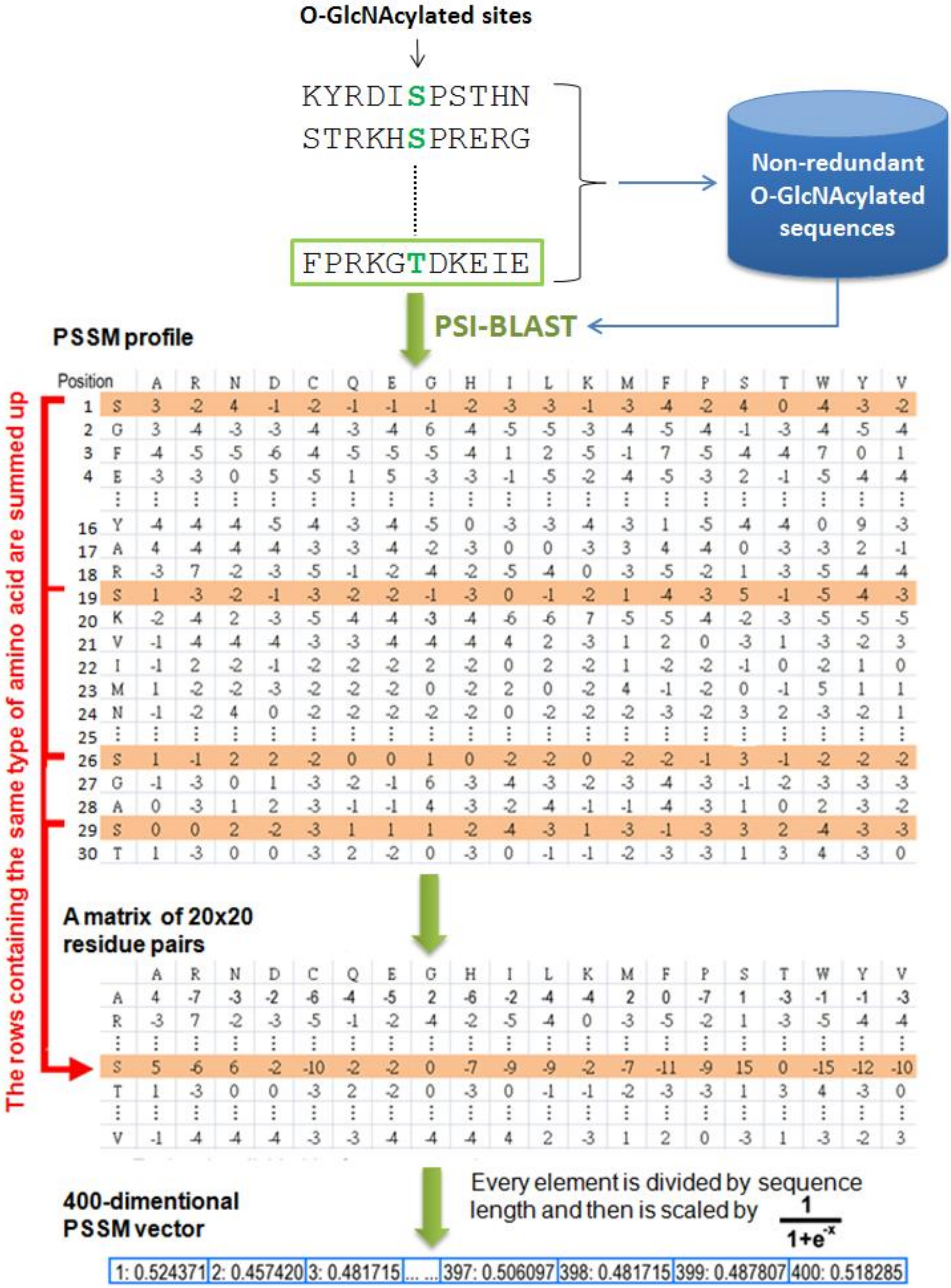
Supplementary Figures S1-S2 and Tables S1-S4



Position		A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
1	S	3	-2	4	-1	-2	-1	-1	-1	-2	-3	-3	-1	-3	-4	-2	4	0	-4	-3	-2
2	G	3	-4	-3	-3	-4	-3	-4	6	-4	-5	-5	-3	-4	-5	-4	-1	-3	-4	-5	-4
3	F	-4	-5	-5	-6	-4	-5	-5	-5	-4	1	2	-5	-1	7	-5	-4	-4	7	0	1
4	E	-3	-3	0	5	-5	1	5	-3	-3	-1	-5	-2	-4	-5	-3	2	-1	-5	-4	-4
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	Y	-4	-4	-4	-5	-4	-3	-4	-5	0	-3	-3	-4	-3	1	-5	-4	-4	0	9	-3
17	A	4	-4	-4	-4	-3	-3	-4	-2	-3	0	0	-3	3	4	-4	0	-3	-3	2	-1
18	R	-3	7	-2	-3	-5	-1	-2	-4	-2	-5	-4	0	-3	-5	-2	1	-3	-5	-4	-4
19	S	1	-3	-2	-1	-3	-2	-2	-1	-3	0	-1	-2	1	-4	-3	5	-1	-5	-4	-3
20	K	-2	-4	2	-3	-5	-4	-4	-3	-4	-6	-6	7	-5	-5	-4	-2	-3	-5	-5	-5
21	V	-1	-4	-4	-4	-3	-3	-4	-4	-4	4	2	-3	1	2	0	-3	1	-3	-2	3
22	I	-1	2	-2	-1	-2	-2	-2	2	-2	0	2	-2	1	-2	-2	-1	0	-2	1	0
23	M	1	-2	-2	-3	-2	-2	-2	0	-2	2	0	-2	4	-1	-2	0	-1	5	1	1
24	N	-1	-2	4	0	-2	-2	-2	-2	-2	0	-2	-2	-2	-3	-2	3	2	-3	-2	1
25	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
26	S	1	-1	2	2	-2	0	0	1	0	-2	-2	0	-2	-2	-1	3	-1	-2	-2	-2
27	G	-1	-3	0	1	-3	-2	-1	6	-3	-4	-3	-2	-3	-4	-3	-1	-2	-3	-3	-3
28	A	0	-3	1	2	-3	-1	-1	4	-3	-2	-4	-1	-1	-4	-3	1	0	2	-3	-2
29	S	0	0	2	-2	-3	1	1	1	-2	-4	-3	1	-3	-1	-3	3	2	-4	-3	-3
30	T	1	-3	0	0	-3	2	-2	0	-3	0	-1	-1	-2	-3	-3	1	3	4	-3	0

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-7	-3	-2	-6	-4	-5	2	-6	-2	-4	-4	2	0	-7	1	-3	-1	-1	-3
R	-3	7	-2	-3	-5	-1	-2	-4	-2	-5	-4	0	-3	-5	-2	1	-3	-5	-4	-4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S	5	-6	6	-2	-10	-2	-2	0	-7	-9	-9	-2	-7	-11	-9	15	0	-15	-12	-10
T	1	-3	0	0	-3	2	-2	0	-3	0	-1	-1	-2	-3	-3	1	3	4	-3	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
V	-1	-4	-4	-4	-3	-3	-4	-4	-4	4	2	-3	1	2	0	-3	1	-3	-2	3

1: 0.524371	2: 0.457420	3: 0.481715	... ...	397: 0.506097	398: 0.481715	399: 0.487807	400: 0.518285

Figure S1. The encoding scheme of position-specific scoring matrix (PSSM) with 400-dimensional vector.

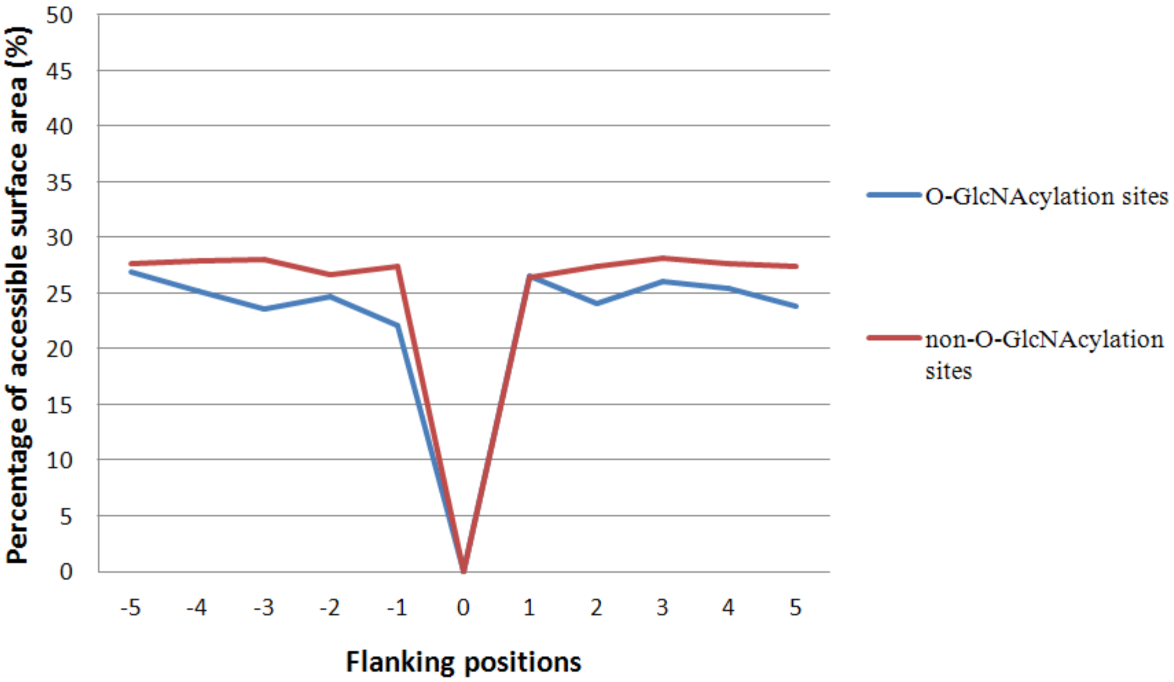


Figure S2. Comparison of solvent-accessible surface area between O-GlcNAcylated and non-O-GlcNAcylated sites.

Table S1. The amino acids group of MDDLogo used in this study.

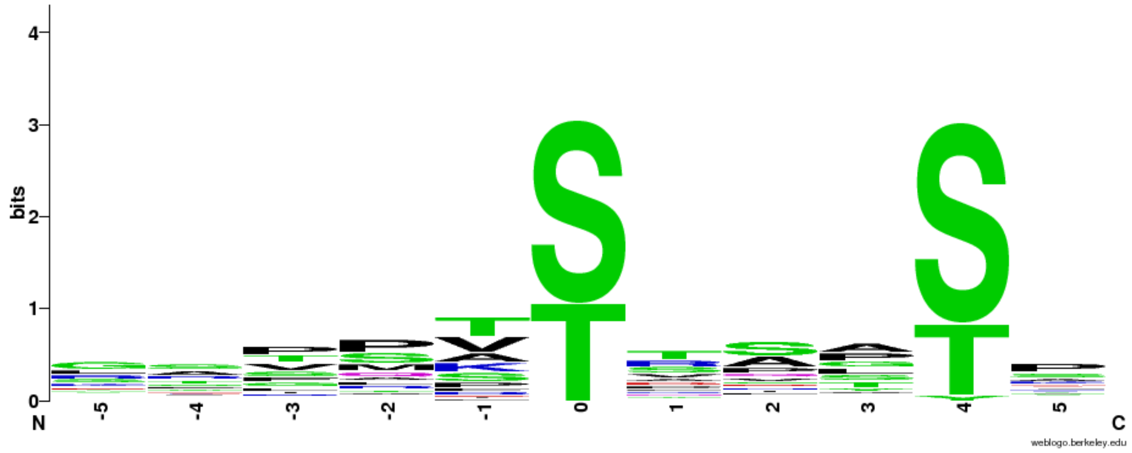
Chemical properties	Amino acids
Polar group	Glycine (G), Serine (S), Threonine (T), Cysteine (C), Glutamine (Q), Asparagine (N)
Acidic group	Aspartic acid (D), Glutamic acid (E)
Basic group	Lysine (K), Arginine (R), Histidine (H)
Hydrophobic group	Alanine (A), Valine (V), Leucine (L), Isoleucine (I), Proline (P), Methionine (M)
Aromatic group	Phenylalanine (F), Tyrosine (Y), Tryptophan (W)

Table S2. The detailed independent testing results between our method and other three O-GlcNAcylation prediction tools.

Predictive tools	TP	FP	TN	FN	Sn	Sp	Acc	MCC
Single SVM	683	21713	43282	365	0.65	0.67	0.67	0.08
Two-layered SVMs	836	3934	61061	212	0.80	0.94	0.94	0.36
YinOYang	482	12349	52646	566	0.46	0.81	0.80	0.08
O-GlcNAcScan	430	5199	59796	618	0.41	0.92	0.92	0.15
O-GlcNAcPRED	587	22748	42247	461	0.56	0.65	0.65	0.05

Table S3. The phosphorylation and O-GlcNAcylation sites that located in the consensus sequence of subgroup OG1 with S/T residues at +4.

Entropy plot of OG1 sequence logo



Gene Symbol	O-GlcNAc site	Consensus Suence[a]	PubMed ID[b]	Modification on +4 position[c]
ABLM3_MOUSE	419	LDVRS S TPTSY	19458039	O-GlcNAc
AKT1_HUMAN	308	GATMK T FCGTP	22629392 22322096	O-GlcNAc Phosphothreonine
AKT1_HUMAN	473	HFPQF S YSASG	22322096	Phosphoserine
BPTF_HUMAN	1749	ENCAK S TVTTT	20068230	O-GlcNAc
BPTF_HUMAN	2094	KMVLT T KVGSP	18669648 19690332 20068231 21406692 19413330	Phosphoserine
BSN_MOUSE	1395	GSPAT T ASPTR	16452088	O-GlcNAc
BSN_MOUSE	1407	YMTPT S PAGSE	16452088 22645316	O-GlcNAc
BSN_MOUSE	2054	DLRHP T DLLSH	16452088 19458039 22645316	O-GlcNAc
BSN_MOUSE	2070	RYSSV S NIYSD	22645316 18034455	O-GlcNAc Phosphoserine
BSN_MOUSE	2941	LAGPT T VPATK	16452088 19458039 22645316	O-GlcNAc
CEBPB_MOUSE	180	GSLST S SSSSP	15985551	Phosphoserine

			17110376 17264204 17601773 19478079 21183079 21659604 21659605 22262865 15985552	(GSK-3_beta)
CRTC2_HUMAN	173	RTSSD S ALHTS	19690332 18669648	Phosphothreonine
CTND2_MOUSE	453	GTFRT S TAPSS	16452088 16452087	O-GlcNAc Phosphoserine
DEMA_MOUSE	285	LHSGT S KSSSL	16452087 17114649 18388127 19737024 20415495 21183079	Phosphoserine
HCFC1_HUMAN	579	TTIVK T MAVTP	20068231	Phosphothreonine
K1C18_HUMAN	30	GARPV S SAASV	16083285 9524113 10623656 14694114 18669648 19415658 20044836 20068231 15368451 19007248	Phosphoserine (CDK1)
K1C18_HUMAN	49	SRISV S RSTSF	10623656 18669648 20044836 20068231 17081983 7523419 1374067 17924679	Phosphoserine (PKC_epsilon;CAMK2A;PRKCA;PRKCE;RPS6KA3;AURKA)

					19007248 15368451 16424149	
KCC4_HUMAN	356	ESHKA	S	RDPSP	17192257 19690332 20363803 19413330	Phosphoserine (CAMK4)
LMNA_HUMAN	612	VGGPI	S	SGSSA	17081983 18669648 19691289 20068231	Phosphoserine
MYC_HUMAN	58	FELLP	T	PPLSP	10551811 10866684 11018017 11243879 11904304 14563837 15103331 15150404 15175348 15210690 16452238 16483932 1651323 16537924 16855632 17157259 17634226 18220336 18408012 18669648 18701488 19028839 19131971 20068231 20140016 21406692 21659604	Phosphoserine (MAPK1;GSK3;CDK2)

					21808024 22025562 22115753 8386367 17081983 17558397 19966300 20713526	
MYH6_RAT	622	MATLF	S	TYASA	18988896	O-GlcNAc
NCOR1_HUMAN	1487	RRTPV	S	YQNTM	17081983	Phosphothreonine
NFL_MOUSE	48	YSAPV	S	SSLSV	16452088	O-GlcNAc
NFM_MOUSE	46	SRGSP	S	TVSSS	16452088	O-GlcNAc
NFM_MOUSE	47	RGSPS	T	VSSSY	16452088	O-GlcNAc
P121A_HUMAN	693	TKPQA	T	SAPSP	18669648 19413330	Phosphoserine
RRP1B_HUMAN	731	LKTPT	S	SPASS	17081983 18669648 19664995 19651622 19691289 20068231 19690332 21659604 22115753	Phosphoserine
SHAN2_MOUSE	1292	PEVMS	T	VSGTR	16452088	O-GlcNAc
SKT_MOUSE	357	SSLPV	S	RSISP	19367708	Phosphoserine
SP1_RAT	699	KRFMR	S	DHLSK	16332679 15561949 12900380 9111324 16365142 16052526	O-GlcNAc Phosphoserine
SPTB2_HUMAN	2324	STPAS	S	RAQTL	17081983 18669648 19651622 19362540	Phosphothreonine
SRBS1_MOUSE	1201	LPYSS	S	PSRSA	20139300 21183079	Phosphoserine

SRF_HUMAN	309	APVSA	S	VSPSA	1512232 12684542	O-GlcNAc
SYN1_MOUSE	564	GAPQA	T	RQASI	16452087 19737024 20415495	Phosphoserine
SYN1_RAT	562	GPPQA	T	RQASI	3118371	Phosphoserine (CaMK2)
TAU_RAT	711	KSPVV	S	GDTSP	15249677 15375165 15719129 15992372 16397900 16452087 16675464 17085446 17537547 18245083 18587388 19054758 19451179 20308788 21489990 8245007	Phosphoserine (CK1;PDPK1)
VIME_HUMAN	34	SYVTT	S	TRTYS	18083107 20090780 21253578	Phosphotyrosine
YTHD3_MOUSE	201	IGGDL	T	AAVTK	22645316	O-GlcNAc
ZC3HE_HUMAN	369	ESVTK	T	TNYST	19413330	Phosphoserine

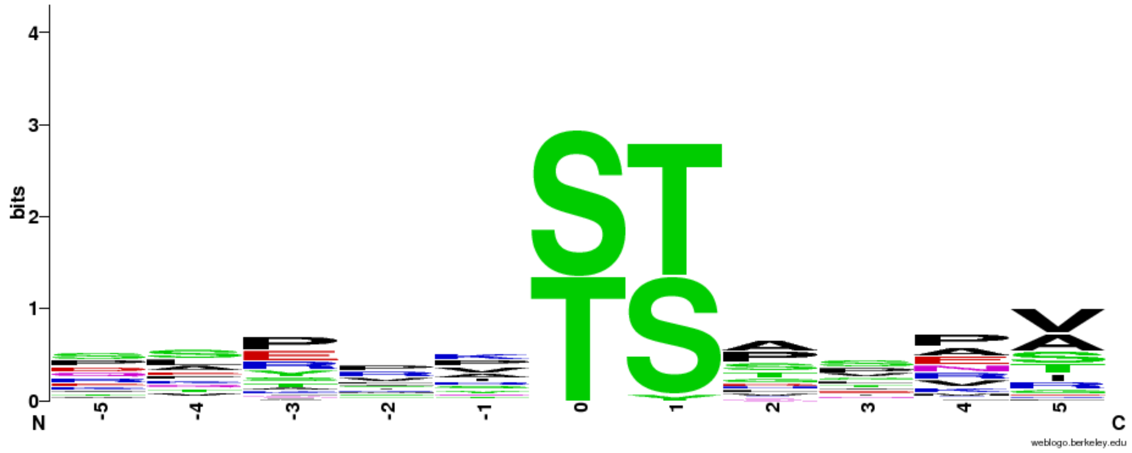
[a] Residues colored by blue is the annotated O-GlcNAylation sites in the training data.

[b] PubMed ID that reported the modification on +4 position to the central residue

[c] Reported phosphorylation or O-GlcNAcylation on +4 position

Table S4. The phosphorylation and O-GlcNAcylation sites that located in the consensus sequence of subgroup OG2 with S/T residues at +1.

Entropy plot of OG2 sequence logo



Gene Symbol	O-GlcNAc site	Consensus Suence	PubMed ID	Modification on +1 position
BSN_MOUSE	1418	RSPST S STIHS	16452088	O-GlcNAc
BSN_MOUSE	2067	PLRRY S SVSNI	16452087 20415495 21183079 6452088 19458039	O-GlcNAc Phosphoserine
BSN_MOUSE	2188	RQLLP S TATVR	16452088 22645316 19458039	O-GlcNAc
BSN_MOUSE	2317	EEPFS T TAPAV	16452088 19458039 22645316	O-GlcNAc
DLGP1_RAT	525	SSPPR T TTTVR	19458039	O-GlcNAc
EMSY_HUMAN	556	GGKII S SNIVS	20068230	O-GlcNAc
EMSY_MOUSE	499	TYVKT T SGSII	21606357	O-GlcNAc
FOXO1_HUMAN	317	TFRPR T SSNAS	18420577	O-GlcNAc
H31_HUMAN	11	QTARK S TGGKA	18066052 18243098 19407812 20948546 21357687 21514157 12560483	Phosphothreonine (DAPK3;PKC;CHEK1)

HCFC1_HUMAN	800	PITII T TKVMT	20068230 23335398	O-GlcNAc
K2C8_HUMAN	13	KSYKV S TSGPR	19007248	Phosphothreonine
KPCG_RAT	689	DARSP T SPVPV	18295358	O-GlcNAc
LT_SV40	111	SEEMP S SDDEA	9499386	O-GlcNAc Phosphoserine
MYH6_RAT	1470	QSELE S SQKEA	18988896	O-GlcNAc
NU153_HUMAN	1112	QQEPV T STSLV	20068230	O-GlcNAc
NUP98_HUMAN	567	PKALQ T TGTAK	20068230	O-GlcNAc
P53_HUMAN	149	QLWVD S TPPPG	11285227 12628923 17906639 21127074	Phosphothreonine
PCLO_MOUSE	2686	DVKRQ T TANEV	22645316	O-GlcNAc
PCLO_MOUSE	2889	AQQPA T TLPED	16452088 19458039 119458039 15648052 22645316	O-GlcNAc
PCLO_MOUSE	2993	STPQY S TARMT	16452088	O-GlcNAc
PHB_HUMAN	258	RSRNI T YLPAG	19238206 22410782	Phosphotyrosine
RBP2_HUMAN	1399	VFTPK T SPENV	20068230 18669648 19651622 20068231	O-GlcNAc Phosphoserine
RPRD2_HUMAN	399	EKSAV S TSVPT	20068230 19413330	O-GlcNAc Phosphothreonine
SON_HUMAN	252	AAPVP T TTLVL	20068230	O-GlcNAc
SP1_RAT	641	KVYGK T SHLRA	16332679 15561949 12900380 9111324 16365142 16052526	O-GlcNAc Phosphoserine
SRBS1_MOUSE	1200	DLPYS S SPSRS	16452088 15648052	O-GlcNAc Phosphoserine

					18056256 19367708 19737024 21183079 21659604	
SYN1_MOUSE	55	SAERA	S	TAAPV	16452088 19004831 16452087	GlcNAc Phosphothreonine
TPR_HUMAN	1676	PTPVV	S	TPSKV	18669648 19413330 17446864 18847512	Phosphothreonine
UBP2L_HUMAN	782	RSSVA	T	TSGKA	20068230	O-GlcNAc
ZFR_HUMAN	148	YSYVR	S	TAPAV	20068230	O-GlcNAc