

Figure S1. MALDI-MS/MS spectra of a tryptic peptide with m/z 1398.9 derived from a ~ 55 kDa band from a protein extract of *C.amadis*.

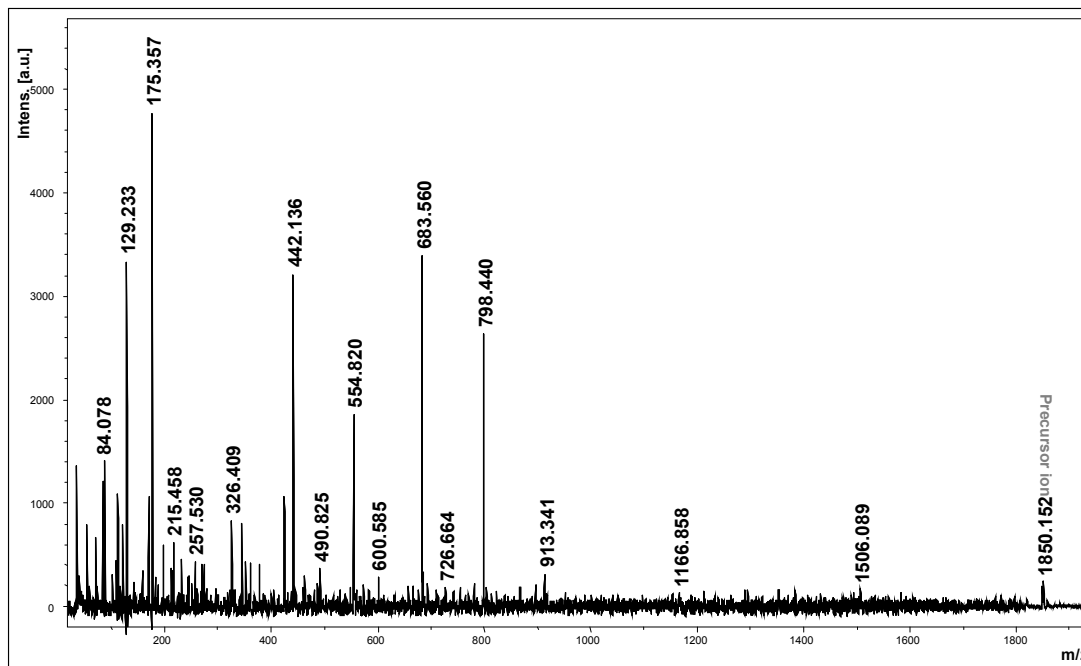


Figure S2. MALDI-MS/MS spectra of a tryptic peptide with m/z 1850.3 derived from a ~55kDa band from a protein extract of *C. amadis*.

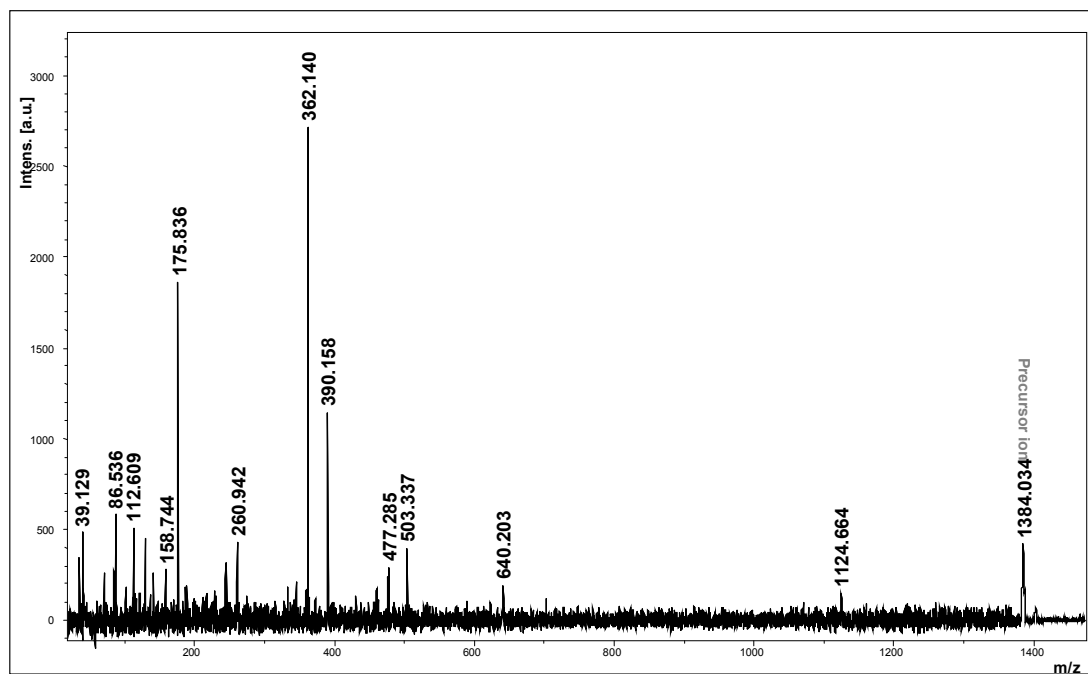


Figure S3. MALDI-MS/MS spectra of a tryptic peptide with m/z 1383.9 derived from a ~ 55 kDa band from a protein extract of *C.amadis*.

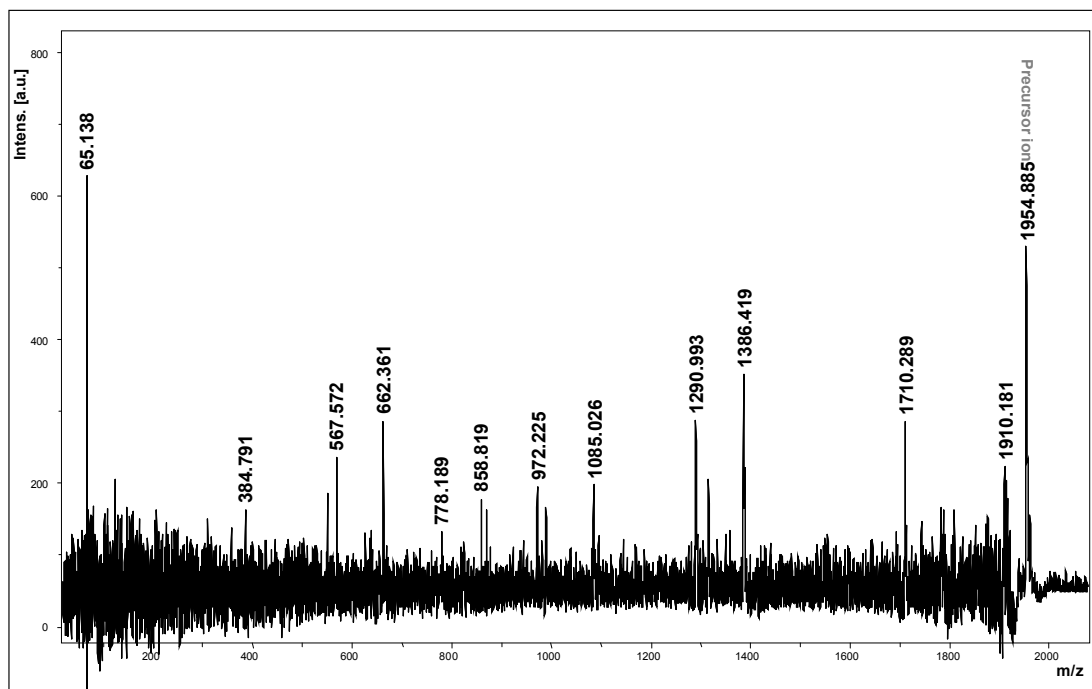


Figure S4. MALDI-MS/MS spectra of a tryptic peptide with m/z 1954.3 derived from a ~55kDa band from a protein extract of *C.amadis*.

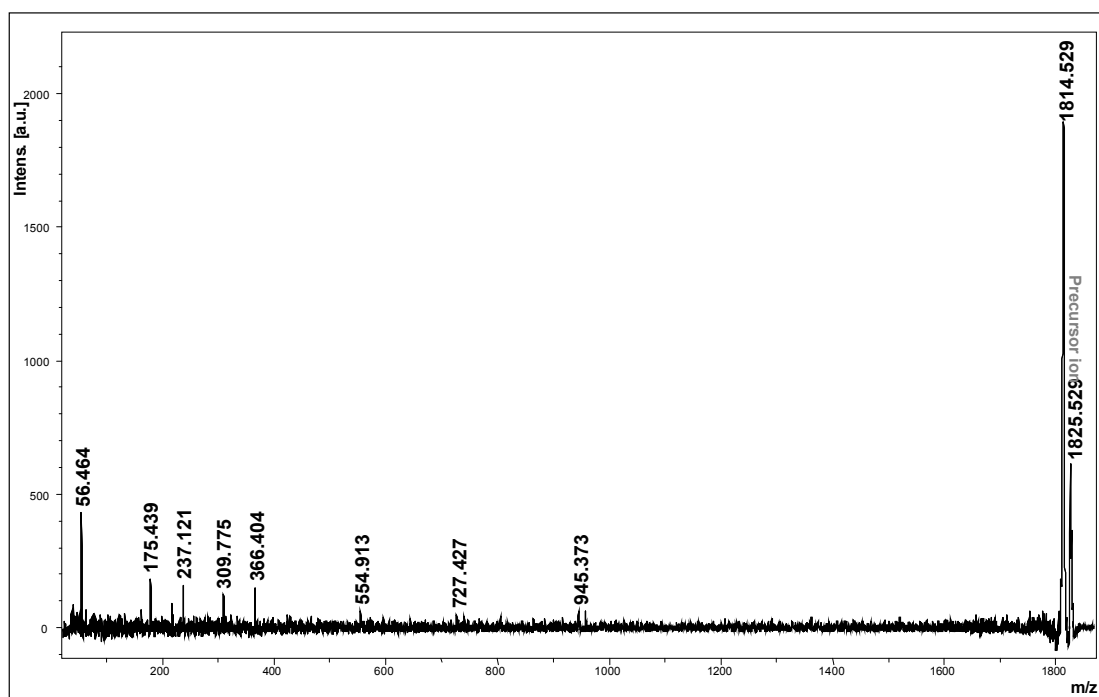


Figure S5. MALDI-MS/MS spectra of a tryptic peptide with m/z 1825.2 derived from a ~55kDa band from a protein extract of *C.amadis*.

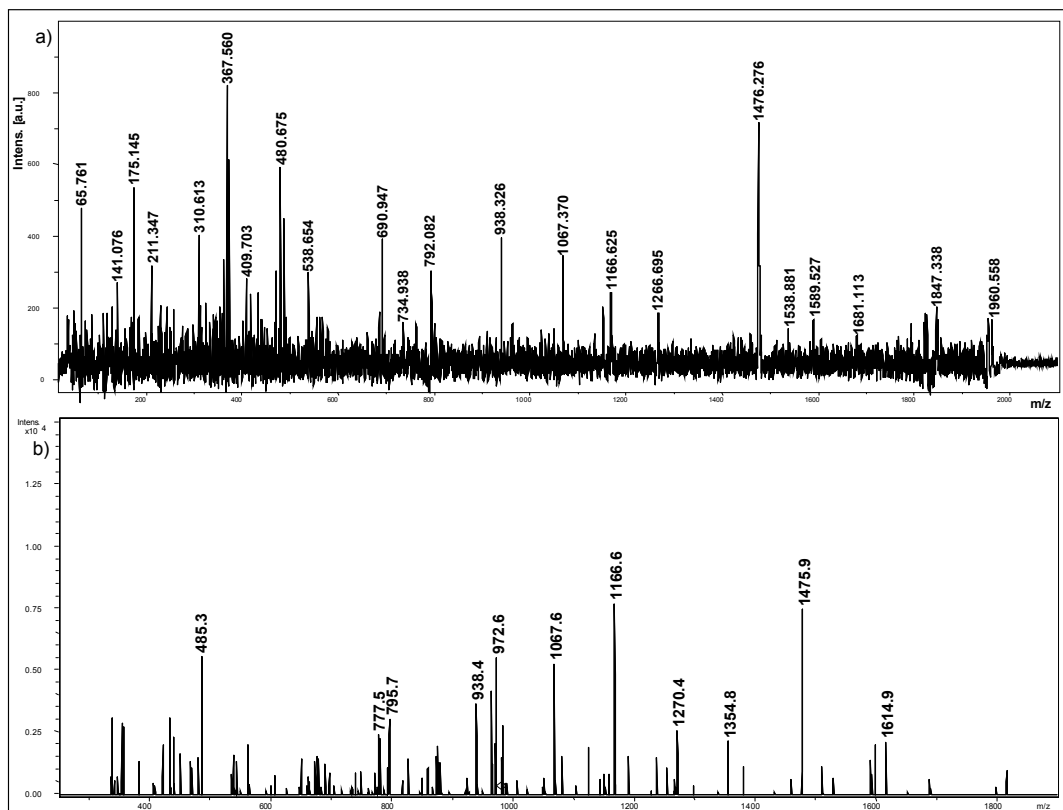


Figure S6. a) MALDI-MS/MS spectra of a tryptic peptide with m/z 1960.5 and b) ESI-MS/MS spectra obtained by CID of doubly charge ion (m/z 980.6) derived from a ~55kDa band from a protein extract of *C.amadis*.

Table S1. Diverse tetrapeptides derived from different proteins and the resulting BLAST analysis over a non-redundant database.^a

SR No.	Tetrapeptide	Protein	Accession number	E-value	BLAST hit	SR No	Tetrapeptide	Protein	Accession number	E-value	BLAST hit
1	ADQR	Ribonuclease	NP_774106	81660	-	26	IPER	Catalase	NP_001743	36802	-
2	AAIK	"	"	197267	-	27	VVWK	Amidating enzyme	AAF67216	24791	-
3	DHVK	Hemoglobin	AAA29796	60859	-	28	TSVK	Synaptotagmine	P21521	178718	-
4	HSVR	"	"	81660	-	29	ALQR	RAD51	CAA45563	90553	-
5	EHQK	"	"	45357	-	30	GCQR	"	"	32360	-
6	IPVK	Myoglobin	NP_976312	81660	-	31	DIAR	GroEL	AAR17695	76049	-
7	EFFK	TIM	P47721	33803	-	32	LQER	"	"	55645	-
8	EVTR	Rhodopsin	BAB88652	81660	-	33	ECIR	Kinesin	BAC06512	22824	-
9	YEVK	PDI	AAA85099	33803	-	34	YADR	"	"	45462	-
10	AAAK	"	"	355156	-	35	DVFR	"	"	41092	-
11	LGEK	"	"	147018	-	36	NSDR	Barnase	AAC53661	70660	-
12	GLQR	Hydroxylase	AAO43431	109569	-	37	TGTR	Shibire	NP996468	100183	-
13	DFHK	"	"	33803	-	38	WFGR	Glutaminyl cyclase	AAA30549	10262	-
14	HCDK	"	"	25193	-	39	WYFR	"	"	3911	-
15	LPGK	DNA polymerase	P00581	147018	-	40	SDYR	Peptide isomerase	AAB34914	39975	-
16	GNLK	Hydroxylase	AAO43431	109569	-	41	ALFR	Carboxylase	AAK97782	78895	-
17	EEHK	Hemoglobin	AAA29796	45357	-	42	QYAR	"	"	36802	-
18	AIVK	Phosphofruktinase	CAB16429	109569	-	43	CIAR	"	"	37142	-
19	IVTK	Dynamin	CAA42068	109569	-	44	QWER	Beta-lactamase	AAA93528	8462	-
20	IVEK	DNA gyrase	CAA91553	90553	-	45	LPLR	"	"	81849	-
21	LMAR	"	"	39975	-	46	GLSK	Alpha-synuclin	NP_000336	246503	-
22	NVYK	Shibire	NP996468	54133	-	47	GGGR	Prion protein	NP991149	147359	-
23	LNFK	Synaptotagmine	P21521	78895	-	48	SLQR	Tubulin	BAA36504	79624	-
24	ENGR	Amidating enzyme	AAF67216	71312	-	49	AEVK	GFP	AAA69675	176344	-
25	FHYK	Catalase	NP_001743	21799	-	50	TLGR	Lysozyme	NP631919	176344	-

a) The reported E-values are based on the database on 22nd December 2006.

The derived confidence parameter (P) is $0/50 = 0$

Table S2. Diverse pentapeptides derived from different proteins and the resulting BLAST analysis over a non-redundant database.^a

SR No.	Pentapeptide	Protein	Accession number	E-value	BLAST hit	SR No	Pentapeptide	Protein	Accession number	E-value	BLAST hit
1	ICLSK	Ribonuclease	NP_774106	9715	-	26	QSEQK	Synaptotagmine	P21521	14456	-
2	DIQFR	"	"	4022	-	27	DVGHR	Amidating enzyme	AAF67216	10773	-
3	GDVEK	Cytochrome C	NP_061820	17491	-	28	AWGEK	"	"	5984	-
4	IFIMK	"	"	2234	-	29	EFGPR	"	"	8029	-
5	ELCMK	Hemoglobin	AAA29796	2234	-	30	GFAVK	Catalase	NP_001743	26026	-
6	VGTSK	"	"	42254	-	31	VWPHK	"	"	1579	-
7	ENFTK	"	"	7241	-	32	MLQGR	"	"	3480	-
8	SLAPK	TIM	P47721	31491	-	33	IEGSK	GroEL	AAR17695	18869	-
9	FDEGR	PDI	AAA85099	7241	-	34	IHQIK	"	"	3610	-
10	DDTWR	Chymotrypsin	CAA74031	1665	-	35	RPLNK	Kinesin	BAC06512	10010	-
11	VAAGK	"	"	56695	-	36	GQDTR	"	"	8562	-
12	GCNTR	"	"	7241	-	37	MAGCR	Glutaminyl cyclase	AAA30549	3032	-
13	GICSR	Hydroxylase	AAO43431	9715	-	38	HLASK	"	"	23960	-
14	LEGIK	"	"	17491	-	39	IVGGK	Peptide isomerase	AAB34914	21068	-
15	VGHER	"	"	9715	-	40	LELSK	"	"	30147	-
16	DHFIR	"	"	2997	-	41	DTGER	Carboxylase	AAK97782	10973	-
17	LQLNR	DNA polymerase	P00581	9715	-	42	LELQK	"	"	16440	-
18	VLATR	Phosphofructkinase	CAB16429	23470	-	43	SNGEK	Subtilisin	AAC63365	23096	-
19	HMADR	Dynamin	CAA42068	2997	-	44	AAVDK	"	"	36228	-
20	MCTAK	"	"	4022	-	45	LSASK	Beta-lactamase	AAA93528	53780	-
21	ASMDK	DNA gyrase	CAA91553	10773	-	46	SIPTK	"	"	13680	-
22	YSEGR	"	"	10773	-	47	DQLGK	Alpha-synuclin	NP_000336	16899	-
23	ASFLR	Shibire	NP996468	19396	-	48	QLIFK	Hexokinase	P50506	6560	-
24	EEMLR	"	"	3324	-	49	EDIFK	Tubulin	BAA36504	6280	-
25	GVDMK	Synaptotagmine	P21521	8029	-	50	GSQGR	Protease	CAD36507	15416	-

a) The reported E-values are based on the database on 22nd December 2006.

The derived confidence parameter (P) is $0/50 = 0$

Table S3. Diverse hexapeptides derived from different proteins and the resulting BLAST analysis over a non-redundant database.^a

SR No.	Hexapeptide	Protein	Accession number	E-value	BLAST hit	SR No	Hexapeptide	Protein	Accession number	E-value	BLAST hit
1	YIPGTK	Cytochrome C	NP_061820	1109	+	26	SFGAGR	Amidating enzyme	AAF67216	9635	-
2	SLEHAK	Hemoglobin	AAA29796	2679	-	27	FSADGK	"	"	3988	-
3	DPFFIK	"	"	459	-	28	TDQGIK	Catalase	NP_001743	2215	+
4	SEDEMK	Myoglobin	NP_976312	616	-	29	VANYQR	"	"	917	-
5	ASEDLK	"	"	3594	-	30	TDFSGR	RAD51	CAA45563	2215	-
6	ALELFR	"	"	1488	-	31	QMHLAK	"	"	683	-
7	NLLENK	TIM	P47721	1488	-	32	QIVVQK	GroEL	AAR17695	1651	-
8	LLCTVK	Rhodopsin	BAB88652	1488	-	33	GASANK	Kinesin	BAC06512	9635	-
9	AAHVLK	PDI	AAA85099	3594	-	34	ISVCVR	"	"	1230	-
10	EDSPIK	"	"	1488	-	35	LTMVLR	"	"	917	-
11	GYPTLK	"	"	1488	-	36	QFEQSK	Barstar	P11540	1230	-
12	DSLLPK	Chymotrypsin	CAA74031	2679	-	37	MEWDER	Protease	CAD36507	65	-
13	FANWIR	Trypsin	AAA29340	190	-	38	YADALR	"	"	917	-
14	LLDQNR	Hydroxylase	AAO43431	1109	-	39	ELGLLK	Glutaminyl cyclase	AAA30549	3988	-
15	EFHLPR	DNA polymerase	P00581	459	-	40	DHSSER	"	"	1651	-
16	GLLENR	Phosphofructkinase	CAB16429	1996	-	41	SILHTK	Peptide isomerase	AAB34914	2215	-
17	LVPEER	"	"	1996	-	42	WAHNAK	Carboxylase	AAK97782	509	-
18	MACDEK	Dynamain	CAA42068	342	+	43	DVISEK	Subtilisin	AAC63365	2215	-
19	QVETIR	"	"	1996	-	44	EVGEVR	Beta-lactamase	AAA93528	2215	-
20	NYQEAK	BSA	CAA76847	826	-	45	WDGKPR	"	"	380	-
21	ILNYEK	DNA gyrase	CAA91553	1651	+	46	MDVFMK	Alpha-synuclin	NP_000336	157	+
22	VGSEIK	"	"	3988	-	47	GGGFGR	Prion protein	NP991149	5352	-
23	EVEEFK	Shibire	NP996468	917	-	48	VPDDIK	Hexokinase	P50506	1230	-
24	NIHGIR	"	"	917	+	49	DQDLSK	"	"	2215	-
25	MPPNAK	Synaptotagmine	P21521	683	-	50	YIVGFK	Subtilisin	AAC63365	917	-

a) The reported E-values are based on the database on 22nd December 2006.

The derived confidence parameter (P) is $6/50 = 0.12$

Table S4. Diverse octapeptides derived from different proteins and the resulting BLAST analysis over a non-redundant database.^a

SR No.	Octapeptide	Protein	Accession number	E-value	BLAST hit	SR No	Octapeptide	Protein	Accession number	E-value	BLAST hit
1	ADLIAYLK	Cytochrome C	NP_061820	32	+	26	HDQIGEVK	Synaptotagmine	P21521	20	+
2	QDGIDLYK	He moglobin	AAA29796	13	+	27	IAIMQNGK	"	"	15	+
3	TTLDEPTK	"	"	32	+	28	THAHGLGK	Amidating enzyme	AAF67216	65	+
4	HAWQEIGK	"	"	5.5	+	29	GNLYVFHR	"	"	8.3	+
5	GHPETLEK	Myoglobin	NP_976312	24	+	30	DAQIFIQK	Catalase	NP_001743	15	+
6	ALAPEYAK	PDI	AAA85099	43	+	31	TGAQIVYK	GroEL	AAR17695	36	+
7	DTESADAK	"	"	78	-	32	FTAAPLVK	Kinesin	BAC06512	65	+
8	ILEFFGLK	"	"	18	-	33	MLEQPQYR	"	"	2.5	+
9	QLMPTWDK	"	"	1.7	+	34	NSSHVPFR	"	"	20	+
10	IHPNYNSR	Trypsin	AAA29340	5.5	+	35	LPDNYITK	Barnase	AAC3661	11	+
11	AESSELVR	Hydroxylase	AAO43431	78	+	36	YFVCNYYK	Protease	CAD36507	1.1	+
12	YDVPALTK	DNA polymerase	P00581	24	+	37	YFPHWDDR	Glutaminyl cyclase	AAA30549	0.59	+
13	DTDMGLLR	"	"	13	+	38	TTIDNLNK	"	"	27	+
14	VPLEDAVR	Phosphofruktinase	CAB16429	32	+	39	HYPGVYMK	Peptide isomerase	AAB34914	2.5	+
15	FSSFDEIR	Dynamain	CAA42068	13	+	40	WGDDEACR	Carboxylase	AAK97782	3.4	+
16	QMIFQFIR	"	"	1.7	+	41	MHYMSMYR	"	"	0.24	+
17	DDSPDLPK	BSA	CAA76847	24	+	42	FWLEDQLR	Beta-lactamase	AAA93528	2.5	+
18	TEEMLVEK	DNA polymerase	P00581	10	+	43	FALPENMR	Hexokinase	P50506	6.2	+
19	SPYGEQCR	"	"	7.4	+	44	DFDTTQSK	"	"	27	+
20	VIDFTGER	PDI	AAA85099	18	+	45	LAELIGER	"	"	48	+
21	LGNSEVVK	DNA gyrase	CAA91553	87	+	46	TTVLDVMR	Tubulin	BAA36504	15	+
22	ILANEEIR	"	"	20	+	47	NVMVSTGR	"	"	27	+
23	GYIGVVNR	Shibire	NP996468	27	+	48	EAFMEQFR	"	"	3.4	+
24	NLVDSYMK	"	"	8.3	+	49	EAFSLFDK	Calmodulin	MCKKA	27	+
25	AAEAASQR	Synaptotagmine	P21521	156	+	50	NAIHNHIR	Huntingtin	NP002102	8.3	+

a) The reported E-values are based on the database on 22nd December 2006.

The derived confidence parameter (P) is 48/50 = 0.96

Table S5. Diverse decapeptides derived from different proteins and the resulting BLAST analysis over a non-redundant database.^a

SR No.	Decapeptide	Protein	Accession number	E-value	BLAST hit	SR No	Decapeptide	Protein	Accession number	E-value	BLAST hit
1	HGTCSEGLADR	Ribonuclease	NP_774106	0.66	+	26	HWSDMLASPR	Synaptotagmine	P21521	0.052	+
2	HMFEHYPPAMK	Hemoglobin	AAA29796	0.008	+	27	TLVLQVHYAK	Amidating enzyme	AAF67216	0.41	+
3	ENYTPADVQK	"	"	0.20	+	28	DLFNAIATGK	Catalase	NP_001743	0.98	+
4	HMFEHYPPMR	"	"	0.002	+	29	LCENIAGHLK	"	"	0.30	+
5	SHQHLTEAEK	"	"	0.37	+	30	VGAATEVEMK	GroEL	AAR17695	0.73	+
6	TYQVCATQLK	TIM	P47721	0.20	+	31	TLQASTAVNR	Kinesin	BAC06512	1.8	+
7	DAASIVAWLK	PDI	AAA85099	0.49	+	32	SHAIFQILR	"	"	0.23	+
8	LISLEEDMTK	"	"	0.20	+	33	SIGGDIFSNR	Barnase	AAC3661	0.73	+
9	DGAGLSDEEK	"	"	1.6	+	34	TTDHYQFTFK	"	"	0.13	+
10	LWTNGPIADK	Chymotrypsin	CAA74031	0.15	+	35	SISDLHQTLK	Barstar	P11540	0.73	+
11	TLDNDIAIMR	Trypsin	AAA29340	0.42	+	36	DPEETDEILK	Synaptotagmine	P21521	0.23	+
12	SGGPPSEQLR	"	"	0.11	+	37	IVLHPGYKPK	Peptide isomerase	AAB34914	0.23	+
13	QALFQQEMAR	Hydroxylase	AAO43431	0.11	+	38	VSSYTNWMSK	"	"	0.07	+
14	QTNGTADPER	Epsin	AAW34263	0.49	+	39	LFGFSLSDLK	Carboxylase	AAK97782	0.98	+
15	TCAAFTDFGR	"	"	0.27	+	40	GFLDPQAWSK	"	"	0.17	+
16	CVFISIFNEK	"	"	0.084	+	41	QTMSTMSAAK	Subtilisin	AAC63365	0.41	+
17	EESDQAGLDK	Enolase	CAJ03181	0.89	+	42	HPNWTNTQVR	"	"	0.039	+
18	MGSEVYHALK	"	"	0.20	+	43	DNFDEMDTSR	Tubulin	BAA36504	0.07	+
19	NLVSASQDGK	G-protein	CAA35532	2.1	+	44	LSEQQIAEFK	Calmodulin	MCCKA	0.41	+
20	AGVLAGHDNR	"	"	1.2	+	45	WCDDGTPGGK	Lysozyme	NP631919	0.13	+
21	EGLTAVISVK	DNA gyrase	CAA91553	2.4	+	46	GAHHTGGLK	Huntingtin	NP002102	0.55	+
22	GISNIPINLR	Shibire	NP996468	0.41	+	47	CVEEILGYLK	"	"	0.17	+
23	HFQPGDASIK	"	"	0.41	+	48	EGALDAVLER	MutS	AAC43637	1.3	+
24	LSPPSPGGVR	"	"	1.8	+	49	QSQHMTEVVR	p53	BAC16799	0.13	+
25	SLPYADAMNK	Synaptotagmin	P21521	0.23	+	50	EVHVCACPGR	"	"	0.13	+

a) The reported E-values are based on the database on 22nd December 2006.

The derived confidence parameter (P) is $50/50 = 1$

Table S6. Diverse unidecapeptides derived from different proteins and the resulting BLAST analysis over a non-redundant database.^a

SR No.	Unidecapeptide	Protein	Accession number	E-value	BLAST hit	SR No	Unidecapeptide	Protein	Accession number	E-value	BLAST hit
1	TGPNLHGLFGR	Cytochrome C	NP_061820	0.063	+	26	GQQTILFAWAK	Amidating enzyme	AAF67216	0.021	+
2	EFSHEISHHGK	Hemoglobin	AAA29796	0.014	+	27	HMNGYGSHTFK	Catalase	NP_001743	0.009	+
3	VGLSASTGLYK	Concannavalin	CVJB	0.37	+	28	LSQEDPDYGIR	"	"	0.021	+
4	HCMITTLCCGK	Rhodopsin	BAB88652	0.002	+	29	SELICLTTGSK	RAD51	CAA45563	0.13	+
5	SGKPQEYGGGR	PDI	AAA85099	0.11	+	30	ELVEGLEGDEK	GroEL	AAR17695	0.093	+
6	DYPCYVTGWGR	Chymotrypsin	CAA74031	0.002	+	31	AEVDVITIPSR	Kinesin	BAC06512	0.069	+
7	VSAYIDWINEK	"	"	0.006	+	32	THTMGGDFSGK	"	"	0.052	+
8	GGLIVFHNGHK	DNA polymerase	P00581	0.047	+	33	SDEEYDEMVEK	"	"	0.007	+
9	ENCIDTLVLSR	"	"	0.026	+	34	EADINYTSGFR	Barnase	AAC3661	0.029	+
10	IDNPAAIAELK	Dehydrogenase	CAA38581	0.084	+	35	ILYSSDWLIYK	"	"	0.005	+
11	YCGAGCTQAVK	Enolase	CAJ03181	0.035	+	36	ADVTDAANMLK	Protease	CAD36507	0.069	+
12	IYAMHWASDSR	G-protein	CAA35532	0.003	+	37	YPGSPGSFAAR	Glutamyl cyclase	AAA30549	0.13	+
13	LIVWDSYTTNK	"	"	0.008	+	38	HLVLACHYDSK	"	"	0.016	+
14	FATGSDDATCR	"	"	0.063	+	39	MASTPHPPGAR	"	"	0.039	+
15	VLALDAAAFK	Eps15	AAA02912	0.27	+	40	DQLLPQEASPK	Carboxylase	AAK97782	0.069	+
16	AQLEEQLQEVK	"	"	0.035	+	41	QLTPPTQPTFR	"	"	0.021	+
17	SEDVPPALPPK	"	"	0.063	+	42	GGGWGGSSSGR	Prion protein	NP991149	0.41	+
18	FVPIVAEEDGR	K ⁺ -channel	BAA23630	0.035	+	43	AGWGAAGGHR	"	"	0.069	+
19	LQNLLSGPWP	GPCR	NP_683765	0.019	+	44	ITDHFISELEK	Hexokinase	P50506	0.021	+
20	VDYYLLHETWR	"	"	0.001	+	45	LSICGIAAICK	"	"	0.052	+
21	SGLEISNLPK	DNA gyrase	CAA91553	0.23	+	46	IHEDIFDIIDR	Tubulin	BAA36504	0.005	+
22	VSVDAAEADK	"	"	0.23	+	47	DYGIFQINSFK	Lysozyme	NP631919	0.012	+
23	VLNQQLTNHIR	Shibire	NP996468	0.021	+	48	FHVGDWMTIR	Huntingtin	NP002102	0.003	+
24	AIMMLIINNAK	"	"	0.009	+	49	EGYDPDLALR	MutS	AAC43637	0.029	+
25	TLVFAIFDFDR	Synaptotagmine	P21521	0.016	+	50	LSDADVDQMIK	Calmodulin	MCKKA	0.029	+

a) The reported E-values are based on the database on 22nd December 2006.

The derived confidence parameter (P) is 50/50 = 1

Table S7. Diverse random (nontryptic) nonapeptides derived from different proteins and the resulting BLAST analysis over a non-redundant database.^a

SR No.	Nonapeptide	Protein	Accession number	E-value	BLAST hit	SR No	Nonapeptide	Protein	Accession number	E-value	BLAST hit
1	SWVMTCAAYA	G-protein	CAA35532	0.29	+	26	IGCGQSACG	Protease	CAD36507	5.5	+
2	PSGSFVACG	"	"	10	+	27	NNKYFVCNY	"	"	0.22	+
3	GLDNICSIY	"	"	1.3	+	28	YKGTMGDEP	"	"	1.3	+
4	SLKTREGNV	"	"	10	+	29	YQLGRPCSQ	"	"	1.3	+
5	RVSRELPGH	"	"	4.1	+	30	CRSSCQHIR	"	"	0.71	+
6	TGYLSCCR	"	"	0.71	+	31	PALHSQGFT	Subtilisin	AAC63365	5.5	+
7	LDDNQIVTS	"	"	4.1	+	32	GSNVKVAVI	"	"	18	+
8	SGDMTCALW	"	"	0.53	+	33	DSGIDSSHP	"	"	7.4	+
9	DIETGQQCT	"	"	1.7	+	34	DLKVAGGAS	"	"	43	+
10	AFTGHTGDV	"	"	7.4	+	35	MVPSETNPF	"	"	0.95	+
11	VRSADTGKD	Lysozyme	NP631919	13	+	36	QDNNSHGTH	"	"	2.3	+
12	YGIFQINSF	"	"	0.95	+	37	VAGTVAALN	"	"	32	+
13	KWCDDGTPG	"	"	0.71	+	38	NSVFVLGVA	"	"	13	+
14	GKNLCKVAC	"	"	3.1	+	39	PSASLYAVK	"	"	13	+
15	SDLLNDDLK	"	"	5.5	+	40	VLGADGSGQ	"	"	32	+
16	ASVGCALKI	"	"	13	+	41	KELGTVMRS	Calmodulin	MCCKA	4.1	+
17	VKMDGLKSW	"	"	0.95	+	42	LGQNPSESE	"	"	7.4	+
18	ETWDSYCNG	"	"	0.29	+	43	LTDMINEVD	"	"	0.95	+
19	RKMSRWVKG	"	"	0.71	+	44	VNSDGSIDF	"	"	5.5	+
20	CEQRKQSLR	"	"	1.7	+	45	PEFLTMMAR	"	"	0.39	+
21	AWSSGDLTF	Protease	CAD36507	3.1	+	46	KMKDTDSEA	"	"	3.1	+
22	LGAVQMWAD	"	"	0.95	+	47	EIAEAFKVF	"	"	7.1	+
23	EIVDFQYGV	"	"	1.3	+	48	DRNGDGKIS	"	"	7.4	+
24	WTDGTGHI	"	"	0.53	+	49	AAELRHLLT	"	"	7.4	+
25	QQVFAGASR	"	"	10	+	50	SIGEKLSDA	"	"	18	+

a) The reported E-values are based on the database on 22nd December 2006.

The derived confidence parameter (P) = 50/50 = 1