

SUPPLEMENTARY INFORMATION

Global Identification of Peptidase Specificity by Multiplex Substrate Profiling

Anthony J. O'Donoghue¹, A. Alegra Eroy-Reveles^{1,4}, Giselle M. Knudsen¹, Jessica Ingram², Min Zhou¹, Jacob B. Statnekov¹, Alexander L. Greninger³, Daniel R. Hostetter¹, Gang Qu⁵, David A. Maltby¹, Marc O. Anderson⁴, Joseph L. DeRisi³, James H. McKerrow², Alma L. Burlingame¹, Charles S. Craik^{1,*}

¹Dept. of Pharmaceutical Chemistry, UCSF,

²Sandler Center for Drug Discovery, UCSF,

³Howard Hughes Medical Institute and Dept. of Biochemistry and Biophysics, UCSF,

⁴Dept. of Chemistry and Biochemistry, San Francisco State University,

⁵Dept. of Electrical and Computer Engineering, University of Maryland

Contents:

Supplementary Tables 1-5

Supplementary Figures 1-5

SUPPLEMENTARY TABLES

Supplementary Table 1: Calculation of Pearson correlation coefficients to compare MSP-MS and COFRADIC assays. Recombinant mouse cathepsin E purchased from R&D Systems was used in both assays. The negative and positive selection values of each amino acid was determined using the Z score calculator in iceLogo. Cleavage at 233 sites in the 1200 minute MSP-MS assay was compared to 1223 sites in the mouse proteome that were identified by COFRADIC (<http://merops.sanger.ac.uk/cgi-bin/substrates?id=A01.010>). Nle in the MSP-MS assay was compared to Met in the COFRADIC assay.

		P4	P3	P2	P1	P1'	P2'	P3'	P4'
Pearson Correlation		-0.09	0.12	0.40	0.87	0.78	0.55	0.20	0.27
MSP-MS									
	Gly	0.42	1.34	-2.58	-2.85	-1.84	-1.76	-1.26	0.87
	Ala	1.26	-0.33	1.34	-1.50	1.11	0.83	-0.58	-0.42
	Ser	2.86	0.17	0.49	-2.28	-0.63	0.00	0.40	1.39
	Pro	1.32	1.01	-0.65	-2.68	-1.31	-1.79	-1.90	0.85
	Val	-0.61	0.80	1.99	-2.74	3.32	1.73	0.15	-0.28
	Thr	1.19	2.35	-0.52	-1.57	-0.98	1.01	0.10	-1.17
	Cys	-	-	-	-	-	-	-	-
	Leu	1.47	1.50	1.37	4.09	1.71	0.40	0.33	0.56
	Ile	-0.18	0.74	3.15	-0.94	2.60	1.26	3.96	1.29
	Asn	-0.28	-1.54	0.05	0.09	-1.87	-1.44	-1.43	-1.29
	Asp	0.40	0.80	0.49	-1.24	-2.17	-0.79	-1.86	0.16
	Gln	0.96	-1.61	0.51	-1.15	-1.19	-0.04	0.28	0.11
	Lys	0.15	-0.21	-0.75	-2.08	-2.39	-1.27	0.18	0.42
	Glu	-0.23	0.47	0.85	0.22	-1.84	-0.13	-0.58	0.06
	Nle	-0.64	0.17	-0.33	5.74	3.85	-0.04	-0.12	-1.61
	His	-0.21	0.56	-2.05	-1.55	-2.14	-1.54	-1.38	-1.13
	Phe	0.40	0.59	2.06	8.56	1.71	0.35	-1.07	-1.44
	Arg	1.41	1.17	-0.60	-1.99	-1.73	1.36	0.80	-0.58
	Tyr	0.35	0.37	-0.48	1.46	2.60	-0.48	0.33	-0.99
	Trp	0.27	1.68	0.91	3.22	1.68	-1.02	0.20	-0.66
COFRADIC									
	Gly	3.03	-2.78	-4.69	-6.92	-7.26	-3.57	-0.99	6.61
	Ala	2.63	-3.53	3.71	-3.20	8.79	9.98	0.25	5.76
	Ser	-2.21	-2.61	1.05	-9.13	-4.49	1.05	-1.52	-1.52
	Pro	3.38	-5.48	-8.58	-8.70	-8.12	-8.70	-6.97	-0.99
	Val	-0.36	5.64	8.07	-7.97	18.68	11.64	4.37	1.37
	Thr	-0.42	0.81	-0.42	-5.74	-4.38	3.66	2.42	1.68
	Cys	-1.20	-1.58	0.90	0.33	-3.69	0.71	-2.35	-2.16
	Leu	-3.77	0.74	-6.30	34.67	5.70	-2.24	1.28	-3.32
	Ile	-1.22	3.32	3.46	-6.58	16.25	4.01	4.01	0.43
	Asn	-0.39	-1.61	3.24	-5.24	-2.36	-2.06	2.18	1.73
	Asp	3.00	4.94	10.38	-3.72	-2.30	-1.65	4.17	0.16
	Gln	-1.03	-1.03	1.99	-5.64	-4.98	2.51	1.46	2.25
	Lys	-1.69	-0.12	-3.01	-8.19	-5.78	1.68	4.33	4.21
	Glu	2.50	4.34	6.93	0.67	-2.25	3.37	4.99	1.96
	Met	-0.69	2.22	0.86	10.76	2.80	-0.69	2.22	0.09
	His	-0.58	-2.73	-2.55	-5.59	-5.59	-5.59	-5.59	-5.23
	Phe	-0.02	3.11	-4.63	34.07	5.19	-5.23	-3.00	-3.89
	Arg	0.26	-3.60	-4.56	-7.46	-8.30	-8.30	-8.30	-8.30
	Tyr	-0.54	1.20	-2.97	4.15	2.93	-3.15	-3.84	-2.28
	Trp	-2.32	-0.25	-3.87	-2.58	-3.87	-3.87	-3.87	-3.87

Supplementary Table 2: Screening of proteolytic activity using viral proteases expressed in a bacterial IVTT system. A selection of viral proteases were incubated with a peptide mixture consisting of three peptides from the MSP-MS library and six peptides corresponding to viral polypeptide processing sites. * indicates a commercial preparation of enzyme.

Peptide Substrate	Peptide Sequence	HRV14*	HRV14	EV71	Polio	HAV	Klasse
TDP-18	YnDSIRHQ↓GPFQnL	Yes	Yes	Yes	Yes	Yes	Yes
TDP-67	nEFHWRInQ↓GKKAP	No	No	No	No	No	No
TDP-101	RKWQ↓SPQVDLYDKS	No	No	No	No	No	No
Hrv14_2C-3A	SLETLFQ↓GPVYKDL	Yes	Yes	Yes	Yes	No	No
Hrv14_3C-3D	QYFVEKQ↓GQVIARH	Yes	Yes	Yes	No	No	No
Klassevirus_2B-3C	QAAGGRQ↓GVRDYND	No	No	No	No	No	No
Klassevirus_3C-3D	SNLATPQ↓SVLHFHP	No	No	No	No	No	No
CVB3_2A	MTNTGAF↓GQQSGAA	No	No	No	No	No	No
Hrv2_2A	RPIITTA↓GPSDMYV	No	No	No	No	No	No

Supplementary Table 3: Complete list of proteins identified in conditioned media from a mouse pancreatic ductal adenocarcinoma. Proteomic analysis was performed on conditioned media from three biological samples. Proteins observed in 2 or more replicates are reported. Proteins are ordered by relative abundance as estimated from spectral counting. Proteolytic enzymes and inhibitors and highlighted in green and red, respectively. Raw data files are available from ProteomeCommons.org Tranche using the following hash:
wM5aFPz1/VVUCSdmdZMX/MAPB2abMc0ggtXHzwZc+6ZQG/79fVnOtSgt8PsY8daOdU346QzNgpOlluScmJACXO/IBQAAAAAACrQ==

Acc #	Protein Name	Protein MW	Replicates Observed	Unique Peptides	Peptide Count	Percent Coverage	Best Database Score	Best Expect Value
P10493	Nidogen-1	136538.9	3	31	77	36.1	9.27	1.50E-16
P04186	Complement factor B	85005.3	3	31	97	50.3	10.06	5.00E-18
Q61468	Mesothelin	69423.9	3	26	130	44.3	8.59	2.70E-15
Q06890	Clusterin	51656.1	3	28	181	44	9.93	8.80E-18
Q92111	Serotransferrin	76724.6	3	28	90	47.5	7.85	6.30E-14
Q61508	Extracellular matrix protein 1	62832.8	3	26	75	62.4	11.4	1.60E-20
Q07797	Galectin-3-binding protein	64491.7	3	20	85	45.6	10.41	1.10E-18
Q61398	Procollagen C-endopeptidase enhancer 1	50168.5	3	20	72	59.4	9.55	4.40E-17
Q61147	Ceruloplasmin	121152.4	3	22	58	29.5	8.57	2.90E-15
Q02819	Nucleobindin-1	53409.1	3	23	70	56.9	7.65	1.50E-13
P09803	Cadherin-1	98256.7	3	23	77	29.6	12.44	2.00E-22
Q8BPP5	EGF-containing fibulin-like extracellular matrix protein 1	54953.4	3	15	44	45.6	15.72	1.70E-28
Q5JCT0	Core2-GlcNAc-transferase 3	50698.9	3	17	51	45.5	8.28	1.00E-14
P35951	Low-density lipoprotein receptor	95135.6	3	15	41	28	9.7	2.30E-17
P12023	Amyloid beta A4 protein	86722.8	3	17	36	27.1	8.52	3.60E-15
Q00493	Carboxypeptidase E	53256.3	3	16	51	50.4	9.58	3.90E-17
P52480	Pyruvate kinase isozymes M1/M2	57845.4	3	15	29	43.5	11.8	3.00E-21
P10923	Osteopontin	32459.7	3	13	62	54.4	10.42	1.10E-18
P22777	Plasminogen activator inhibitor 1	45170.6	3	17	43	48.3	7.69	1.20E-13
P60710	Actin, cytoplasmic 1	41737.1	3	14	34	54.7	8.04	2.70E-14
Q8BND5	Sulfhydryl oxidase 1	82786	3	12	29	22.1	9.45	6.80E-17
P70275	Semaphorin-3E	89544	3	14	28	24.4	10.16	3.40E-18
P07214	SPARC	34450.4	3	11	50	56	9.16	2.30E-16
Q61581	Insulin-like growth factor-binding protein 7	28969.7	3	11	42	56.2	8.77	1.20E-15
Q62181	Semaphorin-3C	85289.8	3	13	23	24.1	9.03	4.10E-16
P10605	Cathepsin B	37280.2	3	12	39	38.3	8.62	2.40E-15
P17742	Peptidyl-prolyl cis-trans isomerase A	17971.5	3	11	38	58.5	10.05	5.40E-18
Q05793	Perlecan	398297.7	3	11	26	4	8.32	8.30E-15
Q62165	Dystroglycan	96905.5	3	10	26	19.3	8.36	7.20E-15
Q61282	Aggrecan core protein	221942.7	3	10	25	9	8.74	1.40E-15
Q9DAU7	WAP four-disulfide core domain protein 2	18031.8	3	8	44	39.1	12.17	6.20E-22
P47879	Insulin-like growth factor-binding protein 4	27807.2	3	9	40	40.9	10.36	1.40E-18
P21460	Cystatin-C	15531	3	8	38	55	8.7	1.60E-15
P25785	Metalloproteinase inhibitor 2	24328.4	3	11	23	50	7.72	1.10E-13
P63101	14-3-3 protein zeta/delta	27771.4	3	10	25	47.3	7.92	4.60E-14
P01887	Beta-2-microglobulin	13823.1	3	7	21	74.8	11.94	1.60E-21
P07356	Annexin A2	38676.5	3	9	25	29.8	6.42	2.80E-11
P06151	L-lactate dehydrogenase A chain	36498.9	3	9	29	31	6.76	6.50E-12
O88968	Transcobalamin-2	47586.2	2	10	14	29.8	9.07	3.40E-16
Q62266	Cornifin-A	15765.4	3	8	18	71.5	7.38	4.60E-13
Q61001	Laminin subunit alpha-5	404057.7	3	7	10	3	8.59	2.70E-15
Q9JJH1	Ribonuclease 4	17024.8	3	7	22	51.4	7.81	7.60E-14
Q91XA2	Golgi membrane protein 1	44310.6	3	8	26	24.4	6.98	2.60E-12
Q08423	Trefoil factor 1	9670.2	3	5	52	55.2	8.36	7.00E-15
Q60997	Deleted in malignant brain tumors 1 protein	226817.4	3	5	9	5.3	10.35	1.50E-18
P10853	Histone H2B type 1-F/J/L	13936.3	3	7	34	38.1	7.27	7.40E-13
P09411	Phosphoglycerate kinase 1	44550.9	3	6	11	19.2	7.85	6.20E-14
Q03404	Trefoil factor 2	14172.4	3	4	14	69	10.76	2.60E-19
P62806	Histone H4	11367.4	3	6	15	52.4	6.75	6.70E-12
P26040	Ezrin	69407.4	3	7	15	15	6.3	1.80E-11
P02468	Laminin subunit gamma-1	177300.2	2	5	8	3.5	10.05	3.20E-18
P08228	Superoxide dismutase [Cu-Zn]	15942.8	3	5	11	66.2	7.45	3.40E-13
P17182	Alpha-enolase	47141.2	3	5	13	17.3	8.17	1.60E-14
Q61937	Nucleophosmin	32560.3	3	5	17	27.4	9.5	5.40E-17
P62962	Profilin-1	14957.4	3	5	9	35	6.83	4.80E-12
P24369	Peptidyl-prolyl cis-trans isomerase B	23713.7	3	6	12	31.9	6.4	3.00E-11
Q00724	Retinol-binding protein 4	23206.3	3	5	11	29.4	9.21	1.90E-16
Q00993	Tyrosine-protein kinase receptor UFO	98191.8	3	5	10	8	5.46	1.70E-09
P10648	Glutathione S-transferase A2	25533.2	3	5	10	25.2	7.69	1.20E-13
Q3TNNW5	Diffuse panbronchiolitis critical region protein 1 homolog	54310.2	3	4	21	12.6	6.71	8.10E-12
O88593	Peptidoglycan recognition protein 1	20489.4	3	3	5	39	8.34	7.80E-15
P08226	Apolipoprotein E	35866.9	3	6	10	21.2	5.39	2.30E-09

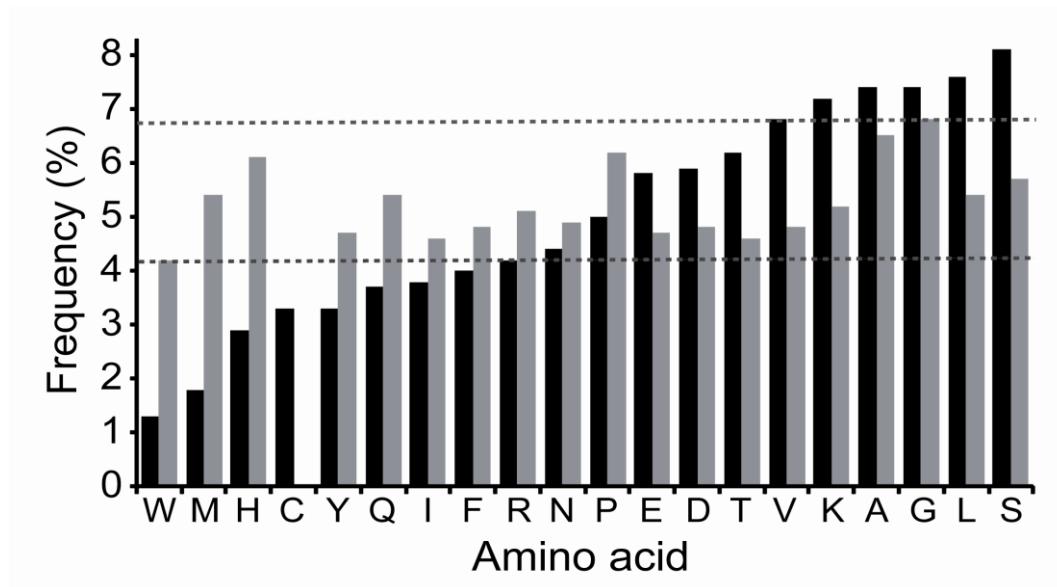
P63017	Heat shock cognate 71 kDa protein	70871.6	3	5	11	9.3	6.7	8.60E-12
Q9R098	Hepatocyte growth factor activator	70568.5	3	4	5	9.3	10.71	3.20E-19
P35700	Peroxisiredoxin-1	22176.7	3	6	9	27.6	4.28	1.70E-07
P70269	Cathepsin E	42938	3	5	11	9.6	6.03	1.50E-10
P15116	Cadherin-2	99797.2	3	4	9	6.8	7.17	1.10E-12
P70296	Phosphatidylethanolamine-binding protein 1	20830.6	2	3	4	34.8	8.22	1.30E-14
Q9Z1P8	Angiotensinogen-related protein 4	45538.7	3	3	9	13.7	8.57	2.90E-15
Q64695	Endothelial protein C receptor	27212.3	3	3	12	23.6	7.85	6.20E-14
P20152	Vimentin	53688.1	3	4	7	9	5.93	2.30E-10
P06797	Cathepsin L1	37547.7	3	4	9	15.3	7.26	4.90E-13
Q01768	Nucleoside diphosphate kinase B	17363.2	3	3	5	36.2	6.39	3.10E-11
P30412	Peptidyl-prolyl cis-trans isomerase C	22794.4	3	4	9	33	6.15	8.80E-11
P05064	Fructose-bisphosphate aldolase A	39356.3	3	4	7	12.4	5.23	4.50E-09
Q9R069	Basal cell adhesion molecule	67670.7	3	3	7	7.2	7.97	3.70E-14
P18760	Cofilin-1	18559.7	3	3	7	31.9	6.54	1.70E-11
P10810	Monocyte differentiation antigen CD14	39204.3	3	3	6	11.7	7.36	5.10E-13
P10126	Elongation factor 1-alpha 1	50114.3	2	3	5	16.2	7.55	2.20E-13
P97467	Peptidyl-glycine alpha-amidating monooxygenase	108963.5	3	3	5	7.3	7.25	8.20E-13
P50228	C-X-C motif chemokine 5	14190.1	2	4	6	33.3	5.09	3.00E-09
Q9EPL2	Calsyntenin-1	108900.6	2	4	5	5.4	4.5	6.00E-09
Q00731	Vascular endothelial growth factor A	25283.6	3	3	6	27.6	5.98	1.80E-10
P22752	Histone H2A type 1	14135.6	3	4	27	35.4	6.15	8.80E-11
Q9ESY9	Gamma-interferon-inducible lysosomal thiol reductase	27812.7	3	3	7	21	8.06	2.60E-14
Q62179	Semaphorin-4B	91393.1	3	4	9	7.2	5.15	6.30E-09
P11499	Heat shock protein HSP 90-beta	83281.9	3	4	7	6.1	6.24	6.00E-11
P50543	Protein S100-A11	11082.8	3	3	12	37.8	6.3	4.60E-11
Q8BMS2	Spondin-2	35965	2	3	4	12.4	6.64	1.10E-11
Q9R1Q9	V-type proton ATPase subunit S1	51008	2	3	4	13.4	7.06	1.80E-12
P15864	Histone H1.2	21266.8	3	3	7	11.8	5.49	1.50E-09
Q9DBJ1	Phosphoglycerate mutase 1	28832.2	2	3	4	16.9	5.98	9.60E-11
P63028	Translationally-controlled tumor protein	19462.4	3	3	6	21.5	6.25	5.70E-11
P16045	Galectin-1	14866	3	3	7	28.9	5.4	2.20E-09
Q9EQU5	Protein SET	33378	3	2	4	15.9	7.1	1.50E-12
P26350	Prothymosin alpha	12254.2	3	2	8	26.1	6.68	9.20E-12
P21107	Tropomyosin alpha-3 chain	32863.1	3	3	9	14.4	6.09	1.10E-10
Q00623	Apolipoprotein A-I	30587.7	3	3	5	12.5	4.74	7.10E-09
Q35639	Annexin A3	36384.4	3	3	5	10.5	5.44	1.80E-09
Q9CQ01	Ribonuclease T2	29609.1	2	2	3	14.3	6.08	2.00E-11
P02463	Collagen alpha-1(IV) chain	160680.5	2	2	3	1.4	6.62	1.20E-11
Q8R3G9	Tetraspanin-8	25582.5	3	2	7	10.2	5.9	2.60E-10
P15379	CD44 antigen	85617.8	3	2	7	2.7	5.62	8.30E-10
P01326	Insulin-2	12364.5	2	2	3	17.3	7.11	1.50E-12
P07724	Serum albumin	68693.2	3	2	4	4.4	5.49	1.50E-09
P40240	CD9 antigen	25258.2	3	1	5	11.1	10.54	6.60E-19
A2ASQ1	Agrin	207540.8	2	2	2	1.4	6.54	1.70E-11
P14069	Protein S100-A6	10050.7	3	2	22	29.2	7.67	1.30E-13
P62204	Calmodulin	16837.7	3	2	4	20.8	5.35	2.70E-09
P35762	CD81 antigen	25828.8	3	1	3	9.7	9.91	9.60E-18
Q9QZF2	Glypican-1	61360.3	2	2	3	6.3	5.59	1.30E-10
Q9CYN9	Renin receptor	39092.3	3	2	6	8	7.44	3.60E-13
P68433	Histone H3.1	15404.2	3	3	5	16.9	4.24	3.00E-07
Q9D7Z6	Calcium-activated chloride channel regulator 1	100072.1	2	2	3	2.3	5.01	1.10E-09
Q99JZ0	Syntenin-2	31564.9	2	3	3	10.6	3.54	2.10E-06
P34884	Macrophage migration inhibitory factor	12504.4	2	2	4	18.3	5.35	2.70E-09
Q80XD8	Proline-rich acidic protein 1	16798.1	3	1	3	23.5	8.7	1.70E-15
Q9DCK3	Tetraspanin-4	26053.8	2	1	2	10.5	8.08	2.30E-14
Q99JR5	Tubulointerstitial nephritis antigen-like	52665.1	3	2	4	4.7	4.3	1.10E-07
Q9Z0K8	Pantetheinase	57091.9	2	2	4	3.7	5.33	3.20E-10
P0CG49	Polyubiquitin-B	34368.8	2	2	5	7.2	5.27	2.10E-09
P48678	Prelamin-A/C	74238.2	3	2	4	3.3	4.19	3.70E-07
P57780	Alpha-actinin-4	104978	2	2	3	3.2	3.87	1.60E-08
P15535	Beta-1,4-galactosyltransferase 1	44411.5	3	2	5	5	4.36	7.30E-09
P52793	Ephrin-A1	23802.4	3	3	5	15.1	2.87	2.50E-07
Q9CR36	Gastroskin-1	21887.1	3	2	4	15.9	4.76	2.00E-09
P29533	Vascular cell adhesion protein 1	81318.2	2	1	2	2	6.43	2.70E-11
P35456	Urokinase plasminogen activator surface receptor	35428.7	3	1	5	4	6.22	6.40E-11
P32261	Antithrombin-III	52004	3	2	5	4.5	3.26	1.10E-07
P07091	Protein S100-A4	11721.5	3	1	3	17.8	6.13	9.60E-11
O08543	Ephrin-A5	26339.3	2	1	2	7.5	5.93	2.30E-10
P09405	Nucleolin	76723.7	3	2	4	3.4	3.43	1.90E-06
Q35887	Calumenin	37064	2	1	2	4.1	5.74	1.70E-10
Q6ZWY8	Thymosin beta-10	5025.7	3	1	3	34.1	5.7	5.90E-11
Q35660	Glutathione S-transferase Mu 6	25605.6	2	1	2	7.8	5.67	2.30E-10
P62259	14-3-3 protein epsilon	29174.2	2	1	3	9.4	5.32	2.90E-10

Q3UR85	Myelin gene regulatory factor	123288.7	2	1	2	1.9	5.18	3.00E-10
Q61207	Sulfated glycoprotein 1	61422.8	3	2	4	3.8	3.61	3.00E-07
P17751	Triosephosphate isomerase	32192.1	2	2	2	7	2.65	5.40E-07
P27773	Protein disulfide-isomerase A3	56678.8	2	1	2	3.2	5.02	1.10E-08
O89093	C-C motif chemokine 20	10826	2	2	4	15.5	2.48	5.50E-04
O09131	Glutathione S-transferase omega-1	27497.9	3	1	3	4.2	4.94	1.50E-08
O08992	Syntenin-1	32379.7	2	1	2	4.7	4.89	1.90E-08
P52795	Ephrin-B1	37859.7	2	1	2	3.2	4.72	3.30E-08
Q9Z0J0	Epididymal secretory protein E1	16442.2	2	1	2	10.7	4.43	2.00E-09
O35188	Fractalkine	42099	3	1	6	2	4.02	7.80E-07
P16056	Hepatocyte growth factor receptor	153550	3	1	3	0.8	3.9	7.10E-08
P47955	60S acidic ribosomal protein P1	11475	2	1	2	14	3.8	1.30E-07
P55292	Desmocollin-2	99962.3	3	1	3	0.9	3.79	1.30E-08
P37889	Fibulin-2	131819.9	2	1	2	1.1	3.61	3.60E-09
Q60605	Myosin light polypeptide 6	16930.2	2	1	2	10.6	3.23	1.60E-07
P15119	Mast cell protease 2	26732.1	2	1	2	4.1	3.03	1.60E-06
P01325	Insulin-1	12160.3	2	1	22	6.5	2.15	6.80E-06
P70168	Importin subunit beta-1	97185	2	1	2	1	2.11	3.50E-05
P29699	Alpha-2-HS-glycoprotein	37326	2	1	2	2	1.42	3.00E-05

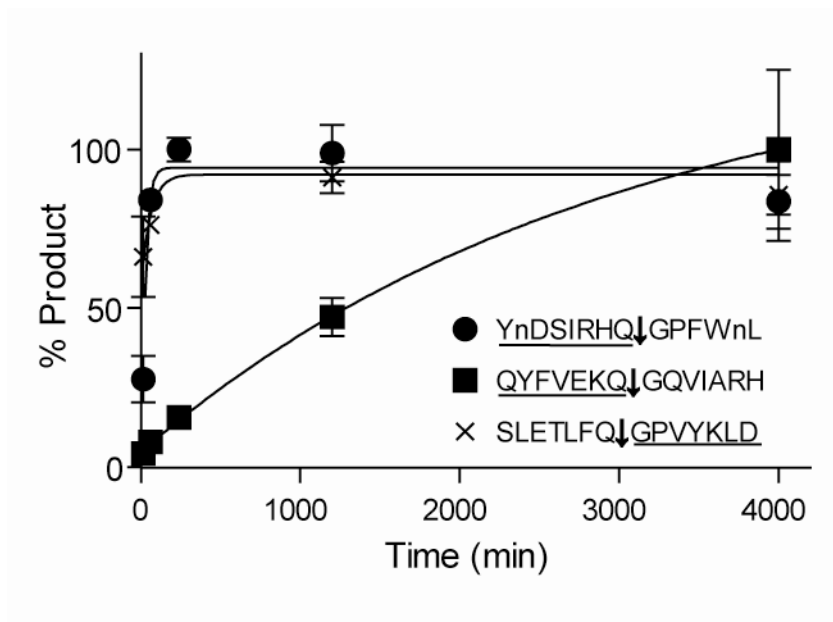
Supplementary Table 4: Complete list of **F* motifs in the MSP-MS library. The first observed time of motif cleavage by cathepsin E is listed. “Never” indicates that cleavage was not observed by 1200 minutes.

**F↓* Sequences	Cleavage first observed (minutes)	**F↓* Sequences	Cleavage first observed (minutes)
nYFK	Never	DH FY	30
HPFK	Never	IVFI	30
GSFK	Never	TnFI	30
AnFR	Never	WAFR	15
IFFR	Never	nQFN	15
TAFR	Never	EQFT	15
TYFH	Never	WKFG	15
nEFH	Never	YEFG	15
QRFH	Never	VIFF	15
AHFH	Never	LFFW	15
ELFD	Never	IHFK	5
ETFD	Never	HNFH	5
FKFD	Never	PSFN	5
SRFD	Never	VNFQ	5
NIFE	Never	AWFS	5
VGFE	Never	GRFG	5
HGFE	Never	RLFF	5
KIFN	Never	GPFW	5
HLFN	Never	YDFW	5
HTFQ	Never	PDFY	5
SKFQ	Never	GIFY	5
YHFQ	Never	SAFA	5
QVFS	Never	nEFA	5
KQFT	Never	LIFV	5
EnFG	Never	PnFV	5
DnFP	Never	LYFI	5
ENFP	Never	VnFL	5
FQFP	Never	ANFL	5
PKFA	Never	nSFn	5
nTFI	Never	KRFn	5
PAFn	Never	VnFn	5
NDFT	1200	nAFn	5
KSFG	1200	KVFn	5
IWFD	240		
WHFS	240		
HHFT	60		
GPFH	30		

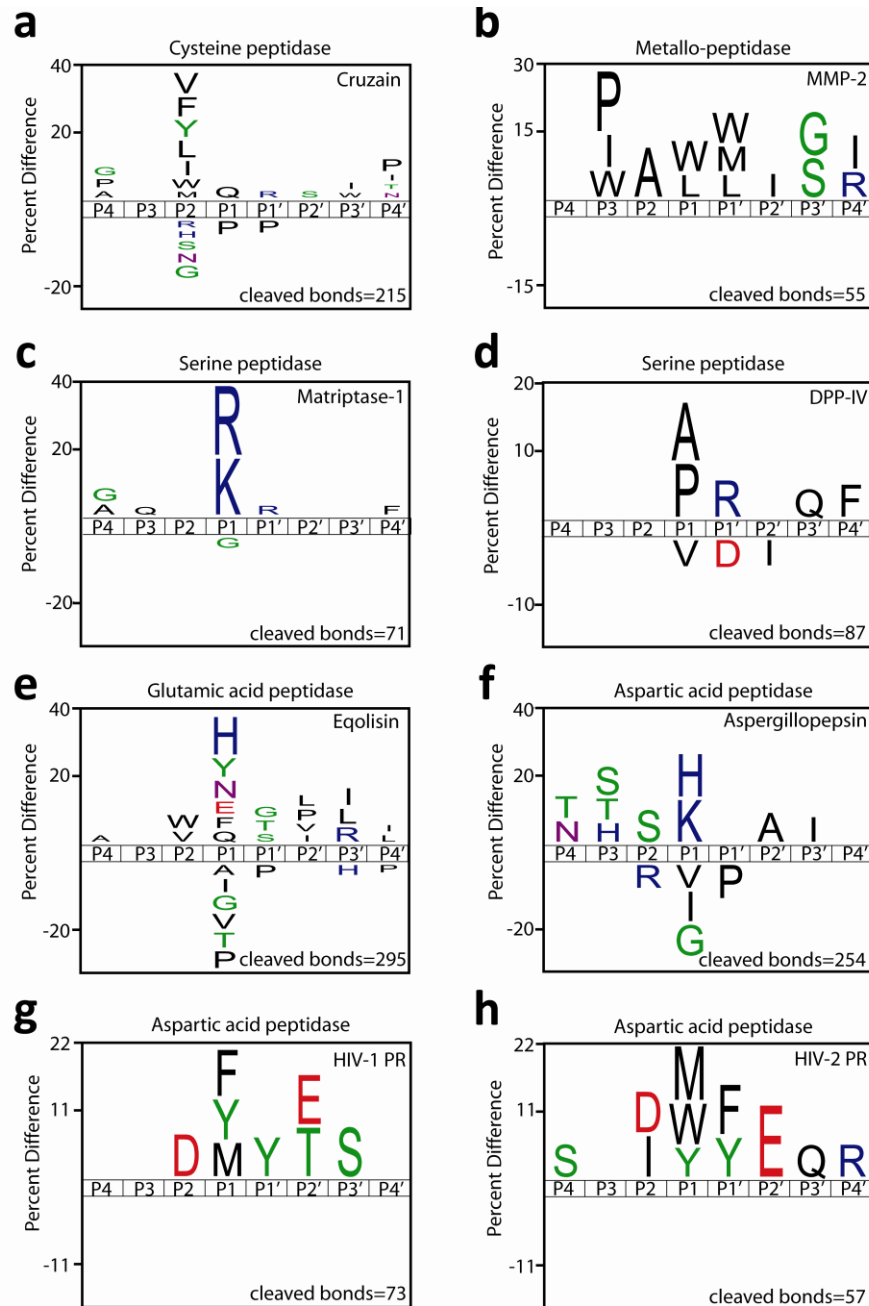
SUPPLEMENTARY FIGURES



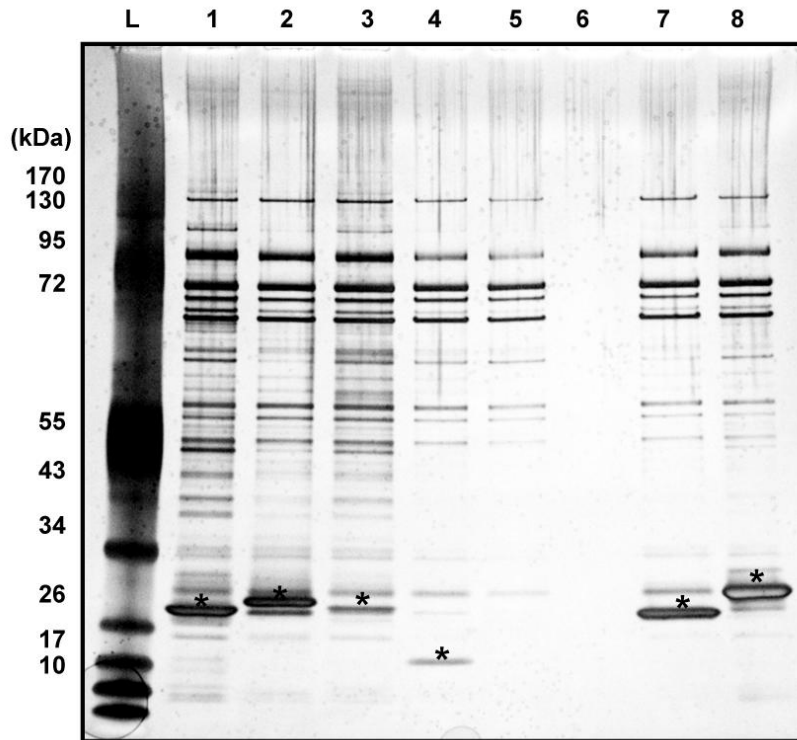
Supplementary Figure 1: Comparison of amino acid frequency between the synthetic tetradecapeptide library and vertebrate proteome. The vertebrate proteome data (black bars) was generated from SwissProt and grey bars represent the amino acid frequency of the 124-member tetradecapeptide library. The upper and lower frequency of amino acids in the tetradecapeptide library are illustrated with a broken line.



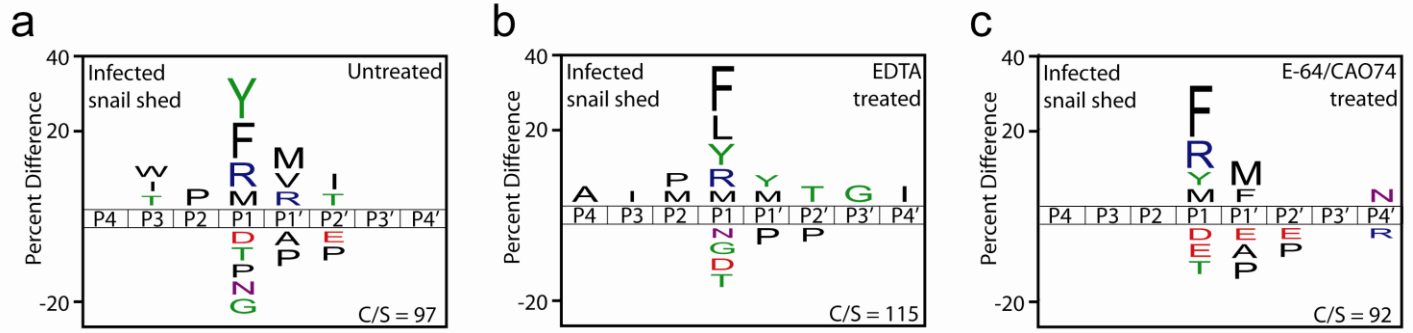
Supplementary Figure 2: Progress curve of three HRV 3C substrates in a MSP-MS assay containing eight peptides. No cleavage was observed for five of the peptides (see Supplementary Table 2). This assay was performed in triplicate and the fragments that were used for quantitation are underlined.



Supplementary Figure 3: Substrate profiles of representative exo- and endo-peptidases from multiple families. IceLogos were generated for amino acids in the P4 to P4' position that are enriched or de-enriched after 1200 minutes incubation of each peptidase with peptide library. **(a)** Cruzain from *Trypanosoma cruzi* showed preference for hydrophobic residues in the P2 position **(b)** human MMP-2 preferred Pro at P3, Ala at P2 and Gly or Ser at P3' **(c)** human matriptase-1 specificity was dominated by Arg and Lys at P1. **(d)** human DPP-IV had a preference of Pro and Ala at P1, **(e)** the fungal glutamic acid peptidase eqolisin preferred bulky residues at P1 and small residues at P1', **(f)** the fungal aspartic acid protease, aspergillopepsin had broad specificity but preferred positively charged residues at P1 and **(g-h)** HIV-1 and HIV-2 protease had similar specificity with a preference for bulky residues at P1 and P1' and Glu at P2'. All cleavage sites used to generate iceLogos are listed in the Supplementary "MSP-MS cleavages" file.



Supplemental Figure 4: Silver stain gel showing partially purified viral 3C proteases. IVTT protease preparations were separated by SDS-PAGE on 4-20% acrylamide Bis-TRIS gels run in MOPS-SDS buffer, then silver stained for imaging. Lanes are: 1) HRV14 3C; 2) Poliovirus 3C; 3) Klassevirus 3C; 4) Poliovirus 2A; 5) Klassevirus 2A, 6) Empty, 7) Enterovirus 71 3C and 8) Hepatitis A Virus 3C. Protease bands are indicated with an asterisk (*). Klassevirus 2A failed to express in the IVTT system. Stock enzyme concentrations for each sample were estimated as 7.5 μ M HRV14 3C, 7.5 μ M Poliovirus 3C, 1 μ M Klassevirus 3C, 0.7 μ M Poliovirus 2A, 11 μ M Enterovirus and 14 μ M Hepatitis A Virus 3C.



Supplementary Figure 5: Comparison of untreated *Schistosoma mansoni* infected fresh water snails with EDTA and E-64/CAO74 treated samples. Substrate signature illustrates all cleavage sites observed within 240 minutes of incubation with the peptide library.