Table SI. Peptides containing *N*-glycosylation sequons identified by mass spectrometry. Proteins covalently linked to the yeast polysaccharide cell wall were prepared, glycans were released by endoglycosidase H, proteins were digested with trypsin or AspN, detected by LC-ESI-MS/MS and identified using Mascot.

Glycosylation site	Observed (*m/z*)	Observed (Da)	Predicted (Da)	Δmass (Da)	Mascot value	Expect	pI †	Sequence
CCW14_87*	1105.4781	2208.9416	2208.9481	-0.0064	0.00072		3.55	A.DAAYSAFKSSCSEQ**N**ASLG.D
CCW14_87	1003.9399	2005.8652	2005.8687	-0.0034	8.2e-08		3.55	A.DAAYSAFKSSCSEQNASLG.D
CRH1_177*	587.2590	1758.7551	1758.7559	-0.0008	0.0054		5.41	K.FH**N**YTLDWAMDK.T
CRH1_177#	519.5700	1555.6800	1555.6800	0.0000	8.40e-04		5.41	K.FHNYTLDWAMDK.T
CRH1_201*	1103.9957	2205.9768	2205.9810	-0.0041	0.00015		6.34	R.VLS**N**TSSEGYPQSPMYLM.M
CRH1_201	1002.4566	2002.8986	2002.9016	-0.0029	2.9e-07		6.34	R.VLSNTSSEGYPQSPMYLM.M
CRH2_28*	800.8329	1599.6512	1599.6545	-0.0033	0.0034		3.85	A.ATFC**N**ATQACPE.D
CRH2_28	699.2941	1396.5737	1396.5751	-0.0014	0.00011		3.85	A.ATFCNATQACPE.D
CRH2_96*	1095.5126	2189.0106	2189.0124	-0.0018	1.8e-05		3.55	K.DYSSKLGNANTFLG**N**VSEA.D
CRH2_96#	993.9700	1985.9300	1985.9300	0.0000	3.90e-06		3.55	K.DYSSKLGNANTFLGNVSEA.D
CRH2_233**#	694.3300	2079.9800	2079.9800	0.0000	4.00e-02		9.14	R.TLYK**N**ETY**N**ATTQK.Y
CRH2_233*	626.6443	1876.9110	1876.9054	0.0055	0.002		9.14	R.TLYKNETY**N**ATTQK.Y
CRH2_310*	1014.4757	2026.9368	2026.9371	-0.0003	0.0021		9.13	K.**N**GTSAYVYTSSSEFLAK.D
CRH2_310#	913.4300	1824.8400	1824.8400	0.0000	1.30e-08		9.13	K.NGTSAYVYTSSSEFLAK.D
CWP1_45	973.9697	1945.9248	1945.9269	-0.0021	1.5e-12		6.23	R.SGSDLQYLSVYSDNGTLK.L
ECM33_83*	1064.5200	3190.5382	3190.5405	-0.0023	0.0031		6.02	G.DLGSAALASIQEIDGSLTIF**N**SSSLSSFSA.D
ECM33_197*#	960.9200	1919.8300	1919.8300	0.0000	2.00e-03		3.49	S.DSLQFSSNGD**N**TTLAF.D
ECM33_210*	816.4229	1630.8312	1630.8315	-0.0003	4.3e-05		6.23	F.DNLVWAN**N**ITLR.D
ECM33_210	714.8821	1427.7496	1427.7521	-0.0025	2.2e-08		6.23	F.DNLVWANNITLR.D
ECM33_329*	1040.4481	2078.8816	2078.8851	-0.0035	1e-06		6.18	R.GGANFDSSSS**N**FSCNALK.K
ECM33_329	938.9090	1875.8034	1875.8057	-0.0024	3.5e-10		6.18	R.GGANFDSSSSNFSCNALK.K
EXG2_50*	848.4091	1694.8036	1694.8039	-0.0004	0.01		6.23	K.FASYYA**N**DTITVK.G
GAS1_40*	923.4253	1844.8360	1844.8370	-0.0010	0.0007		9.17	K.FFYSN**N**GSQFYIR.G
GAS1_57*	1098.8142	3293.4208	3293.4266	-0.0058	9.4e-11		3.55	R.GVAYQADTA**N**ETSGSTVNDPLANYESCSR.D
GAS1_57	1031.1222	3090.3448	3090.3472	-0.0025	1.4e-12		3.55	R.GVAYQADTANETSGSTVNDPLANYESCSR.D
GAS1_95*	922.5202	2764.5388	2764.5374	0.0013	0.003		5.46	R.DIPYLKKLNTNVIRVYAI**N**TTL.D
GAS1_95	854.8272	2561.4597	2561.4581	0.0017	0.0056		5.46	R.DIPYLKKLNTNVIRVYAINTTL.D
GAS1_149**	1404.3256	4209.9550	4209.9543	0.0007	2e-05		3.75	K.TVVDTFA**N**YTNVLGFFAGNEVTN**N**YTNTDASAFVK.A
GAS1_149**	1065.1405	3192.3997	3192.4048	-0.0051	5.3e-05		3.75	V.DTFA**N**YTNVLGFFAGNEVTN**N**YTNT.D
GAS1_253*	1223.5809	2445.1472	2445.1522	-0.0050	3.7e-05		6.34	K.**N**LSIPVFFSEYGCNEVTPR.L

GAS1_253	1122.0434	2242.0722	2242.0728	-0.0006	1.6e-08	6.34	K.NLSIPVFFSEYGCNEVTPR.L
GAS3_201*	847.0911	2538.2514	2538.2536	-0.0022	0.13	10.45	R.DMKQYISKHA**N**RSIPVGYSAA.D
GAS3_201	779.3984	2335.1733	2335.1743	-0.0010	0.093	10.45	R.DMKQYISKHANRSIPVGYSAA.D
GAS3_269	578.7716	1155.5286	1155.5295	-0.0009	0.031	3.88	Y.DKL**N**STFE.D
GAS3_350*	995.1586	2982.4539	2982.4557	-0.0018	0.0021	9.53	K.DDFVNLESQLK**N**VSLPTTKESEISS.D
GAS3_350[#]	927.4700	2779.3800	2779.3800	0.0000	4.00e-02	9.53	K.DDFVNLESQLKNVSLPTTKESEISS.D
GAS5_24*	805.8911	1609.7677	1609.7683	-0.0006	0.013	3.85	A.ASSS**N**SSTPSIEIK.G
GAS5_24	704.3501	1406.6856	1406.6889	-0.0033	1.6e-05	3.85	A.ASSSNSSTPSIEIK.G
GAS5_60*	788.0484	2361.1235	2361.1237	-0.0002	1.7e-05	3.49	G.ERFYIRGVDYQPGGSS**N**LT.D
GAS5_60*	1121.5385	3361.5937	3361.5984	-0.0047	4.2e-05	3.49	R.GVDYQPGGSS**N**LTDPLADASVCDRDVPVLK.D
GAS5_60	1053.8453	3158.5141	3158.5190	-0.0050	7.5e-08	3.49	R.GVDYQPGGSSNLTDPLADASVCDRDVPVLK.D
MKC7_346	821.9844	1641.9543	1641.9553	-0.0010	0.00019	3.88	E.DNLTTLTTTKIPVLL.D
PLB1_52*	891.4071	1780.7997	1780.8003	-0.0006	0.03	4.43	R.EASGLSD**N**ETEWLK.K
PLB1_489*	1217.1494	2432.2842	2432.2839	0.0004	2.7e-05	3.88	R.**N**LTDLEYIPPLIVYIPNSR.H
PLB2_47*	848.4067	1694.7988	1694.7999	-0.0011	1.6e-05	10.55	R.**N**ASGLSTAETDWLK.K
PLB2_47*	912.9444	1823.8742	1823.8749	-0.0007	5.5e-05	10.55	D.DTSLVR**N**ASGLSTAET.D
PRY3_101*	884.4063	2650.1971	2650.1976	-0.0005	0.00058	6.45	K.Y**N**YSNPGFSESTGHFTQVVWK.S
SAG1_79*	1135.5709	2269.1272	2269.1325	-0.0053	2e-09	11.66	K.LL**N**SSQTATISLADGTEAFK.C
SAG1_79	1034.0334	2066.0522	2066.0531	-0.0009	1.1e-12	11.66	K.LLNSSQTATISLADGTEAFK.C
SAG1_109*	752.3167	1502.6188	1502.6195	-0.0007	0.0001	3.85	Y.E**N**TTFTCTAQN.D
SAG1_109	650.7751	1299.5357	1299.5401	-0.0044	5e-07	3.85	Y.ENTTFTCTAQN.D
SAG1_248*	723.3017	1444.5889	1444.5936	-0.0047	0.00087	3.49	N.DWWFPQSY**N**.D
TOS1_417*	627.8141	1253.6136	1253.6139	-0.0003	0.033	7.55	K.AAVIF**N**SSDK.T
TOS1_417	526.2738	1050.5331	1050.5346	-0.0015	2.5e-05	7.55	K.AAVIFNSSDK.T

* GlcNAc-modified asparagine (Bold **N** in sequence). [#] Previously identified (Schulz and Aebi 2009). All methionines are oxidized and all cysteines are alkylated.
[†]pI of Acceptor Site ±5 amino acids calculated with bisector method and EMBOSS pK_i table excluding amine and carboxy terminal groups.

Table SII. Site-specific *N*-glycosylation occupancy of yeast strains expressing *T. brucei* Stt3p orthologues with and without yeast *STT3*. The relative *N*-glycosylation occupancy at a given site was determined from the abundance of GlcNAc-modified and unmodified versions of the same sequon-containing peptide as measured by LC-ESI-MS.

Glycosylation site	Strain					
	Wild type	pTbSTT3A	pTbSTT3B	pTbSTT3C	*Δstt3*	
					pTbSTT3B	pTbSTT3C
SAG1_248	1.00±0.00*	1.00±0.00	1.00±0.00	1.00±0.00	0.99±0.02	1.00±0.00
PRY3_101	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00
PLB1_52	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.72±0.15	1.00±0.00
PLB1_489	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00
GAS1_57	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.90±0.10	1.00±0.00
ECM33_83	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00
CRH2_96	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00
GAS1_149	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.99±0.02
ECM33_197	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.91±0.01	0.99±0.01
GAS3_269	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.99±0.01
CRH1_177	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.97±0.02	0.99±0.01
EXG2_50	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.96±0.05
GAS1_40	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.94±0.07
CRH2_310	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.91±0.16
GAS3_350	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.01	0.88±0.08
CRH2_28	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.98±0.02	0.87±0.06
GAS5_24	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.98±0.2	0.80±0.02
GAS1_95	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.67±0.13
CRH1_201	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.97±0.04	0.66±0.04
GAS1_253	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.91±0.04	0.63±0.11
GAS5_60	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.72±0.06	0.63±0.11
ECM33_210	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.99±0.01	0.61±0.05
CRH2_233	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.57±0.10
MKC7_346	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.39±0.17	0.43±0.04
PLB2_47	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.57±0.23	0.00±0.00
GAS3_201	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.00±0.00
ECM33_329	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.41±0.08	0.00±0.00
SAG1_109	0.97±0.02	1.00±0.00	0.93±0.06	0.87±0.10	0.69±0.07	0.30±0.02
SAG1_79	0.80±0.03	0.84±0.05	0.85±0.07	0.75±0.05	0.91±0.05	0.00±0.00
CWP1_45	0.02±0.02	0.02±0.01	0.01±0.01	0.00±0.01	0.00±0.00	0.00±0.00
CCW14_87	0.00±0.00	0.00±0.00	0.01±0.00	0.01±0.00	0.01±0.01	0.08±0.01
TOS1_417	0.00±0.00	0.00±0.00	0.23±0.08	0.00±0.00	0.43±0.17	0.03±0.04

* Error is range. Values are the average of three independent measurements

Table SIII. Isobaric glycoforms of sVSG221 detected by ES-MS

The molecular weights of different glycoforms of sVSG221 were calculated according to the indicated compositions. The -, trace, +, ++, +++ scores indicate the relative abundances of those glycoforms observed in sVSG preparations.

Protein[a]	I-cP[b]	GlcN[b]	EtNP[b]	HexNAc	Hex	Mr	WT	*STT3A,B,C*$^{+/-}$	*STT3 A-*	*STT3 A+*	*STT3 B-*	*STT3 B+*
1	1	1	1	2	9	48663	-	-	-	-	-	trace
1	1	1	1	2	10	48825	-	-	-	-	-	+
1	1	1	1	2	11	48987	-	-	-	-	-	+++
1	1	1	1	2	12	49146	-	-	-	-	-	+++
1	1	1	1	3	11	49188	-	-	-	-	-	+
1	1	1	1	2	13	49312	-	-	-	-	-	-
1	1	1	1	3	12	49350	-	-	-	-	-	++
1	1	1	1	2	14	49465	-	-	-	-	-	+
1	1	1	1	3	13	49513	-	-	-	-	-	+
1	1	1	1	3	14	49675	+	-	-	+	-	+
1	1	1	1	4	17	50363	trace	trace	trace	trace	trace	-
1	1	1	1	4	18	50525	++	++	++	+	++	-
1	1	1	1	6	16	50611	-	-	-	-	-	+
1	1	1	1	4	19	50687	+++	+++	+++	++	+++	-
1	1	1	1	5	18	50726	trace	trace	trace	trace	trace	-
1	1	1	1	6	17	50773	trace	trace	trace	trace	trace	-
1	1	1	1	7	16	50812	-	-	-	+	-	+
1	1	1	1	4	20	50849[c]	+++	+++	+++	+++	+++	-
1	1	1	1	5	19	50889	+	+	+	+	+	-
1	1	1	1	4	21	51012	++	++	++	++	++	-
1	1	1	1	5	20	51053	+++	+++	+++	++	+++	-
1	1	1	1	4	22	51174	+	+	+	+	+	-
1	1	1	1	8	17	51178	-	-	-	-	-	+
1	1	1	1	5	21	51211	+	+	+	++	+	-
1	1	1	1	4	23	51337	+	+	+	+	+	-
1	1	1	1	8	18	51340	-	-	-	-	-	+
1	1	1	1	5	22	51372	+	+	+	+	+	-
1	1	1	1	4	24	51497	trace	trace	trace	++	trace	-
1	1	1	1	5	23	51534	-	-	-	trace	-	-
1	1	1	1	4	25	51659	-	-	-	+++	-	-
1	1	1	1	4	26	51821	-	-	-	++	-	-
1	1	1	1	4	27	51983	-	-	-	+	-	-

a Protein Mr is based on the amino acid sequence of the VSG221 precursor (accession no. P26332) minus residues 1- 27 (signal peptide) and 460-476 (GPI attachment signal peptide) and allows for four disulfide bonds (Mr = 46284)

b Components specific to the GPI anchor and common to all glycoforms; I-cP *myo*-inositol-1,2 cyclic phosphate; EtNP, ethanolamine phosphate.

c The most abundant glycoform of wild-type sVSG221 is expected to contain a GPI anchor of composition of Man3Gal5, a C-terminal N-linked glycan of Man9GlcNAc2 and an internal N- linked glycan of Man3GlcNAc2; i.e., HexNAc (GlcNAc) = 4 and Hex (Man + Gal) = 20.

Table SIV. VSG221 glycopeptide ions identified by ES-MS. The masses of the principal glycopeptides observed after sVSG pronase digestion of the different cell lines and conditions are listed along with their relative intensities and their peptide and glycan assignments. The ions corresponding to the Asn263 external site are in yellow, those corresponding to the Asn428 external site are in pink and those corresponding to the Ser433 GPI site are in blue. Isobaric Asn263 and Asn 428 glycopeptides are in green.

				Ion intensity					
Ion	**Mass**	**AA**	**Glycan**	**wt**	**sKO**	**stt3A -Tet**	**stt3A +Tet**	**stt3B -Tet**	**stt3B + Tet**
[M+H]1+	1254.46	NET	Hex3HexNAc2	263	58	439.4	173.4	1095.3	881.7
[M+2H]2+	1287.54	NT	Hex4HexNAc2					92.2	
[M+2H]2+	1388.59	NTT	Hex4HexNAc2					43.5	
[M+2H]2+	1410.56	RNET	Hex3HexNAc2	324	170.8	529.2	202.6	854.1	396.7
[M+2H]2+	1416.51	NET	Hex4HexNAc2	13		44.7	18.3	424.7	196.3
[M+2H]2+	1423.36	S-EtNP	Hex5HexN-InocP					410.2	20.6
[M+2H]2+	1438.61	NET	Hex4HexNAc2						24.7
[M+2H]2+	1449.53	NT	Hex5HexNAc2	0.8				262.11	
[M+2H]2+	1457.54	NET	Hex3HexNAc3	146	57.3	222.4	31.1	1137.8	363.5
[M+2H]2+	1489.65	TNTT	Hex4HexNAc2						
[M+2H]2+	1490.56	NT	Hex4HexNAc3	0.5	0.4			37.1	24.5
[M+2H]2+	1550.58	NTT	Hex5HexNAc2	0.8				18.5	
[M+2H]2+	1578.56	NET	Hex5HexNAc2	0.7		5.5		32.8	38.4
[M+2H]2+	1585.42	S-EtNP	Hex6HexN-InocP	146	107.2	286.2	11.8	1884	89
[M+2H]2+	1591.61	NTT	Hex4HexNAc3	0.4		4.9		28.5	14.2
[M+2H]2+	1611.58	NT	Hex6HexNAc2	33.1	25.4			349	
[M+2H]2+	1613.64	RNET	Hex3HexNAc3	132	65.2	182.3	3.2	604.3	170.6
[M+2H]2+	1619.59	NET	Hex4HexNAc3	116	73.9	231.2	75.6	1387.4	323.8
[M+2H]2+	1635.65	RNET+Na	Hex3HexNAc3						
[M+2H]2+	1641.7	NET+Na	Hex4HexNAc3	54.2	22.6	70.4	52.3	316.5	350.2
[M+2H]2+	1651.63	TNTT	Hex5HexNAc2					3.9	
[M+2H]2+	1685.67	NET+3Na	Hex4HexNAc3						1.7
[M+2H]2+	1712.62	NTT	Hex6HexNAc2	69.7	38.9	1.7	2.5	535.3	
[M+2H]2+	1740.62	NET	Hex6HexNAc2			119.6	4.2	595.1	122.2
[M+2H]2+	1747.48	S-EtNP	Hex7HexN-InocP	170	327.7	600	133.8	3472.8	514.5
[M+2H]2+	1762.74	NET+Na	Hex6HexNAc2			0.4			167.1
[M+2H]2+	1773.62	NT	Hex7HexNAc2	50.3	35.4	108.4	54.8	488.3	
[M+2H]2+	1775.77	RNET	Hex4HexNAc3	143	91.7	318	85.4	788	224.4
[M+2H]2+	1781.64	NET	Hex5HexNAc3	44.5	32.2	100.7		786.4	35.2
[M+2H]2+	1803.78	NET+Na	Hex5HexNAc3	33.9				258.6	244.8
[M+2H]2+	1813.67	TNTT	Hex6HexNAc2	20.1	10.1			1.2	
[M+2H]2+	1814.66	NT	Hex6HexNAc3					1.23	
[M+2H]2+	1874.67	NTT	Hex7HexNAc2	135	65.8	182.9	60.2	922.4	
[M+2H]2+	1896.72	RNET	Hex6HexNAc2	80.1	25.4			233.5	194.6
[M+2H]2+	1902.86	NET	Hex7HexNAc2			117.7	1.7	0.6	0.4
[M+2H]2+	1909.53	S-EtNP	Hex8HexN-InocP	241	175.1	724.3	142.3	2491.9	438.1
[M+2H]2+	1935.67	NT	Hex8HexNAc2	131	112.6	192.2	107.7	1301.7	88.6
[M+2H]2+	1937.74	RNET	Hex5HexNAc3			131.8			
[M+2H]2+	1943.7	NET	Hex6HexNAc3			0.6		425.2	
[M+2H]2+	1956.73	NTT	Hex5HexNAc4			0.4			
	1957.6	NT+Na	Hex8HexNAc2	87	38.3			616.5	286.1
[M+2H]2+	1965.86	NET+Na	Hex6HexNAc3						155.5
[M+2H]2+	1975.72	TNTT	Hex7HexNAc2	42.3	18		60.7	274.8	
[M+2H]2+	1978.88	NTT+Na	Hex5HexNAc4						253.5
[M+2H]2+	2036.72	NTT	Hex8HexNAc2	262	126.6	357.9	197.6	1728.9	
[M+2H]2+	2065.01	NET	Hex8HexNAc2				42.5	0.8	
[M+2H]2+	2071.58	S-EtNP	Hex9HexN-InocP	125	31.9	349.4	57.1	827.8	152.6
[M+2H]2+	2097.72	NT	Hex9HexNAc2	212	159.2	307.5	180.3	2127	
[M+2H]2+	2198.77	NTT	Hex9HexNAc2	420	250	533.5	280	3084.5	
	2220.7	NTT+Na	Hex9HexNAc2	280	96.5	207.2	294.7	925.7	218.3
[M+2H]2+	2227.01	NET	Hex9HexNAc2		8.1		58.5		
[M+2H]2+	2299.82	TNTT	Hex9HexNAc2	183	50.3			1025.2	
[M+2H]2+	2586.16	NT	Hex7HexNAc6	8	1	55.8		178.3	124.1
[M+2H]2+	2608.16	NT+Na	Hex7HexNAc6		0.5				149.5
[M+3H]3+	2686.99	NTT	Hex7HexNAc6			72.8		207	103
[M+2H]2+	2709.2	NTT+Na	Hex7HexNAc6	7.5	1.2	19	30.3	46.9	146.7

Table SV. Distribution of the isoelectric point per class of glycan

Summary statistics for the distribution of pI of the sequon ± 5 amino acids for each class of glycan. The -, +, +/-, values indicate whether the glycan is resistant, sensitive or partially sensitive to digestion with EndoH, respectively.

	EndoH Sensitivity			
Site Type	-	+	+/-	No glycan
Count[a]	95	26	55	4
Mean[b]	4.8	9.5	7.3	6.0
Variance	3.0	4.7	5.8	7.1
Standard deviation	1.7	2.2	2.4	2.7

[a] Absolute number of sequons distributed according to their sensitivity to EndoH and/or its occupancy from a total set of 180 sequons identified.
[b] Arithmetic mean of the pI value of the sequon ± 5 amino acids

Table SVI. Statistical analysis of the pI distribution per class of glycan

Statistics analysis for the pI distribution of the sequon ± 5 amino acids for each class of glycan. The -, +, +/-, values indicate whether the glycan is resistant, sensitive or partially sensitive to digestion with EndoH, respectively.

Glycan PI value sets		Wilcoxon[b]		Kolmogorov-Smirnov[c]	
A	B	W Parameter	P value[a]	D parameter	P value[a]
EndoH (-)	EndoH (+)	149.5	7.3×10^{-12}	0.7789	3.5×10^{-11}
EndoH (-)	EndoH (+/-)	1025	5.8×10^{-10}	0.5426	2.5×10^{-9}
EndoH (+)	EndoH (+/-)	345	1.8×10^{-4}	0.4923	3.8×10^{-4}

a P-value is an estimate of the probability of the two sets of pI values (A and B) could be drawn from the same distribution.

b The Wilcoxon test, normally applied to test the membership of normally distributed data, is included since the datasets are sufficiently large for the P value to be meaningful for non-normally distributed data.

c The Kolmogorov-Smirnov statistic is suitable for testing membership of any distribution.

Fig. S1

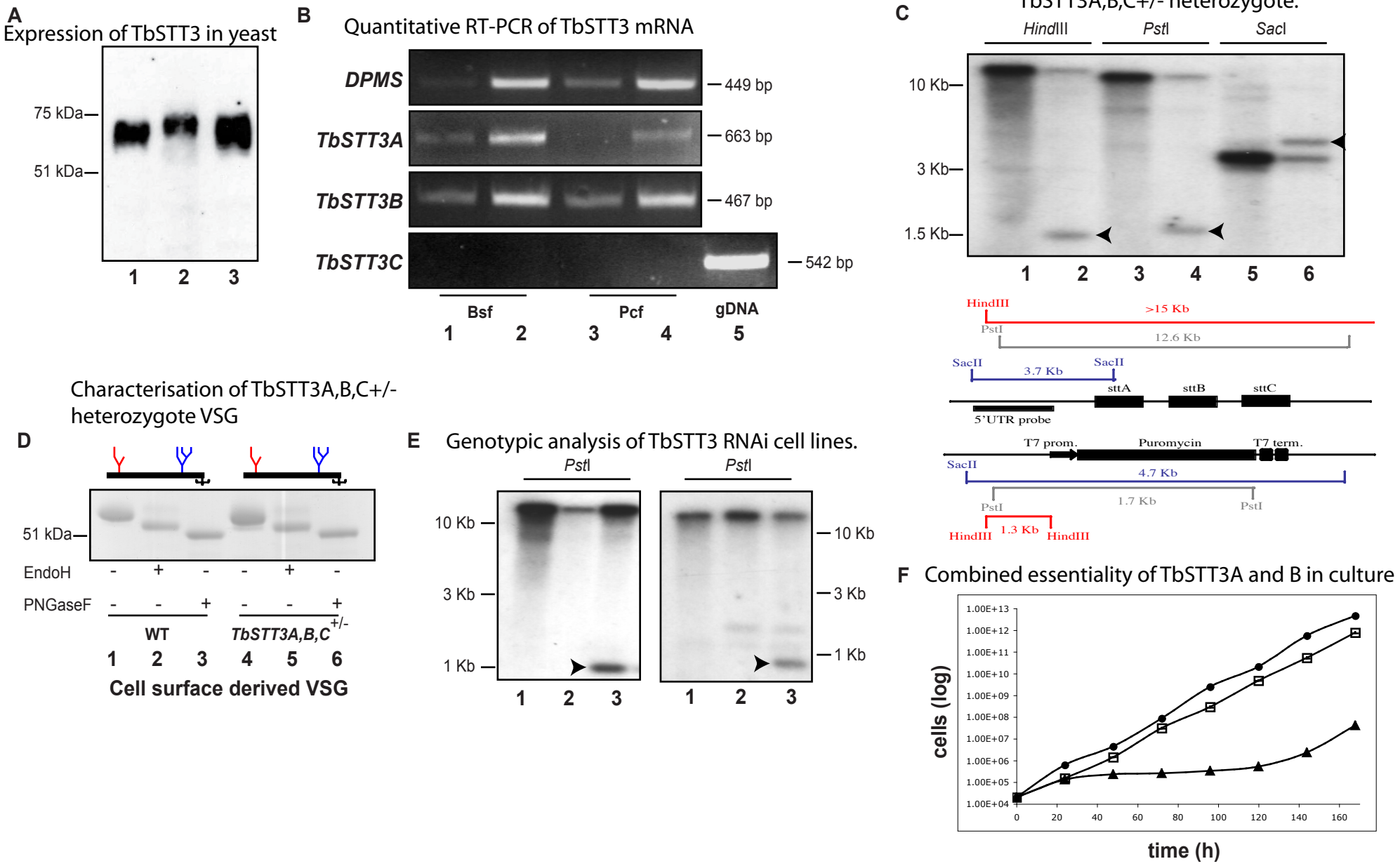

A
Expression of TbSTT3 in yeast
75 kDa
51 kDa
1 2 3
B
Quantitative RT-PCR of TbSTT3 mRNA
DPMS
TbSTT3A
TbSTT3B
TbSTT3C
449 bp
663 bp
467 bp
542 bp
Bsf
Pcf
gDNA
1 2 3 4 5
C
Creation of the T.brucei bsf
TbSTT3A,B,C+/- heterozygote.
HindIII
PstI
SacI
10 Kb
3 Kb
1.5 Kb
1 2 3 4 5 6
HindIII
>15 Kb
PstI
12.6 Kb
SacII
3.7 Kb
SacII
sttA
sttB
sttC
5'UTR probe
T7 prom.
Puromycin
T7 term.
SacII
4.7 Kb
PstI
1.7 Kb
PstI
HindIII
1.3 Kb
HindIII
D
Characterisation of TbSTT3A,B,C+/-
heterozygote VSG
51 kDa
EndoH - + - - + -
PNGaseF - - + - - +
WT
TbSTT3A,B,C+/-
1 2 3 4 5 6
Cell surface derived VSG
E
Genotypic analysis of TbSTT3 RNAi cell lines.
PstI
PstI
10 Kb
3 Kb
1 Kb
10 Kb
3 Kb
1 Kb
1 2 3
1 2 3
F
Combined essentiality of TbSTT3A and B in culture
cells (log)
1.00E+13
1.00E+12
1.00E+11
1.00E+10
1.00E+09
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
0 20 40 60 80 100 120 140 160
time (h)

Fig. S2

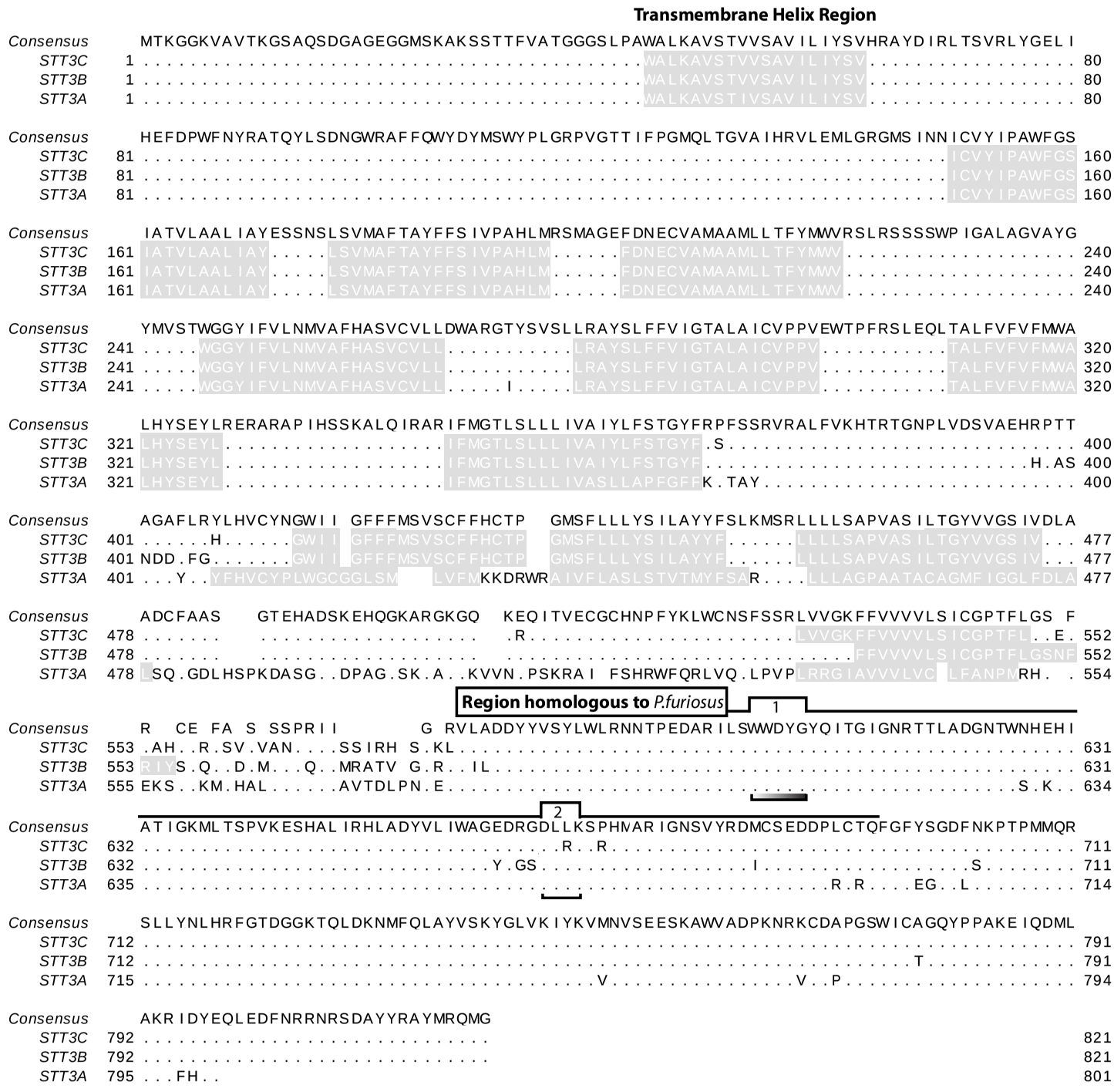

Transmembrane Helix Region
Region homologous to P.furiosus
1
2

Fig. S3

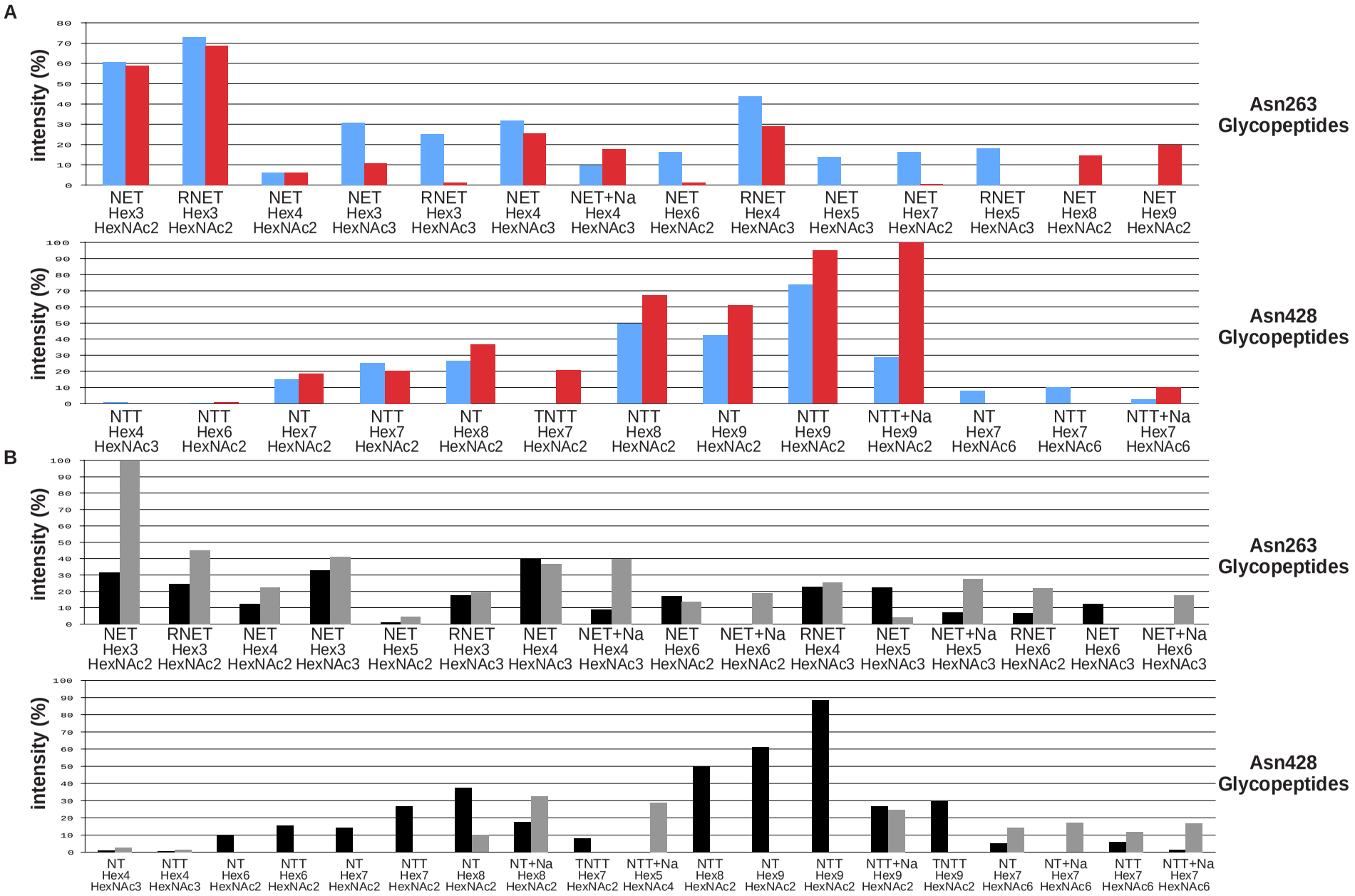
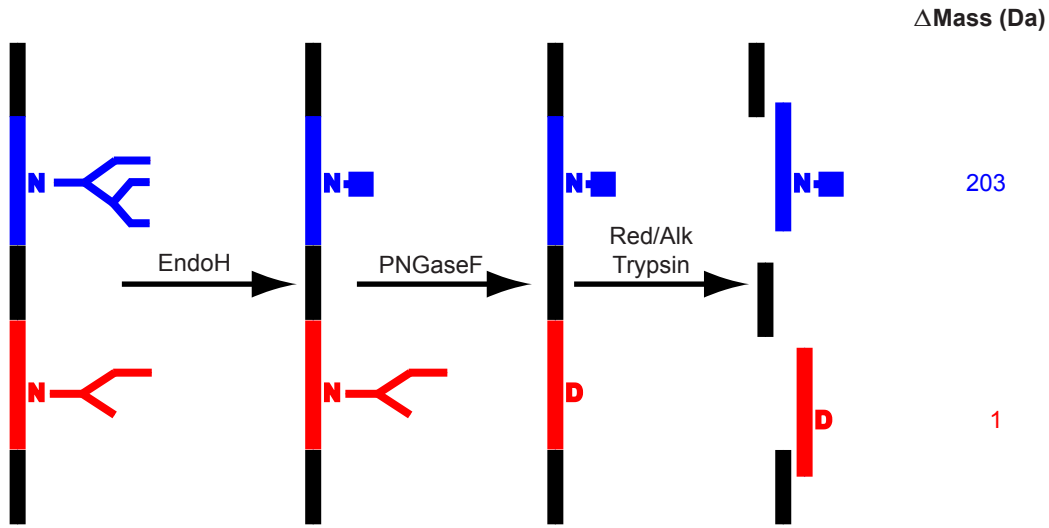

A
intensity (%)
NET Hex3 HexNAc2
RNET Hex3 HexNAc2
NET Hex4 HexNAc2
NET Hex3 HexNAc3
RNET Hex3 HexNAc3
NET Hex4 HexNAc3
NET+Na Hex4 HexNAc3
NET Hex6 HexNAc2
RNET Hex4 HexNAc3
NET Hex5 HexNAc3
NET Hex7 HexNAc2
RNET Hex5 HexNAc3
NET Hex8 HexNAc2
NET Hex9 HexNAc2
Asn263 Glycopeptides
intensity (%)
NTT Hex4 HexNAc3
NTT Hex6 HexNAc2
NT Hex7 HexNAc2
NTT Hex7 HexNAc2
NT Hex8 HexNAc2
TNTT Hex7 HexNAc2
NTT Hex8 HexNAc2
NT Hex9 HexNAc2
NTT Hex9 HexNAc2
NTT+Na Hex9 HexNAc2
NT Hex7 HexNAc6
NTT Hex7 HexNAc6
NTT+Na Hex7 HexNAc6
Asn428 Glycopeptides
B
intensity (%)
NET Hex3 HexNAc2
RNET Hex3 HexNAc2
NET Hex4 HexNAc2
NET Hex3 HexNAc3
NET Hex5 HexNAc2
RNET Hex3 HexNAc3
NET Hex4 HexNAc3
NET+Na Hex4 HexNAc3
NET Hex6 HexNAc2
NET+Na Hex6 HexNAc2
RNET Hex4 HexNAc3
NET Hex5 HexNAc3
NET+Na Hex5 HexNAc3
RNET Hex6 HexNAc2
NET Hex6 HexNAc3
NET+Na Hex6 HexNAc3
Asn263 Glycopeptides
intensity (%)
NT Hex4 HexNAc3
NTT Hex4 HexNAc2
NT Hex6 HexNAc2
NTT Hex6 HexNAc2
NT Hex7 HexNAc2
NTT Hex7 HexNAc2
NT Hex8 HexNAc2
NT+Na Hex8 HexNAc2
TNTT Hex7 HexNAc2
NTT+Na Hex5 HexNAc4
NTT Hex8 HexNAc2
NT Hex9 HexNAc2
NTT Hex9 HexNAc2
NTT+Na Hex9 HexNAc2
TNTT Hex9 HexNAc2
NT Hex7 HexNAc6
NT+Na Hex7 HexNAc6
NTT Hex7 HexNAc6
NTT+Na Hex7 HexNAc6
Asn428 Glycopeptides

Fig. S4 Scheme describing the proteomic approach for the identification of EndoH-sensitive (blue) and EndoH-resistant (red) N-glycosylation sites.



EndoH sensitive sites are cleaved with endoH, leaving a 203Da GlcNAc (blue square) tags attached to the glycosylated asparagine (N) residues. Subsequent treatment with PNGaseF cleaves all remaining N-glycans and converts the glycosylated asparagine (N) residues to aspartate (D) residues, wich are 1 Da heavier. Subsequent reduction, S-alkylation and digestion with trypsin liberates the modified peptides for anakysis by LC-MS/MS and the original glycosylation status can be inferred from the 203 Da and 1 Da mass shiflts in the precursor ions, the MASCOT assignments and and manual inspection of the MS/MS spectra.

Supplementary Figure Legends

Fig. S1.

(*A*) Expression of T. *brucei* Stt3p paralogues in *S.cerevisiae*. Whole cell protein extracts of yeast cells with wild type OTase expressing in addition TbStt3Ap, TbStt3Bp or TbStt3Cp (lanes 1-3, respectively) were separated by SDS-PAGE, electroblotted to nitrocellulose and probed with α-C-myc antibody.

(*B*) Quantitative RT-PCR of TbSTT3 mRNA. of mRNA levels of *DPMS*, *TbSTT3A*, *TbSTT3B* and *TbSTT3C* were quantified by RT-PCR amplification in bloodstream and procyclic form *T. brucei*, using total RNA from each lifecycle stage as the template (lane 1, 6 ng of bloodstream form total RNA; lane 2, 48 ng of bloodstream form total RNA; lane 3, 6 ng of procyclic form total RNA; lane 4, 48 ng of procyclic form total RNA; lane 5, 30 ng of *T. brucei* gDNA). The primers used were specific for the *STT3A*, *STT3B* and *STT3C* genes, as well as for the gene encoding *DPMS* (1) constitutively expressed in a similar way in both lifecycle stages.

(*C*) Creation of the T.brucei bsf TbSTT3A,B,C$^{+/-}$ heterozygote. Schematic representation of the replacement of one *TbSTT3* locus with drug resistance gene *PAC* (bottom) and Southern blot of genomic DNA (top) from wild type (lanes 1, 3 and 5) and the *TbSTT3A,B,C*$^{+/-}$ cell line (lanes 2, 4 and 6) digested with *Hind*III (lanes 1 and 2), *Pst*I (lanes 3 and 4) and *Sac*II (lanes 5 and 6) and probed with a 5'UTR *TbSTT3A,B,C* fragment. The additional bands in lanes 2, 4 and 6 (marked with an arrowhead) are consistent with the replacement of one of the two *TbSTT3A,B,C* loci with the *PAC* construct.

(*D*) Characterisation of TbSTT3A,B,C$^{+/-}$ heterozygote VSG. Purified mature sVSG WT (lanes 1-3) and *TbSTT3A,B,C*$^{+/-}$ (lanes 4-6) cells digested or not, as indicated, with Endo H or PNGase F were subjected SDS-PAGE and stained with Coomassie blue.

(*E*) Genotypic analysis of TbSTT3 RNAi cell lines. Southern blot of genomic DNA from wild type (lane 1), *TbSTT3A,B,C*$^{+/-}$ (lane 2) and the *TbSTT3A,B,C*$^{+/-}$-*STT3A*RNAi cell line (lane 3, left panel) or *TbSTT3A,B,C*$^{+/-}$-*STT3B*RNAi cell line (lane 3, right panel) digested with *Pst*I and probed with a *TbSTT3A* ORF fragment (left panel) or *TbSTT3B* ORF fragment (left panel). In both panels, the extra band in lane 3 (marked with arrowheads) are consistent to respective RNAi fragment integrated in the cell line genome.

(*F*) Combined essentiality of TbSTT3A and B in culture. Growth at 37^{o}C of *T. brucei* wild type cells (closed circles), *TbSTT3A,B,C*$^{+/-}$-*STT3All*RNAi minus Tet (open squares), *TbSTT3A,B,C*$^{+/-}$-*STT3All* RNAi plus Tet (closed triangles). Cells were counted in triplicate and mean values ± S.D. are shown.

Fig. S2. Analysis of TbSTT3 predicted amino acid sequences. Alignment of TbSTT3A, TbSTT3B and TbSTT3C extracted from HOGENOM family alignment HBG269471, A '.' indicates residues identical to the consensus, gaps shown as blanks. The solid line marks the region homologous to *Pirococcus furiosus* (2) identified using the sequence of STT3A as input to the FFAS03 fold recognition server (3). The conserved WWDYG and DxxK motifs are indicated in both panels with numbers 1 and 2, respectively. Transmembrane helices predicted by the Phobius server (4) are shown as white letters on a light grey background. Alignments visualized with Jalview (5).

Fig. S3. Relative abundances of N-linked Pronase glycopeptide ions.

(A) Intensities of ES-MS glycopeptide ions, relative to the most abundant glycopetide ion in the sample spectrum, for Asn263 (upper panel) and Asn428 (lower panel) Pronase glycopeptides of sVSG221 from *TbSTT3A,B,C*$^{+/-}$-*STT3A*RNAi cells grown minus Tet (blue bars) and plus Tet (red bars).

(*B*) Intensities of ES-MS glycopeptide ions, relative to the most abundant glycopetide ion in the sample spectrum, for Asn263 (upper panel) and Asn428 (lower panel) Pronase glycopeptides of sVSG221 from *TbSTT3A,B,C*$^{+/-}$-*STT3B*RNAi cells grown minus Tet (black bars) and plus Tet (grey bars).

The data used to generate these graphs are in Table SIV. The glycopeptide compositions were assigned by deduced glycopeptide mass and their MS/MS daughter ion spectra (not shown), as described in (Manthri et al, 2008).

Fig. S4. Scheme describing the proteomic approach for the identification of EndoH-sensitive and EndoH-resistant N-glycosylation sites. Glycoprotein enriched fractions from ricin- and ConA-affinity chromatography are digested sequentially with EndoH, leaving a 203 Da GlcNAc tag on the EndoH-sensitive sites, and subsequently with PNGaseF, leaving a 1 Da Asn→Asp tag on the EndoH-resistant sites.

Supplementary Material and Methods

Immunoblot analysis. Cells were grown at 23^{o}C to mid-log phase, corresponding to an O.D.$_{600nm}$ of 1. Cell pellets were lysed with glass beads and cell debris was removed by centrifugation at 3 000 rcf at 4 ^{o}C for 3 min. Proteins were denatured by incubation in reducing SDS sample buffer including 4M urea. Proteins were separated by SDS-PAGE, electroblotted to nitrocellulose membrane, and probed with α-C-myc antibody (Calbiochem).

Semiquantitative RT-PCR. In order to assess the amount of mRNA, RT-PCR reactions were performed using AccessQuick RT-PCR System (Promega). As template 6 ng and 48 ng of total RNA from bloodstream and procyclic form of the parasite was used. The reaction was carried out in a GeneAmp PCR System 2700 from Applied Biosystems. Tb927.5.890 mRNA 663bp fragment was amplified with primers 5' TGTGGCAAGTTTGCTTGC 3' and 5' ATTGGGTAGATCAGTCACG 3'. Tb927.5.900 mRNA 467bp fragment was amplified with primers 5' GATGATTTCTTTGGTTACC 3' and 5' CTCAGAATATATCCGGAA 3'. Tb927.5.910 mRNA 542bp fragment was amplified with primers 5' GGCCTTCCTACGTCATC 3' and 5' TCATCGGCAAGAACCAAC 3'. As a control of a similar mRNA levels in both life stages of the parasite, primers 5' AATGGATGCGGACCTTCAGCACCCAC 3' and 5' TAGAACCGTGAGCGCGGTGCCATAC 3' amplifying a 449bp length product of dolichyl-phosphate-mannose synthetase (Tb10.70.2610) were used (1). The cycling parameters used were 48 °C for 45 min, 94 °C for 2 min, 25 cycles of 94 °C for 40 s, 60 °C for 1 min 15 s, and 68 °C for 1 min followed by a final 7 min extension time at 68 °C.

Southern Blotting. Aliquots of genomic DNA isolated from 100 ml of bloodstream-form *T. brucei* cultures (~2 x 10^8 cells) were digested with various restriction enzymes, run on agarose gels and transferred to positively charged nylon membranes (Roche) for hybridization with [α-32P] dCTP-labeled DNA probes (Amersham Rediprime II).

Statistical analysis of glycan site dataset. After manual analysis and verification of the MASCOT searches, a dataset was obtained comprising 180 characterised sequons. pI calculations of the region comprising the sequon and the five adjacent amino acids

up and downstream were performed with the Bisector method as implemented in BioPerl, with th non-terminal amino-acid pKi values from the EMBOSS table. Both Wilcoxon and Kolgomorov-Smirnov tests confirmed that the distribution of pI for the three glycan site types are significantly different (Table S5).

Supplementary references

1. Mazhari-Tabrizi R, *et al*. (1996) Cloning and functional expression of glycosyltransferases from parasitic protozoans by heterologous complementation in yeast: the dolichol phosphate mannose synthase from Trypanosoma brucei brucei. *Biochem J* 316 (Pt 3): 853-858.

2. Igura M, *et al.* (2008) Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *Embo J* 27(1): 234-243.

3. Jaroszewski L, Rychlewski L, Li Z, Li W & Godzik A. (2005) FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Res* 33, W284-W288 http://ffas.burnham.org/ffas-cgi/cgi/ffas.pl

4. Käll L, Krogh A, Sonnhammer EL. (2007) Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res* 35 W429-W432

5. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, doi:10.1093/bioinformatics/btp033