

Online Supporting Information S2. The independent testing dataset S_T consists of a positive dataset S_T^+ and a negative dataset S_T^- . The former contains 81 SNO sites and the latter 100 non-SNO sites. None of the sites and peptide fragments included here occurs in the learning dataset S_L . See the text of the paper for further explanation.

S_T^+ contains 81 SNO sites and their corresponding peptide fragments in proteins

Uniprot ID	Site	Sequences
Q08211	1029	NALIHKSSVNC PFSSQDMKYP
P97447	71	WHDNCFRCAK CLHPLA SETFV
Q99NB1	41	SRSTATRLPG CV PAAAQPGSY
O08756	91	KFGRIDVAVN CAGI AVAIKTY
Q8BGK2	294	LASGSNWTEL CQ RAMFHGGES
Q9Z2I9	270	MVEDSDGKVL CM DAKINFDSN
O95573	450	LLGGNIRLLL CGG APLSATTQ
P13804	53	AATRLGGEVS CL VAGTKCDKV
P46060	169	GKILAAALTE CHR KSSAQGKP
Q13907	172	NPDPNEIKSY CY VSKEELKEL
O00468	405	CPEPCRFNAV CL SRRGRPRCS
O55143	998	RNYLEQPGKE CV QPATKSSCS
P23297	86	VVLVAALTVA CNN FFWENSXX
Q8JZX4	343	ECEKYGKVGK CV IFEIPGAPD
Q9WU79	162	EEAERKEMES CT SEAERDGSG
O60502	896	RILEFYSKLG CFE IAKMEGFP
P68366	54	GGDDSFTTFF CE TGAGKHVPR
P21817	2326	KGYPDIGWNP CG GERYLDFLR
P11142	574	DEDKQKILDK CNE IINWLDKN
Q9NPA8	40	KELLRAKLIE CG WKDQLKAHC
P23528	147	ANCYEEVKDR CTL AEKLGGSA
O75390	211	DSMDLIAKLP CV AAKIYRNLY
Q99439	61	QKGLKDGTIL CTL MNKLQPGS
P56382	19	GLSYIRFSQI CA KAVRDALKT
Q99JW4	94	KAMNNSWHPE CF RCDLCQEVL
P62878	94	ISRWLKTRQV CP LDNREWEFQ
Q9UBF2	296	APAVSVLQLF CSS PKPALRYA
Q8BMS1	747	ESAYGTQFT PC QLLLDHANNS
P50914	54	QVRRQAMPFK CM QLTDFILKF
Q61425	211	CKTLGKHPVS CK DTPGFIVNR
P10518	162	EVALAYAKAG CQ VVAPSDMMD
P07237	53	YLLVEFYAPW CG HCKALAPEY
P61982	112	SLLDNYLIK NC SETQYESKVF
Q9Z0X1	255	NDGSQITFEK CL IATGGTPRS
P50247	195	TKSKFDNLYG CR ESLIDGIKR
P97370	249	GLNKKEVTVE CHI AGTRNLKN
P63244	249	LCFSPNRYWL CA ATGPSIKIW
Q9JHI5	348	RQYVYNVAKA CD EGHIIPKDC
O70572	176	TSKNADVLL CG DLNMHPKDL
Q9CQ60	33	AQLVAQRAAS CL EGDRGRFAL
P51174	342	HKLAELKTHI CV TRAFVDSCL

Q9WTI7	763	KFLRVKRSACIQSWWRGTLG
P16858	154	IVSNASCTTNCLAPLAKVIHD
Q8R081	401	MNCDRVFNVFCLYGNVEKVKF
Q9D172	219	EAIKALGAKHCVKGVTEAHVD
Q80UJ7	181	HKWRRMYMGE CQGPVVRTDFE
P35754	83	VFIGKDCIGGCSDLVSLQQSG
Q9R0P5	23	RIFYDMKVRK CSTPEEIKKRK
P04350	354	WIPNNVKTAVCDIPPRGLKMA
P48444	441	HDSRRNTLEWCLPVIDAKNKS
Q9CQR4	74	VDSISTMALMCTERGAPGVS
O75369	604	IEYNDQNDGSCDVKYWPKEPG
P10415	158	VAFFEFGGVMCVESVNREMS
A2AJB2	26	RHPGLEEFACQSHAFMKGVF
Q924X2	86	KVDISMGLVDCIQRCLPERYG
Q9CQ48	45	PPGTRAQDIQCGLQSRHVALA
P08670	328	EYRRQVQSLTCEVDALKGTNE
P07900	597	VVSNRLVTSPCCIVTSTYGWT
P51125	408	AKAKEERQEKCGEDEDTPAE
Q9NP79	155	ARWKATYIHNCLKNGETPQAG
Q9DBC7	18	TSEEERSLRECELYVQKHNIQ
Q91W50	129	APGQSPTGSVCYERNGEVFYL
Q06830	173	FQFTDKHGEVCPAGWKPGSDT
P53618	248	NPSEARFIRCIYNLLQSSSP
Q9D819	242	TKKTDGKGISCMNTTVSESPF
Q9D023	54	WAPIMKWGLVCAGLADMARPA
P28650	183	SKAARTGLRICDLLSDFDEFS
Q8VDG5	308	DLRSRHTAFICDKNXXXXXXXX
Q9WVQ5	146	HQEMIKGIRKCTSGGYRYDD
O09110	29	SKRKKDLRISCVSKPPVSNPT
Q8BH95	225	VEKLVEEAIQCAEKIASNSKI
P63330	20	LDQWIEQLNECKQLSESQVKS
Q8BW75	365	LTKEERLRKLCELYAKVLNSQ
P55060	344	FEDQNTLTSICEKVIVPNMEF
O95801	374	FLVCVGSSPFCKNFLRGRKVY
Q9DC70	164	EPRYVSMGSCANGGGYYHYS
P62910	91	NVKELEVLLMCNKSYCAEIAH
Q60864	370	ALEEKKNKGNECFQKGDYPQAM
P19338	543	FEDAKEALNSCNKREIEGRAI
O94874	708	THSMLHAPGRVVPQIIAFLNS
O35295	250	DSKRFFFDVGCNKYGVFLRVS

\mathcal{S}_T^- contains 100 non-SNO sites and their corresponding peptide fragments in proteins

Uniprot ID	Site	Sequences
Q99KB8	20	LRSLSALGATCARRGLGQALL
Q9CXJ4	259	KLVISQGLRSC TQVIGSLVSL
Q99JW4	100	WHPECFRCDL CQEVLADIGFV
Q8K274	150	VDQFGFDVVTCCGYLPQVNDW
Q3UZY0	740	LQNWFOHWRFC SQRAAQORFO
O15523	466	EDFLYHEGYACTSIHGDRSQR

P70695	129	DPLDGSSNIDCLASIGTIFAI
Q92616	2558	PAKLSSLFVKCLQNPSSDIRL
Q9D6R2	351	LTKDLGGNAKCSDFTEEICRR
O75828	150	SLQCLRAFENCSEDLQERFHS
P63323	108	GKPRKVVGCSVVVKDYGKES
O35737	267	SDRFGRDLNYCFSGMSDHRYG
Q6R5N8	735	DVYGNKLQCTCDNLWFKNWSM
Q14690	510	GDEVKCRVLLCDPEAKKLMMT
P78527	1525	LELAFAFGGLCERLVSLLLNP
Q71FD7	337	GRNFHENCYRCEDCSVLLSVE
O75369	622	EPGEYAVHIMCDDEDIKDSFY
Q8BLF1	380	LDHFEDGFHGCMIFTSWPTNF
Q6NSR8	115	PGTHRCILMVCEQTEVFASAC
O70433	221	TCFCDLYAKKAGCTNPISGL
O00233	59	GIGMNEPLVDCGYPRSDVDL
P48725	1299	PRPPGPRDSQCVQLDEEVELL
Q71FD7	362	GCYPLNDHLFCCKPCHLKRSAA
O15519	426	SHSSPSLYLQCLSQKLRQERK
Q9BQG0	305	KMQFWPASYLCFRLLGAALPL
Q9DCW4	71	IIAVSCGPSQCQETIRTALAM
Q8BRK8	302	AVKEVCEKFECTESEVMNSLY
P15532	145	QPEELVEYKSCAQNWIYEXXX
P40227	508	YCVKKQLLHSCTVIATNILLV
Q99798	27	VRQYHVASVLCQRAKVAMSHF
Q791T5	340	LVGDLMAVNNCGLRAGLPPYS
P23368	120	IVYTPTVGLACSQYGHIFRRP
P27635	12	GRRPARCYRYCKNKPYPKSRF
O70433	51	EECGTPIGCDCKDLSYKDRHW
Q8VDC0	849	QMAVLINNKACGKIPVPQHVA
Q01433	41	LDVAEPGPSRCRSDSPAFAAV
Q99MN9	383	SVKGARFVRFCDAFNIPLITF
Q921H8	408	GRRAYGVVSMCIGTGMGAAAV
O75694	651	NPATQATNMSCVGTGPEIVYSG
Q8CI94	581	IHEYKRQLLNCLHIITLYNRI
P70398	739	PSLLTENGMKCFERFFKAVNC
Q9Y6I9	92	DNPHMVPPDKCRAVGSILSE
B1AR13	21	DFIFLFPNHICLPALSHPYQR
O00468	682	GGSGSGEDGCEQELCRQGG
Q86VP6	71	NGEVQNLAVKCLGPLVSKVKE
Q9CPR5	187	FYDPRSLEILCKPIPFPLRGQ
O75153	184	KKGLEMDPIDCTPPEYILPGS
P00533	781	NPHVCRLLGICLTSTVQLITQ
A2ASS6	7209	IRPGGNYTITCVGNTPHLRIL
P09528	103	DWESGLNAMECALHLEKSVNQ
P70398	831	DRLKASYDTLQVLDGDKDSIN
P56399	532	LVRAQVPFSSCLEAYGAPEQV
Q9R0Q7	75	KHKRTDRSILCCLRKGESGQS
Q9CQA3	188	DREKLDGLYECILCACSTSC
P50462	61	AHESEIYCKVCYGRRYGPKGI
Q9JHI5	379	ATQVALDGIQCLGGNGYINDF
P41250	444	ENEMAHYACDCWDAESKTSYG
O00468	414	VCLSRGRPRCSCDRVTCDGA

Q3TC72	21	SALLQVQKRPCQPSRNMRLVQ
Q9JHR7	966	VHVLAREMDSCPVVGFEFSON
P16332	558	ALAVDAARARCTVGEITDALK
Q9DCL9	288	LGHCEKIKKACGNFGIPCELR
O89103	440	DARGNPCDSLCFNTDGSFRCG
Q9Y3I1	359	VRSVLSLSAVCRDLFTASNDP
Q9DBC0	21	SLAVARTRPRCVGLALPSSAP
Q8TCG1	663	QADRLIAQHRQRTQAETEAR
O55070	29	ALHDTLALRLCSFNVRSGAS
Q64514	197	KYSDPGPVYDCLVWHDGETWR
Q9CW46	161	LVRPFGSLERCFLVYSERTGH
P23229	978	SVNVNVCVNIRCPRLGLDSKAS
Q99615	164	EKRDFRKVVFCDRALEFAPA
Q8BGD5	558	SHFGKSFIKCHVSSDSFIQL
O75153	899	MTPQELWKNICQEAKNYDFD
O75369	1876	ISCIDNKDGTCTVTYLPTLPG
Q8K4Q0	1168	VQDIPTGADSCVTSLSCDSHR
Q8BIF0	190	SYISYQOKKFCFSIQOGLNAD
Q8BUV3	502	PIGHDIKRGEVLAKGTHMGP
Q9HC38	197	RALLGYADNQCKLELOGVKGG
Q08211	578	YPVQEYFLEDCIQMTHFVPPP
Q91YE3	178	LKLALYIIVPCMKNHGICVVD
Q02248	429	CAAGILSNLTCNNYKNKMMVC
Q91ZA3	708	GKMGKVKLVHCKAGDTVGED
Q9QUR6	25	SVQEYHGHKICDPYSWLEDPD
Q8C460	258	WDLKVMLPGQCHYLGLPVADY
Q8K2M0	143	EAVRAEWERTCGPYHKQRLAE
Q8BGK2	158	LETLIEVSIECGRMTHNHPTG
Q62234	1010	GLGAPSTVSECFKCEEWTIAV
P41216	333	LAHMFETVVECVMLCHGAKIG
P49327	609	AVLAAYWRGQCIKEAHLPPGA
Q9JL8	455	QFAHTVNATACAVPRVLIALL
Q62407	1413	IVYVVEGQPACVTVTFNHVEA
Q9CY45	149	VADPPYLSEECLRKTSETIQF
O00429	345	QLITKFATEYCNTIEGTAKYI
Q9CW46	251	ALSVVYTPTFCQLACGQDGQL
Q8TAT6	355	GDFQNKHPNMCRLSPDGHFGS
Q5UIP0	382	DSNASPQNSCHVATSPGLNP
Q6PAR5	1432	VQYISSFYASCLSGEESYWWM
P35285	120	SIVVAIAGNKCDLTDVREVME
P15170	261	NPKKDIHFMPCSGLTGANLKE
Q6PAR5	702	PGSGSVLLDPCTGSTISETTS
