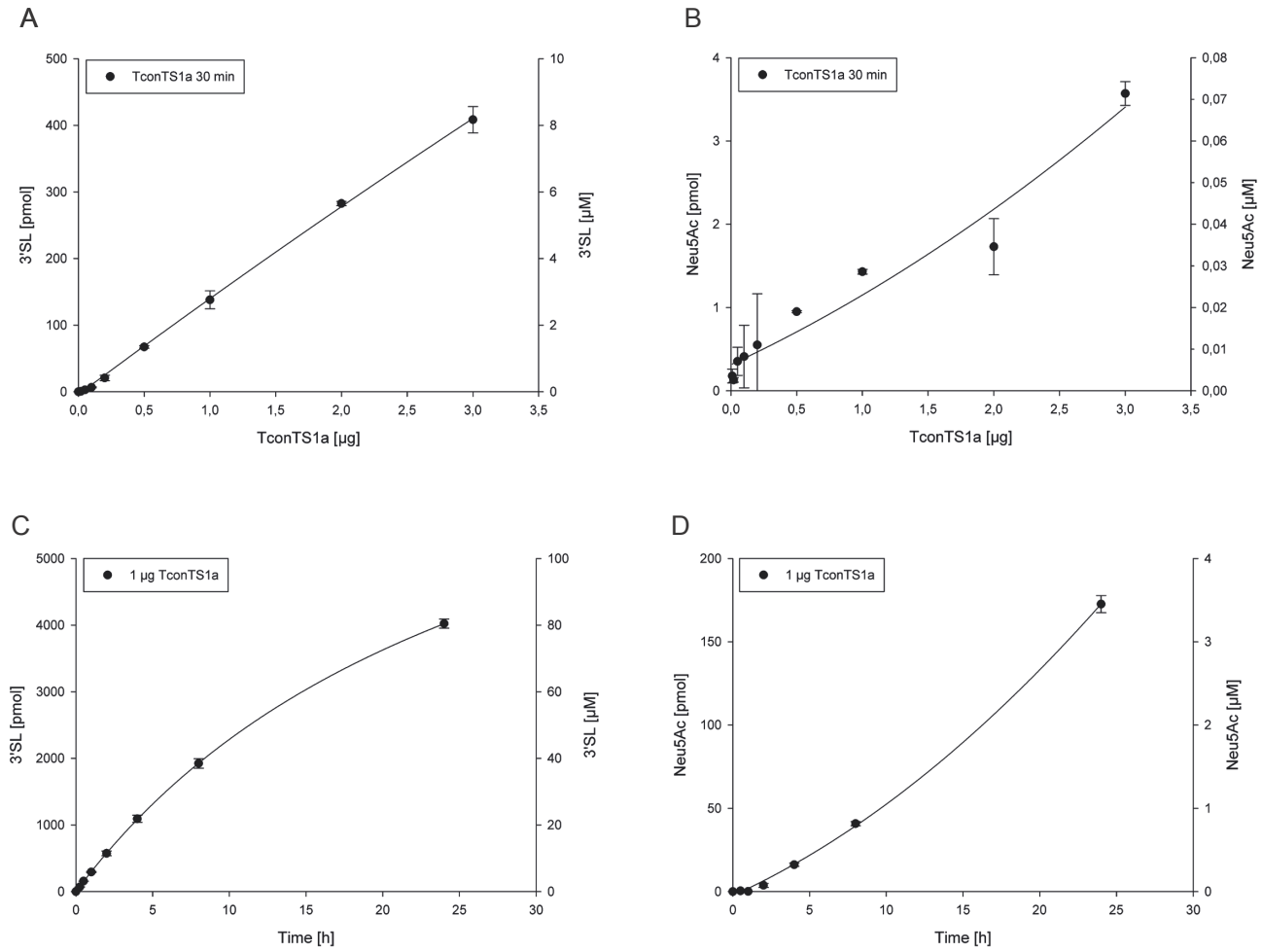
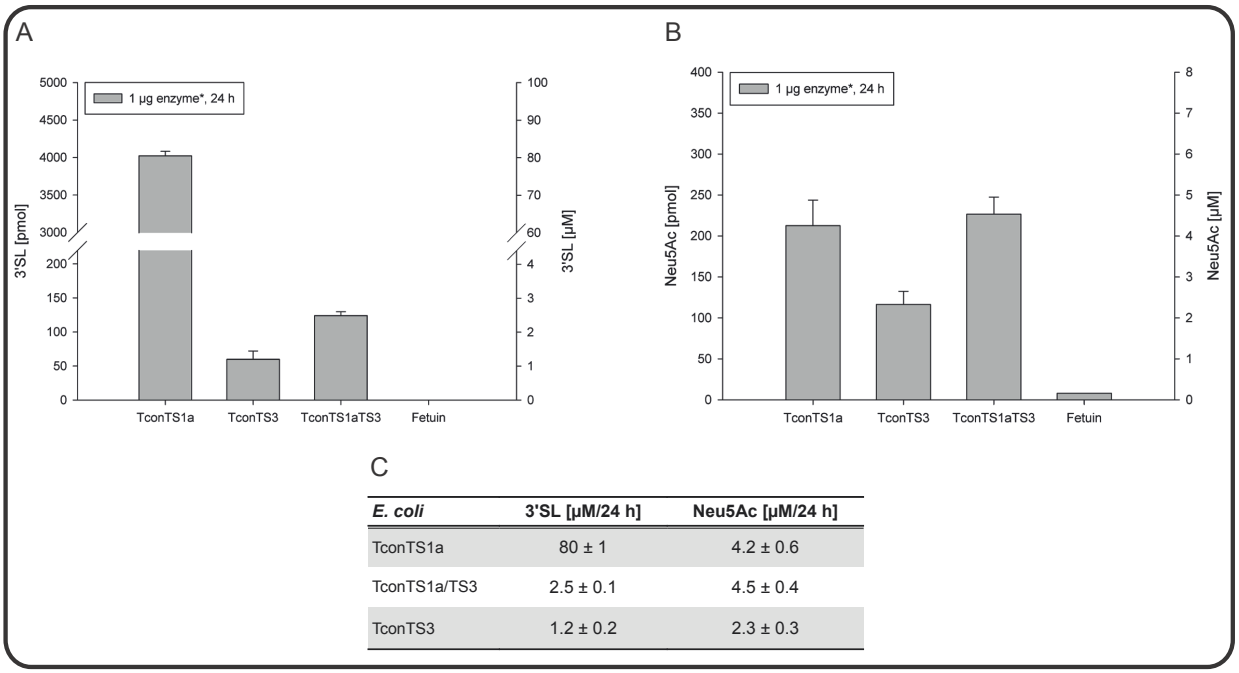
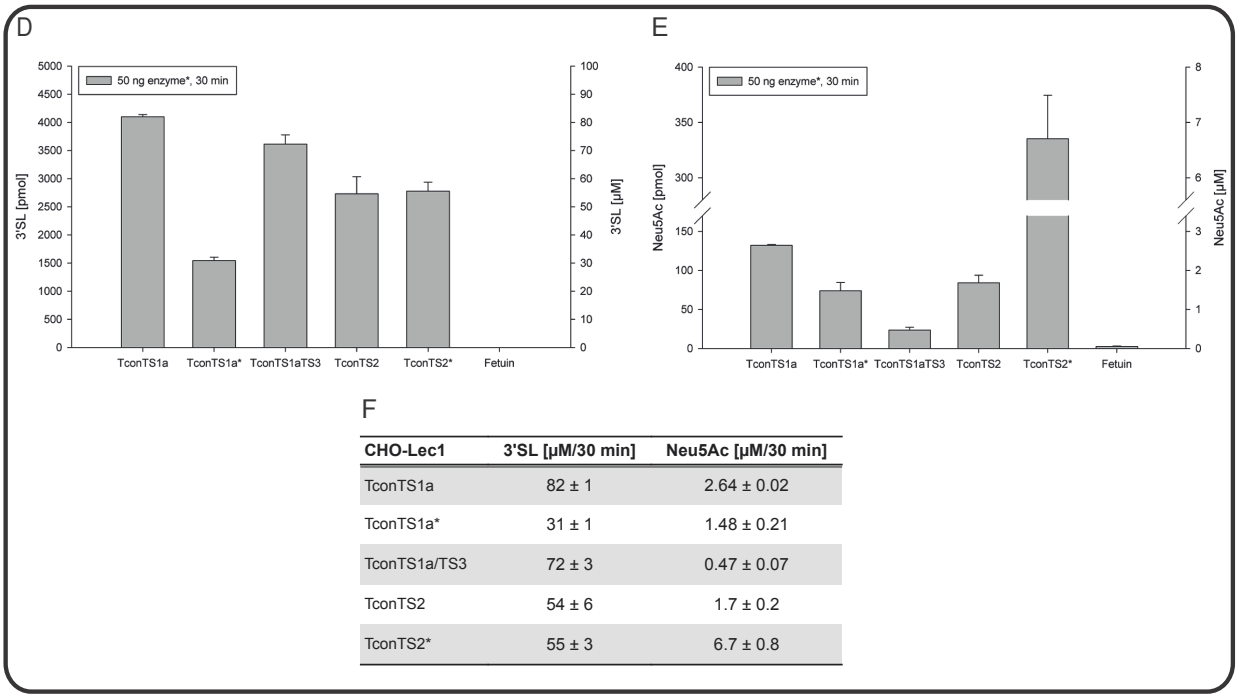

A
TconTS1a 30 min
3'SL [pmol]
3'SL [µM]
TconTS1a [µg]
B
TconTS1a 30 min
Neu5Ac [pmol]
Neu5Ac [µM]
TconTS1a [µg]
C
1 µg TconTS1a
3'SL [pmol]
3'SL [µM]
Time [h]
D
1 µg TconTS1a
Neu5Ac [pmol]
Neu5Ac [µM]
Time [h]

Figure 6



C

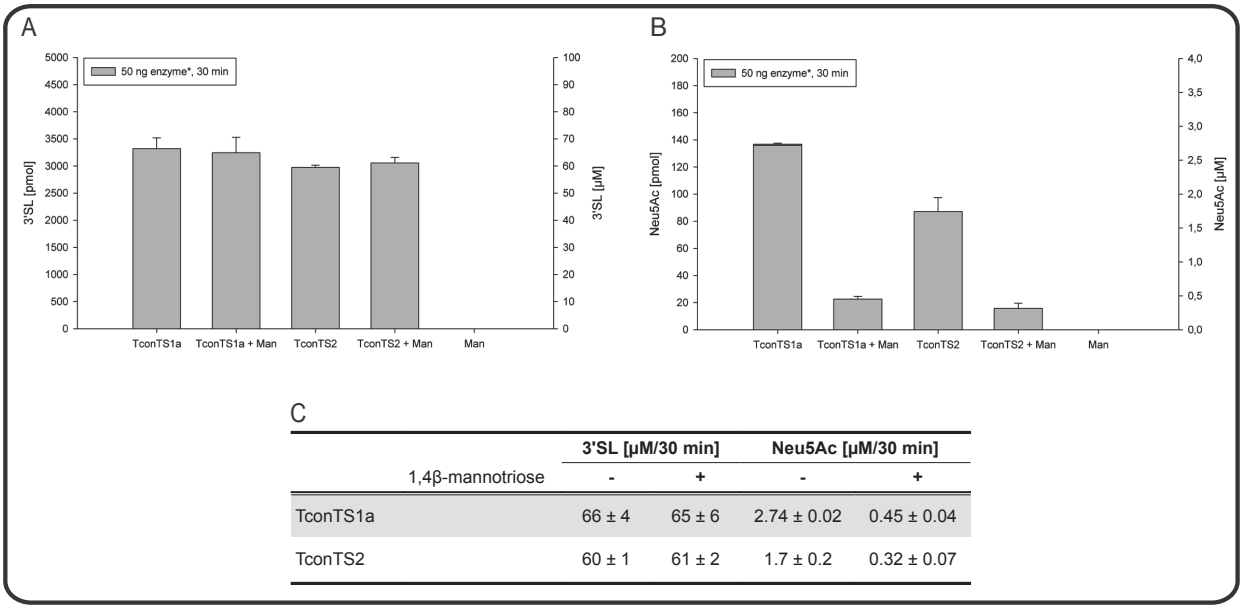
E. coli	3'SL [µM/24 h]	Neu5Ac [µM/24 h]
TconTS1a	80 ± 1	4.2 ± 0.6
TconTS1a/TS3	2.5 ± 0.1	4.5 ± 0.4
TconTS3	1.2 ± 0.2	2.3 ± 0.3



F

CHO-Lec1	3'SL [µM/30 min]	Neu5Ac [µM/30 min]
TconTS1a	82 ± 1	2.64 ± 0.02
TconTS1a*	31 ± 1	1.48 ± 0.21
TconTS1a/TS3	72 ± 3	0.47 ± 0.07
TconTS2	54 ± 6	1.7 ± 0.2
TconTS2*	55 ± 3	6.7 ± 0.8

Figure 7



C

	3'SL [µM/30 min]		Neu5Ac [µM/30 min]	
1,4β-mannotriose	-	+	-	+
TconTS1a	66 ± 4	65 ± 6	2.74 ± 0.02	0.45 ± 0.04
TconTS2	60 ± 1	61 ± 2	1.7 ± 0.2	0.32 ± 0.07

Figure S1

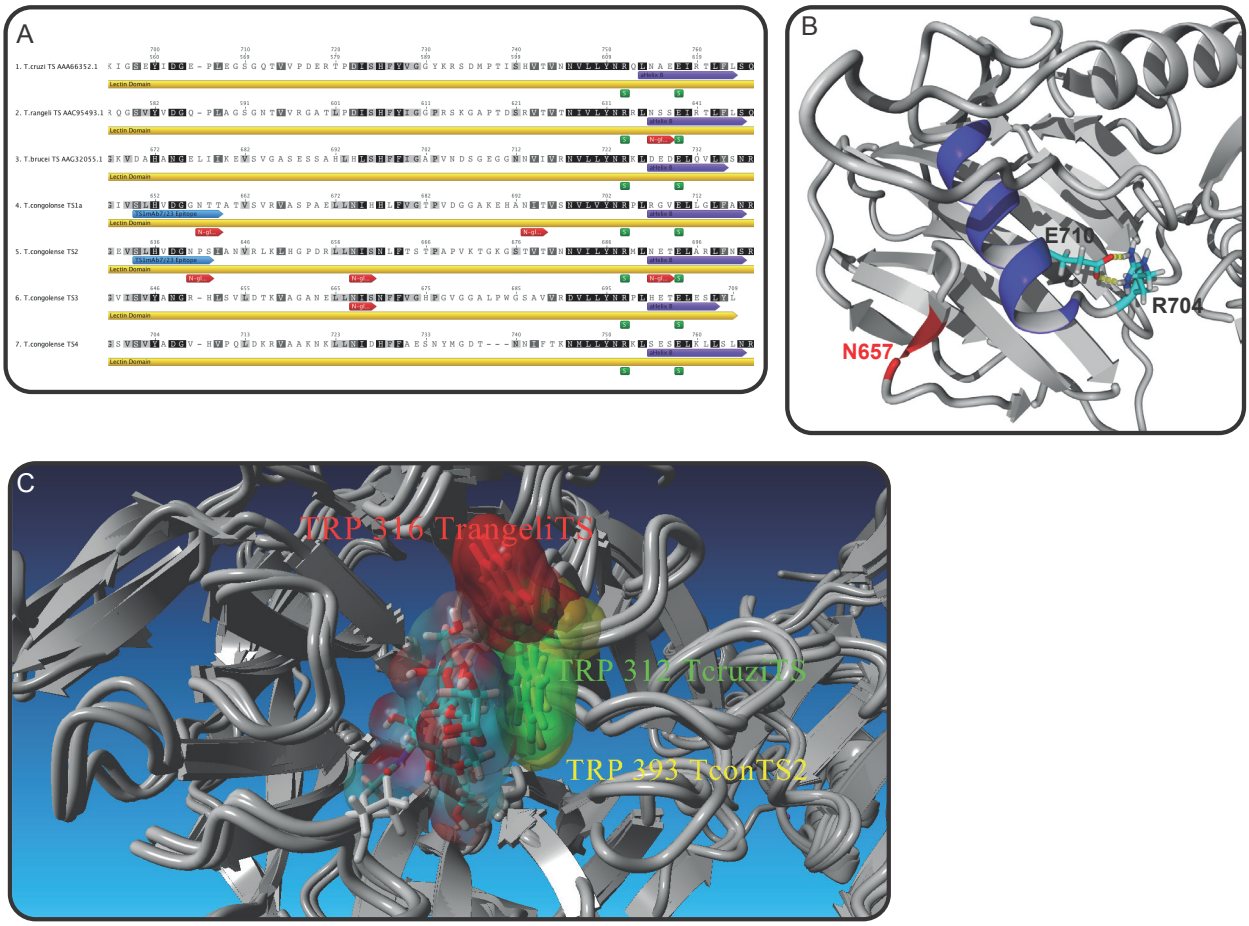

A
B
E710
R704
N657
C
TRP 316 TrangeliTS
TRP 312 TcruziTS
TRP 393 TconTS2

Table S1. List of primers used for cloning and *Eco*105I restriction site insertion.

Gene	Forward primer (restriction enzyme)	Reverse primer (restriction enzyme)
For bacterial expression by *E. coli* Rosetta (DE3) pLacI		
TconTS1a	GCAAGCTTCAGTGCTGCGACCAC ATG (*Hind*III)	CGGGATCCGTCGCTCCCAGGCAC ACG (*Bam*HI)
TconTS2	GCAAGCTTGCCCAGTGCATCTCAA CG (*Hind*III)	GCGGATCCAGACACGGGATGCAC ATC (*Bam*HI)
TconTS3	GCAAGCTTCTGGAAACGGACGA ACG (*Hind*III)	GCGGATCCGAGGTAAAGTGACTC CAGTTC (*Bam*HI)
TconTS4	GCGTCGACATCCTACAAGAAAGCT C (*Sal*I)	GCGGATCCCCTTGCTGCTCTTTTA AGTAAT (*Bam*HI)
For epitope mapping of anti TS mAb 7/23		
TconTS1a-CD (A)	GCAAGCTTCAGTGCTGCGACCAC ATG (*Hind*III)	CGGGATCCGTCACTCGATTCGAAT ATC (*Bam*HI)
TconTS1a-CD (B)		CGGGATCCATCCGAGCTGCCAGG ACCA (*Bam*HI)
TconTS1a-CD (C)		CGGGATCCATCACGATAATACGAG CCCT (*Bam*HI)
TconTS1a-CD (D)		CGGGATCCGTCCTGTGCCTTCCA CACC (*Bam*HI)
TconTS1a-CD (E)		CGGGATCCGTCCACAAGGCGCAC AAGG (*Bam*HI)
TconTS1a-LD (F)	GCAAGCTTGACGAGCTGAAAAGC (*Hind*III)	CGGGATCCGTCGCTCCCAGGCAC ACG (*Bam*HI)
TconTS1a-LD (G)	GCAAGCTTAACTGCCTCCCGGGC (*Hind*III)	
TconTS2-LD	GCAAGCTTCTGGAGGATGAGATG GAGG (*Hind*III)	GCGGATCCAGACACGGGATGCAC ATC (*Bam*HI)
TconTS3-LD	GCAAGCTTCTAGAAGACGAGCTG GAAAGC (*Hind*III)	GCGGATCCGAGGTAAAGTGACTC CAGTTC (*Bam*HI)
TconTS4-LD	GCGTCGACCTCGCTGACGAACTG AAG (*Sal*I)	GCGGATCCCCTTGCTGCTCTTTTA AGTAAT (*Bam*HI)
For *Eco*105I insertion		
TconTS1a*-CD (*Eco*105I)	GCAAGCTTCAGTGCTGCGACCAC ATG (*Hind*III)	GCTACGTAATCGCCCGGGAGGC (*Eco*105I)
TconTS1a*-LD (*Eco*105I)	GCTACGTAAAATATGATCCCGGG (*Eco*105I)	CGGGATCCGTCGCTCCCAGGCAC ACG (*Bam*HI)
TconTS2*-CD (*Eco*105I)	GCAAGCTTGCCCAGTGCATCTCAA CG (*Hind*III)	GCTACGTATTTGTTCAGTTGAC (*Eco*105I)
TconTS2*-LD (*Eco*105I)	GCTACGTAAAGCGCCGGAGCGGC (*Eco*105I)	GCGGATCCAGACACGGGATGCAC ATC (*Bam*HI)
TconTS3*-CD (*Eco*105I)	GCAAGCTTCTGGAAACGGACGA ACG (*Hind*III)	GCTACGTAACCATCCGGTGAGG (*Eco*105I)
TconTS3*-LD (*Eco*105I)	GCTACGTAGATTATACTGAGGG (*Eco*105I)	GCGGATCCGAGGTAAAGTGACTC CAGTTC (*Bam*HI)
TconTS4*-CD (*Eco*105I)	GCGTCGACATCCTACAAGAAAGCT C (*Sal*I)	GCTACGTACCCGGTAGTTGCAG (*Eco*105I)
TconTS4*-LD (*Eco*105I)	GCTACGTAGATGGCAGCGATTGC (*Eco*105I)	GCGGATCCCCTTGCTGCTCTTTTA AGTAAT (*Bam*HI)

Endonuclease restriction sites are underlined.

4.

Summarising discussion

4 Summarising discussion

Trypanosomes are flagellate protozoan parasites belonging to the order of kinetoplastida and are mainly found in Latin America and Sub-Saharan Africa. They develop in mammalian hosts and insect vectors, although some trypanosomes cause fatal diseases in human, comprising American Chagas' disease caused by *T. cruzi* [1] and Human African Trypanosomiasis (HAT or sleeping sickness in human) caused by *T. brucei gambiense* and *T. brucei rhodesiense* [2], as well as Animal African Trypanosomiasis (AAT locally also termed nagana) caused by *T. brucei brucei, T. vivax* and *T. congolense* [3], characterised by similar symptoms and pathogenesis. During their life cycle parasites utilise different very efficient strategies to evade immune response, ensure their survival and effectively establish infection in host and vector [4-6]. These strategies comprise antigenic variation and the expression of trans-sialidases (TS).

TS are generally membrane bound enzymes catalysing the regio and stereo specific transfer of terminal sialic acids (Sia) from host sialo-glycoconjugates to galactose or *N*-acetylgalactosamine residues of parasites surface glycoproteins, including mucins [7] and procyclins such as *T. congolense* GARP [8] and *T. brucei* PARP [9]) as well as glycosylphosphatidylinositol (GPI)-anchors [10-12]. Since trypanosomes are unable to synthesise Sia *de novo*, they utilize TS to cover their glycocalyx with Sia, resulting in a negatively charged surface, protecting parasites from proteolytic digest [5,9,13] and to generate ligand structures for host cell interactions [14]. Both mechanisms are essential for the establishment of infection and pathogenesis of trypanosomiasis, subsequently characterising TS as fundamental virulence factor in trypanosomes [6,15]. In addition, it has been shown that TS is expressed in procyclic insect trypomastigotes [11,16], as well as in bloodstream forms of the parasite [17-19]. However, many studies have been concentrated on *T. cruzi* TS, investigating the catalytic mechanism as well as the architecture of the active centre in detail [20-25], also in respect to develop new strategies for the synthesis of potent TS inhibitors [26]. Nevertheless, no specific and satisfying efficient inhibitor has been established so far. In contrast to Latin American TcruTS, only little was known about African TconTS before this study was initiated [27,28].

The main subjects of this work focused on the identification, expression and biochemical characterisation of TconTS to provide new insight into substrate specificities, as well as enhanced knowledge about the biological function. Primal database search using the partial sequences of TS-form 1 and 2, published by Tiralongo *et al.* [28], revealed 14 related active and 3 inactive TconTS members in total. All active TconTS have been subsequently

cloned and expressed in eukaryotic fibroblasts. Different biochemical and glyco-bioanalytical techniques were employed to characterise these recombinant TconTS variants and revealed significant differences in their overall enzymatic properties. Besides that, homology model structures for each TconTS comprising CD and LD were calculated, which provided a more detailed insight into the catalytic centre. In comparison to previously published data on the active site of TcruTS, striking differences and essential similarities for several TconTS have been observed, which appear to be related to their enzymatic properties determined during the course of this work.

Interestingly, separate phylogenetic analysis of TconTS catalytic domains (CD) and lectin-like domains (LD), surprisingly revealed a close relationship of the more active, as well as the less active enzymes, when aligning LDs compared to a different situation when aligning CDs. This fundamental finding indicated a possibly more relevant role of LD in TconTS enzymatic activities than previously thought. Thus, a major subject of this work has focused on the biochemical characterisation of TconTS-LDs and their possible influence on catalytic activities, since only very limited data has been available in literature and even no carbohydrate binding activities or potential functions had been demonstrated for TconTS-LD. Within this study, for the first time carbohydrate recognising activities were clearly demonstrated for TconTS-LD with distinct ligand specificities. Surprisingly, mannose-containing oligosaccharides have been identified as prominent ligands for at least two TconTS-LDs, namely TconTS1-LD and TconTS2-LD [29] (Chapter 3.4). This was unexpected, since previous studies have clearly demonstrated that mannose residues are no substrates for sialylation by TconTS-CD [28,30]. Thus, the data presented in Chapter 3.4 indicated a different, yet unknown, biological function of TconTS [29]. Along this line, our oligomerisation studies with TconTS1 and TconTS2 are of particular interest, since they have clearly demonstrated that higher order complexes are formed via *N*-linked oligomannosyl oligosaccharide dependent interactions. They are most likely mediated via TconTS-LD, which in turn might also be a natural biological function, since native TS have been found to be glycosylated with oligosaccharides of the high-mannose type [10,31]. Schenkman *et al.* had isolated native TcruTS from cultured trypomastigotes using ConA affinity chromatography, indicating the presence of high-mannose glycans on TcruTS [10]. Furthermore, Engstler *et al.* had isolated native TbruTS from cultured trypomastigotes in a similar way [32].

Based on to the mannose binding activity of LDs observed [29] (Chapter 3.4), we wanted to investigate the potential influence of LD on TconTS enzyme activities. Using a genetic approach, established within the scope of this thesis, the lectin domains between

different active TconTS have been swapped and enzymatic activity data of recombinant TconTS chimeras revealed significant differences in enzyme overall behaviour, TS and SA activities, compared to wild type TconTS (Chapter 3.5).

Binding and inhibition assay established by Bock and Kelm [33], initially utilised for siglec interaction studies, was used as a template method to develop a TconTS-LD binding/inhibition assay, which allows the characterisation of TconTS-LD binding specificity to high-mannosylated glycoproteins as potent ligand structures [29] (Chapter 3.4). Further details will be discussed in following paragraphs.

4.1 Identification and diversity of TconTS gene products

First partial characterisation of isolated native TconTS from procyclic *T. congolense* trypomastigotes cultures was described by Tiralongo *et al.*, who isolated two distinct TS preparations, termed TS-form 1 and TS-form 2, which exhibited different TS and sialidase activities [27,28]. Based on the partial sequence of *T. congolense* TS1, obtained in these studies, open reading frames for 11 TconTS1 variants have been identified in genomic DNA from *T. congolense* STIB 249 which were cloned and sequenced [34] (Chapter 3.1). Corresponding DNA sequences characteristic for these variants were also identified in the Welcome Trust Sanger Institute (WTSI) database of *T. congolense* (strain IL3000). Obtained results revealed 96.3 % amino acid sequence identity among TconTS1 variants (TconTS1a - TconTS1j). Interestingly, the amino acid differences, 25 in total (8 in LD and 17 in CD), are not evenly distributed over the whole protein sequence, but are found concentrated in a few regions of the protein [34] (Chapter 3.1).

Besides the 11 TconTS1 variants 6 additional TS-like genes have been identified from the WTSI database. Three out of these were suggested to encode for inactive TconTS, since they exhibit drastic substitutions of catalytic active amino acid residues (D150 to E or K, E324 to N, Y or S and Y438 to F, L or H) at the catalytic centre besides several others, essential for substrate binding and enzyme activity. Furthermore these genes share only up to ~30 % amino acid sequence similarity compared to TconTS1 and therefore termed as TconTS-Like1, TconTS-Like2 and TconTS-Like3, respectively [35] (Chapter 3.2). In contrast to that, the other three TconTS genes were found to be more closely related sharing about 42 – 49 % overall amino acid sequence similarities. Almost all crucial amino acid residues essential for catalysis and substrate binding were found to be well conserved among all three genes and therefore, these were expected to also encode for active TconTS enzymes, subsequently termed TconTS2, TconTS3 and TconTS4, respectively [35] (Chapter 3.2).

Amino acids, which had been identified to be relevant for TS catalytic activities and substrate interaction [24,25] are mostly conserved among all TconTS variants reported in this work [34,35] (Chapter 3.1 and 3.2). In the following paragraph amino acid residues of TconTS1 will be used as reference to discuss their essential role on enzymatic activities. Catalytically active residues, such as D150 (acid/base catalyst corresponding to D59 in TcruTS) catalysing the cleavage and formation of Sia-Gal gylcosidic linkage, E324 and Y438 (correspond to E230 and Y342 in TcruTS) representing the acid/nucleophile pair involved in the formation of covalent sialyl-enzyme intermediates, are well conserve among all 14 TconTS variants [34,35] (Chapter 3.1 and 3.2). Also R126, R339 and R410 (correspond to R35, R245 and R314 in TcruTS) forming the arginine triad, stabilising the carboxyl group of Sia and maintaining its required positioning for catalysis, as well as D188, W212, Q289, D341 and E454 (correspond to D96, W120, Q195, D247 and E357 in TcruTS), known to be involved in Sia stabilisation through hydrogen bond formation [23-25,36], and other relevant residues at the catalytic centre of TconTS variants are mainly well conserved [34,35] (Chapter 3.1 and 3.2). In addition, Y211 (correspond to Y119 in TcruTS) being part of the acceptor substrate binding-site and involved in stacking interactions (π-π*) with the cyclohexane ring of the Gal or aglycon moiety [24,25,36,37], occurs in all TconTS1 forms [34] (Chapter 3.1), as well as in TconTS3 and TconTS4 (corresponding to Y203 in TconTS3 and Y268 in TconTS4) but is substituted to a proline residue (P196) in TconTS2 [35] (Chapter 3.2), which will be discussed in a subsequent paragraph of this Chapter.

In addition, W312 in TcruTS, which has been demonstrated to be part of the donor substrate binding-site (π-π* interactions to lactose), is substituted by Y408 in all TconTS1 variants and Y463 in TconTS4 [34,35] (Chapter 3.1 and 3.2), but still maintaining potential stacking interactions with the cyclohexane ring of the donor substrate. The consistent presence of all these amino acid residues in TconTS variants essential for enzymatic activity provided strong evidence that these residues are subjects of a high evolutionary conservation pressure. According to that, a mutation, which might negatively influence TS activity and therefore consequently also affect parasite survival, would not lead to an evolutionary dominant species.

However, also drastic amino acid substitutions have been observed in TconTS enzymes. One striking substitution represents the exchange of R144 in TconTS1 (correspond to R53 in TcruTS) to a cysteine residue, only observed for the variant TconTS1g [34] (Chapter 3.1). This arginine residue has been demonstrated to form hydrogen bonds to the C-4 hydroxyl group of Sia in TcruTS [25]. Thus, R144 has been assumed to be involved in donor substrate

binding and possibly contribute to Sia stabilisation in the Michaelis-complex and covalent intermediate, respectively (Chapter 1.5.4, Figure 5). A more detailed discussion regarding the possible influence of the substitution R144C on TconTS1g enzyme activity is provided in the subsequent Chapter 4.2.1.

Another drastic substitution represents the exchange of the well-conserved Y211 in TconTS1, TconTS3 and TconTS4 (corresponding to Y203 in TconTS3 and Y268 in TconTS4) to a proline residue (P196) in TconTS2 [35] (Chapter 3.2). As mentioned before, this tyrosine residue is part of the acceptor substrate binding-site and has been suggested to stabilise the Gal or aglycon moiety through stacking interactions with the cyclohexane ring or the respective aryl substituent (e.g. in MU-Lac and MU-Gal [28]). Mutagenesis experiments on TcruTS exchanging the corresponding Y119 to a serine residue, revealed a drastic decrease in both, TS and sialidase activities. Similar to TconTS2, also in TranSA this tyrosine is replaced by a serine residue (S141), which was first supposed to be one of the main reasons for the lack of TS activity in TranSA. However, Buschiazzo *et al*. demonstrated that the mutation of S141 to tyrosine did not introduce TS activity to TranSA or enhanced its sialidase activity [23]. However, this apparent discrepancy could be explained within the context of the overall structure of the catalytic pocket. For that purpose, of particular importance is the study of Paris *et al.*, who demonstrated the importance of the orientation of W312 in TcruTS for enzymatic activity [38]. As already mentioned earlier, W312 is part of the acceptor substrate binding-site and needs a well-defined orientation for efficient substrate binding and subsequent catalysis. Due to a single substitution of a proline to glutamine (P283 in TcruTS corresponding to Q305 in TranSA) in the subjacent loop in TranSA, W334 (corresponding to W312 in TcruTS) adopts a completely different orientation compared to W312 of TcruTS, because of steric hindrance by Q305 (Chapter 1.5.4, Figure 6). Interestingly, the mutation Q305P in TranSA significantly increased the sialidase activity on MU-Neu5Ac and 3'SL, indicating the importance and indirect role of Q305 on substrate binding and enzyme activity [38]. Furthermore, this proline residue was found to be well conserved in TconTS2 as well as in all other TconTS family members [34,35] (Chapter 3.1 and 3.2), leading to the suggestion that the adjacent tyrosine (Y408 in TconTS1 and Y463 in TconTS4) or tryptophan residues (W393 in TconTS2 and W400 in TconTS3) occupy the positions required for efficient substrate binding, similar to W312 in TcruTS [24].

Typical sialidases and also TS contain the characteristic FRIP (amino acid single letter code) motif or conservative modifications thereof and one of the Asp boxes located early at the N-terminus of CD were also found in all TconTS variants. R35 (numbering of TconTS1)

of the FRIP motif is one of the arginine residues forming the arginine triad, thus positioning FRIP as one of the key elements in Sia binding and influencing enzymatic activity, as it has been proposed for other TS [39-41].

Concerning these findings and all the other amino acid differences in TconTS1-4, it was impossible to predict the exact enzymatic properties for each of the 14 active TconTS forms just by comparing amino acid sequence data with investigations reported for TcruTS and TranSA. Therefore, it was of fundamental importance to determine enzymatic activities and kinetic parameters such as K_M and v_{max} values to investigate potential structure-function relations of TconTS. Since no crystal structure for TconTS or any other African TS is available so far, TconTS structure homology models were calculated based on the crystal structure of TcruTS [24] to get deeper insight into understanding catalytic activities, which were experimentally determined, in the context of structural differences to TcruTS [34,35] (Chapter 3.1 and 3.2).

4.2 Biochemical characterisation of recombinant TconTS

All 14 identified genes encoding for active TconTS (11 TconTS1 variants, TconTS2, TconTS3 and TconTS4) were cloned into modified pDEF, a eukaryotic expression vector, and expressed in fibroblasts as recombinant proteins [34,35] (Chapter 3.1 and 3.2). After purification of recombinant TconTS from tissue culture supernatant of CHO-Lec1 cells, enzymes were incubated with fetuin and lactose as suitable Sia donor and acceptor substrates respectively [34,35] (Methods of Chapters 3.1 and 3.2). To determine enzymatic activities, reaction products, such as 3'-sialyllactose (3'SL) and free Neu5Ac were quantified employing a high performance liquid chromatography (HPLC) based technique specific for carbohydrate analysis termed high performance anion exchange chromatography with pulsed amperometric detection [34] (HPAEC-PAD, description under Methods in Chapter 3.1). This method allows detection and quantification of the TconTS reaction products 3'SL and Neu5Ac from the same chromatogram with a detection limit of about 10 pmol. The kinetic parameters K_M and v_{max} for each substrate were determined employing a dilution series of the appropriate substrate under standard conditions. Substance-specific constants, K_M and v_{max} were used to compare enzymatic activities of different TconTS. Besides fetuin also other molecules have been used as Sia donor substrates, such as 3'SL, pNP-Neu5Ac, MU-Neu5Ac and sialo-glycoconjugates from human blood serum as well as other acceptor substrates besides lactose, such as Gal, MU-Gal and sialidase-treated fetuin (asialofetuin, ASF) [34,35] (Chapter 3.1, 3.2

and 3.3). In addition, sialidase activities of TconTS were also determined in the absence of any suitable acceptor substrates.

4.2.1 The diversity in enzymatic activities of TconTS

Using standard conditions, similar overall TS activities for all 11 TconTS1 variants were determined except for one, termed TconTS1g [34] (Chapter 3.1). Indeed, this variant showed a significantly lower 3'SL production of about 70 % after 24 h incubation using standard conditions, compared to other TconTS1 variants (Figure 1 A, Chapter 3.1). As mentioned in the previous Chapter 4.1, TconTS1g exhibits a unique substitution R144C close to the catalytic centre [34] (Chapter 3.1). It can be suggested that this exchange may lead to destabilisation of Sia in the Michaelis-complex due to the loss of hydrogen bonds formed by R144 with the C-4 hydroxyl group of Sia, which has been previously proposed for TcruTS [25]. A cysteine residue at that position 144, although in principle it is also able to form hydrogen bonds, is expected not to reach the C4-hydroxyl group of Sia for efficient hydrogen bond formation according to the homology model. The relatively low V_{max} value of TconTS1g for fetuin-bound Sia compared to that of TconTS1b and TconTS1e-1 supported this hypothesis (Table 1). However, it remains unclear whether this mutation represents the main reason for the reduced TS activity, and if so, what mechanistic explanation might be reasonable. However, the remaining TS activity of TconTS1g indicated that the substitution R144C has a pronounced effect on the enzyme activity, but it also shows that R144 is not essential for the catalysis of Sia transfer.

Table 1: Kinetic parameters V_{max} and K_M of TconTS1b and TconTS1g.

	Fetuin-bound Sia (donor substrate)		Lactose (acceptor substrate)	
	V_{max} (pmol 3'SL/(min x mg TS))	**K_M** (µM)	**V_{max}** (pmol 3'SL/(min x mg TS))	**K_M** (µM)
TconTS1b	7.9 ± 0.3	359 ± 45	4.3 ± 0.1	1683 ± 101
TconTS1e-1	7.6 ± 0.5	1617 ± 223	2.1 ± 0.1	727 ± 48
TconTS1g	0.08 ± 0.01	3025 ± 490	-	362 ± 50

Data for TconTS1b and TconTS1e-1 were taken from Koliwer-Brandl *et al.* [34] (Chapter 3.1). Kinetic para-meters for TconTS1g were calculated using the Michaelis-Menten kinetics (Figure 1 A and B) employing Prism6. Data points are means ± standard deviation of triplicates.

Interestingly, K_M values of TconTS1g determined for fetuin-bound Sia (3025 µM) and lactose (362 µM, Figure 1 B and C) are more related to those of TconTS1e-1 than TconTS1b (Table 1), indicating a shared structural similarity also influencing the catalytic activities.

Along this line, we noticed that besides the substitution R144C most of the critical amino acid differences comparing TconTS1b with TconTS1e-1 and TconTS1g, are located at the LDs [34] (Chapter 3.1 Figure 3). Thus, the LD of TconTS1e-1 is more related to the LD of TconTS1g than to the LD of TconTS1b. Furthermore, it can be seen that amino acid residues 599-602 represent a significant structural element at the LD comprising the most critical and subsequent changes [34] (Chapter 3.1 Figure 3). A further discussion regarding this significant structure element at TconTS-LD is provided in the following Chapter 4.3.3.

Since all TconTS1 variants, besides TconTS1g, showed similar overall enzyme activities under standard conditions, a more detailed biochemical characterisation of just two TconTS1 variants, namely TconTS1b and TconTS1e-1, which mainly differ in their C-terminal LD, was performed to investigate potential variations in enzymatic activities. In addition also substrates, other then fetuin and lactose have been utilised in further enzyme activity experiments [34] (Chapter 3.1). Indeed significant differences were observed, when fetuin was used as donor substrate under standard reaction conditions, whereas both TconTS1 variants produced different amounts of 3'SL (product of TS activity) in 30 min. Nevertheless, each enzyme generated almost the same amount of 3'SL within 20 h of incubation, when equilibrium was reached.

These findings strongly suggest different substrate affinities of TconTS1 variants for the various sialylated oligosaccharides attached to fetuin, which is also reflected by the distinct K_M values of TconTS1b (359 μM) and TconTS1e-1 (1617 μM) for fetuin-bound Sia [34] (Chapter 3.1). In addition, even no TS or sialidase activity was detected using MU-Neu5Ac with these two recombinant TconTS1 variants. These observations are in contrast to data reported by Tiralongo *et al.*, who determined a K_M for Gal-MU approximately 10 fold higher compared to TconTS1b and TconTS1e-1 [34] (Chapter 3.1), as well as hydrolytic activity on MU-Neu5Ac, when using native TS-form 1 and TS-form 2, preparations isolated from procyclic *T. congolense* trypomastigote cultures [28]. It is important to note that these isolated, native TS-forms do not correspond to TconTS1 and TconTS2, described in this work.

However, it should also be mentioned that the presence of a mixture of different TconTS variants in TS-form 1 and/or TS-form 2 preparations has been proposed [34] (Chapter 3.1), since some TconTS enzymes may exhibit higher affinities for MU-Neu5Ac or pNP-Neu5Ac and efficiently transfer Sia to MU-Gal or hydrolyse the donor substrate. Nevertheless, no sialidase activities were detected neither for TconTS1b nor TconTS1e-1 after 30 min incubation using fetuin and lactose as substrates [34] (Chapter 3.1). Only when

enzymes were incubated for 20 h clear sialidase activities were observed, indicated by the production of free Neu5Ac in the reaction. Interestingly, free Neu5Ac was already detected very early in TS reactions with 3'SL as donor and ASF as acceptor substrate (Sia transfer from 3'SL to asialofetuin). This finding together with the occurrence of free Neu5Ac only in extended reactions using fetuin and lactose, suggested that free Sia as a side product generated by the hydrolysis of 3'SL, if it is present in sufficient amounts accumulated at later stages during standard TS reaction with fetuin as donor and lactose as acceptor substrates. Without any suitable acceptor substrate, both, TconTS1b and TconTS1e-1, showed similar hydrolytic activity on fetuin as well as on 3'SL. Compared to TconTS1b and TconTS1e-1, more drastic differences in enzyme activities have been observed for the other three TconTS gene family members TconTS2, TconTS3 and TconTS4.

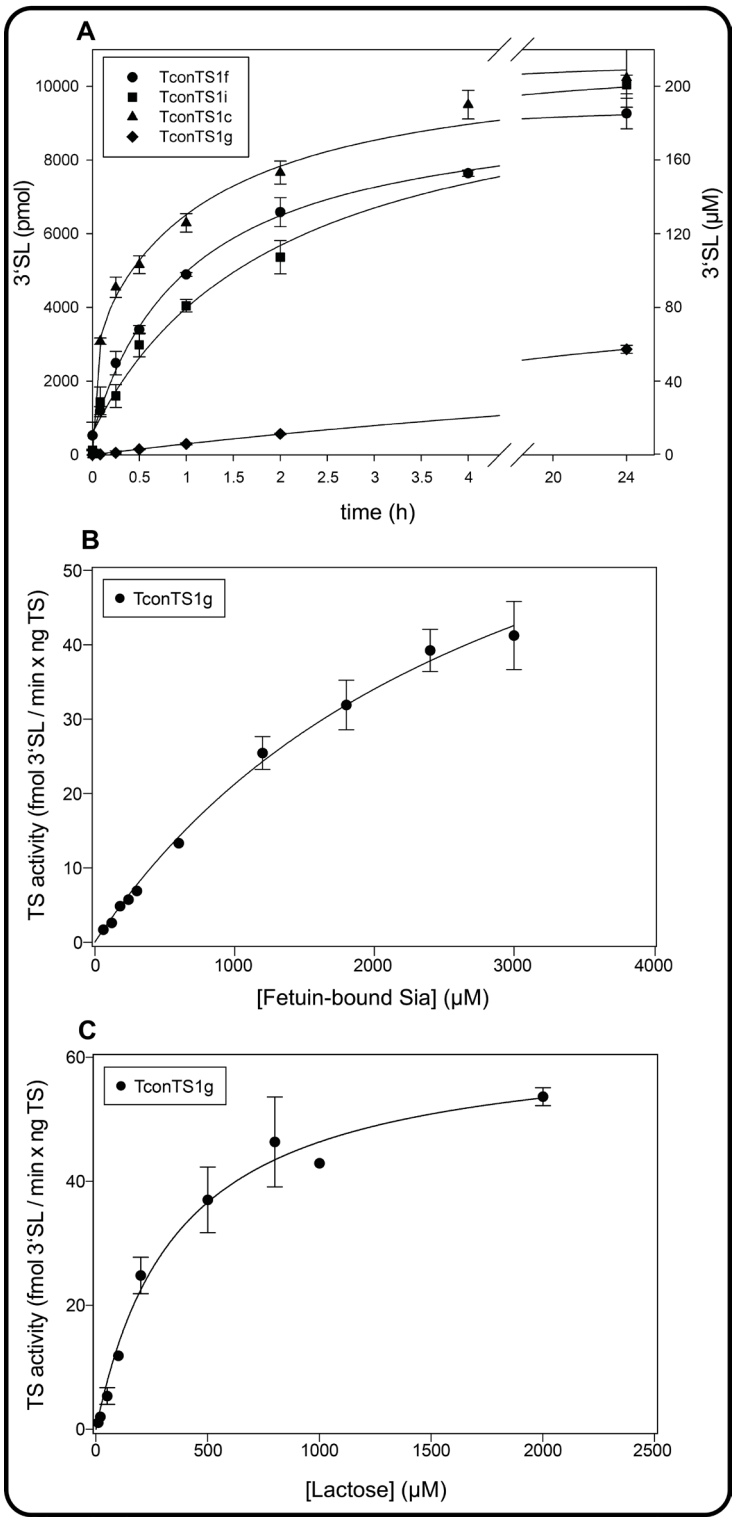


Figure 1: Time dependency of TconTS1 variants and Michaelis-Menten kinetics of TconTS1g reactions. **A)** Time dependent enzymatic activities of TconTS1c, f, g and i were determined using standard conditions (100 µg fetuin, 2 mM lactose, 50 ng TconTS incubated at 37°C) and different incubation times. **B)** Varying concentrations of fetuin-bound Sia were used as donor substrate and incubated for 30 min with 100 ng TconTS1g under standard conditions. **C)** 600 µM fetuin-bound Sia was used as donor substrate with varying amounts of lactose and 100 ng TconTS1g under standard conditions. Data points are means ± standard deviation of triplicates. 3'SL production was quantified employing HPAEC-PAD analysis [34] (Chapter 3.1).

As discussed in Chapter 4.1, all four TconTS1 through TconTS4 contain

almost all amino acid residues critical for enzymatic activity, but more drastically differ in their overall amino acid sequence similarities. To determine enzymatic activities, recombinant TconTS2, TconTS3 and TconTS4 were incubated with fetuin and lactose as donor and acceptor substrates under standard conditions. TS activities were detected for all three enzymes, clearly presenting them as active TS, although significant variations in catalytic activities were observed. High 3'SL production demonstrated that TconTS1 and TconTS2 are the most active enzymes with about 100- to 1000-fold higher specific Sia transfer activity compared to TconTS3 and TconTS4 under the same reaction conditions [35] (Chapter 3.2). Surprisingly, TconTS2 exhibited the highest specific TS activity among all characterised TconTS. Tyrosine Y211 in TconTS1, which is assumed to be part of the acceptor substrate binding-site (demonstrated for TcruTS [25]), is substituted to a proline residue (P196) in TconTS2 (Figure 2). This drastic change was suggested to decrease TS activity in TconTS2 due to the loss of substrate interaction, whereas experimental data demonstrated the contrary.

However, since all other amino acid residues involved in Sia transfer activity of TconTS2 are conserved relative to those of the also highly active TconTS1 (except the less drastic substitution of Y408 in TconTS1 to W393 in TconTS2) further, yet unknown reasons have to be responsible for this unexpected observation. One possible explanation arose from results obtained by Amaya *et al*. regarding the Y119 in TcruTS (corresponding to Y211 in TconTS1 and P192 in TconTS2), which has been demonstrated to contribute to Sia stability in the covalent intermediate through hydrogen bond formation to the C-9 hydroxyl group, mediated by subsequent conformational change of the catalytic centre upon substrate binding (allosteric effect) [25,37]. Due to the presence of P192, such a hydrogen bond formation cannot occur in TconTS2 but might be partially compensated by hydrogen bond formation of the conserved residues W197 and Q274 in TconTS2 (corresponding to W120 and Q195 in TcruTS), which have also been demonstrated to contribute to the hydrogen bond network of the glycerol side chain of Sia in TcruTS [25]. Nevertheless, this substitution in the highly active TconTS2 indicates that a tyrosine at this position is not required for enzymatic activity, but may even contribute to the high TS activity of TconTS2.

Along this line, it seems obvious to look for other structural elements, which might be responsible for the relatively high catalytic activity of TconTS2. In addition, similar reasons might also explain the relatively low TS activities of TconTS3 and TconTS4, since both enzymes comprise the well conserved composition of all critical amino acid residues at the catalytic centre [35] (Chapter 3.2).

Another aspect regarding the specific enzymatic activities of TconTS deals with significant environmental changes during parasites life cycle in mammalian host and tsetse vector, which is for example characterised by an shift from pH 7 (mammalian blood stream) up to pH 10 (tsetse midgut [4]). Therefore, it was important to investigate pH dependency of TconTS enzymatic activities, which may provide further insight into the catalytic properties of TconTS1 through TconTS4 and moreover, how these may relate to their potential role in different host systems.

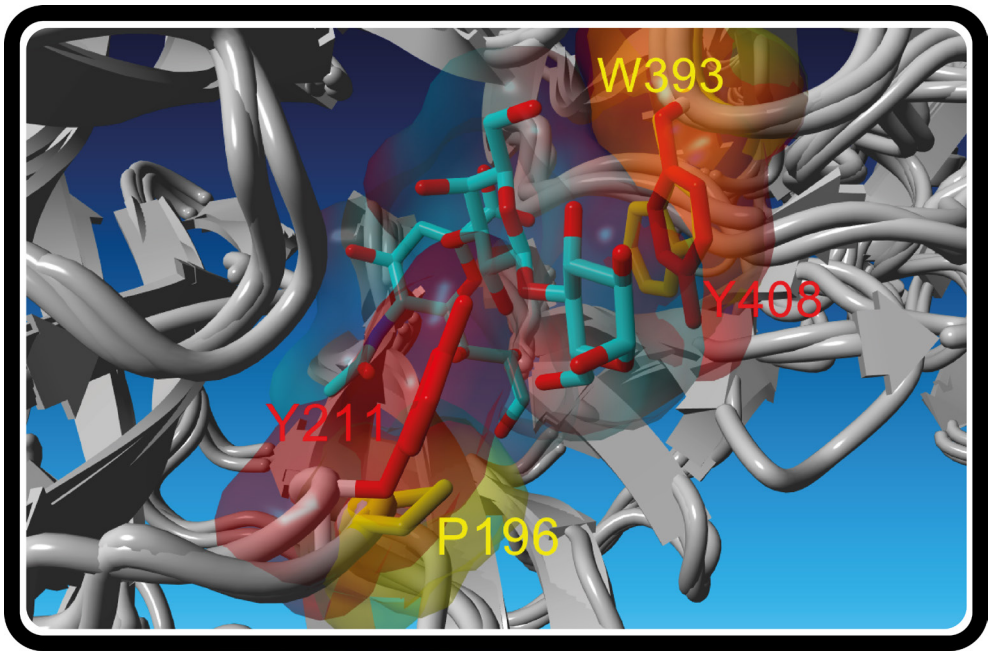


Figure 2. Structural comparison of TconTS1 (red coloured amino acid residues) and TconTS2 (yellow coloured amino acid residues). The superimposed homology models of TconTS1 in complex with 3'-sialyllactose (coloured in light blue) and TconTS2 were generated using the crystal structure of TcruTS (PDB code: 1SOI) as template employing the software Yasara as described under Methods in chapter 3.1. A detailed illustration of the catalytic centres of both enzymes is shown. The conserved tryptophan residue W393 in TconTS2 (yellow, corresponds to 334 in TcruTS) is substituted to a tyrosine (red Y408) and occupies the same donor substrate binding-site in TconTS1 compared to TconTS2. The conserved tyrosine at position 211 (red Y211, corresponds to Y119 in TcruTS) in TconTS1, which stabilise the acceptor substrate via stacking interactions is substituted to a proline (yellow P196) in TconTS2. Molecular Van der Waals surfaces of the illustrated residues and 3'-sialyllactose are shown in transparent spheres, respectively.

Significant differences in TS and sialidase activities of TconTS were determined using fetuin and lactose as donor and acceptor substrates under standard conditions at different pH values (Chapter 3.3). TconTS1, TconTS3 and TconTS4 exhibit maximum TS activities at pH 7 and a more or less drastic decrease of their activities towards higher pH values. These findings fit to possible roles as active TS in trypomastigote bloodstream forms, since the

present pH value of 7.4 in blood is very close to the optimal pH of 7.0 determined for TconTS1, TconTS3 and TconTS4 (Chapter 3.3 Figure 1). Interestingly, in contrast to that, TconTS2 showed a maximum TS activity at pH 9, indicating a potential predominant role as TS of procyclic parasites in tsetse midgut with its alkaline environment of about pH 10. Furthermore, in the absence of lactose, TconTS4 showed relative significantly enhanced sialidase activity (appr. by the factor of 7) at higher pH values (maximum sialidase activity at pH 9) compared to the sialidase activity determined for the other enzymes, although the amount of free Sia produced by TconTS4 is still less compared to that produced by TconTS1 and TconTS2 under same conditions, respectively. However, if lactose is present, sialidase activity of TconTS4 is almost abolished at pH 9. Nevertheless, this has suggested that TconTS4 might be a more prominent sialidase in tsetse midgut, if expressed in procyclic trypomastigotes. Nevertheless, it should be noted that the precise expression patterns of TconTS in the parasites life cycle, including procyclic trypomastigotes and blood stream parasites, is still unknown so far.

In summary, it has been demonstrated that TconTS1 and TconTS2 exhibit the highest TS activities, whereas TconTS3 and TconTS4 have the lowest specific activities. Comparing Sia transfer with sialidase activities, TconTS2 has the lowest transfer/hydrolysis ratio and TconTS4 the highest among TconTSs under the reaction conditions used [34,35] (Chapter 3.1 and 3.2). Furthermore, it might be possible that the low Sia transfer activity observed for TconTS3 and TconTS4 on substrates tested, is due to higher specific affinities to other donor and/or acceptor substrates, such as blood glycoproteins and glycolipids, as well as parasites own glycoconjugates [35] (Chapter 3.2). In this context, it should be noted that the *T. congolense* procyclic glutamic acid/ alanine-rich protein (GARP), which is expressed on the parasites surface, has been demonstrated to be a suitable substrate for TconTS trans-sialylation [30], and was found to be relatively strong associated to isolated native TS-forms [28]. Thus, additional natural substrates for TconTS reaction have to be identified and characterised also to investigate the biological role of TconTS3 and TconTS4.

4.2.2 Cooperatively mediated specific sialidase activities of TconTS

Besides critical amino acids at the catalytic centre, which are mostly well conserved among TconTSs, other structural features have to exist, which modulate TS and sialidase activities in TconTS and possibly their orthologous from other trypanosomes.

One first surprising evidence for cooperativity between the different TconTS was obtained, when mixtures of different recombinant TconTS were tested for enzymatic activities

with MU-Neu5Ac as substrate. As mentioned earlier in this Chapter 4.2, recombinant TconTS1 and TconTS2 exhibit relatively poor hydrolysing activities on the synthetic donor substrate MU-Neu5Ac. A similar behaviour was observed for TconTS3 and TconTS4, which did not show any hydrolytic activities on MU-Neu5Ac. However, mixtures of TconTS1 or TconTS2 together with any of the other recombinant TconTS (e.g. TconTS3 or TconTS4) revealed relatively good hydrolytic activities (TT Gbem, personal communication). In contrast, a mixture of TconTS3 and TconTS4 was not able to hydrolyse MU-Neu5Ac under the same reaction conditions, leading to the suggestion that TconTS1 and TconTS2 are generally able to cleave synthetic substrates such as MU-Neu5Ac, but need the presence of another different TconTS, which somehow induce this or other activities of TconTS1 and TconTS2. This consideration is in good agreement with the hypothesis that not only one TconTS form is expressed at each stage of the parasites life cycle, but a composition of different active TS and possibly inactive proteins is responsible for the oveall TS and sialidase activities [19,35] (Chapter 3.2). Interestingly, as mentioned earlier in this Chapter, Tiralongo *et al.* demonstrated that native TS-form 1 as well as TS-form 2 isolated from *T. congolense* procyclic trypomastigotes were able to hydrolyse MU-Neu5Ac [28].

Tiralongo *et al.* also described an anti TS specific monoclonal antibody (anti-*T.congolense* TS mAb 7/23), which was generated against partially purified TS forms and recognised both, TS-form 1 and TS-form 2, but precipitated only 75% of the TS activity present in *T. congolense* cell culture supernatnats [28]. In addition, also no cross reactivity with recombinant TcruTS or TbruTS has been found [28]. It is important to note that this anti-TconTS mAb specifically recognises recombinant TconTS1 and TconTS2, but not TconTS3 and TconTS4 [34,35] (Chapter 3.1 and 3.2). In summary, these findings have suggested that at least the native TS-form 1, isolated by Tiralongo *et al.* might have been a mixture of TconTS1 and TconTS2, since the anti-TconTS mAb 7/23 obviously recognises an epitope shared by both enzymes, TconTS1 and TconTS2.

In summary, these observations together with the fact that the isolated native TS-forms 1 and 2 are able to hydrolyse MU-Neu5Ac provided good evidence that both isolated TS-forms most likely represent mixtures of TconTS1 and TconTS2 enzymes and that the *T. congolense* cell culture supernatnats possibly also contained TconTS3 and/or TconTS4. This would eb in good agreement with the hypothesis of a cooperatively regulated sialidase activity for MU-Neu5Ac and possibly also other synthetic and yet unknown natural substrates.

4.3 The lectin domain of TconTS and its influence on enzyme activities

The majority of studies available in literature regarding the catalytic mechanism and structure-function relationships between trypanosomal TS have been focused almost exclusively on the catalytic domain, mainly that of TcruTS [23-25,36,37,42,43]. In contrast, only very limited data was available addressing potential links between the diversity in catalytic activities of different TS and their lectin-like domain (LD) of these unusual enzymes. Obviously, differences in TconTS enzymatic activities observed for TconTS1 through TconTS4 in this work, cannot be exclusively explained by the composition of critical amino acid residues at the catalytic centre. For example, all 12 amino acid residues at the catalytic centre known to be essential for substrate binding and catalysis are completely identical in TconTS1 and TconTS4 [35] (Chapter 3.2).

This observation may lead to the expectation that both enzymes also exhibit similar catalytic activities, whereas TconTS1 and TconTS4 drastically differ in their enzymatic properties [34,35] (Chapter 3.1 and 3.2). When using fetuin and lactose as Sia donor and acceptor substrates under the same standard reaction conditions, TconTS1 was found to be highly active (TS activity = 4000 nmol/(min x mg TS), sialidase activity = 60 nmol/(min x mg TS)), whereas TconTS4 showed drastically lower TS (9 nmol/(min x mg TS)) and sialidase (2,7 nmol/(min x mg TS)) activities, corresponding to reductions by the factors of ~450 and ~22, respectively. These observations either indicated important roles of other amino acid residues at or near the catalytic centre, which are not directly involved in substrate binding and catalysis, but indirectly by possibly maintaining the flexibility of loops containing critical amino acids and thus contribute to the highly dynamic processes at the CD during catalysis, and/or point to different fundamental regions elsewhere in the protein, which have not been considered regarding differences in enzymatic activity so far. The later hypothesis is supported by the surprising results of a more detailed phylogenetic analysis of TconTS1 through TconTS4 and the TconTS-Like genes, together with TcruTS, TranSA, TbruTS and TvivTS genes [35] (Chapter 3.2). Besides the close relation of African active TS on one hand, and American active TS on the other, as well as the distant relation to TconTS-Like genes encoding for inactive TconTS, results surprisingly also revealed the grouping of highly active TconTS2 together with the less active TconTS3 and its more distant relation to the also highly active TconTS1. Interestingly, further phylogenetic analysis only considering TconTS-LDs, grouped the highly active TconTS1 and TconTS2 together, as well as the less active TconTS3 and TconTS4 in another branch [35] (Figure 3 B, Chapter 3.2). These surprising findings, for the first time indicated a functional relevance of LD and suggested their potential involvement

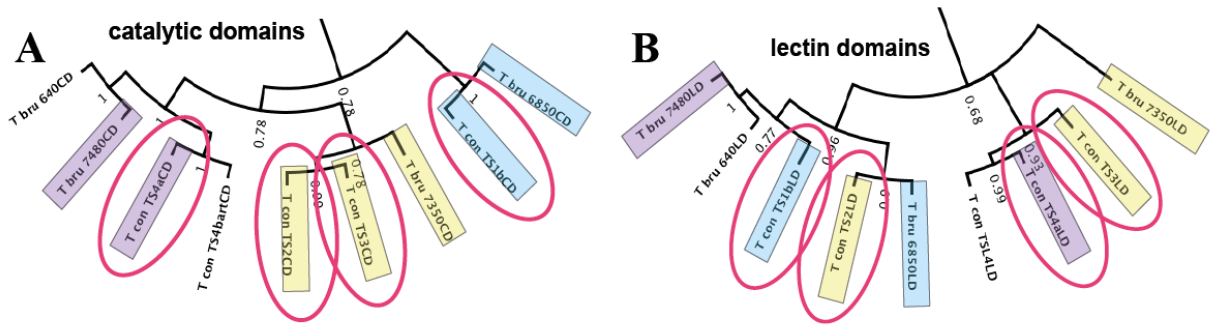


Figure 3. Phylogenetic analysis on DNA genes of catalytic and lectin domain of TconTS1-4. **A**) Phylogenetic analysis on TconTS-CDs. The two more inactive TconTS3 and TconTS4 group together with the highly active TconTS2, which is in turn surprisingly distantly related to the also very active TconTS1. **B**) Phylogenetic analysis on TconTS-LDs. TconTS-LD alignment shows TconTS1 and TconTS2 group together as well as TconTS3 with TconTS4. Active TconTS are labelled with red cycles. Phylogenetic trees were calculated as described under Methods in Chapter 3.2 [35]. The graphic was originally generated and provided by Prof. Dr. Sørge Kelm.

in modulating enzymatic activities. Since only very limited data were available about structure-function relations of the LD of trypanosomal TS, including their possible lectin activities, recombinant LDs from TconTS1 through TconTS4 were cloned, expressed and biochemically characterised [29] (Chapter 3.4).

4.3.1 TconTS-LD specific binding to different carbohydrate ligands

Specific binding of recombinant TconTS-LDs to several carbohydrate structures could be demonstrated for the first time, employing specific techniques optimised to analyse and characterise protein-ligand interactions, such as glycan array and STD NMR spectroscopy [29] (Chapter 3.4). Using the glycan array approach as a rapid and broad screening method to identify glycan structures as suitable TconTS-LD ligands, a full library of more than 300 different glycans was tested. A brief description regarding the principle of this method is provided in Chapter 3.4 [29].

Testing all four TconTS-LDs, predominantly recombinant TconTS2-LD was found to bind several galactosyl and lactosyl oligosaccharides, as well as some oligomannosyl glycans [29] (Chapter 3.4 Figure 2). In addition, TconTS1-LD was also found to bind galactosyl glycans but with significantly lower binding activity and decreased number of glycan structures, whereas no binding was determined neither for TconTS3 nor TconTS4 on the glycan array. Besides the observed lectin activity of TconTS2-LD for several galactosyl and lactosyl oligosaccharides, especially binding to mannose containing glycans was unexpected. Interestingly, it had been described that mannose is not a suitable acceptor substrate for trans-sialylation by TbruTS [11] and TconTS [28,30]. Thus, lectin activities of TconTS-LD to

mannosyl glycans indicate possible non-identified biological functions distinct from that of TconTS-CD.

The missing lectin activity observed for TconTS3-LD and TconTS4-LD might either indicate that they are not binding carbohydrates or that suitable ligand structures were not present on the glycan array. In addition, it should be noted that, in contrast to common C-type lectins [44-46], TconTS-LDs do not require divalent ions, such as Ca^{2+}, for efficient lectin activity, suggesting a different mode of ligand binding compared to those mannose binding lectins. In conclusion, these preliminary data already provided evidence for a novel, yet unknown function of TconTS.

STD NMR experiments were employed to further characterise TconTS-LD binding to mannosyl and lactosyl oligosaccharides using recombinant TconTS2-LD and α-D-lactose (Lac) as well as α1-3,α1-6-mannotriose as ligand structures [29] (Chapter 3.4). Results revealed that both ligands were clearly bound by TconTS2-LD, indicated by the relatively strong STD NMR signals for both molecules compared to the control spectra (off-resonance spectra) [29] (Chapter 3.4 Figure 3). Furthermore, the STD NMR signals for the β-GlcH2 and GalH4 protons of Lac provided good evidence that both monosaccharide units interact with TconTS2-LD. Similar observations were made for α1-3,α1-6-mannotriose [29] (Chapter 3.4 Figure 3). In summary, these data provided evidence for more extended carbohydrate binding sites in TconTS-LD. An important observation was made when adding an equimolar amount of Lac to the TconTS2-LD/α1-3,α1-6-mannotriose complex. No evidence for a competition of the two ligands was found, since (I) no reduction of the STD NMR signals of α1-3,α1-6-mannotriose occurred and (II) additional signals from Lac were detected. These surprising findings are very good evidence that both ligands bind simultaneously to thermodynamically distinct binding sites on TconTS2-LD. It will be interesting to extend such ligand competition experiments to further confirm and characterise these distinct binding sites and and the binding epitopes for both ligands in detail with series of related ligand molecules.

4.3.2 TconTS-LD binding to high-mannose N-glycans as a potential biological function?

Microtitre plate-based binding and inhibition assays had been developed by Bock and Kelm [33] to characterise siglec-ligand interactions. The principle of this method is generally similar to the classical ELISA (enzyme-linked immunosorbent assay), which is commonly used in biochemical applications for specific qualitative and quantitative analysis of antigens [47]. The primary antibody (capture antibody) in the so called sandwich ELISA is commonly

immobilised to a solid phase (surface of a microtitre plate) and utilised to "capture" the target antigen, which is in turn detected via a second antigen specific antibody, mostly conjugated to a selected detection system (Enzyme, fluorophore etc.). When using the siglec-binding assays, this primary antibody was replaced by a Sia containing binding partner, such as the highly sialylated glycoprotein fetuin, to quantify the binding of siglecs in the presence and absence of a potential inhibitor [33,48].

Utilising the principle of this method, a TconTS-LD binding and inhibition assay was established to further characterise the lectin activity of LDs to glycoproteins [29] (Chapter 3.4). The concept comprises the usage of immobilised, high-mannose glycans containing glycoproteins as potential TconTS-LD model binding structures, since mannosyl glycans have been demonstrated as relatively effective TconTS2-LD ligands. One major difference of this binding assay, in contrast to the glycan array performed in this work, was the usage of more complex *N*-linked oligosaccharides. Since glycoproteins expressed in CHO-Lec1 cells are glycosylated with *N*-linked oligosaccharides of the high-mannose type [49] recombinant proteins produced by these cells seemed to eb good candidates for TconTS binding and therefore, we used these as TconTS-LD binding target structures.

TconTS-LD binding assay results demonstrated specific binding activity of TconTS2-LD to immobilised Siglec-2Fc applied as high-mannose glycan carrying glycoprotein. The specificity of this interaction for glycan binding was confirmed by competitive inhibition with high-mannose oligosaccharides, which were enzymatically released from glycoproteins expressed by CHO-Lec1 cells [29] (Chapter 3.4 Figure 4). Interestingly, in contrast to that finding no clear inhibition of TconTS2-LD binding has been observed when using the monosaccharide α-methyl-mannopyranoside, which is a well known inhibitor for concanavalin A (ConA) [50], even at relatively high concentrations of about 50 mM [29] (Chapter 3.4 Figure S4). This is a good indicator for the proposed extended binding site of TconTS2-LD interacting with more than the final mannose residues. It also points towards a preference for oligo- or polyvalent ligands reported for other lectins, which have been postulated to prevent unspecific interactions leading to enhanced ligand selectivities [46,51].

Recombinant TconTS expressed in CHO-Lec1 cell lines are also glycosylated with *N*-linked oligosaccharides of the high-mannose type, which was demonstrated by ConA lectin blot analysis [29] (Chapter 3.4 Figure 5). *N*-glycosylation site predictions revealed eight to nine potential sites differently distributed in TconTS1 through TconTS4 with one site being conserved [29] (Chapter 3.4 Figure S3). However, it remains unclear which of the predicted *N*-glycosylation sites are used for glycosylation in post-translational modification processes.

Furthermore, it should be noted that native TbruTS was shown to contain high-mannose oligosaccharides, since it binds to ConA [52], although the glycosylation pattern of these enzymes has remained unknown. Along this line, it can be assumed that also native TS from *T. congolense* carries high-mannose glycans.

Size exclusion chromatography experiments with TconTS1 and TconTS2 produced in CHO-Lec1 demonstrated the presence of dimeres and oligomeres for both enzymes, where for TconTS1 almost equal quantities of monomers and dimers were found, whereas predominantly monomers were detected for TconTS2 [29] (Chapter 3.4 Figure 6). In agreement to these findings, Tiralongo *et al.* observed high molecular weights for native TS-forms isolated from *T. congolense* trypomastigotes [28]. Although the TS-form 1 and TS-form 2 characterised by Tiralongo *et al.* do not directly correspond to TconTS1 and TconTS2, their data are in perfect agreement with the oligomerisation potential of TS from *T. congolense* described in this work. Furthermore, TconTS oligomerisation appears to be *N*-glycan dependent and therefore possibly LD mediated, since EndoH treatment of TconTS1 significantly shifted the amount of dimeric to monomeric structures. Although the biological relevance of TconTS oligomerisation has remained unclear so far, our experimental data have provided evidence for a potential role in enzyme stability and/or enhanced substrate affinities as well as enzyme activities.

Additional *in vitro* experiments addressing stability of TS activity of TconTS1 and TconTS2 after storage for a long time period at 37°C revealed that even after four month at 37°C TconTS1 retained almost full TS activity, whereas TconTS2 completely lost its activity (data not shown). In the light of the observed oligomerisation, it can be suggested that the relatively high oligomerisation potential of TconTS1 may support its stability and provides one possible explanation for the retained TS activity after incubation for four month at 37°C. In addition, the relatively low oligomerisation potential of TconTS2 is also in agreement with this hypothesis and would explain the loss of TS activity possibly due to enzyme degradation. Formation of TconTS homo- (or hetero-) dimers and oligomers thus might play a role for the half-life of shed TconTS in the mammalian host and insect vector, although no experimental data regarding this theory are available in literature so far. In addition, also the possible formation of TconTS hetero-dimers and/or association to other native molecules, such as GARP (*T. congolense* procyclic glutamic acid/ alanine-rich protein [53]), have to be considered in respect to potential oligomerisation of TconTS on the parasite and its biological relevance.

4.3.3 Structural properties of TconTS-LD

Our results obtained from glycan array, STD NMR spectroscopy, TconTS-LD binding/inhibition assay and oligomerisation experiments clearly demonstrated the lectin activity of TconTS-LD binding to several different carbohydrate ligands [29] (Chapter 3.4). In addition, clear binding preferences of TconTS-LDs towards polymeric and complex ligand structures were observed. Accordingly, the specific structural arrangement of TconTS-CD and LD relative to each other might provide the simultaneous binding of both domains to the same ligand, which could consequently enhance the specific binding affinities and possibly also catalytic activities. The fact that both, the catalytic binding pocket of CD as well as the proposed carbohydrate binding-site of LD [23,54] are localised on the same site of the enzyme (Figure 4 A), supports this hypothesis of simultaneous binding. To investigate this possibility, the direct distance between the catalytic tyrosine residue (Y438 in TconTS1, Chapter 3.2 Table 2) in active centres of TconTS-CD and a phenylalanine residue in the predicted carbohydrate binding-site of TconTS-LD was determined from homology models of TconTS1 through TconTS4 and was found to range from 40.5 to 42.6 Å [29] (Figure 4 A and Chapter 3.4).

It should be noted that the phenylalanine residue assumed to be involved in carbohydrate binding [54] is well-conserved among TconTS (Figure 4 B-E) with an RMSD (root-mean-square deviation) value of 1.1, 1.3 and 1.6 Å for TconTS2, TconTS3 and TconTS4 relative to TconTS1, respectively. Interestingly this phenylalanine was also found in TbruTS and is substituted by tyrosine in both, TcruTS and TranSA. One interesting and significant observation was made when comparing the overall topology of the proposed binding-site of TconTS-LDs employing the homology models generated in this work. As shown in Figure 4, the predicted binding pocket of TconTS2-LD is more open compared to that of TconTS1, TconTS3 and TconTS4 (Figure 4 B-F), due to an overlying loop (labelled in green) covering the phenylalanine-containing groove (unpublished data).

Therefore the flexibility of this loop, defined through its amino acid sequence, might be a critical factor for effective lectin activities of TconTS-LD assuming that this location participates to the carbohydrate binding-site. Table 2 illustrates the amino acid sequence alignment of TconTS1 through TconTS4 of the loop region mentioned above. It can be seen that amino acid residues flanking the loop are very diverse among TconTSs (Table 2).

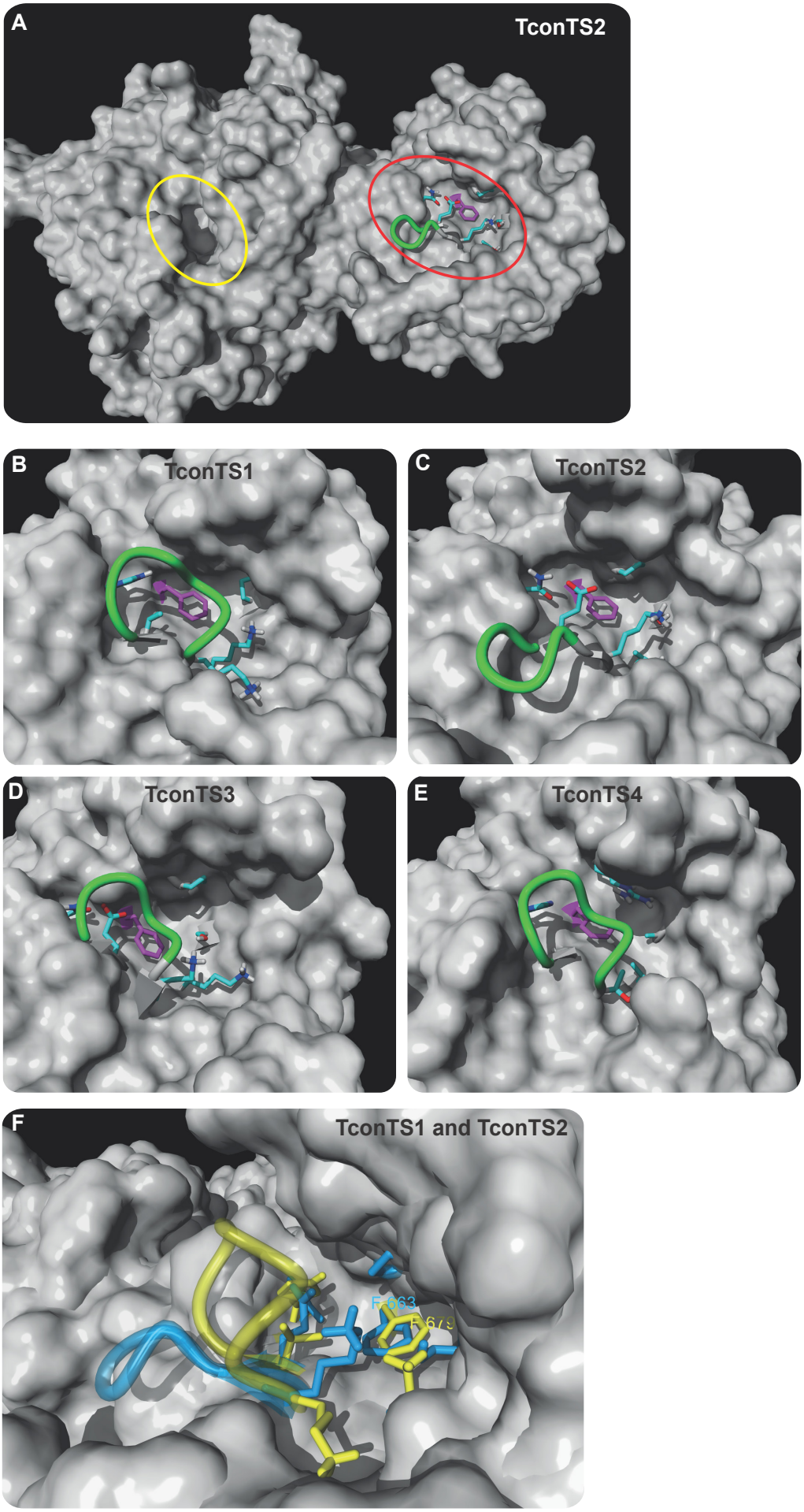

A
TconTS2
B
TconTS1
C
TconTS2
D
TconTS3
E
TconTS4
F
TconTS1 and TconTS2

Figure 4. Proposed carbohydrate binding-site of TconTS-LD. **A)** Homology model of TconTS2. The active centre of the catalytic domain (CD) is marked with a yellow ellipse. The proposed carbohydrate binding-site of the lectin domain (LD) is marked with a red ellipse. **B-E)** Homology models of TconTS(1-4)-LD showing only the proposed carbohydrate binding-site. A flexible loop element proposed to influence lectin activity is labelled in green. **F)** Structural comparison of TconTS1-LD (yellow) and TconTS2-LD (blue). Significantly different configurations of a loop element covering the proposed binding site are illustrated. Homology models were generated using the crystal structure of TcruTS (PDB code: 1SOI) as template employing the software Yasara as described under Methods in Chapter 3.1 [34]. The molecular surface of each structure model is shown in grey. Amino acid side chains suggested to be involved in carbohydrate recognition are presented.

Table 2: Comparison of amino acid residues from TconTS1 through TconTS4 forming a loop structure close to the predicted carbohydrate binding-site of TconTS-LD

Enzyme	Amino acid residue				
TconTS1b	R599	E600	G601	K602	V603
TconTS1e-1	R599	K600	D601	Q602	V603
TconTS2	N584	E585	-	E586	S587
TconTS3	Y592	E593	G594	E595	T596
TconTS4	S654	K655	G656	E657	N658

Conserved amino acid residues are grey shaded

One striking observation comprises the missing glycine residue in TconTS2, which shortened the loop but will not drastically reduce its overall flexibility. In addition, only TconTS2 and TconTS3 contain two glutamic acid residues in close relation (E585 and E586 in TconTS2, E593 and E595 in TconTS3), whereas both, TconTS1 variants and TconTS4 exhibit a relatively drastic substitution from a negatively charged glutamic acid to a positively charged lysine residue (K600 in TconTS1e-1, K602 in TconTS1b and K655 in TconTS4). This change of the electrostatic potential may influence the flexibility of the loop due to repulsive and/or attractive interactions to adjacent amino acid side chains. However, these three amino acids have been suggested to play a critical role in the sterical orientation of the loop and thus, possibly are indirectly involved in lectin activity of TconTS-LD.

In this context, it should be mentioned again that experimental results concerning the specific enzymatic activities of the two TconTS1 variants, TconTS1b and TconTS1e-1 revealed significant differences [34] (Chapter 3.1). Interestingly, the only drastic differences in the amino acid sequence of these two TconTS1 variants are located exactly in the region of the critical loop (599-602) mentioned above and have been proposed to be involved in recognition of larger substrate molecules, such as glycoproteins [34] (Chapter 3.1 Figure 3).

These findings could explain the observed variations in TS activity of TconTS1b and TconTS1e-1 and would support the hypothesis concerning the importance of the loop structure in carbohydrate binding activity of LD and consequently catalytic activity of TconTS.

It will be interesting, to perform mutagenesis experiments in future to further characterise TconTS-LD binding and to challenge the hypothesis mentioned above.

4.3.4 The influence of TconTS-LD on enzymatic activity

The ability of TconTS-LD to bind carbohydrate structures as well as their structural significant features and ligand specificities has been discussed in the previous Chapters. Based on that, the question arises, whether TconTS-LD plays a direct role in influencing enzyme activities including TS/sialidase rates. To address this question a genetic approach was used to recombine TconTS-domains from differ, on the DNA level, without the disruption of any critical structure elements (Chapter 3.5). In order to introduce a unique restriction site two additional amino acids were inserted into a loop segment, which was proposed to be the most potent location for such a mutation. This loop element follows the domain-connecting α-helix and is neither resolved in the crystal structure of TcruTS [24] nor conserved in trypanosomal TS, demonstrating its flexibility and suggesting that it is not important for the overall folding of the TS (Chapter 3.5 Figure 2).

To test this option in theory, homology models of domain swapped TconTS were calculated to investigate potential problems of a practical recombination of TconTS-domains. The *in silico* results revealed domain swapped TconTS, exhibiting high overall structure similarity to wild type TconTS, without the loss or drastic change of any critical structure elements (data not shown). In addition, an extended molecular surface area and a well-defined hydrogen bond network at the interface between TconTS-CD and LD of domain swapped enzymes was observed, which was also found in wild type TconTS1 through TconTS4 and is well-conserved in the TS family (Chapter 3.5 Figure 1 and Tables 1-3). It has been reported that this extended contact site together with these hydrogen-bonds is significantly stronger pronounced compared to related microbial sialidases [24,43], possibly indicating an enhanced interplay of both, TconTS-CD and LD.

Two out of twelve theoretically possible domain swapped TconTS enzymes were selected for cloning and expression, comprising CD of TconTS1a and LD of TconTS3 (termed TconTS1aTS3) as well as CD from TconTS2 and LD of TconTS4 (termed TconTS2TS4). The idea was to recombine the CD of a relatively high active TconTS

(TconTS1 or TconTS2) to the LD of a lower active enzyme (TconTS3 or TconTS4) to investigate possible changes in TS and sialidase activities. TconTS1aTS3 has been successfully cloned and expressed in fibroblasts (CHO-Lec1). Due to time consuming stable transfection of CHO-Lec1 cells, TconTS1aTS3 has also been expressed in *E. coli* (Chapter 3.5 Figure 3). Thus, preliminary enzyme activity data of TconTS1aTS3 were determined with this bacterial protein, clearly demonstrating its TS activity. Interestingly, this activity was significantly lower compared to that of TconTS1a wild type expressed by *E. coli*, but still approximately two fold higher than TconTS3 wild type (Chapter 3.5 Figure 6 A and C). In contrast, the amount of free Neu5Ac produced by TconTS1aTS3 was similar to that of TconTS1a wild type and again double as high as for TconTS3 wild type (Chapter 3.5 Figure 6 B and C). Thus, TconTS1aTS3 exhibited a rather similar TS activity to TconTS3 and sialidase activity to TconTS1a. In summary, it can be concluded that the exchange of LD modified the catalytic activity of TconTS1a supporting the correlated hypothesis mentioned above regarding a potential influence of LD on enzyme activity. The mechanistic reasons for that are still not clear, but it is proposed that the lectin activity of LD to glycoproteins, which serve as substrates for Sia transfer activity, enhances the affinity to these substrates and therefore supports the catalytic activity of the enzyme. However, it should be kept in mind that the enzyme activities of TconTS expressed by bacteria was found to be orders of magnitudes lower compared to that of enzymes expressed in fibroblasts [34] (Chapter 3.1 and Chapter 3.5).

Interestingly a complete different situation regarding the enzymatic behaviour of domain swapped TconTS1aTS3 was found, when the enzyme was expressed by CHO-Lec1 cells. Experimental data revealed no significant reduction of TS activity for TconTS1aTS3, indicated by the amount of 3'SL produced, which is almost the same as that generated by TconTS1a wild type (Chapter 3.5 Figure 6 D and F). In contrast, sialidase activity of TconTS1aTS3 was drastically reduced of about 85 % compared to TconTS1a wild type (Chapter 3.5 Figure 6 E and F,).

In summary, it seems that TS activity of TconTS1aTS3 is drastically reduced in contrast to the retained sialidase activity, when TS is expressed in bacteria, whereas the exact contrary enzymatic behaviour was observed when TconTS1aTS3 was expressed in CHO-Lec1 cells. In addition, the general overall enzymatic properties (TS/sialidase ratio) of TconTS1a wild type remains approximately constant, independent from the expression system used in this work. Considering only these results, it can be concluded that another factor, besides the domain-exchange, influences enzyme activities, which could be either the folding and/or

glycosylation of the enzyme, since glycoproteins expressed by CHO-Lec1 fibroblasts are glycosylated with high-mannose *N*-glycans, whereas proteins expressed by *E. coli* Rosetta pLacI completely lack glycosylation.

Surprisingly, the situation becomes even more complex, when enzymatic activities of the control TconTS1a* were considered. TconTS1a* is similar to the wild type but additionally comprises the insertion if the two amino acids (Tyr-Val) introduced to facilitate the domain swap (Chapter 3.5 Figure 2). It has been observed that both, TS and sialidase activities, of TconTS1a* expressed by CHO-Lec1 fibroblasts were drastically reduced compared to TconTS1a wild type (Chapter 3.5 Figure 6 D-F). Interestingly, only TS activity of TconTS1a* is significantly lower compared to that of domain swapped TconTS1aTS3, whereas its sialidase activity was two fold higher then that of TconTS1aTS3.

Currently, it remains unclear what kind of effect might be induced by the insertion in TconTS1a* causing the observed variations in TS and sialidase activities. Referring to this, further data regarding the enzymatic activities of TconTS2 and TconTS2* (Chapter 3.5 Figure 6 D-F), additional *in silico* data (Chapter 3.5 Figure S1 A), as well as Western blot analysis (Chapter 3.5 Figure 3 D) using anti TS mAb 7/23 [28,34,35] (Chapter 3.1 and 3.2) provided cumulative evidence for important roles of *N*-glycosylation at a defined location in TconTS-LD strerically close to the loop segment carrying the inserted *Eco*105I endonuclease restriction site (Chapter 3.5 Figure S1 B). It was proposed that the inserted two amino acids possibly interfere with *N*-glycosylation of the N657, which is sterically in close vicinity to the loop segment comprising the insertion (Chapter 3.5 Figure S1 B). Along this line, it was assumed that sufficient *N*-glycosylation of this site might be essential for proper folding of the corresponding region. In addition to this, it has previously been reported that single point mutations of amino acid residues (corresponding to E710 and R704 in TconTS1 variants) forming a conserved salt bridge close to the *N*-glycosylation site mentioned above, significantly influence enzymatic activities of TcruTS [43,55]. However, it has remained unclear what kind of mechanistic effect might be responsible for this change in TS activity. Interestingly, this salt bridge is far away from the catalytic centre of TcruTS, thus indicating another functional structure element contributing to the hypothesis that LD influences TconTS enzymatic activity.

In conclusion, the relatively high TS activity of domain swapped TconTS1aTS3 clearly demonstrated the efficiency and general usability of the TconTS domain-swap strategy, which was established and demonstrated in this work. Based on the data obtained in this study, additional optimisations appear possible for even better comparability of TconTS catalytic

activities when swapping TconTS domains and possibly also LDs from other trypanosomal TS. For example, it would be of fundamental interest in future experiments to recombine TconTS2-LD with CDs from TconTS1, TconTS3 and TconTS4 as well as the domain swap between the eleven TconTS1 variants, since they represent TconTS1 enzymes comprising natural mutations in CD and LD [34] (Chapter 3.1).

4.3.5 TconTS-LD activity opens new strategies towards novel potent TS inhibitor synthesis

Potent TS inhibitors exhibiting selective inhibitory properties in the low µM to nM range are not available so far. However, besides all current full synthetic compounds, which have only moderate TS inhibitory effects, some natural occurring molecules have been identified and successfully tested as relatively potent TcruTS inhibitors [56]. This class comprises the family of gangliosides as well as *N*-linked-type oligosaccharides [56]. For example terminal α2-3-linked Sia containing gangliosides, such as GM3 (Neu5Acα2,3-Galβ1,4-Glc-ceramide) and GSC-31 (Neu5Acα2,3-Galβ1,4-GlcNAcβ1,3-Galβ1,4-Glc-R) showed IC_{50} values in the lower µM range using TcruTS [56]. Furthermore, a tetra-antennary, complex *N*-linked-type oligosaccharide, comprising four terminal sialylated saccharide chains of the type (Neu5Acα2,3-Galβ1,4-GlcNAc) exhibit also inhibitory effects in the low µM range [56]. Therefore, it was proposed that the enhanced inhibitory properties of GM3 and the *N*-linked-type oligosaccharide compared to other synthetic compounds, is mainly due to their ability to occupy both, the donor and the acceptor substrate binding-sites in the catalytic centre of TcruTS [56]. According to these findings, a new set of synthetic compounds has been synthesised, which should be able to target both binding sites similar to the natural ligands mentioned above [57]. Against all expectations, these novel TS inhibitors surprisingly showed only minor TS inhibitory properties on TcruTS [57]. Thus, it is still of fundamental importance to fully understand the mechanistic of trypanosomal TS activity for more potent inhibitor synthesis.

However, the observed lectin activity of TconTS-LDs on several glycan structures, demonstrated in this work (Chapter 3.4) [29], is suggested to bring new light into potent TS inhibitor synthesis. The fact that TconTS-LD showed binding preferences towards oligosaccharide ligands, strongly suggested that the enhanced inhibitory effect of oligomeric glycans on TS activity is due the additional binding of the ligand to the LD of TS, which is in agreement with data observed from STD NMR experiments and supports the hypothesis of an extended binding site on TconTS-LD [29] (Chapter 3.4). Such a simultaneous binding of a

multivalent ligand to TS would possibly significantly increase affinity to TconTS and consequently also its inhibitory potential. In theory it would be possible for a relatively long ligand to bind both, CD and LD at the same time, since the carbohydrate binding pocket of CD and LD are located on the same site of the enzyme and targeted in the same direction (Figure 4 A). Furthermore, the distance between the active centre of the catalytic domain and the proposed extended binding-site of the lectin domain of TconTS is in the range of approximately 40 Å, as it can be seen from Figure 4 A. This value equals approximately the length of an oligosaccharide comprising six to seven monosaccharide units, where each monosaccharide is about 7 Å long. Therefore it might be of significant importance to consider the aspect of TconTS-LD lectin activity for future strategies to establish a class of novel and more potent TS inhibitors. The impact of 1-4β-mannotriose on the sialidase activity of TconTS1a and TconTS2 (Chapter 3.5) is an example for the potential of addressing the LD to modulate the enzymatic activities. Furthermore, blocking of this LD site might have significant impact on the supramolecular organisation of the glycoproteins on the parasites surface.

Another aspect refers to the hypothesis that trypanosomal TS and siglecs share ligand-binding properties for the same type of Sia containing glycan structures. Indeed such an correlation between trypanosomes and siglecs from host cells has been reported previously [14,58,59]. It has been demonstrated that trypanosomal TS generates sialo-glycoconjugates on the parasites surface, which were specifically bound by Siglec-E expressed on host phagocytic and antigen-presenting cells (macrophages), respectively [60]. However, it can be hypothesised that TS generates sialo-glycans on the parasite surface, which might also specifically interact with other related receptors for example from the siglec family [61], such as Siglec-1 (sialoadhesin, Sn [62]), Siglec-2 (CD22 [63]) and Siglec-4 (myelin associated glycoprotein, MAG [64]), which are expressed on macrophages, oligodendrocytes and other cell types [34] (Chapter 3.1). Indeed specific interaction of sialoadhesin (Sn), expressed on macrophages, to sialo-glycoconjugates of cultured *T.cruzi* trypomastigotes and epimastigotes has been demonstrated [65]. With respect to these findings and the hypothesis made above it can be further concluded that potent siglec inhibitors in general may also serve as promising TS inhibitors, since both proteins obviously recognise (compete) the same ligands on host cells [59]. Therefore, it will be interesting to test known siglec inhibitors, such as the novel C-4 and C-9 disubstituted Neu5Ac derivatives reported by Kelm *et al.*, which have been demonstrated to be relatively strong Siglec-2 inhibitors with affinities in the sub-µM range [48], for their inhibition potential on TconTSs.

In summary, it might be of significant importance to consider both aspects regarding TconTS ligand properties discussed in this Chapter, (1) the extended length to simultaneously bind to TconTS-CD and LD and (2) novel C-4, C-9 and/or other substitutions at the terminal Neu5Ac residue. These new strategies may provide a novel generation of potent TS inhibitors comprising affinities potentially in the lower μM to nM range.

4.4 References

1. Clayton J. Chagas disease 101. Nature. 2010 Jun 24;465(7301):S4–5.
2. World Health Organization. Control and surveillance of human African trypanosomiasis. World Health Organ Tech Rep Ser. 2013;(984):1–237.
3. Connor RJ. The impact of nagana. Onderstepoort J Vet Res. 1994 Dec;61(4):379–83.
4. Dyer NA, Rose C, Ejeh NO, Acosta-Serrano A. Flying tryps: survival and maturation of trypanosomes in tsetse flies. Trends Parasitol. Elsevier; 2013 Apr;29(4):188–96.
5. Nagamune K, Acosta-Serrano A, Uemura H, Brun R, Kunz Renggli C, Maeda Y, et al. Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors. J Exp Med. Rockefeller Univ Press; 2004 May 17;199(10):1445–50.
6. Muiá RP, Yu H, Prescher JA, Hellman U, Chen X, Bertozzi CR, et al. Identification of glycoproteins targeted by *Trypanosoma cruzi* trans-sialidase, a virulence factor that disturbs lymphocyte glycosylation. Glycobiology. Oxford University Press; 2010 Jul;20(7):833–42.
7. Acosta-Serrano A, Almeida IC, Freitas-Junior LH, Yoshida N, Schenkman S. The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. Mol Biochem Parasitol. 2001 May;114(2):143–50.
8. Utz S, Roditi I, Kunz Renggli C, Almeida IC, Acosta-Serrano A, Bütikofer P. Trypanosoma congolense procyclins: unmasking cryptic major surface glycoproteins in procyclic forms. Eukaryotic Cell. 2006 Aug;5(8):1430–40.
9. Acosta-Serrano A, Vassella E, Liniger M, Kunz Renggli C, Brun R, Roditi I, et al. The surface coat of procyclic *Trypanosoma brucei*: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. Proc Natl Acad Sci USA. National Acad Sciences; 2001 Feb 13;98(4):1513–8.
10. Schenkman S, Jiang MS, Hart GW, Nussenzweig V. A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. Cell. 1991 Jun 28;65(7):1117–25.
11. Engstler M, Reuter G, Schauer R. The developmentally regulated trans-sialidase from *Trypanosoma brucei* sialylates the procyclic acidic repetitive protein. Mol Biochem Parasitol. 1993 Sep;61(1):1–13.
12. Greganova E, Bütikofer P, Acosta-Serrano A. The protease resistant surface (PRS) glycoconjugate from *Trypanosoma congolense* has an inositol-acylated glycosylphosphatidyl-inositol anchor, containing a significant proportion of myristate at the sn-2 position. Mol Biochem Parasitol. 2010 May;171(1):50–4.
13. Liniger M, Acosta-Serrano A, Van Den Abbeele J, Kunz Renggli C, Brun R, Englund PT, et al. Cleavage of trypanosome surface glycoproteins by alkaline trypsin-like enzyme(s) in the midgut of *Glossina morsitans*. Int J Parasitol. 2003 Oct;33(12):1319–28.
14. Schauer R, Kamerling JP. The chemistry and biology of trypanosomal trans-sialidases: virulence factors in Chagas disease and sleeping sickness. Chembiochem. WILEY-VCH Verlag; 2011 Oct 17;12(15):2246–64.
15. Burgos JM, Risso MG, Brenière SF, Barnabé C, Campetella O, Leguizamón MS. Differential distribution of genes encoding the virulence factor trans-sialidase along *Trypanosoma cruzi* Discrete typing units. PLoS ONE. Public Library of Science; 2013;8(3):e58967.
16. Schenkman S, Eichinger D, Pereira ME, Nussenzweig V. Structural and functional properties of *Trypanosoma* trans-sialidase. Annu Rev Microbiol. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1994;48(1):499–523.
17. Scudder P, Doom JP, Chuenkova M, Manger ID, Pereira ME. Enzymatic characterization of β-

D-galactoside α2,3-trans-sialidase from *Trypanosoma cruzi*. J Biol Chem. 1993 May 5;268(13):9886–91.

18. Nok AJ, Balogun EO. A bloodstream *Trypanosoma congolense* sialidase could be involved in anemia during experimental trypanosomiasis. J Biochem. 2003 Jun;133(6):725–30.

19. Coustou V, Plazolles N, Guegan F, Baltz T. Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. Cellular Microbiology. 2012 Feb 13;14(3):431–45.

20. Campetella OE, Uttaro AD, Parodi AJ, Frasch AC. A recombinant *Trypanosoma cruzi* trans-sialidase lacking the amino acid repeats retains the enzymatic activity. Mol Biochem Parasitol. 1994 Apr;64(2):337–40.

21. Buschiazzo A, Campetella O, Frasch AC. *Trypanosoma rangeli* sialidase: cloning, expression and similarity to *T. cruzi* trans-sialidase. Glycobiology. 1997 Dec;7(8):1167–73.

22. Cremona ML, Sánchez DO, Frasch AC, Campetella O. A single tyrosine differentiates active and inactive *Trypanosoma cruzi* trans-sialidases. Gene. 1995 Jul 4;160(1):123–8.

23. Buschiazzo A, Tavares GA, Campetella O, Spinelli S, Cremona ML, París G, et al. Structural basis of sialyltransferase activity in trypanosomal sialidases. EMBO J. 2000 Jan 4;19(1):16–24.

24. Buschiazzo A, Amaya MF, Cremona ML, Frasch AC, Alzari PM. The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. Mol Cell. 2002 Oct;10(4):757–68.

25. Amaya MF, Watts AG, Damager I, Wehenkel A, Nguyen T, Buschiazzo A, et al. Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. Structure. 2004 May;12(5):775–84.

26. Sartor PA, Agusti R, Leguizamon MS, Campetella O, de Lederkremer RM. Continuous nonradioactive method for screening trypanosomal trans-sialidase activity and its inhibitors. Glycobiology. 2010 Jul 1;20(8):982–90.

27. Tiralongo E, Martensen I, Grötzinger J, Tiralongo J, Schauer R. Trans-sialidase-like sequences from *Trypanosoma congolense* conserve most of the critical active site residues found in other trans-sialidases. Biol Chem. 2003 Aug;384(8):1203–13.

28. Tiralongo E, Schrader S, Lange H, Lemke H, Tiralongo J, Schauer R. Two trans-sialidase forms with different sialic acid transfer and sialidase activities from *Trypanosoma congolense*. J Biol Chem. 2003 Jun 27;278(26):23301–10.

29. Waespy M, Gbem TT, Elenschneider L, Jeck A-P, Day CJ, Hartley-Tassell L, et al. Carbohydrate recognition specificity of trans-sialidase lectin domain from *Trypanosoma congolense*. (in press) PLoS Negl Trop Dis. 2015;:1–49. doi:10.1371/journal.pntd.0004120

30. Engstler M, Schauer R, Brun R. Distribution of developmentally regulated trans-sialidases in the *Kinetoplastida* and characterization of a shed trans-sialidase activity from procyclic *Trypanosoma congolense*. Acta Tropica. 1995 May;59(2):117–29.

31. Engstler M, Schauer R. Sialidases from African trypanosomes. Parasitol Today (Regul Ed). 1993 Jun;9(6):222–5.

32. Engstler M, Reuter G, Schauer R. Purification and characterization of a novel sialidase found in procyclic culture forms of *Trypanosoma brucei*. Mol Biochem Parasitol. 1992 Aug;54(1):21–30.

33. Bock N, Kelm S. Binding and inhibition assays for Siglecs. Methods Mol Biol. 2006;347:359–75.

34. Koliwer-Brandl H, Gbem TT, Waespy M, Reichert O, Mandel P, Drebitz E, et al. Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*. BMC Biochem. 2011;12(1):39.

35. Gbem TT, Waespy M, Hesse B, Dietz F, Smith J, Chechet GD, et al. Biochemical diversity in the *Trypanosoma congolense* trans-sialidase family. PLoS Negl Trop Dis. 2013;7(12):e2549.

36. Oliveira IA, Goncalves AS, Neves JL, Itzstein von M, Todeschini AR. Evidence of ternary complex formation in *Trypanosoma cruzi* trans-sialidase catalysis. Journal of Biological Chemistry. 2014 Jan 3;289(1):423–36.

37. Demir O, Roitberg AE. Modulation of catalytic function by differential plasticity of the active site: case study of *Trypanosoma cruzi* trans-sialidase and *Trypanosoma rangeli* sialidase †. Biochemistry. 2009 Apr 21;48(15):3398–406.

38. París G, Cremona ML, Amaya MF, Buschiazzo A, Giambiagi S, Frasch AC, et al. Probing molecular function of trypanosomal sialidases: single point mutations can change substrate specificity and increase hydrolytic activity. Glycobiology. 2001 Apr;11(4):305–11.

39. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure. 1995 Nov 15;3(11):1197–205.

40. Montagna G, Cremona ML, Paris G, Amaya MF, Buschiazzo A, Alzari PM, et al. The trans-sialidase from the African trypanosome *Trypanosoma brucei*. Eur J Biochem. 2002 Jun;269(12):2941–50.

41. Freire-de-Lima L, Oliveira IA, Neves JL, Penha LL, Alisson-Silva F, Dias WB, et al. Sialic acid: a sweet swing between mammalian host and *Trypanosoma cruzi*. Front Immunol. Frontiers; 2012;3:356.

42. Watts AG, Damager I, Amaya ML, Buschiazzo A, Alzari P, Frasch AC, et al. *Trypanosoma cruzi* trans-sialidase operates through a covalent sialyl-enzyme intermediate: Tyrosine is the catalytic nucleophile. J Am Chem Soc. 2003 Jun;125(25):7532–3.

43. Amaya MF, Buschiazzo A, Nguyen T, Alzari PM. The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T.cruzi* trans-sialidase. Journal of Molecular Biology. 2003 Jan;325(4):773–84.

44. Rini JM. Lectin structure. Annu Rev Biophys Biomol Struct. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 1995;24(1):551–77.

45. Turner MW. Mannose-binding lectin: the pluripotent molecule of the innate immune system. Immunol Today. 1996 Nov;17(11):532–40.

46. Weis WI, Drickamer K. Structural basis of lectin-carbohydrate recognition. Annu Rev Biochem. 1996;65:441–73.

47. Engvall E, Perlmann P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. Immunochemistry. 1971 Sep;8(9):871–4.

48. Kelm S, Madge P, Islam T, Bennett R, Koliwer-Brandl H, Waespy M, et al. C-4 modified sialosides enhance binding to Siglec-2 (CD22): towards potent Siglec inhibitors for immunoglycotherapy. Angew Chem Int Ed Engl. 2013 Mar 25;52(13):3616–20.

49. Robertson MA, Etchison JR, Robertson JS, Summers DF, Stanley P. Specific changes in the oligosaccharide moieties of VSV grown in different lectin-resistnat CHO cells. Cell. 1978 Mar;13(3):515–26.

50. Dani M, Manca F, Rialdi G. Calorimetric study of concanavalin A binding to saccharides. Biochim Biophys Acta. 1981 Jan 30;667(1):108–17.

51. Shibata S, Goldstein IJ, Baker DA. Isolation and characterization of a Lewis b-active lectin from *Griffonia simplicifolia* seeds. J Biol Chem. 1982 Aug 25;257(16):9324–9.

52. Pontes de Carvalho LC, Tomlinson S, Vandekerckhove F, Bienen EJ, Clarkson AB, Jiang MS, et al. Characterization of a novel trans-sialidase of *Trypanosoma brucei* procyclic trypomastigotes and identification of procyclin as the main sialic acid acceptor. J Exp Med. 1993 Feb 1;177(2):465–74.

53. Bütikofer P, Vassella E, Boschung M, Renggli CK, Brun R, Pearson TW, et al. Glycosylphosphatidylinositol-anchored surface molecules of *Trypanosoma congolense* insect forms are developmentally regulated in the tsetse fly. Mol Biochem Parasitol. 2002 Jan;119(1):7–16.

54. Luo Y, Li SC, Li YT, Luo M. The 1.8 Å structures of leech intramolecular trans-sialidase complexes: evidence of its enzymatic mechanism. Journal of Molecular Biology. 1999 Jan 8;285(1):323–32.

55. Smith LE, Eichinger D. Directed mutagenesis of the *Trypanosoma cruzi* trans-sialidase enzyme identifies two domains involved in its sialyltransferase activity. Glycobiology. 1997 Apr;7(3):445–51.

56. Vandekerckhove F, Schenkman S, Pontes de Carvalho L, Tomlinson S, Kiso M, Yoshida M, et al. Substrate specificity of the *Trypanosoma cruzi* trans-sialidase. Glycobiology. 1992 Dec;2(6):541–8.

57. Streicher H, Busse H. Building a successful structural motif into sialylmimetics-cyclohexene-phosphonate monoesters as pseudo-sialosides with promising inhibitory properties. Bioorganic & Medicinal Chemistry. 2006 Feb 15;14(4):1047–57.

58. Erdmann H, Steeg C, Koch-Nolte F, Fleischer B, Jacobs T. Sialylated ligands on pathogenic *Trypanosoma cruzi* interact with Siglec-E (sialic acid-binding Ig-like lectin-E). Cellular Microbiology. Blackwell Publishing Ltd; 2009 Nov;11(11):1600–11.

59. Jacobs T, Erdmann H, Fleischer B. Molecular interaction of Siglecs (sialic acid-binding Ig-like lectins) with sialylated ligands on *Trypanosoma cruzi*. Eur J Cell Biol. 2010 Jan;89(1):113–6.

60. Zhang JQ, Biedermann B, Nitschke L, Crocker PR. The murine inhibitory receptor mSiglec-E is expressed broadly on cells of the innate immune system whereas mSiglec-F is restricted to eosinophils. Eur J Immunol. WILEY-VCH Verlag; 2004 Apr;34(4):1175–84.

61. Crocker PR, Clark EA, Filbin M, Gordon S, Jones Y, Kehrl JH, et al. Siglecs: a family of sialic-acid binding lectins. Glycobiology. 1998 Feb;8(2):v.

62. Crocker PR, Kelm S, Dubois C, Martin B, McWilliam AS, Shotton DM, et al. Purification and properties of sialoadhesin, a sialic acid-binding receptor of murine tissue macrophages. EMBO J. 1991 Jul;10(7):1661–9.

63. Stamenkovic I, Seed B. The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion. Nature. Nature Publishing Group; 1990 May 3;345(6270):74–7.

64. Kelm S, Pelz A, Schauer R, Filbin MT, Tang S, de Bellard ME, et al. Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. Curr Biol. 1994 Nov 1;4(11):965–72.

65. Monteiro VG, Lobato CSS, Silva AR, Medina DV, de Oliveira MA, Seabra SH, et al. Increased association of *Trypanosoma cruzi* with sialoadhesin positive mice macrophages. Parasitol Res. Springer-Verlag; 2005 Nov;97(5):380–5.

5.
Outlook

5 Outlook

Main achievements of this work comprise the detailed biochemical characterisation of trans-sialidases from *T. congolense* and the carbohydrate-binding activities of TconTS-LD**.** The results obtained are novel and provide important and enhanced knowledge about TS in general and their biological role in *T. congolense* during parasite's developmental life cycle in host and vector. Thus, based on this study new strategies can be developed for anti-TS drug design and possibly other new therapy forms against Trypanosomiasis in human and animal.

The diversity of active and non-active TconTS genes, described in this study, presents the relatively complex situation in the African *T. congolense*. Furthermore, biochemical characterisations of TconTS, performed in this work demonstrated striking differences in enzymatic activities of TconTS utilizing several combinations of Sia acceptor and donor substrates and different physical reaction conditions. These findings together with recent mRNA isolation experiments from blood stream forms of *T. congolense* provided evidence for a well-defined Sia transfer machinery of the parasite in both the mammalian host and insect vector. Since the previously reported anti TS mAb 7/23, which was further biochemically characterised in this work, recognises TconTS1 as well as TconTS2, it cannot be efficiently used to investigate the expression pattern of the different TconTS proteins *in vivo*. Therefore, it would be of significant importance to generate antibodies specific for each TconTS enzyme as potential tools for their detection and differentiation during the development of trypanosomiasis and establishment of disease control measures.

Glycosylation of trypanosomal TS was previously demonstrated by ConA affinity purification of native TS from *T. cruzi* and *T. brucei* procyclic trypomastigotes. 8-9 potential *N*-glycosylation sites were predicted for TconTS in contrast to only 2-4 for TcruTS and TranSA, indicating a possible, more complex glycosylation pattern in the African trypanosomal TS. Oligomerisation experiments of TconTS performed in this study demonstrated the high-mannose *N*-glycan dependency of this process. In addition to that, experimental data obtained, provided good evidence for the direct influence of *N*-glycosylation on enzymatic activities. However, only limited data is available regarding the glycosylation pattern of trypanosomal TS and their direct or indirect influence on enzyme activities so far. Additional MALDI-MS and mutagenesis experiments should provide further insight into glycosylation of TconTS and the role in enzymatic catalysis.

Finally, carbohydrate binding of TconTS lectin domain to mannosyl oligosaccharides was demonstrated for the first time and presented in this work. Distinct lectin activities of TconTS1 through TconTS4 to different carbohydrate ligands together with experimental data

obtained from domain swap experiments provided strong evidence for a direct role of TconTS-LD on enzymatic activities. However, ligand specificities as well as the location and topology of the binding sites of TconTS-LD still remain to be investigated. Therefore, it is of fundamental importance to further characterise the specific TconTS-LD-ligand interactions through subsequent biophysical methods, such as SPR (surface plasmon resonance), ITC (isothermal titration calorimetry), STD NMR and crystallisation experiments. It can be suggested that enhanced knowledge about TconTS-LD ligand affinities and specificities provide new insight into TS enzymatic activities, their biological functions and possibly represents TconTS-LD as a biotechnological tool for specific carbohydrate detection and/or cell traffic inhibition in several applications.

6. Appendix

6.1 Curriculum Vitae: Mario Waespy

Personal data

Date of birth	30/07/1983
Place of birth	Bogotá, Columbia
Nationality	German

Education

11/2011 – 11/2015	**Doctoral thesis** in Biochemistry and Molecular Biology at the Centre for Biomolecular Interactions Bremen (CBIB) University of Bremen, Germany	
	Advisor:	Prof. Dr. Sørge Kelm
	Title:	Structure and functional relations of trans-sialidases from *Trypanosoma congolense*
03/2014 - 05/2014	**Scientific research** stay at the Institute for Glycomics, Griffith University Goldcoast, Queensland, Australia	
	Advisor:	Dr. Joe Tiralongo, Dr. Thomas Haselhorst
	Host:	Prof. Dr. Mark von Itzstein
	Objective:	Functional characterisation of the trans-sialidase lectin domain from *Trypanosoma congolense*
01/2011 - 07/2011	**Diploma thesis** at the Centre for Biomolecular Interactions Bremen (CBIB) University of Bremen, Germany	
	Advisor:	Prof. Dr. Sørge Kelm
	Title:	Charakterisierung von Reaktionsprodukten der Trans-Sialidasen aus *Trypanosoma congolense*
10/2003 - 10/2010	**Chemistry studies** at the University of Bremen	
	Major Subject:	Biochemistry, Organic Chemistry, Physical Chemistry, Inorganic Chemistry and Analytical Chemistry
07/2000 - 06/2003	**Abitur** (general qualification for university entrance) at Gymnasium	

Schulzentrum Bremen Neustadt

Conferences and Workshops

Poster presentation: *Binding activity and specificity of Trans-sialidase lectin domain from Trypanosoma congolense.* **Mario Waespy,** Thaddeus T. Gbem, Leroy Elenschneider, Joe Tiralongo, Thomas Haselhorst, Jonathan A. Nok, Sørge Kelm, Joint Meeting of the Society of Glycobiology (SFG) and the Japanese Society for Carbohydrate Research (JSCR), November 16th - 19th **2014**, **Honolulu, Hawaii**

Poster presentation: *The lectin domain of Trypanosoma congolense trans-sialidase and its influence on enzyme activity*. **Mario Waespy**, Thaddeus Gbem, Bettina Hesse, Shanmugam Solaiyappan, Joe Tiralongo, Thomas Haselhorst and Sørge Kelm, 9th International Symposium on Glycosyltransferases, June 18th to 21st **2014**, **Porto, Portugal**

- **Poster award price** for one of the best posters on the 9th International Symposium on Glycosyltransferases, GlycoT 2014, from 18th to 21st june 2014 in Porto, Portugal

Workshop: 8th Workshop Biochromatographie: *Bio unter Hochdruck*, 25th of September **2013**, **Hannover**, **Germany**, Dionex Thermo Scientific.

Workshop: *Ionenchromatographie als Routinemethode im Alltag,* 28th of Mai **2013**, **Bremen, Germany**, Thermo Scientific.

Workshop: *HPLC/UHPLC-Methodenentwicklungsseminar mit der Kinetex Core-Shell Technologie,* 1st of October **2012**, **Hamburg, Germany**, Phenomenex.

Workshop: Glycoanalytic Roundtable, *Herausforderungen der Glykoanalytik über LC, IC und MS-Kopplungen*, 6th of September **2012**, **University Bremen, Germany**, Dionex, Thermo Scientific.

Poster presentation: *Characterization of two TS1 variants from Trypanosoma congolense.* **Mario Waespy**, Thaddeus T. Gbem, Sørge Kelm, 8th International Symposium on Glycosyltransferases, June 5th - 9th **2012**, **Hannover, Germany**.

Teaching at the University of Bremen

Supervision in biochemical courses	Basics of Biochemistry and Molecular Biology 2011 - 2014
	Chemie für Biologen 2012 - 2014
	Recombinant Proteins 2013 and 2014
	Glycobiochemistry 2012 and 2013

6.2 List of publications

Mario Waespy, Thaddeus T. Gbem, Leroy Elenschneider, André-Phillipe Jeck, Christopher J. Day, Lauren Hartley-Tassell, Nicolai Bovin, Joe Tiralongo, Thomas Haselhorst, Sørge Kelm (2015). Carbohydrate recognition specificity of trans-sialidase lectin domain from *Trypanosoma congolense*. *(Manuscript in press) PLOS Neglected Tropical Disease (ID: PNTD-D-15-01085). doi:10.1371/journal.pntd.0004120*

Sørge Kelm, Paul Madge, Tasneem Islam, Ryan Barnett, Hendrik Koliwer-Brandl, **Mario Waespy**, Mark von Itzstein, Thomas Haselhorst (2013). C-4 modified sialosides enhance binding to Siglec-2 (CD22): towards potent siglec inhibitors for immunoglycotherapy. *Angewandte Chemie International Edition English,* **52** (13), 3616 – 3620.

Thaddeus T. Gbem, **Mario Waespy**, Bettina Hesse, Frank Dietz, Joel Smith, Gloria D. Chechet, Jonathan A. Nok, Sørge Kelm (2013). Biochemical diversity in the *Trypanosomal congolense* trans-sialidase family. *PLOS Negl. Trop. Dis.* **7** (12): e2549. doi:10.1371/journal.pntd.0002549.

Hendrik Koliwer-Brandl, Thaddeus T. Gbem, **Mario Waespy,** Olga Reichert, Phillip Mandel, Eric Drebitz, Frank Dietz, Sørge Kelm (2011). Biochemical characterisation of trans-sialidase TS1 variants from *Trypanosoma congolense. BMC Biochemistry,* **12** (1), 39.

6.3 List of manuscripts in preparation

Moein Amin, Paul D. Madge, Mauro Pascolutti, Andrea Maggioni, **Mario Waespy,** Bernadette Bellette, Robin J. Thomson, Sørge Kelm, Mark von Itzstein, Thomas Haselhorst (2015). The design of high-affinity Siglec-2 (CD22) ligands using cell-based NMR spectroscopy. *Manuscript submitted to Nature Communications (ID: NCOMMS-15-12784)*.

Jessica Nüsse, Ursula Mirastschijski, **Mario Waespy,** Janina Oetjen, Nadine Brandes, Frederico Paroni, Sørge Kelm, Frank Dietz (2015). Who is who in the world of hepatoma-derived growth factor? Differential tasks of isoforms. *Manuscript submitted to PLOS ONE (ID: PONE-D-15-29878).*

Mario Waespy, Thaddeus T. Gbem, Nilima Dinesh, Shanmugam Solaiyappan, Sørge Kelm (2015). The influence of *Trypanosoma congolense* trans-sialidase lectin domain on enzyme activities. *Manuscript in preparation.*

Thaddeus T. Gbem, **Mario Waespy,** Jonathan A. Nok, Sørge Kelm (2015). Diverse expression and different pH optima of trans-sialidases from *Trypanosoma congolense* is a direct response to changing environments during life cycle. *Manuscript in preparation.*