

Supplementary Information

Supplementary Information S4. The positive dataset $\mathbb{S}_{\text{Hyl}}^+$ contains 44 hydroxylysine peptide fragments and the negative dataset $\mathbb{S}_{\text{Hyl}}^-$ contains 528 non-hydroxylysine peptide fragments. None of the sequences included has $\geq 40\%$ pairwise sequence identity with any other in a same subset. See the text of the paper for further explanation.

Positive dataset		
Uniprot ID	Position	Fragments
Q3Y5Z3	28	EDPPLP K GACAGW
Q3Y5Z3	60	DGTPGE K GEGKDP
Q3Y5Z3	63	PGEKGE K GDPGLV
Q3Y5Z3	72	PGLVGP K GDTGET
Q15848	101	PGIQGR K GEPGEG
Q4ZJN1	127	QGQKGD K GVEGPT
P30754	183	PGTPGS K GDRGQS
P30754	351	SGEPGA K GPTGDT
P30754	546	PGVAGA K GQGGPP
P30754	567	PGERGP K GVGGSK
P30754	933	KGEAGG K GAKGDK
P30754	939	KGAKGD K GWTGLP
P84698	18	NGWECC K GYPNV
P84698	25	GYPNV K ACSGCT
P02457	419	SGAPGP K GNSGEP
P02457	1097	QGDRGM K GHRGFS
P04258	95	PGPPGM K GPAGMP
P12108	181	NCPPGP K GPQGLQ
P23805	63	ECPHGE K GDPGSP
P23805	87	VGPIGP K GDNGFV
P23805	141	KGETGP K GGVGAP
Q9Y383	266	RSRSHS K NPKRSR
Q9Y383	269	SHSKNP K RSRSRE
P30943	41	VNMAMD K SAKAPV
P01170	120	CKNFYW K GFTSCX
P26368	15	RQLNEN K QERDKE
P26368	276	LNDDQV K ELLTSF
P02745	33	RAPDGK K GEAGRP
P04142	47	IRDGIV K AGPAIE
P20908	795	PGPRGV K GADGIR
P20908	804	DGIRGL K GTKGEK
P20908	807	RGLKGT K GEKGED
P20908	819	DGFPGF K GDMGIK
P20908	846	DGPEGP K GRGGPN
P20908	882	PGRQGP K GSIGFP
P20908	897	PGANGE K GGRGTP
P12111	2337	NGTTGP K GIRGRR
Q9SE35	174	GSYSGS K GSKRRI
O18495	27	EAGWLR K AAKSVG
O18495	30	WLRKAA K SVGKFY
O18495	34	AAKSVG K FYYKHK
O18495	38	VGKFYY K HKYYIK
O18495	40	KFYYKH K YYIKAA
O18495	44	KHKYYI K AAWQIG

Negative dataset

Uniprot ID	Position	Fragments
Q3Y5Z3	129	VPIRFTKIFYNQQ
Q3Y5Z3	144	YDGTGKFLCNIP
Q3Y5Z3	164	HITVYLKDVKVSL
Q3Y5Z3	167	VYLKDVKVS LYKN
Q3Y5Z3	172	VKVS LYKNDKALL
Q3Y5Z3	175	SLYKNDKALLFTH
Q3Y5Z3	187	HDQFQDKNVDQAS
Q3Y5Z3	202	VLLYLEKGDQVWL
Q15848	33	VLLPLPKGACTGW
Q15848	178	KVSLFKKDKAMLF
Q4ZJN1	49	DGRDGA KGDKGDA
Q4ZJN1	52	DGAKGD KGDA GEP
Q4ZJN1	66	HPGGPG KDGI RGE
Q4ZJN1	88	VEAKGI KGDP GSR
Q4ZJN1	99	SRGSPG KHGP KGS
Q4ZJN1	103	PGKHGP KGS IGPT
Q4ZJN1	124	TGPQGQ KGDKGEV
Q4ZJN1	148	TGPLGP KGLPGPM
Q4ZJN1	159	PMGP IGP GPRGE
Q4ZJN1	184	RGMRGW KGDRGEK
Q4ZJN1	190	KGDRGE KGKVGEA
Q4ZJN1	192	DRGEKG KVGEAPL
Q4ZJN1	201	EAPLVP KSAFTVG
Q4ZJN1	213	GLTVIS KFPPDA
Q4ZJN1	222	PPDAPI KFDKILY
Q4ZJN1	225	APIKFD KILYNEL
Q4ZJN1	240	YNVATG KFTCHVA
Q4ZJN1	268	VQVSLV KNGVKVL
Q4ZJN1	272	LVKNGV KVLHTKD
Q4ZJN1	277	VKVLHT KDSYMSS
Q4ZJN1	296	GIVQEL KLGDEVW
P30754	62	PAGIAG KSGDDGR
P30754	96	PGMPGP KGHRGFR
P30754	108	RGLSGS KGEQGKS
P30754	113	SKGEQG KSGNQGP
P30754	146	ERGRDG KSGLPGL
P30754	299	HAGEAG KRGS PGS
P30754	383	SDGRPG KDGRPGI
P30754	395	IRKDG KQGEQGP
P30754	425	ARGE PGKNGAPGE
P30754	443	EQGDAG KDGETGA
P30754	573	KGVGGS KGSRGDI
P30754	584	DIGPRG KAGERGK
P30754	612	SGLAAS KGERGDM
P30754	657	AGPAGT KGDIGFP
P30754	686	PRGLAG KRGLRGA
P30754	738	AGKQGV KGARGSP
P30754	749	SPGLVG KQGDRGS
P30754	765	PGRDGT KGERGED
P30754	810	MGGQGM KGDGGPP
P30754	927	AGIMGM KGEAGGK
P02457	43	GLTYND KDVWKPE
P02457	47	NDKDVW KPEPCQI
P02457	107	AGVEGP KGD TGPR
P02457	160	SYGYDE KSAGVAV
P02457	341	PGAAGA KGETGPQ

P02457	386	DGQPGAKGATGAP
P02457	541	SPGPDGKTGPPGP
P02457	646	PPGEAGKPGEQGV
P02457	923	PGPPGEKGS PGAD
P02457	973	PSGEPGKQG PSGA
P02457	1141	SAGAAGKDGLNGL
P02457	1197	PQPPQEK AHDGGR
P02457	1225	EVD TTLKSLSQQI
P02457	1242	SPEGTRKNPARTC
P02457	1252	RTCRDLKMCHGDW
P02457	1259	MCHGDWKSGEYWI
P02457	1277	CNLDAIKVYCNME
P02457	1299	QATIAQKNWYLSK
P02457	1305	KNWYLSKNPKEKK
P02457	1308	YLSKNPKEKKHVV
P02457	1310	SKNPKEKKHVWFG
P02457	1311	KNPKEKKHVWFG
P02457	1360	NVTYHCKNSVAYM
P02457	1374	HDTGNLKKALLLQ
P02457	1375	DTGNLKKALLLQG
P02457	1413	HTGAWGKTVIEYK
P02457	1419	KTVIEYKTTKTSR
P02457	1422	IEYKTTKTSRLPI
P02459	30	HGQDVQKAGSCVQ
P02459	73	IICEDMKDCLSP
P02459	113	GEPGDIKDIVGPK
P02459	119	KDIVGPKGPPGPQ
P02459	140	RGDRGDKGEGAP
P02459	143	RGDKGEGKAPGPR
P02459	190	AGGFDEKAGGAQM
P02459	884	TGVTGPKGARGAQ
P02459	1174	PVGPSGKDGANGI
P02459	1231	GLGQREKGPDP LQ
P02459	1333	NPASVPKKNWWSS
P02459	1334	PASVPKKNWWSSK
P02459	1340	KNWWSSKSKDKKH
P02459	1342	WWSSKSKDKKHIW
P02459	1435	FTYTVLKD GCTKH
P02459	1440	LKDGCTKHTGKWG
P02459	1444	CTKHTGKWGKTMI
P04258	8	YEAYDVKSGVAGG
P04258	239	DGSPGGKGEMGPA
P04258	436	NDGAPGKNGERGG
P04258	551	PGPEGGKGAAGPP
P12108	87	DGLTGAKGSRGPW
P12108	207	EPGQQGKQGPKGD
P12108	211	QGKQGPKG DVGVS
P12108	250	TGPAGYKGMVGTI
P12108	463	SGEPGPKGQQGIQ
P12108	529	IIDVVLKMMQEQL
P12108	543	EVAVSAKRAALGG
P12108	634	GHALAGKDGERGP
Q05707	7	MKIFQRKMRYWLL
Q05707	51	SIQISWKAPRGKF
Q05707	56	WKAPRGKFGGYKL
Q05707	61	GKFGGYKLLVTPT
Q05707	71	TPTSGGKTNQLNL
Q05707	83	LQNTATKAI IQGL
Q05707	104	QIIAYNKDKESKP

Q05707	106	IAYNKD K ESKPAQ
Q05707	109	NKD K ES K PAQGQF
Q05707	118	QGQFRI K DLEKRK
Q05707	122	RIKDLE K RKDPKP
Q05707	124	KDLEKR K DPKPRV
Q05707	127	EKRKDP K PRVKVV
Q05707	131	DPKPRV K VVDRGN
Q05707	148	SSPEEV K FVCQTP
Q05707	194	FDVGSE K TRIGLA
Q05707	218	LNAFST K DEVIEA
Q05707	231	VRNL P Y K GGNTLT
Q05707	250	IFENS F K PEAGSR
Q05707	261	SRTGV S K IGILIT
Q05707	270	ILITD G K SQDDII
Q05707	295	LFAIGV K NADVNE
Q05707	346	EQDREI K ASAHAI
Q05707	382	APGNVE K YRVVYY
Q05707	394	YPTRG G K PDEVVV
Q05707	410	VSSTVL K NLM S LT
Q05707	462	ENSMRV K WD A VP G
Q05707	488	GLAGDE K EMKIGE
Q05707	491	GDEKEM K IGETHT
Q05707	591	ITTFPL K GLT P LT
Q05707	655	ADEGLH K LMWIPV
Q05707	665	IPVYGG K TEEVVL
Q05707	672	TEEVVL K EEQDSH
Q05707	754	TSSLRV K WDISDS
Q05707	796	QNNLLL K P L L P DT
Q05707	805	LPDTEY K VT V T P I
Q05707	825	SVSAP G K TL P SS G
Q05707	857	PPSSPV K G Y RIVY
Q05707	864	GYRIVY K PVSV P G
Q05707	898	GMDYNV K IFASQA
Q05707	916	ALTGMV K TL F LG V
Q05707	928	VTNLQA K HVEMTS
Q05707	960	SLQDR Q K QESTVG
Q05707	984	QPDSEY K ISVYTK
Q05707	990	KISVYT K LQEIEG
Q05707	1004	SVSIME K TQSLPT
Q05707	1023	PTIPPA K EVCKAA
Q05707	1027	PAKEVC K AAKADL
Q05707	1030	EVCKAA K ADLVFM
Q05707	1050	GDENFN K IISFLY
Q05707	1064	TVGALN K IGTDGT
Q05707	1086	DPRT E F K LNAYKT
Q05707	1091	FKLNAY K TKETLL
Q05707	1093	LNAYKT K ETLLDA
Q05707	1101	TLLDAI K HISYKG
Q05707	1106	IKHISY K GGNTKT
Q05707	1111	YKGGNT K TGKA I K
Q05707	1114	GNTKT G KA I KYVR
Q05707	1117	KTGKA I K YVRDTL
Q05707	1136	TRRGIP K VIVVIT
Q05707	1152	SQDDVN K ISREM Q
Q05707	1183	LVSIG S K PSARHV
Q05707	1199	DDF D AF K KIEDEL
Q05707	1200	DFDA F K KIEDELI
Q05707	1222	TCPVVH K DGIDLA
Q05707	1231	IDLAG F K MMEMFG

Q05707	1241	MFGLVEKDFSSVE
Q05707	1266	PCYQLHKDALVSQ
Q05707	1311	LWEILNKNSDPLV
Q05707	1326	ILDNGGKTLTYFN
Q05707	1351	EGPEIRKIFYGSF
Q05707	1359	FYGSFHKLHIVVS
Q05707	1370	VSETLVKVVIDCK
Q05707	1376	KVVIDCKQVGEKA
Q05707	1381	CKQVGEKAMNASA
Q05707	1399	GVEVLGKMVRSRG
Q05707	1431	SWANTDKCCELPG
Q05707	1558	QRGLPGKDGSSGP
P23805	29	MTTFSQKILANAC
P23805	122	REGPSGKQGSMP
P23805	222	AEVNALKQRTIL
P23805	243	NAFSQYKKAVLFP
P23805	244	AFSQYKKAVLFPD
P23805	257	GQAVGEKIFKTAG
P23805	260	VGEKIFKTAGAVK
P23805	266	KTAGAVKSYSDAE
P23805	279	QLCREAKGQLASP
P23805	303	MVRAQEKNAAYLSM
P23805	355	EIFPDGKWNDVPC
P23805	363	NDVPCSKQLLVIC
Q9Y383	29	TTRQRIKFSDDR
Q9Y383	37	SDDRVCKSHLLNC
Q9Y383	61	DLGECLKVHDLAL
Q9Y383	76	DYEIASKEQDFFF
Q9Y383	105	RRTEVAKKRLAET
Q9Y383	106	RTEVAKKRLAETQ
Q9Y383	122	SAEVAAKAERVHE
Q9Y383	135	LNEEIGKLLAKVE
Q9Y383	139	IGKLLAKVEQLGA
Q9Y383	154	NVEESQKVMDEVE
Q9Y383	161	VMDEVEKARAKKR
Q9Y383	165	VEKARAKKREAE
Q9Y383	166	EKARAKKREAEV
Q9Y383	186	SSFQQQKLRVCEV
Q9Y383	213	ADHFGGKLHLGFI
Q9Y383	224	FIEIREKLEELKR
Q9Y383	229	EKLEELKRVVAEK
Q9Y383	235	KRVVAEKQEKRNQ
Q9Y383	238	VAEKQEKRNQERL
Q9Y383	245	RNQERLKRREERE
Q9Y383	257	EREEREKLRRSRS
Q9Y383	289	SMSRERKRRTRSK
Q9Y383	295	KRRTRSKSREKRH
Q9Y383	299	RSKSREKRHRHS
Q9Y383	332	RSRERSKRRSSKE
Q9Y383	337	SKRRSSKERFRDQ
Q9Y383	365	PRDRDRKDKKRSY
Q9Y383	367	DRDRDKKRSYES
Q9Y383	368	RDRDKKRSYESA
P30943	44	AMDKSAKAPVITI
P30943	61	GCSRAPKEYTGAK
P30943	67	KEYTGAKAGKDD
P30943	71	GAKAGGKDDEMMV
P30943	78	DDEMMVKAQSVKI
P30943	83	VKAQSVKIEVSTG

Q00433	98	VKVA STK VTVSES
Q00433	108	SESDAA KKL QEFI
Q00433	109	ESDAA KKL QEFIT
Q00433	118	EFITFE K GIDGPF
P01170	61	LSQEW S KRAVEEL
P01170	111	LPPRER K AGCKNF
P01170	115	ERKAGC K NFYWKG
P26368	20	NKQERD K ENRHRK
P26368	26	KENRHR K RSHSRS
P26368	63	DRRRRS K PLTRGA
P26368	70	PLTRGA K EEHGGL
P26368	84	RSPRHE K KKKVRK
P26368	85	SPRHE K KKKVRKY
P26368	86	PRHE K KKKVRKYW
P26368	87	RHE K KKKVRKYWD
P26368	90	KKK K VRKYWDVPP
P26368	108	ITPMQY K AMQAAG
P26368	195	VQINQD K NFAFLE
P26368	225	FQGQSL K IRRPD
P26368	260	VPDSA HKL FIGGL
P26368	286	TSFGPL K AFNLVK
P26368	292	KAFNLV K DSATGL
P26368	300	SATGL S KGYAFCE
P26368	328	GMQLGD K KLVLQR
P26368	329	MQLGD K KLVLQRA
P26368	340	RASVGAK N ATLVS
P26368	408	VRDECS K YGLVKS
P26368	413	SKYGLV K SIEIPR
P26368	431	EVPGCG K IFVEFT
P26368	444	SVFDCQ K AMQGLT
P26368	453	QGLTGR K FANRVV
P26368	462	NRVVVT K YCDPDS
P02745	32	CRAPDG K KGEAGR
P02745	110	GSPGNI K DQPRPA
P02745	195	FCDTTN K GLFQVV
P02745	219	DQVWVE K DPKKGH
P02745	222	WVEKDP K KGHIYQ
P02745	223	VEKDP K KGHIYQG
P02747	14	LPHLGL K LLLLLL
P02747	115	GEEGRY K QKFQSV
P02747	117	EGRY K QKFQSVFT
P02747	161	TGKFTC K VPGLYY
P02747	188	LYRSGV K VVTFCG
P02747	198	FCGHTS K TNQVNS
P04142	5	XXMNF A KILSFVF
P04142	29	APEPRW K IFKKIE
P04142	32	PRWKIF K KIEKMG
P04142	33	RWKIF K KIEKMGR
P04142	36	IFKKIE K MGRNIR
P04142	59	EVLGSA K AIGKXX
P02453	79	VICDEL K DCPNAK
P02453	85	KDCPN A KVPTDEC
P02452	67	CVCDNG K VLCDDV
P02452	430	GGPPGP K GNSGEP
P02454	45	PNGETW K PDVCLI
P02454	67	CDGVLC K EDLDCP
P02454	77	DCPNP Q KREGECC
P02454	896	VGKEGG K GPRGET
P02454	1299	QPSVP Q KNWYISP

P02461	7	MMSFVQKGSWLLL
P02461	503	NGIPGEKGPAGER
P02461	1212	AGIGGEKAGGFAP
P02461	1228	DEPMDFKINTDEI
P02461	1239	EIMTSLKSVNGQI
P02461	1286	DPNQGCKLDAIKV
P02461	1313	PLNVPRKHWWTDS
P02461	1387	QASGNVKKALKLM
P02461	1391	NVKKALKLMGSNE
P02461	1401	SNEGEFKAEGNSK
P02461	1407	KAEGNSKFTYTVL
P20908	8	DVHTRWKARSALR
P20908	44	QPADLLKVLDFHN
P20908	57	LPDGITKTTGFCA
P20908	69	ATRRSSKGPDVAY
P20908	79	VAYRVTKDAQLSA
P20908	88	QLSAPTQQLYPAS
P20908	107	SILTTVKAKKGSQ
P20908	109	LTTVKAKKGSQAF
P20908	110	TTVKAKKGSQAFL
P20908	146	YEDHTGKPGPEDY
P20908	164	INLSDGKWHRIAL
P20908	174	IALSVHKKNVTLI
P20908	175	ALSVHKKNVTLIL
P20908	184	TLILDCKKKTTKF
P20908	185	LILDCKKKTTKFL
P20908	186	ILDCKKKTTKFLD
P20908	189	CKKKTTKFLDRSD
P20908	288	DPEDLGKEPTPSK
P20908	294	KEPTPSKKPVEAA
P20908	295	EPTPSKKPVEAAK
P20908	301	KPVEAAKETTEVP
P20908	329	TSEGAGKEEDVGI
P20908	455	RGEKGQKGEPAL
P20908	535	GGDAGSKGPMVSA
P20908	582	PGSGGLKGEPGDV
P20908	1074	VGALGLKGNEGPP
P20908	1167	IGEPGQKGSKGDK
P20908	1410	QGEKGAKGEAGLE
P20908	1473	PGLPGLKGDSGPK
P20908	1619	GSLNSLKLEIEQM
P20908	1626	LEIEQMKRPLGTQ
P20908	1640	NPARTCKDLQLCH
P20908	1668	CSRDSFKVYCNFT
P20908	1685	TCVFDPKKSEGAR
P20908	1686	CVFPDKKSEGARI
P20908	1697	RITSWPKENPGSW
P20908	1708	SWFSEFKRGKLLS
P20908	1711	SEFKRGKLLSYVD
P20908	1762	ATGSYDKALRFLG
P20908	1792	VDGCATKKGQYQKT
P20908	1793	DGCATKKGQYQKT
P20908	1797	TKKGQYQKTVLEID
P20908	1806	LEIDTPKVEQVPI
P20908	1826	FGEASQKFGFEVG
P12111	33	QQQADVKNAAAD
P12111	67	FLYDVVKSLAVGE
P12111	99	LNTYRTKQEVLSH
P12111	120	GTNQTGKGLEYIM

P12111	132	MQSHLT	KAAGSRA
P12111	155	LTDGHS	KDGLALP
P12111	166	LPSAEL	KSADVNV
P12111	186	ADEGAL	KEIASEP
P12111	233	GDTETL	KDITAQD
P12111	271	LVNLLE	KLPIGTQ
P12111	302	LDTYST	KAQVLGA
P12111	310	QVLGAV	KALGFAG
P12111	370	YGVVAL	KQASVFS
P12111	414	FGDLQE	KLLPYIV
P12111	430	QRHIVL	KPPTIVT
P12111	443	QVIEVN	KRDIVFL
P12111	470	IRDFIA	KVIQRLE
P12111	505	FNTHPT	KREVITA
P12111	514	VITAVR	KMKPLDG
P12111	516	TAVRKM	KPLDGSA
P12111	551	AAEGIP	KLLVLIT
P12111	573	QPAQEL	KRSSIMA
P12111	637	PEVHSN	KRDIIFL
P12111	652	GSANVG	KTNFPYV
P12111	783	GASQAN	KAELEQI
P12111	835	LAQPES	KRDILFL
P12111	861	VRDFLY	KIIDELN
P12111	869	IDELNV	KPEGTRI
P12111	885	QYSDDV	KVESRFD
P12111	896	FDEHQS	KPEILNL
P12111	904	EILNLV	KRMKIKT
P12111	907	NLVKRM	KIKTGKA
P12111	909	VKRMKI	KTGKALN
P12111	912	MKIKTG	KALNLGY
P12111	930	QRYIFV	KSAGSRI
P12111	964	GPASNL	KQSGVVP
P12111	976	PFIFQA	KNADPAE
P12111	1001	AAESLP	KIGDLHP
P12111	1014	QIVNLL	KSVHNGA
P12111	1028	APVSGE	KDVVFL
P12111	1048	SGFPLL	KEFVQRV
P12111	1088	LNSYMN	KQDVVNA
P12111	1156	NPSVVV	KRGGAVP
P12111	1231	SPGVGG	KRDVVFL
P12111	1281	QFSDDP	KVEFLLN
P12111	1305	VQRLRP	KGGRQIN
P12111	1325	VSRNIF	KRPLGSR
P12111	1347	VLISSG	KSDDEVD
P12111	1380	DQEELV	KISLSPE
P12111	1403	LPSLEQ	KLLTPIT
P12111	1418	TSEQIQ	KLLASTR
P12111	1472	LNIGPS	KVRVGVV
P12111	1491	FPEFYL	KTYRSQA
P12111	1518	SPLNTG	KALEFVA
P12111	1551	VLVLGG	KSQDDVS
P12111	1635	PPSRPE	KKKADIV
P12111	1636	PSRPEK	KKADIVF
P12111	1637	SRPEKK	KADIVFL
P12111	1694	TDEFFL	KDFSTKR
P12111	1699	LKDFST	KRQIIDA
P12111	1708	IIDAIN	KVVYKGG
P12111	1712	INKVVY	KGGRHAN
P12111	1720	GRHANT	KVGLEHL

P12111	1754	FVITGGKSVEDAQ
P12111	1772	LTQRGVQVFAVG
P12111	1788	DSEEVGKIASNSA
P12111	1833	GVTDAAKACNLDV
P12111	1856	NVFVAQKGFESKV
P12111	1861	QKGFESKVDAILN
P12111	1913	QPEMLEKFRNMRS
P12111	1931	LTEDTLKVYLNKF
P12111	1936	LKVYLNKFRQSSP
P12111	1946	SSPDSVKVVIHFT
P12111	2027	LDNIAEKACCGVP
P12111	2035	CCGVPCKCSGQRG
P12111	2052	IGSIGPKGIPGED
P12111	2121	DGEDGDKGLPGSS
P12111	2170	SQERGPKGETGDL
P12111	2196	GPGETGKNGGFGR
P12111	2355	PGIVGQKGDPGYP
P12111	2384	ALIQSIKDKCPCC
P12111	2386	IQSIKDKCPCCYG
P12111	2464	IRFADSKRKSVLL
P12111	2466	FADSKRKSVLLDK
P12111	2472	KSVLLDKIKNLQV
P12111	2474	VLLDKIKNLQVAL
P12111	2483	QVALTSKQQSLET
P12111	2500	VARNTFKRVNRNGF
P12111	2510	NGFLMRKVAVFFS
P12111	2532	LREAVLKLSDAGI
P12111	2639	FQFNEMKKYIAYL
P12111	2640	QFNEMKKYIAYLV
P12111	2656	DMSPDPKASQHFA
P12111	2684	ASMPPVKVEFSLT
P12111	2695	LTDYGSKEKLVDF
P12111	2697	DYGSKEKLVDFLS
P12111	2737	PNPRDLKIVVLML
P12111	2763	RVILQAKCKGYFF
P12111	2765	ILQAKCKGYFFVV
P12111	2777	VLGIGRKVNIKEV
P12111	2781	GRKVNIKEVYTFA
P12111	2796	PNDVFFKLVDKST
P12111	2800	FFKLVDKSTELNE
P12111	2834	LSPDIRKQCDWFQ
P12111	2846	QGDQPTKNLVKFG
P12111	2850	PTKNLVKFGHKQV
P12111	2854	LVKFGHKQVNVPN
P12111	2875	NPVTTTKPVTITTK
P12111	2895	PVTTTTKPVITIIN
P12111	2906	INQPSVKPAAAKP
P12111	2911	VKPAAAKPAPAKP
P12111	2921	AKPVAAKPVATKM
P12111	2926	AKPVATKMATVRP
P12111	2937	RPPVAVKPATAAK
P12111	2943	KPATAAKPVAAKP
P12111	2986	TKPATTKPMVKMS
P12111	2990	TTKPMVKMSREVQ
P12111	3006	ITENSAKLHWERA
P12111	3035	DQSLVLKQNLTVT
P12111	3077	HGSFSTKKSQPPP
P12111	3078	GSFSTKKSQPPPP
P12111	3113	TETDICKLPKDEG

P12111	3116	DICKLPKDEGTCR
P12111	3127	CRDFILKWYYDPN
P12111	3135	YYDPNTKSCARFW
P12111	3151	CGGNENKFGSQKE
P12111	3156	NKFGSQKECEKVC
P12111	3160	SQKECEKVCAPVL
P12111	3168	CAPVLAKPGVISV
Q02388	164	TAAQRLKGQGVKL
Q02388	184	ADPEELKRVASQP
Q02388	263	GPVTGYKVQYTPL
Q02388	459	GLEPPQKVVLPSD
Q02388	727	SAHGPEKSQLVSG
Q02388	1253	CPVYCPKGQKGEP
Q02388	1256	YCPKGQKGEPGEM
Q02388	1399	LPGTAMKGDKGDR
Q02388	1447	PGKKGEKGDSGDG
Q02388	1533	TGRQGEKGEPGRP
Q02388	1718	GPGAREKGEPGDR
Q02388	1786	RSGLDGKPGAAGP
Q02388	1843	KPGLNGKNGEPGD
Q02388	1952	LETAGIKASALRE
Q02388	1981	ERRRGPKGDSGEQ
Q02388	1992	EQGPPGKEGPIGF
Q02388	2166	MAGPEGKPGLQGP
Q02388	2519	AGLKGDKGDSAVI
Q02388	2550	RGLDGDGKPRGDN
Q02388	2780	AGPRGEKGEEALT
P19999	28	TCEETLKTCSVIA
P19999	94	SRAIEVKLANMEA
P19999	106	AEINTLKSLELT
P19999	108	INTLKSLELTNK
P19999	114	KLELTNKLHAFSM
P19999	122	HAFSMGKKS GKFF
P19999	123	AFSMGKKS GKFF
P19999	126	MGKKS GKFFVTN
P19999	127	GKKS GKFFVTNH
P19999	140	ERMPFSKV KALCS
P19999	142	MPFSKV KALCSEL
P19999	162	RNAEENKAIQEVA
P19999	169	AIQEVAKTS AFLG
P19999	199	LTYSNWKDEPN
P19999	200	TYSNWKDEPN
P35248	23	DLGAEMKTL SQR
P35248	236	MEALNGKLQRLEA
P35248	277	EPFEDAKEMCRQA
P35248	308	LVTAHSKAAFLSM
P35248	322	DVGTEGKFTYPTG
P35248	362	NGQWNDKACGEQR
Q9SE35	36	LSTEDSKSAQLIS
Q9SE35	110	LRILSSKKS GSY
Q9SE35	111	RILSSKKS GSY
Q9SE35	121	YYSYGTKKS GSY
Q9SE35	122	YSYGTKKS GSY
Q9SE35	132	YSGYSTKKSASRR
Q9SE35	133	SGYSTKKSASRRI
Q9SE35	177	SGSKGSKRRILSS
O18495	51	AAWQIGKHALGDM
O18495	67	EFQDFMKEVEQAR
