

Online Supporting Information S5. Seven negative subsets for studying Arg-methylation. Each subset contains 185 negative samples randomly taken from the 1,296 negative samples in [Online Supporting Information S3](#) except for the 6th subset, which contains 186 samples. None of the samples in one subset occurs in any other subset.

(1) Subset 1 (185 samples)

Sample #	Protein ID	Site	Sequences
1	Q9NR22	229	VAIEDRQYKDF
2	Q9UBL3	324	LFSAQLRPPHG
3	Q92934	72	VEIRSRHSSYP
4	Q7Q2B7	222	IDLPERVDVII
5	P11911	9	GLEALRALPLL
6	P11911	118	CGTYLRVRNPV
7	Q09472	86	LSELLRSGSSP
8	Q09472	354	HKCQRREQANG
9	Q09472	408	WKNCTRHDPCV
10	Q09472	580	ITQDLRNHLVH
11	Q09472	604	ALKDRRMENLV
12	Q09472	649	LEEKRRTRLQK
13	Q09472	1015	TETEERSTELK
14	Q09472	1076	ESLPFRQPVDP
15	Q09472	1324	GEVTVRVVHAS
16	Q09472	1410	RPKCLRTAVYH
17	Q09472	1462	IPKPKRLQEWY
18	P03372	28	LEPLNRPQLKI
19	P03372	158	NSDNRROGGRE
20	P03372	234	IDKNRRKSCQA
21	P03372	256	MKGGIRKDRRG
22	P03372	503	QQQHQLAQLL
23	P03372	515	ILSHIRHMSNK
24	Q01844	614	GFGGRRGGPG
25	Q01844	632	EQMGGRRGGRG
26	P22509	76	GGGRGRGGGRG
27	P22509	224	VIEDARHPHKY
28	P35637	213	YGQQDRGGGRGR
29	P35637	269	KFGGPRDQGSR
30	P35637	274	RDQGSRHDSEQ
31	P35637	328	NLYTDRETGKL
32	P03347	299	RDYVDRFYKTL
33	P03347	305	FYKTLRAEQAS
34	P03347	429	KDCTERQANFL
35	P03347	452	NFLQSRPEPTA
36	P03347	502	PLTSLRSLFGN
37	P68431	54	LREIRYQKST
38	P68431	73	FQRLVREIAQD
39	P68431	117	AIHAKRVTIMP
40	P68431	130	IQLARRIRGER
41	P62805	24	HRKVLRDNIQG

42	P62805	68	LENVIRDAVY
43	P62805	96	LKRQGRRTLYGF
44	P17096	84	PGRKPRGRPKK
45	P31942	19	SDGTVRLRGLP
46	P31942	21	GTVRLRGLPFG
47	P31942	105	YDPPRRLLGQR
48	P31942	222	PLNPIRVHIDI
49	O60506	303	TAAQARRRLMS
50	O60506	305	AQARRRLMSGK
51	O60506	446	PPTRGRGRGGR
52	Q07666	13	AARMSRSSGRS
53	Q07666	43	PPLPHRSRGGG
54	Q07666	291	VPVRGRGAAPP
55	Q07666	331	GATVTRGVPPP
56	P08199	394	KVRAARTLLAK
57	P08199	604	RIVTDRETGSS
58	O75925	63	KELYRRRFPQK
59	O75925	257	ITSLVRLSTTV
60	O75925	579	LLHSSRFFPYT
61	Q92804	101	GSQGGRAPSYD
62	Q92804	164	HTQDDRRDVS
63	Q92804	165	TQDDRRDVSRY
64	Q92804	185	SQGGGRGRGGY
65	Q92804	321	SFATRPEFMR
66	Q92804	372	NMNFARRNSCN
67	Q92804	415	GRGGDRGGYGG
68	Q92804	431	GYGGDRSSGGG
69	Q92804	459	GYGGDRGGGYG
70	Q92804	580	GKMGGRNDYRN
71	Q92804	584	GRNDYRNDQRN
72	P09867	82	HKVDGRVVEPK
73	P09867	88	VVEPKRAVSRE
74	O00267	17	EEDSERSSDGE
75	O00267	30	EVDEERSAAG
76	O00267	130	LQNLWRDQREE
77	O00267	368	GNRYSRKGFLF
78	O00267	571	KVVTVRHQAVT
79	O00267	641	FVCKTRHLVLA
80	O00267	652	GGSKPRDVTNF
81	O00267	783	YGSGSRTPMYG
82	P69697	7	EPVDPRLPEWK
83	P69697	56	KRRQRRRPPQG
84	Q15554	28	SGRARRGRHEP
85	Q15554	30	RARRGRHEPGL
86	Q15554	47	GAGEARLEEAV
87	Q15554	69	ALRAFRRGSRYG
88	Q15554	72	AFRGSRYGDFR
89	Q15554	192	TTQKLRNDLLN
90	Q15554	356	QPKNKRMTISR
91	P15056	188	KALMMRGLIPE
92	P15056	260	LFQGFRCQTCG
93	P15056	347	IPQPFRLPAED

94	P15056	360	NQFGQRDRSSS
95	P15056	426	GPQREBKSSSS
96	Q5TAQ9	161	PALRERELGSS
97	Q96EY1	187	PEELFRKIFGE
98	Q96EY1	272	GPFVMRSTCRR
99	Q96EY1	276	MRSTCRRCGGR
100	Q96EY1	281	RRCGGRGSIII
101	Q96EY1	303	AKQKKRVMIPV
102	Q96EY1	364	LGGTARAQGLY
103	P06746	89	KLEKIRQDDTS
104	P06746	126	TLEDLRKNEDK
105	Q92997	215	ASRLMRRHKRR
106	Q92997	220	RRHKRRRRKQK
107	Q54G21	660	KMMESRNEAIT
108	Q54G21	708	INQLKRDYNTL
109	Q54G21	827	TLETERSALSS
110	Q54G21	841	LVENTRLETES
111	Q9R1E0	19	FEPLPRQRST
112	Q9R1E0	313	STFRPRTSSNA
113	Q9R1E0	551	TSAMNRLTPVK
114	Q9R1E0	586	CNGYGRMGVLH
115	Q13283	106	NNQALRRFMQT
116	Q13283	107	NQALRRFMQTF
117	Q13283	283	PASQPRPESKP
118	P31943	230	GAGFERMRRGA
119	P04264	30	INYQRRTTSSS
120	P04264	65	GGFGSRSLVNL
121	P04264	199	FIDKVRFLQQ
122	P04264	343	SMDNNRSLDL
123	P08729	48	GASRPVAVRS
124	P08729	305	DLRNRNEISE
125	P15720	10	KRSSFRHGSKM
126	P15720	48	FFGGDRHVPRR
127	P15720	96	NIVSPRTPPPM
128	P15720	162	GGSGSRPGSRS
129	P25274	128	FGYGGRAADYK
130	Q8IZS5	61	EKKVIRHDMPD
131	Q8IZS5	67	HDMPDRTLAAH
132	Q8IZS5	77	HQQKFRLPASA
133	P11940	153	QEAAERAIEKM
134	P11940	166	MLLNDRKVFVG
135	P11940	186	AELGARAKEFT
136	P11940	287	QMKQDRITRYQ
137	P11940	331	MMEGGRSKGFG
138	Q9JMB7	6	MTGRARARARG
139	Q9JMB7	36	GYIPPRPQQSP
140	Q9JMB7	51	LVGRGRQGMV
141	Q9JMB7	84	ERGRRRDFHD
142	Q9JMB7	147	EARRLSALLF
143	Q8CDG1	177	FGLTARDPPRL
144	Q86WR7	140	AGKEHRKQDAE
145	Q86WR7	204	QKISERMAGNE

146	Q86WR7	221	PFREG R PGEWR
147	P09651	265	YSGGS R GYGSG
148	P22626	185	HNAE V RKALSR
149	P51991	239	GGN F RGGNFG
150	Q99729	30	EAS R RGWTGA
151	P14678	228	PPPG M RGPPPP
152	Q14524	179	VKIL A RGFCLH
153	Q14524	340	CPE G YRCLKAG
154	Q14524	367	FLAL F R L MTQD
155	P23246	399	FGPI E RAVVIV
156	P68363	105	ANN Y ARGHYTI
157	P68363	264	LVP Y PR I HFP L
158	P68363	320	CCL L YRGDVVP
159	P07437	77	TMDS V RS G PFG
160	P07437	162	EE Y PD R IMNTF
161	P07437	282	GS Q Q Y RAL T V P
162	P07437	359	CDIP P RGLKMA
163	Q86V81	20	LN R S Q RGGRGG
164	Q86V81	174	VPLD G RP M NIQ
165	Q86V81	197	AQ S V N RGGMTR
166	Q86V81	202	RGG M TRNRGAG
167	O08583	56	GGG P M R NRPAI
168	O08583	203	GM T R N RGSGGF
169	P39748	70	MGM F Y R TIRMM
170	P31269	207	HAR S T R KKRCP
171	P31269	233	N M Y L T R DRRYE
172	P31269	241	RYE V ARLLNLT
173	P04049	100	CCA V F R LLHEH
174	P04049	143	TH N F A RK T FLK
175	P04049	231	VSS Q H R Y S TPH
176	P04049	282	IED A I R SHSES
177	P04049	318	PA Q RE R APVSG
178	P04049	333	NK I R P R G Q R DS
179	P46783	8	MP K KN R IAIYE
180	P46783	80	GI Q Y L R D Y L H L
181	P46783	96	PAT L R R SRPET
182	P46783	105	ET G R P R P KGLE
183	P46783	116	GER P AR L TRGE
184	P46783	128	DR D T Y RRSAVP
185	O76922	87	SV R GR L LITDL

(2) Subset 2 (185 samples)

Sample #	Protein ID	Site	Sequences
1	Q9NR22	13	RCL L L R RKMAE
2	Q9NR22	14	CL L L R RKMAEN
3	Q9UBL3	211	EC M T T R Q RPGK
4	Q9UBL3	231	TMS K ER D VFLV
5	Q9UBL3	315	TT K K A RS D PLF
6	Q9UBL3	374	PG D LY R ACLYE
7	Q9UBL3	411	G Y SM V RASHGV

8	Q9UBL3	592	KDLTYRPMSDM
9	Q92934	116	GRELR RMSDEF
10	Q92934	131	KKGLP RPKSAG
11	Q7Q2B7	84	ELVLF R FASRG
12	Q7Q2B7	88	FRFAS R GDCQQ
13	Q7Q2B7	95	DCQQF R SLLHK
14	Q7Q2B7	341	EKDLH R IQIDV
15	Q7Q2B7	447	KNPYF R YTGAP
16	P11911	47	LGEEA R LTCEN
17	P11911	96	VNKNH R GLYWC
18	P11911	137	EGTKN R IITAE
19	Q09472	1104	LSTIK R KLDTG
20	Q09472	1133	AWLYN R KTSRV
21	Q09472	1197	YSYQN R YHFCE
22	Q09472	1392	PPNQ R VYISY
23	Q09472	1494	QATED R L TSAK
24	P03372	37	KIPLER R PLGEV
25	P03372	142	SGYTV R EAGPP
26	P03372	241	SCQAC R LRKCY
27	P03372	436	TSSRF R MMNLQ
28	Q01844	392	VKMNK R TGQPM
29	P22509	56	GGGGF R GRGGG
30	P22509	84	GRGGK R GNQSG
31	P35637	521	EHRQD R RERPY
32	P03347	275	LNKIV R MYSPT
33	P03347	403	EGHTA R NCRAP
34	P68431	129	DIQLA R RIRGE
35	P62805	40	IRRLA R RGGVK
36	P62805	79	TEHAK R KTVTA
37	P31942	93	IFRSS R SEIKG
38	P31942	129	YYGAG R GSMYD
39	P31942	261	NNMQH R YIELF
40	P31942	301	YGSVG R MGMGN
41	P31942	323	GGYGR R GGGGS
42	O60506	103	VMKTY R QREKQ
43	O60506	229	NNHEI R SGKHI
44	O60506	451	RGRGG R GGYGY
45	O60506	504	GGRGA R GAAPS
46	O60506	557	GARGG R GGNVG
47	O60506	565	NVGKK R KADGY
48	Q07666	10	DDPAARMSRSS
49	Q07666	204	GKGSMDKAKE
50	Q07666	436	VKGAY R EHPYG
51	P08199	343	AVVDV R TGTNR
52	P08199	421	DALEI R LVSQD
53	P08199	474	GQRQER T GKNS
54	Q28165	17	GAAGG R GSGPG
55	Q28165	25	GPGRR R HLVPG
56	Q28165	78	EPPRP R APPGA
57	Q28165	200	CGSVN R VTILC
58	Q28165	248	KVIPK R TNRPG
59	O75925	276	TAEIG R NY SMA

60	O75925	528	LIQDYRHPFHM
61	Q96PU8	106	RILGPRGLTAK
62	Q92804	221	RDYGPRTDADS
63	Q92804	320	VSFATRRPEFM
64	Q92804	341	GRGGYRGRGGF
65	Q92804	349	GGFQGRGGDPK
66	Q92804	388	RPEDSRPSGGD
67	Q92804	406	GERGYRGRGGR
68	Q92804	422	GYGGDRSGGGY
69	P09867	31	TDESLRSHFEQ
70	P09867	196	SSQRGRSGSGN
71	P09867	300	YFAKPRNQGGY
72	O00267	31	VDEERRSAAGS
73	O00267	65	EEDDDRPPKKP
74	O00267	319	DRIKARMSLKD
75	O00267	346	DAEKIRSLGGD
76	O00267	516	LKVLPRDLQLC
77	O00267	582	RKKDNRFVAL
78	O00267	711	IGQTVRISQGP
79	O00267	798	LQDGSRTPHYG
80	O00267	831	PNTPSRAEEY
81	O00267	1049	ILGEDREATGV
82	O00267	1066	EDGIVRMDLDE
83	P69697	78	PTSQSRGDPTG
84	Q15554	13	SDGSGRAAGR
85	Q15554	77	RYGDFRQIRDI
86	Q15554	102	VSRLLRVMQCL
87	P15056	239	THNFVRKTFFT
88	Q5TAQ9	11	SSTDGRTDLAN
89	Q5TAQ9	151	ALPRPRWQALP
90	Q5TAQ9	177	EACGARVVFVQR
91	Q5TAQ9	182	RVFVQRFRLQH
92	Q96EY1	102	ILGVPRNASQK
93	Q96EY1	332	IFITFRVQKSP
94	P06746	102	INFLTRVSGIG
95	P06746	137	LNHHQRIGLKY
96	P06746	258	RRIDIRLIPKD
97	Q92997	145	SAQREPRRRD
98	Q92997	167	AKGERRREPGG
99	Q60487	105	MKEDGRLASK
100	Q54G21	180	VANGDRVSIKY
101	Q54G21	230	GVIGMRKSAKR
102	Q54G21	793	EESLKRLKLDL
103	Q54G21	1425	IDNSPRQTTTE
104	Q9R1E0	249	GKSPRRRAASM
105	Q9R1E0	487	AQPNSRVLGQN
106	Q9R1E0	534	ASVNGRTLPHV
107	Q13283	159	EEPEERQQTPE
108	Q13283	296	QIPPQRPQRDQ
109	P31943	16	FVVKVRGLPWS
110	P31943	52	YTREGRPSGEA
111	P31943	192	SRAEVRTHYDP

112	P31943	217	RPGAGR GY NSI
113	P04264	37	TSSSTR RR SGGG
114	P04264	386	QITAGR HG D ^S V
115	P08729	30	RLSSAR PG GLG
116	P08729	46	GLGASR PR VAV
117	P08729	52	PRVAVR S AYGG
118	P08729	64	VGAGIR EV TIN
119	P08729	187	DEINHR T AAEN
120	P08729	276	MAKCSR AE EAE
121	P08729	313	ISEMNR AI QRL
122	P08729	317	NRAIQ RL QAEI
123	Q36736	14	SAKLD RL DQEF
124	P02687	41	LDSLGR FF FGSD
125	P15720	78	RSQHGR PG DDN
126	P15720	166	SRPGSR SG SPV
127	P25274	9	KRPSQ RH GSKY
128	P25274	49	FFSSDR G APKR
129	Q8IZS5	31	MVKED RE ATEG
130	Q8IZS5	142	QDESG RE ETLN
131	Q8IZS5	177	LEEFQ RS QKTI
132	P11940	89	IMWSQ RD PSLR
133	P11940	206	DMDDE RL KDLF
134	P11940	278	QTELK RK FEQM
135	P11940	311	DDERL RK EFSP
136	Q9JMB7	83	AERGG RR RDFH
137	Q9JMB7	145	LMEAR RL RSAL
138	Q9JMB7	160	EDLIG R CHAFD
139	Q8CDG1	9	VRPLF R GP ^T PV
140	Q8CDG1	88	RPPSK RE VPPL
141	Q8CDG1	111	MVRKD RE EPRS
142	Q8CDG1	156	LLPLG RA ASSI
143	Q86WR7	6	MPVTH RK S ^D AS
144	Q86WR7	22	TSPSC RL RAFS
145	Q86WR7	173	PDP PR ELRAP
146	Q86WR7	176	PRREL R APSPP
147	Q86WR7	265	TNGA A REPRRT
148	P22626	15	LERKK RE KEQF
149	P22626	60	DPASK RS R ^G FG
150	P22626	228	PGSNF R GGSDG
151	P22626	238	GYGSG R GFGDG
152	P51991	9	KPPPG R PQ ^P DS
153	P51991	113	KRAVS RE DSVK
154	P51991	143	EEYNL R DYFEK
155	P51991	226	GNFMG R GGNFG
156	Q99729	248	GGRGN R NRGNR
157	Q99729	250	RGNRN R GNRGS
158	P14678	221	PPPGM R PPPPG
159	P23246	7	SRDRF RS RGGG
160	P23246	18	GGGFH R RGGGG
161	P23246	25	GGGG R GGGLHD
162	P68363	215	YDICR R NLDIE
163	P68363	243	ITASL R FDGAL

164	P68363	308	VKCDP R HGKYM
165	P07437	213	YDICF R TLKLT
166	P07437	241	VTTCL R FPGQL
167	P07437	276	APLTS R GSQQY
168	P07437	306	AACDP R HGRYL
169	P07437	380	QELFK R ISEQF
170	Q86V81	63	RPAI A RGAAGG
171	Q86V81	79	PAPYS R PKQLP
172	Q86V81	144	YDRSG R SLGTA
173	O08583	189	IDT Q R R PAQSI
174	P39748	211	EFHLS R ILQEL
175	P39748	239	YCESI R GIGPK
176	P39748	332	RLSKS R QGSTQ
177	P39748	355	LSSAK R KEPEP
178	P31269	128	GLPSS R PYGIK
179	P31269	141	PLSAR R GDCPT
180	P04049	59	TSNTI R VFLPN
181	P04049	223	RESVS R MPVSS
182	P46783	55	QSLKS R GYVKE
183	O76922	86	GSVR G RRLITD
184	O76922	123	FKVLK R PNWTI
185	O76922	146	ATRL R RSFLYE

(3) Subset 3 (185 samples)

Sample #	Protein ID	Site	Sequences
1	Q9NR22	368	NAKNV R DLDFI
2	Q9UBL3	181	LTWQS R TQDEH
3	Q9UBL3	380	ACLYE R VLLAL
4	Q9UBL3	417	ASHGV R KGAWY
5	Q9UBL3	437	PD T A A R L GWSQ
6	Q9UBL3	622	EVDG R RSPPE
7	Q92934	112	AQRY G RELRRM
8	Q92934	142	TATQ M RQSSSW
9	Q92934	149	SSSW T R V FQSW
10	Q92934	157	QSWW D RNLGRG
11	Q7Q2B7	40	QGM A V R LLQEG
12	Q7Q2B7	67	T T HL S R G H T SH
13	Q7Q2B7	299	NL V AL R DAAMK
14	P11911	162	LLFRK R WQNEK
15	Q09472	15	PPSAK R PKLSS
16	Q09472	370	NLPH C RTMKNV
17	Q09472	504	SPQGM R PMNSM
18	Q09472	626	ESANN R AEYYH
19	Q09472	1187	LCTI P R D ATYY
20	Q09472	1297	RLP S T R LG T FL
21	Q09472	1342	PGMK A R F VDSG
22	P03372	183	SAKET R YCAVC
23	P03372	269	MLKHK R QRDDG
24	Q01844	486	GGPM G RMGGRG
25	P22509	38	GFGG G RGGFGG

26	P22509	58	GGFRGRGGGGG
27	P22509	74	QSGGGRGRGGG
28	P22509	106	GVFICRGKEDA
29	P35637	216	QDRGGGRGGGS
30	P35637	234	GGGYNRSSGGY
31	P35637	372	SFATRRAFDFNR
32	P35637	518	SRGEHRQDRRE
33	P03347	20	RWEKIRLRPGG
34	P03347	22	EKIRLRPGGKK
35	P68431	41	VKKPHRYRPGT
36	P68431	43	KPHRYRPGTVA
37	P68431	50	GTVALREIRRY
38	P68431	64	TELLIRKLPFQ
39	P68431	84	FKTDLRFQSSA
40	P62805	18	KGGAKRHRKVL
41	P62805	37	KPAIRRLARRG
42	P62805	46	RGGVKRISGLI
43	P17096	24	DGTEKRGRGRP
44	P17096	28	KRGRGRPRKQP
45	P31942	174	RMRDGRGMGGH
46	P31942	205	RGLPFRATEND
47	O60506	381	IHFDERDGAVK
48	O60506	526	AGYSQRGGPGS
49	O60506	554	GVRGARGGRRGG
50	O60506	577	QPDSKRRQTNN
51	O60506	578	PDSKRRQTNNQ
52	Q07666	31	AHPSVRQTPSR
53	Q07666	282	VPEPSRGRGVP
54	Q07666	284	EPSRGRGVPVR
55	Q07666	346	GAPAPRARTAG
56	P08199	348	RTGTNRKFGYV
57	P08199	554	EGRTIRLELQG
58	P08199	599	GSVRARIVTDR
59	Q28165	23	GSGPGRRRHLV
60	O75925	248	PKRPSRPINIT
61	O75925	347	LTIPCRALTCS
62	O75925	417	TWAPMRSKKEV
63	O75925	507	ASPVSRTPSLP
64	Q96PU8	53	EISRVRKDMYN
65	Q96PU8	139	KEEQNRGKPNW
66	Q96PU8	205	LNGTYRDANIK
67	Q96PU8	256	IMPLIRQIQTA
68	Q96PU8	320	ATKVRRHDMRV
69	Q96PU8	324	RRHDMRVHPYQ
70	Q92804	216	NFGGHRDYGPR
71	Q92804	334	GSGGRRRGRGG
72	Q92804	335	SGGRRRGRGGY
73	Q92804	411	RGRGGRRGGDRG
74	P09867	122	EEHHLRDYFEQ
75	P09867	140	EIMTDRGSGKK
76	P09867	284	GNFGGRSSGPY
77	O00267	71	PPKKPRHGGFI

78	O00267	190	KIGEERATAIS
79	O00267	198	AISLMRKFIA Y
80	O00267	610	GPHSGREGEIR
81	O00267	615	REGEIRHLFRS
82	O00267	619	IRHLFRSFAFL
83	O00267	767	TSTYGRTPMYG
84	O00267	989	IQVKVRD TYLD
85	P69697	49	GISYGRKKRRQ
86	Q15554	41	GGPAER GAGEA
87	Q15554	274	PREPARQLRNP
88	P15056	199	CCAVYRIQDGE
89	P15056	362	FGQRDRSSSAP
90	P15056	384	IDDLIRDQGFR
91	P15056	435	SSSEDNRMK T
92	Q5TAQ9	65	TSTESRGT DTE
93	Q5TAQ9	94	INDENRVH DRS
94	Q5TAQ9	117	EQPRR RVQRKR
95	Q5TAQ9	125	RKRANRDQ DSS
96	Q5TAQ9	134	SSDDE RALEDW
97	Q5TAQ9	149	TSALP RPRWQA
98	Q5TAQ9	159	ALPALRER ELG
99	Q96EY1	25	PAISGRGAR PP
100	Q96EY1	340	KSPVFR RDGAD
101	Q96EY1	341	SPVFR RDGADI
102	P06746	182	VCGSFR GAES
103	Q92997	73	PCFNGRV VSWL
104	Q92997	105	PPPMERT GGIG
105	Q92997	200	DDSTSRF SSST
106	Q92997	219	MRRHKRRR RQ
107	Q60487	16	AAAARRE PPGA
108	Q60487	96	GVCANRYL AMK
109	Q60487	144	YVALKRTG QYK
110	Q54G21	235	RKSAKRIL VIP
111	Q54G21	613	VQTLHRLV EDN
112	Q54G21	673	VAQMRRQI IDL
113	Q54G21	746	YALKIRAFK KS
114	Q54G21	1031	KDVIKRV TVEY
115	Q9R1E0	222	HSKFIRVQ NEG
116	Q9R1E0	264	KFAKSRG RAAK
117	Q13283	32	PDMLHRF YGKN
118	Q13283	78	CHTKIRH VDAH
119	Q13283	302	PQRDQRV REQR
120	Q13283	304	RDQRVRE QRIN
121	Q13283	317	PQRGPRP IREA
122	Q13283	331	GDIEPRRM VRH
123	Q13283	332	DIEPRRMV RHP
124	Q13283	370	NVVELRINS GG
125	P31943	49	RFIYTR EGRPS
126	P31943	199	HYDPPR KLMAM
127	P04264	12	SRSYGRSG GGF
128	P04264	29	IINYQRRT TSS
129	P04264	239	FINNLR RRV DQ

130	P08729	13	PVFTSRSAAFS
131	P02687	129	FGYGGRASDYK
132	P02687	158	FKLGGDRSRSG
133	P15720	52	DRHVPRRGFGK
134	P15720	112	GLSLTRFSWGG
135	P25274	111	GTVLSRFSWGA
136	P11940	49	RDMITRRSLGY
137	P11940	67	PADAERALDTM
138	P11940	83	KGKPVIRIMWSQ
139	P11940	94	RDP SLRKSGVG
140	P11940	240	FV S FERHEDAQ
141	Q9JMB7	8	GRARARARGRA
142	Q9JMB7	10	ARARARGRARG
143	Q9JMB7	53	GRGRQRGMVVG
144	Q9JMB7	95	LGVNTRQNLDH
145	Q9JMB7	193	NGEHVIRITITL
146	Q8CDG1	21	PSQCVRMPGCW
147	Q8CDG1	31	WPQAPRPLEPA
148	Q8CDG1	50	RGLVFRKPEDS
149	Q8CDG1	206	VL T MERKEKEL
150	Q86WR7	186	PVEHPRLLRSV
151	P22626	62	ASKRSRGGFGV
152	P22626	99	KRAVAREESGK
153	P22626	147	EIITDRQSGKK
154	P51991	167	QSGKRRGFAFV
155	Q99729	28	VPEASRGRGWT
156	P14678	16	QHIDYRMRCIL
157	P14678	73	GLVLLRGENLV
158	P14678	108	GPGIGRAAGRG
159	P14678	209	GIPPGRGTPMG
160	Q14524	34	AEKQARGSTTL
161	Q14524	43	TLQESREGLPE
162	Q14524	222	ALRTFRVLRAL
163	Q14524	225	TFRVLRALKTI
164	Q14524	282	RHKCVRNFTAL
165	P23246	9	DRFRSRGGGGG
166	P23246	267	GGPGRRSEEKI
167	P23246	315	EDEFKRLFAKY
168	P23246	342	IKLESRALAEI
169	P68363	221	NLDIERPTYTN
170	P07437	262	MVPFPRLHFFM
171	Q86V81	29	GGGRGRGRAGS
172	Q86V81	73	GGGRNRPAPYS
173	Q86V81	155	DVHFERKADAL
174	Q86V81	204	GMTRNRGAGGF
175	O08583	38	GSQGGRRGAVQ
176	O08583	188	QIDTQRPAQS
177	O08583	196	AQSINRRGMTR
178	P31269	106	AAPDGRYMRSW
179	P31269	109	DGRYMRSWLEP
180	P31269	248	LNLTERQVKIW
181	P31269	258	WFQNRMRMKMK

182	P04049	316	PVPAQRERAPV
183	O76922	11	NPVIARGRGRG
184	O76922	26	NVEANRGFAPS
185	O76922	189	LVTKSRAGENI

(4) Subset 4 (185 samples)

Sample #	Protein ID	Site	Sequences
1	Q9NR22	95	LKDEVRTLTYR
2	Q9NR22	204	TVIFARDKWLK
3	Q9NR22	218	LMFPDRAALYV
4	Q9NR22	300	CLQIQRNDYVH
5	Q9UBL3	108	DEENGRQLGEV
6	Q9UBL3	160	NTYFLRKQANL
7	Q92934	109	LWAAQRYGREL
8	Q7Q2B7	261	KMYPSRGLHV
9	P11911	125	RNPVPRPFLDM
10	P11911	160	TLLLFRKRWQN
11	Q09472	237	GQTGLRGPQPL
12	Q09472	353	AHKCQRREQAN
13	Q09472	466	PSSIERAYAAL
14	Q09472	754	MGYGPRMQQPS
15	Q09472	883	LHPPPRQTPTP
16	Q09472	915	PQQQPRSQOST
17	Q09472	1055	KPEELRQALMP
18	Q09472	1137	NRKTSRVYKYC
19	Q09472	1166	GYCCGRKLEFS
20	P03372	162	RRQGGRRERLAS
21	P03372	243	QACRLRKCYEV
22	P03372	259	GIRKDRGGRM
23	P03372	277	DDGEGRGEVGS
24	P03372	300	PLMIKRSKNS
25	P03372	394	IGLVWRSMHP
26	P03372	434	LATSSRFRMMN
27	P03372	548	MLDAHRLHAPT
28	Q01844	138	ATAPTTRPDGN
29	Q01844	461	GGLPPREGRGM
30	Q01844	518	GNVQHRAGDWQ
31	P22509	80	GRGGGRGGKRG
32	P22509	213	NLAKKRTNIIP
33	P22509	250	QPDQTRIVALN
34	P35637	371	VSFATRRADFN
35	P35637	472	NYGDDRRGGRG
36	P35637	514	GKMDSRGEHRQ
37	P03347	91	YCVHQRIEIKD
38	P03347	150	QAISPRTLNAW
39	P03347	380	TIMMQRGNFRN
40	P03347	464	PFLQSRPEPTA
41	P03347	476	PEESFRSGVET
42	P68431	70	KLPFQRLVREI
43	P62805	36	TKPAIRRLARR

44	P62805	41	RRLAR R GGVKR
45	P62805	56	IYEET R GV LKV
46	P17096	73	GAAK T RKTTTT
47	P17096	86	RKPRG R PKKLE
48	P31942	116	PGPYD R PIGGR
49	P31942	137	MYDRM R RRGGDG
50	P31942	232	IGADG R ATGEA
51	O60506	60	SDLDE R AIEAL
52	O60506	286	DKKKN R GF CFL
53	O60506	343	KVLFV R NLANT
54	O60506	520	APPRG R AGYSQ
55	Q07666	17	SRSSG R SGSMD
56	Q07666	213	KEEEL R KGGDP
57	Q07666	289	RGVPV R GRGAA
58	Q07666	354	TAGIQ R IPLPP
59	P08199	281	PAPGK R KKEMT
60	P08199	471	GEKGQ R QERTG
61	P08199	551	MEIEG R TIRLE
62	P08199	597	FEGSV R ARIVT
63	Q28165	24	SGPGR R RHLVP
64	Q28165	125	EAIKAR V REME
65	O75925	28	LGYAG R NKHGR
66	O75925	163	SDNSQ R FR ETC
67	O75925	296	TVLLQ R LR AKG
68	O75925	330	ATTSL R VSLLC
69	O75925	632	HTVTN R SSTDT
70	Q96PU8	43	FNHLE R LLDEE
71	Q96PU8	101	FNFV G RILGPR
72	Q96PU8	330	VHPYQ R I VTAD
73	Q96PU8	336	I VTAD R AATGN
74	Q92804	153	SYHSQ R ENYSH
75	Q92804	169	RRDVS R YGEDN
76	Q92804	187	GGGRG R GGYDK
77	Q92804	195	YDKDG R GPMTG
78	Q92804	326	RPEFM R GGGSG
79	Q92804	373	MNFAR R NSCNQ
80	Q92804	395	SGGDF R GRGYG
81	Q92804	441	GYSGD R SRGGY
82	Q92804	526	GYGGD R SRGGY
83	P09867	53	DPNTK R SRGFG
84	P09867	178	HNCEV R KALSK
85	P09867	218	NDNF G RGGNFS
86	O00267	123	DRSGA R RLQNL
87	O00267	133	LWRDQ R EEELG
88	O00267	246	GVGNL R LG YWN
89	O00267	334	RKKFK R PPQRL
90	O00267	416	STGKE R EHNFQ
91	O00267	747	TISVD R QRLTT
92	Q15554	25	SRSSG R ARRGR
93	Q15554	66	FHEAL R AFRGS
94	Q15554	89	QALLV R PLGKE
95	Q15554	200	LLNI R EKNLA

96	Q15554	270	VEKPPREPARQ
97	Q15554	361	RMTISRLVLEE
98	Q15554	490	VMIKDRWRMTK
99	Q5TAQ9	116	EEQPRRRVQRK
100	Q96EY1	8	ARCSTRWLLVV
101	Q96EY1	28	SGRGARPPREG
102	Q96EY1	319	DGQTVRMPVGK
103	Q96EY1	325	MPVGKREIFIT
104	P06746	40	KYNAYRKAASV
105	P06746	112	GPSAARKFVDE
106	P06746	253	KEYPHRRIDIR
107	P06746	299	NEYTIRPLGVT
108	P06746	328	IQWKYREPKDR
109	Q92997	231	VSRIERSSSFS
110	Q60487	63	IHPDGRVDGVR
111	Q60487	84	LQAEDRGVVS
112	Q60487	131	NYNTYRSRKYS
113	Q9R1E0	21	PLPRQRSCTWP
114	Q9R1E0	98	AAAASRGLCGD
115	Q9R1E0	154	TSSSRRNAWGN
116	Q9R1E0	266	AKSRGRAAKKK
117	Q9R1E0	311	NWSTFRPRTSS
118	Q13283	299	PQRPQRDQVR
119	Q13283	307	RVREQRINIPP
120	Q13283	335	PRRMVRHPDSH
121	P31943	81	ETMGHRYVEVF
122	P31943	114	NDGFVRLRGLP
123	P31943	224	YNSIGRGAGFE
124	P31943	232	GFERMRRGAYG
125	P31943	262	SDRFGDLNYC
126	P04264	179	QKVKSBREREQI
127	P04264	240	INNLRRRVDQL
128	P04264	241	NNLRRRVDQLK
129	P04264	403	I SELNRVIQRL
130	P08729	25	RGAQVRLSSAR
131	P08729	88	SLQVRVQEASE
132	P08729	161	QVDGGRLEAEL
133	P08729	226	EINFLRTLNET
134	Q36736	21	DQEFNRKMQEQ
135	P02687	9	KRPSQRSKYLA
136	P02687	63	GHHAARTTHYG
137	P15720	53	RHVPRRGFGKD
138	P25274	63	KDHAARTTHYG
139	P25274	77	QKSGHRPQDEN
140	Q8IZS5	91	GNEYGRNYFDP
141	Q8IZS5	118	REDDDRIFYNR
142	Q8IZS5	221	VTAVIRSRYY
143	Q8IZS5	223	AVIRSRYYRE
144	P11940	44	SIRVCRDMITR
145	P11940	172	KVFGVGRFKSRK
146	Q9JMB7	125	FRLTSRPQWAL
147	Q8CDG1	39	EPAWGRAGPAG

148	Q8CDG1	108	SANMVRKDREE
149	Q8CDG1	115	DREEPRSSLPD
150	Q86WR7	231	RTPAARGPRSG
151	Q86WR7	268	AAREPRTLRSR
152	P51991	16	QPDSGRRRRRR
153	P51991	19	SGRRRRRRGEE
154	P51991	20	GRRRRRRGEEG
155	Q99729	194	PKLNKRRGFVF
156	Q99729	322	GKSQRGGHQN
157	P14678	94	DTGIARVPLAG
158	P14678	112	GRAAGRGIPAG
159	Q14524	219	NVSALRTFRVL
160	Q14524	376	QDCWERLYQOT
161	Q14524	433	EEKEKRFQEAM
162	P23246	32	GLHDFRSPPPG
163	P23246	287	LSLLRRPGEKT
164	P23246	298	YTQRCRLFVGN
165	P23246	407	VIVDDRGRSTG
166	P07437	62	GKYVPRAILVD
167	P07437	390	FTAMFRRKAFL
168	Q86V81	31	GRGRGRAGSQG
169	Q86V81	47	AQAAARVNRGG
170	O08583	50	AARVNRRGGPM
171	P39748	29	KSYFGRKVAID
172	P39748	47	FLIAVRQGGDV
173	P39748	261	IEEIVRRLDPN
174	P39748	320	QFSEERIRSGV
175	P31269	210	STRKKRCPYTK
176	P31269	257	IWFQNRMKMK
177	P04049	73	TVVNVNRGMSL
178	P04049	111	KGKKARLDWNT
179	P04049	218	TMRMRRESVSR
180	P04049	256	LSQRQRSTSTP
181	P46783	65	EQFAWRHFYWY
182	P46783	119	PARLTRGEADR
183	O76922	96	DLVYSRPPGMT
184	O76922	248	NLENFRMQLWP
185	O76922	291	LSDAIRDSDDY

(5) Subset 5 (185 samples)

Sample #	Protein ID	Site	Sequences
1	Q9NR22	251	DMTCIRDVAMK
2	Q9UBL3	213	MTTRQRP GKMT
3	Q9UBL3	621	TEVDGRRSPPW
4	Q92934	20	SSSAERGLGPS
5	Q92934	115	YGRELRRMSDE
6	Q7Q2B7	146	TSTYQRAIYNN
7	Q7Q2B7	239	MLYNERMLETY
8	Q7Q2B7	308	MKEYFRQPIVD
9	Q7Q2B7	326	MAKSI RHTTNF

10	P11911	30	GCQALRVEGGP
11	P11911	55	CENNGRNPNT
12	P11911	120	TYLRVRNPVPR
13	Q09472	200	SIGAGRGRQNM
14	Q09472	362	ANGEVRQCNLP
15	Q09472	424	NAGDKRNQQPI
16	Q09472	613	LVAYARKVEGD
17	Q09472	648	ELEEKRRTRLQ
18	Q09472	651	EKRRTLQKQN
19	Q09472	695	DPSMIRGSVPN
20	Q09472	728	PPIVPRQTPPL
21	Q09472	838	SPVPSRTPTPH
22	Q09472	1067	LEALYRQDPES
23	Q09472	1234	EQFSKRKNDTL
24	Q09472	1281	LKKSARTRKEN
25	Q09472	1305	TFLENRVNDFL
26	Q09472	1311	VNDFLRQNHHP
27	Q09472	1312	NDFLRRQNHPE
28	Q09472	1391	PPPNQRRVYIS
29	Q09472	1405	SVHFFRPKCLR
30	P03372	211	KAFFKRSIQGH
31	P03372	233	TIDKNRRKSCQ
32	P03372	412	NLLLDNRNQGKC
33	P03372	477	KDHIHRVLDKI
34	P03372	555	HAPTSRGGASV
35	Q01844	446	KVSLARKKPPM
36	Q01844	648	DKGEHRQERRD
37	Q01844	651	EHRQERRDRPY
38	P22509	64	GGGGRRGGGFQ
39	P22509	128	VYGEKRVSISE
40	P22509	204	SHRSGRDLINL
41	P22509	310	LEPYERDHAVV
42	P03347	15	GGELDRWEKIR
43	P03347	58	TSEGCROILGQ
44	P03347	361	PGHKARVLAEA
45	P03347	406	TARNCRAPRKK
46	P03347	444	PSYKGRPGNFL
47	P17096	30	GRGRPRKQPPV
48	P31942	80	GKHKEIRIGHRY
49	P31942	200	HFVHMRGLPFR
50	O60506	142	VTTGQRKYGGP
51	O60506	192	PIWDLRLMMDP
52	O60506	203	LTGLNRGYAFV
53	O60506	366	FGKLERVKCLK
54	O60506	448	TRGRGRGGRRG
55	O60506	501	RGRGGRGARGA
56	O60506	546	GAQQQRGRGVR
57	Q07666	186	GNTIKRRLQEET
58	Q07666	302	PPPVPGRGVG
59	P08199	561	ELQGRGSPNA
60	P08199	567	GSPNARSQPSK
61	Q28165	227	DKESVRTSLAL

62	Q28165	240	SLFRGRQIKVI
63	O75925	165	NSQRFRETCFA
64	Q96PU8	51	DEEISRVRKDM
65	Q96PU8	68	GSTEKRS AELP
66	Q96PU8	130	GKGS MRD KKKE
67	Q92804	337	GGRRGRGGYRG
68	Q92804	343	GGYRGRGGFQG
69	Q92804	545	GYGGDRSGGYG
70	P09867	75	AAMNARPHKVD
71	P09867	97	REDSORPGAHL
72	P09867	146	GSGK KRGF AFV
73	O00267	119	VLDEDRSGARR
74	O00267	469	PAQELRKYFKM
75	O00267	552	VGVI VRLERET
76	O00267	555	IVRLE RETFQV
77	O00267	668	APMS PR ISSPM
78	O00267	700	SRGRGRRDNEL
79	O00267	701	RGRGRRDNELI
80	O00267	757	TVGSR RP GGMT
81	O00267	1005	QTGVI RSVTGG
82	P69697	55	KKRRQR RRPPQ
83	Q15554	27	SSGRARRGRHE
84	Q15554	99	EHTVSRLLRVM
85	Q15554	109	MQCLSR IEEGE
86	Q15554	492	IKDRWR TMKRL
87	P15056	95	DALQQREQQLL
88	P15056	178	CGVTVRDSLKK
89	P15056	443	MKTLGR RDSSD
90	Q5TAQ9	98	NRVHDRSEEEE
91	Q5TAQ9	122	RVQRKRANRDQ
92	Q5TAQ9	225	WDWVR RQP VLD
93	Q96EY1	41	GAWLSRKL SVP
94	Q96EY1	64	ALLTLRPGVSL
95	Q96EY1	151	SDEVKRKQYDA
96	Q96EY1	277	RSTCRRCGGRG
97	P06746	149	GDFEKRI PREE
98	P06746	254	EYPHRRIDIRL
99	Q92997	113	GIGDSRPPSFH
100	Q92997	149	ERPRR RDGPEH
101	Q92997	168	KGERRREP GGY
102	Q92997	216	SRLMR RHKRRR
103	Q92997	221	RHKRRRRKQKV
104	Q60487	121	CFFFERLESNN
105	Q60487	133	NTYRSRKYSSW
106	Q54G21	194	LENNQRVGS LF
107	Q54G21	210	SETPFRFV VGE
108	Q9R1E0	29	TWPLR PPEFNQ
109	Q9R1E0	606	GMFIERLD CDM
110	Q13283	17	GREFVRQYYTL
111	Q13283	132	HNDIFRYQDEV
112	Q13283	247	VQEDLRTFSWA
113	Q13283	314	NIPPQRGPRPI

114	P31943	75	ALKKDRETMGH
115	P31943	175	KKHKEIRIGHRY
116	P04264	8	RQFSSRSRGYRS
117	P04264	38	SSSTRRSRGGGG
118	P04264	181	VKSREREQIKS
119	P04264	223	VDTSTRTHNLE
120	P04264	267	MVEDYRNKYED
121	P04264	277	DEINKRTNAEN
122	P08729	136	SAKSSRLPDIF
123	P08729	302	HGDDLNRTRNE
124	P02687	31	FLPRHRDTGIL
125	P02687	96	NIVTPRTPPPS
126	P15720	30	RHGSPRHRDSG
127	P15720	42	LDSLGRFFGGD
128	P25274	150	QGTLSRLFKLG
129	Q8IZS5	132	LFDESROGEPQ
130	Q8IZS5	195	LEQEIIRSTSLH
131	P11940	265	QIYVGRAQKKV
132	Q9JMB7	80	LSLAERGGRRR
133	Q9JMB7	85	RGGRRRDFHDL
134	Q9JMB7	121	STNHFRRLTSRP
135	Q9JMB7	174	LFLPKRLQHKV
136	Q86WR7	24	PSCRLRAFSRG
137	Q86WR7	40	RSSSSRSRSFT
138	Q86WR7	42	SSSRSRSFTLD
139	Q86WR7	172	TPDPPRRELRA
140	Q86WR7	234	AARGPRSGDPG
141	Q86WR7	269	AREPRRTLSRA
142	Q86WR7	273	RRTLSRAAVSV
143	P22626	38	TEESLRNYEQ
144	P22626	89	HSIDGRVVEPK
145	P51991	18	DSGRRRRRRGE
146	P51991	68	DCVVMRDPQTK
147	P51991	74	DPQTKRSRFG
148	P51991	246	GNFGGRGGYGG
149	Q99729	137	HRLDGRVIDPK
150	Q99729	171	TEEKIREYFGE
151	Q99729	321	YGKSQRRGGHQ
152	P14678	25	ILQDGRIFIGT
153	P14678	61	SKQAEREEKRV
154	P14678	65	EREKRVLGLV
155	P14678	132	LAGPVRGVGGP
156	Q14524	104	GKTIFRFSATN
157	Q14524	190	AFTFLRDPWNW
158	Q14524	383	YQOTLRSAGKI
159	Q14524	451	EALTIRGVDTV
160	P23246	358	DDTPMRGRQLR
161	P68363	79	VIDEVRTGTYS
162	P68363	123	VLDRIKRLADQ
163	P68363	214	IYDICRRNLDI
164	P68363	402	LMYAKRAFVHW
165	P07437	156	LISKIREEYPD

166	P07437	320	AVFRGRMSMKE
167	Q86V81	58	GPIRNRPAIAR
168	Q86V81	71	AGGGGRNRPAP
169	Q86V81	141	AVHYDRSGRSL
170	Q86V81	190	IDAQRPAQSV
171	O08583	70	AAGGGRNRPAP
172	O08583	201	RGGMTRNRGSG
173	P39748	103	AKRSEERRAEAE
174	P39748	129	EKFTRKRLVKVT
175	P39748	262	EEIVRRLDPNK
176	P31269	235	YLTRDRRYEVA
177	P04049	215	PSLTMRRMRES
178	P04049	216	SLTMRMRRESV
179	P46783	124	RGEADRD TYRR
180	P46783	153	TEFQFRGGFGR
181	O76922	80	SEGDP RGSVRG
182	O76922	132	TIYQYRVDFTP
183	O76922	145	EATRLRRSFLY
184	O76922	221	NLILRRAMEGL
185	O76922	233	LKLVSRYYYDP

(6) Subset 6 (186 samples)

Sample #	Protein ID	Site	Sequences
1	Q9NR22	100	RTLTYRNSMYH
2	Q9UBL3	300	GRGAKRKQODG
3	Q9UBL3	342	NKDG YRYILAE
4	Q9UBL3	388	LALHDRAPQLK
5	Q92934	39	SGKHHRQAPGL
6	Q92934	70	GAVEI RSRHSS
7	Q92934	98	FRGRSR SAPPN
8	Q92934	161	DRNLGRGSSAP
9	Q7Q2B7	112	SVFNLRTEDSS
10	Q7Q2B7	140	MQDFVRTSTYQ
11	Q09472	202	GAGRGRQNMQY
12	Q09472	397	HCASSRQIISH
13	Q09472	568	STTGIRKQWHE
14	Q09472	603	AALKDRRMENL
15	Q09472	705	NQMMPRITPQS
16	Q09472	1356	ESFPYRTKALF
17	P03372	157	PNSDNRRQGGR
18	P03372	271	KHKRQRDDGEG
19	P03372	287	SAGDMRAANLW
20	P03372	335	EYDPT RPFSEA
21	Q01844	292	GPGENRSMGSP
22	P22509	141	DKIEYRAWNPF
23	P22509	201	VEFSHRSGRDL
24	P22509	261	AHTFLRNGGHF
25	P22509	320	VVG VYRPPPKA
26	P35637	218	RGGRG RGGSGG
27	P35637	441	MNFSWRNECNQ

28	P03347	39	IVWASRELERF
29	P03347	43	SRELERFVAVNP
30	P03347	76	GSEELRSLYNT
31	P03347	214	AAEWDRVHPVH
32	P03347	229	APGQMRPRGS
33	P03347	232	QMREPRGSDIA
34	P62805	20	GAKRHRKVLRD
35	P62805	93	VYALKRQRTL
36	P17096	81	TTTTPGRKPRGR
37	P31942	55	MDYQGRSTGEA
38	P31942	104	FYDPPRLLGQ
39	P31942	110	RLLGQRP GPYD
40	P31942	169	DGFDDMRDGR
41	O60506	130	KALLERTGYTL
42	O60506	171	VGKIPRDLFED
43	O60506	244	SVANNRLFVGS
44	O60506	304	AAQARRRLMSG
45	O60506	413	PPDQKRKERKA
46	O60506	478	DYHNYRGGYED
47	O60506	548	QQQRGRGVRGA
48	O60506	551	RGRGVRGARGG
49	Q07666	36	RQTPSRQPPLP
50	Q07666	159	MKLKERVLIPV
51	Q07666	348	PAPRARTAGIQ
52	Q07666	420	DWNGTRPSLKA
53	Q07666	429	KAPPARPVKGA
54	P08199	385	EKPKGRDSKKV
55	Q28165	76	EEEEPRPRAPP
56	Q28165	127	IKARVREMEEE
57	Q28165	251	PKRTNRPGIST
58	O75925	15	MVMSLRVSELO
59	O75925	201	VQVQLRFCLSE
60	O75925	245	GVEPKRPSRPI
61	O75925	298	LLQRLRAKGIR
62	O75925	309	NPDHSRALIKE
63	O75925	341	PLGKMRLTIPC
64	O75925	622	SSNSLRSHSH
65	Q96PU8	170	EIKLKRAVEEV
66	Q96PU8	319	VATKVRRHDMR
67	Q92804	403	GYGGERGYRGR
68	Q92804	408	RGYRGRGGRGG
69	P09867	14	EPEQLRKLFIG
70	P09867	55	NTKRSRGFGFV
71	P09867	92	KRAVSRSDSOR
72	P09867	232	GFGGSRGGGGY
73	O00267	310	LKMIPRIDYDR
74	O00267	365	IFEGNRYSRKG
75	O00267	484	KVIAGRFEFEDT
76	O00267	733	TESTARVELHS
77	O00267	749	SVDRQRLTTVG
78	O00267	756	TTVGSRRPGGM
79	P69697	57	RRQRRRPPQGS

80	Q15554	21	GRRASRSSGRA
81	Q15554	80	DFRQIRDIMQA
82	Q15554	223	QOKMLRFLESH
83	Q15554	277	PARQLRNPPTT
84	Q15554	332	KNKRPRKDENE
85	P15056	146	PTDVARSNPKS
86	P15056	252	FCDFCRKLLFQ
87	P15056	271	YKFHQRCSTEV
88	P15056	354	ADEDHRNQFGQ
89	P15056	437	SEDRNRMKTLG
90	Q5TAQ9	115	EEEQPRRRVQR
91	Q5TAQ9	168	LGSSARFVYEA
92	P06746	183	CGSFRRGAESS
93	P06746	283	FNKNMRAHALE
94	Q92997	39	KGVLQRPSTYKF
95	Q92997	143	LVSAQRERPRR
96	Q92997	147	QRERPRRRDGP
97	Q92997	157	PEHATRLNGTA
98	Q92997	166	TAKGERRREPG
99	Q92997	222	HKRRRRKQKVS
100	Q92997	228	KQKVSRIERSS
101	Q60487	57	GGFFLRIHPDG
102	Q54G21	339	NESNSRNNSHSN
103	Q54G21	631	VEKENRMVTLK
104	Q9R1E0	147	LAGQPRKTSSS
105	Q9R1E0	153	KTSSSRRNAWG
106	Q9R1E0	177	SSAEKRLTLSQ
107	Q9R1E0	211	WKNSIRHNLSL
108	Q13283	398	KVLSNRPIMFR
109	P31943	29	ADEVQRFSSDC
110	P31943	188	IFKSSRAEVRT
111	P04264	45	GGGGGRFSSCG
112	P04264	251	KSDQSRLDSEL
113	P04264	392	HGDSVRNSKIE
114	P08729	253	SMDNSRSLDL
115	P02687	23	TMDHARHGFLP
116	P02687	52	RGAPKRGSGKD
117	P02687	78	KAQGHRPQDEN
118	P02687	161	GGRDSRSGSPM
119	P15720	6	MASQKRSSFRH
120	P15720	32	GSPRHRDSGLL
121	P15720	63	DIHAARASHVG
122	P15720	73	GSIPQRSQHGR
123	P11940	41	PILSIRVCRDM
124	P11940	176	GRFKSRKEREK
125	P11940	179	KSRKEREAEELG
126	P11940	272	QKKVERQTELK
127	P11940	309	GIDDERLRKEF
128	Q9JMB7	12	ARARGRARGQE
129	Q8CDG1	45	AGPAGRGLVFR
130	Q8CDG1	293	ELKSQRKTDDA
131	Q86WR7	28	LRAFSRGGSL

132	Q86WR7	35	GSLESRSSSSR
133	Q86WR7	218	PTSPFREGRPG
134	Q86WR7	226	RPGEWRTPAAR
135	P22626	12	TVPLERKKREK
136	P22626	82	AAMAARPHSID
137	P22626	129	EEHHLRDYFEE
138	P22626	153	QSGKKRGGFV
139	P22626	190	RKALSROEMQE
140	P51991	17	PDSGRRRRRRG
141	P51991	76	QTKRSRGGFV
142	P51991	161	EVMEDRQSGKK
143	Q99729	110	NTGRSRGGFI
144	Q99729	133	DQKEHRLDGRV
145	P14678	49	DCDEFRKIKPK
146	P14678	181	PPPMGRGAPP
147	P14678	196	PPPGMRPPMGP
148	Q14524	8	NFLLPBGTSSF
149	Q14524	14	GTSSFRFTRE
150	Q14524	15	TSSFRFTRES
151	Q14524	122	FHPIRRAAVKI
152	P23246	236	PKPPHRGGGEP
153	P23246	286	NLSLLRRPGEK
154	P23246	296	KTYTQRCRLFV
155	P23246	360	TPMRGRQLRVR
156	P23246	363	RGRQLRVRFAT
157	P23246	376	AALSVRNLSPY
158	P68363	84	RTGTYRQLFHP
159	P68363	121	DLVLDRIKLA
160	P68363	156	SLLMERLSVDY
161	P68363	229	YTNLNRLISQI
162	P68363	422	EFSEAREDMAA
163	P07437	121	VLDVVRKEAES
164	P07437	251	LNADLRKCLAVN
165	P07437	309	DPRHGRYLTVA
166	P07437	391	TAMFRKAFH
167	Q86V81	17	IIKLNRSQGG
168	Q86V81	23	SQGGRRGGGRG
169	Q86V81	27	GRGGRRGRGRA
170	Q86V81	38	GSQGGRRGGGAQ
171	Q86V81	50	AARVNRGGGPI
172	Q86V81	56	GGGPIRNRPAI
173	O08583	47	VQAAARVNRGG
174	P39748	327	RSGVKRLSKSR
175	P31269	40	LGQPPRQAATL
176	P31269	204	NWLHARSTRKK
177	P31269	236	LTRDRRYEVAR
178	P04049	40	QFGYQRRASDD
179	P04049	41	FGYQRRASDDG
180	P04049	89	KALKVRGLQPE
181	P04049	191	DWSNIRQLLLF
182	P04049	254	GSLSQRQRSTS
183	P46783	95	VPATLRRSRPE

184	P46783	103	RPETGRPRPKG
185	P46783	129	RDTYRRSAVPP
186	O76922	17	GRGRGRKPNNV

(7) Subset 7 (185 samples)

Sample #	Protein ID	Site	Sequences
1	Q9NR22	8	MKHSSRCLLLR
2	Q9NR22	349	DYLTVRRGEEI
3	Q9NR22	350	YLTVRRGEEIY
4	Q9UBL3	398	KISDDRLLTVVG
5	Q9UBL3	461	FSYSWRSKKGT
6	Q7Q2B7	206	NNLTDRIIVIA
7	Q7Q2B7	318	DTFDIRICMAK
8	Q7Q2B7	391	HWYQVRCLLQT
9	Q7Q2B7	417	LAANQRQSYDV
10	Q7Q2B7	535	ISSTGRQQQQQ
11	P11911	111	NNILKRSCGTY
12	Q09472	335	ADPEKRKLIQQ
13	Q09472	1252	CTECGRKMHQI
14	Q09472	1283	KSARTRKENKF
15	Q09472	1292	KFSAKRLPSTR
16	Q09472	1478	KAVSERIVHDY
17	P03372	151	PPAFYRPNSDN
18	P03372	164	QGGRERLASTN
19	P03372	263	DRRGGRMLKHK
20	P03372	352	TNLADRELVHM
21	P03372	363	INWAKRVPGFV
22	Q01844	268	QOSSFRQDHPS
23	Q01844	330	GSAGERGGFNK
24	Q01844	537	QNFARTECNQ
25	P22509	46	FGGGGRGRGGG
26	P22509	48	GGGRGRGGGGG
27	P22509	98	MVEPHRHEGVF
28	P22509	147	AWNPFRSKLAA
29	P22509	230	HPHKYRMLIAM
30	P35637	422	GGGQQRAGDWK
31	P03347	264	GEIYKRWIILG
32	P03347	286	SILDIRQGPKE
33	P03347	294	PKEPFRDYVDR
34	P03347	384	QRGNFRNQRKM
35	P68431	27	ATKAARKSAPA
36	P68431	53	ALREIRRYQKS
37	P17096	56	VTPKRPGRGP
38	P31942	84	ERIGHRYIEIF
39	P31942	90	YIEIFRSSRSE
40	P31942	135	GSMYDRMRRGG
41	P31942	138	YDRMRGGDGY
42	P31942	171	FDDRMRDGRGM
43	O60506	105	KTYRQREKQGT

44	O60506	416	QKRKERKAQRQ
45	O60506	420	ERKAQRQAANK
46	O60506	498	VGARGRGGRGA
47	O60506	512	APSRGRGAAPP
48	O60506	533	GPGSARGVRGA
49	Q07666	56	SRGGARASPAT
50	P08199	391	DSKKVRAARTL
51	P08199	458	AEIDGRSVSLY
52	Q28165	172	MEADARSIYVG
53	O75925	33	RNKHGRKHELL
54	O75925	62	IKELYRRRFPQ
55	O75925	64	ELYRRRFPQKI
56	O75925	479	EPSAKRTCPSL
57	Q96PU8	124	CKIMVBRGK GSM
58	Q96PU8	163	EDAQNRAEIKL
59	Q96PU8	227	AQAAPRIITGP
60	Q92804	175	YGEDNRGYGGS
61	Q92804	383	QCNEPRPEDSR
62	Q92804	397	GDFRGRGYGGE
63	P09867	47	DCVVMRDPNTK
64	O00267	124	RSGARRLQNLW
65	O00267	280	PKSWVRLKRG I
66	O00267	283	WVRLKRG IYKD
67	O00267	315	RIDYDRIKARM
68	O00267	329	DWFAKRKKFKR
69	O00267	338	KRPPQRLFDAE
70	O00267	494	TGLIVRVEENF
71	O00267	577	HQAVTRKKDNR
72	O00267	813	LHDGSRTPAQS
73	Q15554	54	EEAVNRWVLKF
74	Q15554	154	VVESSRKL VKE
75	Q15554	330	ALKNKRPRKDE
76	Q15554	482	YPFVNRTAVMI
77	P15056	158	QKPIV RVFLPN
78	P15056	166	LPNKQRTVVPA
79	P15056	172	TVVPARCGVTV
80	P15056	389	RDQGFRGDGGS
81	P15056	424	SPGPQERERKSS
82	P15056	444	KTLGRDRDSSDD
83	Q5TAQ9	33	GAEEGRETSSG
84	Q5TAQ9	59	DGGPNRTSTES
85	Q5TAQ9	120	RRRVQKRANR
86	Q5TAQ9	184	FVQRFRLQHGL
87	Q5TAQ9	224	VWDWVRRQPVL
88	Q96EY1	18	VVGT PRLPAIS
89	Q96EY1	31	GARPPREGVVG
90	Q92997	148	RERPRRRDGPE
91	Q60487	15	TAAAARREPGG
92	Q60487	68	RVDGVREKTD P
93	Q54G21	12	EEYSNRNAASS
94	Q54G21	320	VAKLGRATGFT
95	Q54G21	672	NVAQMRRQIID

96	Q54G21	759	DEKNARKSDIE
97	Q54G21	800	KLDMVRIKKRN
98	Q54G21	804	VRIKKRNDDKV
99	Q9R1E0	83	SEDFARAPGCV
100	Q9R1E0	324	STISGRLLSPIM
101	Q13283	13	PLLVGREFVRQ
102	Q13283	63	QKEIHRKVMSQ
103	Q13283	320	GPRPIREAGEQ
104	P31943	44	GAQGIIRFIYTR
105	P31943	116	GFVRLRGLPFG
106	P31943	150	VDFQGRSTGEA
107	P31943	206	LMAMQRP GPYD
108	P31943	212	PGPYDRPGAGR
109	P31943	259	GFGSDRFGRDL
110	P04264	86	ARGGGRGSGFG
111	P08729	77	LLAPLRRLDADP
112	P08729	86	DPSLQVRVQEE
113	P08729	149	QIAGLRGQLEA
114	P08729	167	LEAELRSMQDV
115	Q36736	53	YEKFERMIKEH
116	P02687	29	HGFLPRHRDTG
117	P02687	47	FFGSDRGAPKR
118	P02687	112	GLSLSRFSWGA
119	P15720	25	TTDHARHGSPR
120	P25274	43	LDSIGRFFSSD
121	Q8IZS5	105	EEINPRQCATE
122	Q8IZS5	123	RIFYNRLTKLF
123	Q8IZS5	224	VIRSRYYREQ
124	P11940	50	DMITRRSLGYA
125	P11940	290	QDRITRYQGVN
126	Q9JMB7	144	PLMEARRLRSA
127	Q9JMB7	187	VFSQTRNGEHV
128	Q9JMB7	218	YNIIFRRLKI
129	Q8CDG1	83	LDTAFRPPSKR
130	Q8CDG1	140	SVGWSRMLGRG
131	Q8CDG1	144	SRMLGRGSSEV
132	Q8CDG1	181	ARDPPRLPQPP
133	Q8CDG1	255	ECKSMRFGMLK
134	Q8CDG1	326	YNVVFRRVMKL
135	Q86WR7	189	HPRLLRVPTP
136	P22626	21	EKEQFRKLF IG
137	P22626	54	DCVVMRDPASK
138	P22626	95	VVEPKRAVARE
139	P22626	200	EVQSSRSRGGG
140	P22626	266	GYGGGRGGYGG
141	P22626	325	NFGGSRNMGGP
142	P51991	21	RRRRRGEEGH
143	P51991	96	AAMCARPHKVD
144	P51991	214	SAGSORGRGGG
145	P51991	216	GSQRGRGGGSG
146	Q99729	108	DPNTGRSRGFG
147	Q99729	195	KLNKRRGFVFI

148	Q99729	253	RNRGNRGS GGG
149	P14678	18	IDYRMR CILQD
150	P14678	172	QYPPGRGGPPP
151	Q14524	18	FRRFTR E SLAA
152	Q14524	27	AAIEKRMAEKQ
153	Q14524	53	EEEAPRPQLDL
154	Q14524	121	PFHPIRRRAVK
155	Q14524	277	FMGNLRHKCVR
156	P23246	19	GGFHRRGGGGG
157	P23246	44	GLNQNRGPMGP
158	P23246	242	GGGEP RGG RQH
159	P23246	245	EPRGGRQHHP
160	P23246	365	RQLRVRFATHA
161	P68363	64	GKHVPRAVFVD
162	P68363	373	LAKVQRAVCML
163	P68363	390	AEAWARLDHKF
164	P07437	46	DLQLDRISVYY
165	P07437	86	FGQIFRPDNFV
166	Q86V81	189	QIDAQRPAQS
167	O08583	58	GPMRNRPAIAR
168	P39748	73	FYRTIRMMENG
169	P39748	245	GIGPKRAVDLI
170	P39748	322	SEERIRSGVKR
171	P39748	339	GSTQGRLLDDFF
172	P31269	29	ELSVGRYAPGT
173	P31269	166	SPPVDREKQPS
174	P04049	67	LPNKQRTVVNV
175	P04049	164	LLNGFR CQTCG
176	P04049	275	LPVDSRMIEDA
177	P04049	331	EKNKIRPRGQR
178	P46783	98	TLRRSRPETGR
179	P46783	113	GLEGERPARLT
180	O76922	84	PRGSVRGRRLI
181	O76922	143	DVEATRLRRSF
182	O76922	220	LNLILRRAMEG
183	O76922	260	YQTSIRQHEND
184	O76922	278	CHKVMRTETLY
185	O76922	302	QSTFKRAVMGM