Supplemental Data

Global Sequencing of Proteolytic

Cleavage Sites in Apoptosis by

Specific Labeling of Protein N Termini

Sami Mahrus, Jonathan C. Trinidad, David T. Barkan, Andrej Sali, Alma L. Burlingame, and James A. Wells

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Expression and purification of subtiligase. The expression construct for subtiligase was prepared using the *B. subtilis*/*E. coli* shuttle vector pBS42 (ATCC) (Wells et al., 1983). The variant of subtiligase used corresponds to subtilisin BPN' containing point mutations S221C, P225A, M124L, and S125A for ligase activity (Abrahmsén et al., 1991; Atwell and Wells, 1999), and point mutations M50F, N76D, N109S, K213R, AND N218S for increased stability (Chang et al., 1994). Recombinant subtiligase was prepared in *B. subtilis* strain 168 (ATCC). Subtiligase expression and purification was carried out essentially as described (Abrahmsén et al., 1991). The purified enzyme was stored at -80°C in 100 mM BICINE, pH 8.0 and 10 mM DTT or TCEP.

Synthesis of peptide ester substrate. The biotinylated peptide glycolate ester substrate for subtiligase was prepared by solid-phase peptide synthesis using Fmoc chemistry as previously described (Braisted et al., 1997). The peptide was purified using 10 x 50 mm XTerra Prep MS C_{18} ODB colums (Waters) on a Parallex Flex HPLC system (Biotage). Purity and identity were verified by LC/MS analysis using a 4.6 x 50 mm XTerra MS C_{18} column on a 2795 HPLC (Waters) system equipped with a ZQ quadrupole MS detector (Waters).

Cell culture, induction of apoptosis, and cell lysate preparation. Jurkat clone E6-1 (ATCC) cells were grown in RPMI-1640 supplemented with 10% fetal bovine serum, sodium pyruvate, and antibiotics. Cells at a density of 1 x 10^6 cells/ml were treated with etoposide (50 μM) for 0 or 12 hours prior to harvesting. Harvested cells (0.1 to 1 billion) were pelleted at 2,000 x g and 25°C for 5 minutes, washed twice with phosphate buffered saline, and lysed at a typical concentration of 2 x 10^8 cells/ml in 1.0% Triton X-100, 100 mM BICINE pH 8.0, 100 μM Z-VAD-FMK, 100 μM E-64, 1 mM PMSF, 1 mM AEBSF, and 5 mM EDTA. Cell lysates were incubated at room temperature for 1 hour to allow complete inhibition of endogenous protease and esterase activity, and centrifuged at 21,000 x g and 4°C for 15 minutes to pellet insoluble material. Clarified supernatant was immediately used in ligation reactions at a protein concentration of approximately 20 mg/ml, as determined by Bradford assay (Bio-Rad).

Sample biotinylation, denaturation, reduction, alkylation, and gel filtration. Subtiligase (1 μM), biotinylated peptide ester (1 mM), and DTT (2 mM) were added to either control or apoptotic cell lysate. Peptide ester was added from a 10x stock for a final DMSO concentration of 10%. Ligation reactions were typically left to proceed at room temperature for 60 minutes. Samples were then denatured by direct addition of solid

guanidine hydrochloride to a final concentration of 6 M, reduced by addition of neutralized TCEP (2 mM), heated at 95°C for 15 minutes, cooled to room temperature, and alkylated by addition of iodoacetamide (6 mM) and incubation at room temperature in the dark for 1 hour. The alkylation reaction was quenched by addition of DTT (10 mM), the sample was passed through a 0.8 µm filter, and subjected to gel filtration chromatography using a Superdex 30 16/60 column (GE Healthcare) on an ÄKTA FPLC system (GE Healthcare). The mobile phase was 100 mM BICINE pH 8.0, 200 mM NaCl, and 1 M guanidine hydrochloride. Fractions containing protein (corresponding to polypeptides > 5 kDa) were collected and pooled for a final volume of approximately 30 ml.

Trypsinization and recovery of biotinylated N-terminal peptides. The gel-filtered material was supplemented with $CaCl_2$ (20 mM) and digested with sequencing grade modified trypsin (100 µg, Promega) by overnight incubation at 37°C. Trypsinized samples were clarified by centrifugation, supplemented with benzamidine (500 mM), and NeutrAvidin agarose (250 µl bed volume, Pierce) was added for overnight affinity capture of biotinylated N-terminal peptides. NeutrAvidin agarose resin was pelleted and washed with 100 mM BICINE pH 8.0 and AEBSF (1 mM), 100 mM BICINE pH 8.0, 5 M NaCl, and again with a few washes of 100 mM BICINE pH 8.0. More stringent washes using either 1 M or 5 M guanidine hydrochloride were used in some cases. Captured peptides were released from NeutrAvidin agarose resin by overnight treatment with TEV protease (1 µM) (Kapust et al., 2001) in 100 mM BICINE pH 8.0 and DTT (1 mM). Recovered peptides were concentrated and desalted using $ZipTip_{C18}$ pipette tips, or a C_{18} Macrotrap (Michrom) trap column on a 2796 HPLC system (Waters). TEV protease was sometimes depleted from samples using an SCX Macrotrap (Michrom) trap column.

Sample fractionation using strong cation exchange (SCX) chromatography. For larger scale experiments, samples were fractionated by SCX chromatography prior to LC/MS/MS analysis using a 2.1 x 200 mm PolySULFOETHYL Aspartamide column (The Nest Group) at a flow rate of 0.3 ml/min on a 2796 HPLC system (Waters). Buffer A consisted of 25 mM ammonium formate pH 2.8 and 30% acetonitrile, and buffer B consisted of 500 mM ammonium formate pH 2.8 and 30% acetonitrile. Approximately 20 fractions were collected during a 60 minute gradient of 0% to 75% buffer B. Solvent from fractions was removed using an EZ-2 Plus evaporator (GeneVac), and remaining ammonium formate salt was removed by lyophilization. Some samples were also fractionated using a phosphate buffer and KCl salt system, in which case each fraction was subjected to automated desalting using a C_{18} Microtrap (Michrom) trap column on a 2796 HPLC system (Waters) before solvent removal.

Nano-LC-ESI-Qq-TOF MS/MS. Desalted fractionated or unfractionated samples were separated using a 75 µm x 15 cm C_{18} column (LC Packings) at a flow rate of 350 nl/min, with a 60 minute gradient of 3 to 30% acetonitrile in 0.1% formic acid, on a 1100 series HPLC system (Agilent). The LC eluent was coupled to a microion spray source attached to a QSTAR Pulsar, QSTAR XL, or QSTAR Elite mass spectrometer (Applied Biosystems). Peptides were analyzed in positive ion mode. MS spectra were acquired for 1 s. For each MS spectrum, either the single most intense or the two most intense multiply charged peaks were selected for generation of subsequent CID mass spectra, depending on the analysis method used. The CID collision energy was automatically adjusted based upon peptide charge and *m*/*z* ratio. A dynamic exclusion window was applied that prevented the same *m*/*z* from being selected for 3 min after its initial

acquisition. Representative CID spectra are included as **Figures S7-S14**. Additional CID spectra will be made available upon request.

Interpretation of MS/MS spectra. Data were analyzed using Analyst QS software (version 1.1), and MS/MS centroid peak lists were generated using the Mascot.dll script (version 1.6b16). Data were searched against the Swiss-Prot human database (March 2008 release) using Protein Prospector 5.0 (University of California, San Francisco). Initial peptide tolerances in MS and MS/MS modes were 200 ppm and 300 ppm, respectively. The digest protease specified was trypsin, allowing for non-specific cleavage at N-termini in searches for N-terminally labeled semitryptic peptides, and trypsin allowing for non-specific cleavage at 0 N-termini in searches for unlabeled fully tryptic peptide contaminants. Two missed cleavages were typically allowed in searches. An N-terminal SY modification was specified as a fixed modification in searches for N-terminal peptides, but not in searches for unlabeled peptides. Cysteine carbamidomethylation was specified as a fixed modification and methionine oxidation was specified as a variable modification in all searches. High scoring peptide identifications from individual LC/MS/MS runs were then used to internally recalibrate MS parent ion *m*/*z* values within each run. Recalibrated data files were then searched again with an MS peptide tolerance of 100 ppm. Peptides with scores $\geq$ 22 and expectation values $\leq$ 0.05 were considered positively identified. False discovery rates for peptide identifications were estimated by conducting searches using a concatenated database containing the original Swiss-Prot human database, as well as a version of each original database entry where the sequence had been randomized.

Cleavage site predictions. Cleavage site prediction was assessed using 1000 jackknife trials on 473 substrates containing 603 cleavage sequences from both our caspase substrate dataset and the literature substrate dataset (Lüthi and Martin, 2007). A test set consisted of 60 randomly selected true positive cleavage sequences and 3,000 randomly selected true negative peptides derived from all octapeptides in caspase substrates with aspartate at the fourth position that have not been shown to be cleaved by caspases. A training set consisted of all cleavage sequences from respective substrate sets not present in the corresponding test set. Hidden Markov models were constructed using the "hmmbuild" command of HMMer version 2.3.2 (Eddy, 1998) and test peptides were scored using the "hmmpfam" command.

Structural bioinformatics. Secondary structure analysis of cleavage sites was carried out on a set of experimentally determined structures from the Protein Data Bank (Berman et al., 2002) and “good quality” comparative models from ModBase (Pieper et al., 2006). A good quality model has either a N-DOPE score < -0.4 (Shen and Sali, 2006) or is based on $\geq$ 25% sequence identity to the template structure with a “model score” > 0.8 (Melo and Sali, 2007). Such models are likely to have $\geq$ 75% of their C_α atoms within 3.5 Å of the correct positions (Eswar and Sali, 2007). The DSSP algorithm was used to assign the type of secondary structure of each cleavage site, discriminating between α-helix, β-sheet, and loop states. The fraction of solvent accessible surface area of each residue in the cleavage sites was determined by dividing the observed exposed surface area, as also assessed by DSSP, by the maximum exposed surface area of the residue (Rose et al, 1984). Residues were considered exposed if this fraction was > 0.33. A reference control distribution of both solvent accessibility and secondary structure state was determined from the set of all octapeptides with aspartate at the fourth position from 15,787 experimentally determined structures with < 95% sequence identity to each other

(Berman et al., 2002). Domain analysis was performed using domain assignments from the Pfam database (July 2007 release) (Finn et al., 2006). The reference control for this analysis was the set of all octapeptides with aspartate at the fourth found in the human Swiss-Prot database. Statistical significance of differences between caspase cleavage sites and reference controls were assessed using the chi-square test. Molecular graphics were rendered using Pymol 1.0 (DeLano Scientific).

Protein-protein interaction analysis. Data for human and/or rodent binary protein interactions were merged from the HPRD (September 2007 release) (Mishra et al., 2006), IntAct (December 2007 release) (Kerrien et al., 2007), and MINT (December 2007 release) (Chatr-aryamontri et al., 2007) databases. The resulting merged database represented 55,679 unique binary interactions between 12,429 proteins. A network of protein-protein interactions among caspase substrates was constructed by identifying binary interactions in the merged database for which both interacting partners are known caspase substrates (**Table S9**). These included substrates identified in this work, previously reported human substrates, and human orthologs of previously reported rodent substrates (Lüthi and Martin, 2007), but did not include the caspases themselves. The average number of intradataset interactions per protein for caspase substrates was calculated as (total interactions between caspase substrates)/(total caspase substrates) = 1253/602 = 2.08. Ten control sets of 602 protein interactors were obtained by random samplings of proteins from the merged HRDP + IntAct + MINT protein interaction database. The average number of intradataset interactions per protein for each of these control sets was calculated, and the average and standard deviation of these ten values were determined to be 0.21 and 0.04, respectively. This analysis indicates an enrichment of approximately tenfold for intradataset interactions per protein for caspase substrates. Graphical representation of protein interaction networks was carried out using Cytoscape 2.6.0 (Shannon et al., 2003). Overrepresentation of Gene Ontology categories in protein interaction networks was assessed using the BiNGO 2.0 plugin for Cytoscape (Maere et al., 2005). A significance level of 0.05 was used in this analysis, determined by the hypergeometric test and Benjamini & Hochberg false discovery rate (FDR) multiple testing correction.

Immunoblotting. Jurkat cells at a density of 1 x 10^6 cells/ml were treated with etoposide (50 μM) for 0, 2, 4, 8, 12, and 24 hours prior to harvesting. In some cases, cells were treated with both etoposide (50 μM) and the cell-permeable pan-caspase inhibitor Z-VAD(OMe)-fmk (50 μM). Harvested cells were pelleted at 2,000 x g and 25°C for 5 minutes, washed twice with phosphate buffered saline, and lysed at a concentration of 2 x 10^7 cells/ml in 1.0% SDS, phosphate buffered saline, 100 μM Z-VAD-FMK, 100 μM E-64, 1 mM PMSF, 1 mM AEBSF, and 5 mM EDTA. Whole cell lysates were sonicated to shear genomic DNA and normalized to a protein concentration of approximately 2 mg/ml, as determined by BCA assay (Pierce). Cell lysates for each apoptotic timepoint were then analyzed by SDS-PAGE and Western blot. Blots were imaged using SuperSignal West Femto Substrate (Pierce) with a FluorChem SP imager (Alpha Innotech).

Antibodies. Rabbit polyclonal anti-CCTδ (#ab49151), anti-RCOR2 (#ab37113), and anti-LCOR (#ab48339) antibodies were purchased from Abcam. Mouse polyclonal anti-USP5 (#H00008078-A01) antibody was purchased from Abnova. Mouse monoclonal anti-DFF45 (#611036) and mouse polyclonal anti-Ku80 (#611360) antibodies were purchased from BD Transduction Laboratories. Rabbit polyclonal anti-TBLR1 (#A300-

408A), anti-SHARP (#A301-119A), and anti-RBBP7 (#A300-959A) antibodies were purchased from Bethyl Laboratories. Mouse monoclonal anti-caspase-3 (#9668) and rabbit polyclonal anti-HDAC3 (#2632) antibodies were purchased from Cell Signaling Technology. Rabbit polyclonal anti-SMRTe (#06-891) and anti-HDAC6 (#07-732) antibodies were purchased from Millipore. Rabbit polyclonal anti-USP36 antibody (#NB100-40831) was purchased from Novus. Goat polyclonal anti-N-Cor (#sc-1611) and rabbit polyclonal anti-HDAC7 (#sc-11412) antibodies were purchased from Santa Cruz Biotechnology.

SAMPLE	untreated Jurkat cells	
ANALYSIS	1D LC/MS/MS experiment #1	1D LC/MS/MS experiment #2
all N-terminal peptides		
matches to SY-labeled peptides in non-decoy database	102	83
matches to SY-labeled peptides in decoy database	0	0
false positive rate for identification of SY-labeled peptides (%)	0.00	0.00
peptides ambiguous w/ respect to SY label[1]	1	0
P1 Asp N-terminal peptides		
matches to SY-labeled peptides in non-decoy database	0	1
matches to SY-labeled peptides in decoy database	0	0
false positive rate for identification of SY-labeled peptides (%)	0.00	0.00
peptides ambiguous w/ respect to SY label[1]	0	0
total N-termini		
non-homologous N-termini	84	74
non-homologous proteins	80	71
homologous N-termini	2	5
homologous proteins	2	5
N-termini ambiguous w/ respect to SY label[1]		
N-termini ambiguous w/ respect to SY label	1	0
corresponding proteins	1	0
P1 Asp N-termini		
non-homologous N-termini	0	1
non-homologous substrates	0	1
homologous N-termini	0	0
homologous substrates	0	0
possibly P1 Asp N-termini[2]		
homologous N-termini	0	0
homologous substrates	0	0
background N-termini		
non-homologous N-termini	83	73
non-homologous proteins	79	70
homologous N-termini	2	5
homologous proteins	2	5

Table S1. Summary of mass spectrometry results for small-scale 1D LC/MS/MS experiments with untreated Jurkat cells. [1]Peptides and N-termini ambiguous with respect to SY label follow KSY, KYS, KHI, KIH, KHL, KLH, RSY, RYS, RHI, RIH, RHL, or RLH residues in corresponding protein sequences (cases where tryptic cleavage would result in a peptide whose first two amino acids are isobaric with the subtiligase-introduced SY modification). [2]Possibly caspase-derived N-termini are found in multiple proteins and follow an Asp residue in only some of the corresponding protein sequences.

SAMPLE	untreated Jurkat cells		
ANALYSIS	2D LC/MS/MS experiment #1	2D LC/MS/MS experiment #3	combined untreated dataset
all N-terminal peptides			
matches to SY-labeled peptides in non-decoy database	583	451	832
matches to SY-labeled peptides in decoy database	6	2	8
false positive rate for identification of SY-labeled peptides (%)	1.03	0.44	0.96
peptides ambiguous w/ respect to SY label[1]	1	3	4
P1 Asp N-terminal peptides			
matches to SY-labeled peptides in non-decoy database	18	10	24
matches to SY-labeled peptides in decoy database	0	0	0
false positive rate for identification of SY-labeled peptides (%)	0.00	0.00	0.00
peptides ambiguous w/ respect to SY label[1]	0	0	0
total N-termini			
non-homologous N-termini	502	392	703
non-homologous proteins	409	339	553
homologous N-termini	22	17	27
homologous proteins	21	16	25
N-termini ambiguous w/ respect to SY label[1]			
N-termini ambiguous w/ respect to SY label	1	3	4
corresponding proteins	1	3	4
P1 Asp N-termini			
non-homologous N-termini	18	10	24
non-homologous substrates	16	10	22
homologous N-termini	0	0	0
homologous substrates	0	0	0
possibly P1 Asp N-termini[2]			
homologous N-termini	0	0	0
homologous substrates	0	0	0
background N-termini			
non-homologous N-termini	483	380	676
non-homologous proteins	397	329	535
homologous N-termini	22	16	26
homologous proteins	21	16	25

Table S2. Summary of mass spectrometry results for large-scale 2D LC/MS/MS experiments with untreated Jurkat cells and for the combination of all data collected with untreated Jurkat cells. [1]Peptides and N-termini ambiguous with respect to SY label follow KSY, KYS, KHI, KIH, KHL, KLH, RSY, RYS, RHI, RIH, RHL, or RLH residues in corresponding protein sequences (cases where tryptic cleavage would result in a peptide whose first two amino acids are isobaric with the subtiligase-introduced SY modification). [2]Possibly caspase-derived N-termini are found in multiple proteins and follow an Asp residue in only some of the corresponding protein sequences.

SAMPLE	apoptotic (etoposide-treated) Jurkat cells			
ANALYSIS	1D LC/MS/MS experiment #1	1D LC/MS/MS experiment #2	1D LC/MS/MS experiment #3	1D LC/MS/MS experiment #4
all N-terminal peptides				
matches to SY-labeled peptides in non-decoy database	172	159	126	86
matches to SY-labeled peptides in decoy database	1	0	0	0
false positive rate for identification of SY-labeled peptides (%)	0.58	0.00	0.00	0.00
peptides ambiguous w/ respect to SY label[1]	3	1	1	0
P1 Asp N-terminal peptides				
matches to SY-labeled peptides in non-decoy database	78	70	44	37
matches to SY-labeled peptides in decoy database	0	0	0	0
false positive rate for identification of SY-labeled peptides (%)	0.00	0.00	0.00	0.00
peptides ambiguous w/ respect to SY label[1]	0	0	0	0
total N-termini				
non-homologous N-termini	141	138	107	72
non-homologous proteins	134	127	99	68
homologous N-termini	10	7	5	4
homologous proteins	10	7	5	4
N-termini ambiguous w/ respect to SY label[1]				
N-termini ambiguous w/ respect to SY label	3	1	1	0
corresponding proteins	3	1	1	0
P1 Asp N-termini				
non-homologous N-termini	67	62	35	32
non-homologous substrates	65	56	33	30
homologous N-termini	2	1	1	0
homologous substrates	2	1	1	0
possibly P1 Asp N-termini[2]				
homologous N-termini	1	1	1	0
homologous substrates	1	1	1	0
background N-termini				
non-homologous N-termini	73	75	71	40
non-homologous proteins	69	73	66	38
homologous N-termini	5	5	3	4
homologous proteins	5	5	3	4

Table S3. Summary of mass spectrometry results for small-scale 1D LC/MS/MS experiments with apoptotic Jurkat cells. [1]Peptides and N-termini ambiguous with respect to SY label follow KSY, KYS, KHI, KIH, KHL, KLH, RSY, RYS, RHI, RIH, RHL, or RLH residues in corresponding protein sequences (cases where tryptic cleavage would result in a peptide whose first two amino acids are isobaric with the subtiligase-introduced SY modification). [2]Possibly caspase-derived N-termini are found in multiple proteins and follow an Asp residue in only some of the corresponding protein sequences.

SAMPLE	apoptotic (etoposide-treated) Jurkat cells				
ANALYSIS	2D LC/MS/MS experiment #1	2D LC/MS/MS experiment #2	2D LC/MS/MS experiment #3	sum of all other 1D and 2D LC/MS/MS experiments	combined apoptotic dataset
DATASET	1	2	3	4	combined
all N-terminal peptides					
matches to SY-labeled peptides in non-decoy database	527	483	414	530	1099
matches to SY-labeled peptides in decoy database	0	4	7	4	23
false positive rate for identification of SY-labeled peptides (%)	0.00	0.91	1.69	0.75	2.09
peptides ambiguous w/ respect to SY label[1]	1	3	19	3	25
P1 Asp N-terminal peptides					
matches to SY-labeled peptides in non-decoy database	240	179	161	196	418
matches to SY-labeled peptides in decoy database	0	0	1	1	3
false positive rate for identification of SY-labeled peptides (%)	0.00	0.00	0.62	0.51	0.71
peptides ambiguous w/ respect to SY label[1]	0	0	0	0	0
total N-termini					
non-homologous N-termini	462	369	338	421	859
non-homologous proteins	393	317	292	361	659
homologous N-termini	23	11	18	18	42
homologous proteins	23	11	17	18	40
N-termini ambiguous w/ respect to SY label[1]					
N-termini ambiguous w/ respect to SY label	0	3	15	3	20
corresponding proteins	0	3	15	3	19
P1 Asp N-termini					
non-homologous N-termini	210	151	134	156	322
non-homologous substrates	194	139	125	142	282
homologous N-termini	8	4	2	4	11
homologous substrates	8	4	2	4	10
possibly P1 Asp N-termini[2]					
homologous N-termini	3	1	1	2	4
homologous substrates	3	1	1	2	4
background N-termini					
non-homologous N-termini	252	216	194	264	524
non-homologous proteins	225	189	172	235	420
homologous N-termini	12	5	10	10	20
homologous proteins	11	5	9	10	19

Table S4. Summary of mass spectrometry results for large-scale 2D LC/MS/MS experiments with apoptotic Jurkat cells, for the combination of all other 1D and 2D LC/MS/MS experiments with apoptotic Jurkat cells, and for the combination of all data collected with apoptotic Jurkat cells. [1]Peptides and N-termini ambiguous with respect to SY label follow KSY, KYS, KHI, KIH, KHL, KLH, RSY, RYS, RHI, RIH, RHL, or RLH residues in corresponding protein sequences (cases where tryptic cleavage would result in a peptide whose first two amino acids are isobaric with the subtiligase-introduced SY modification). [2]Possibly caspase-derived N-termini are found in multiple proteins and follow an Asp residue in only some of the corresponding protein sequences.

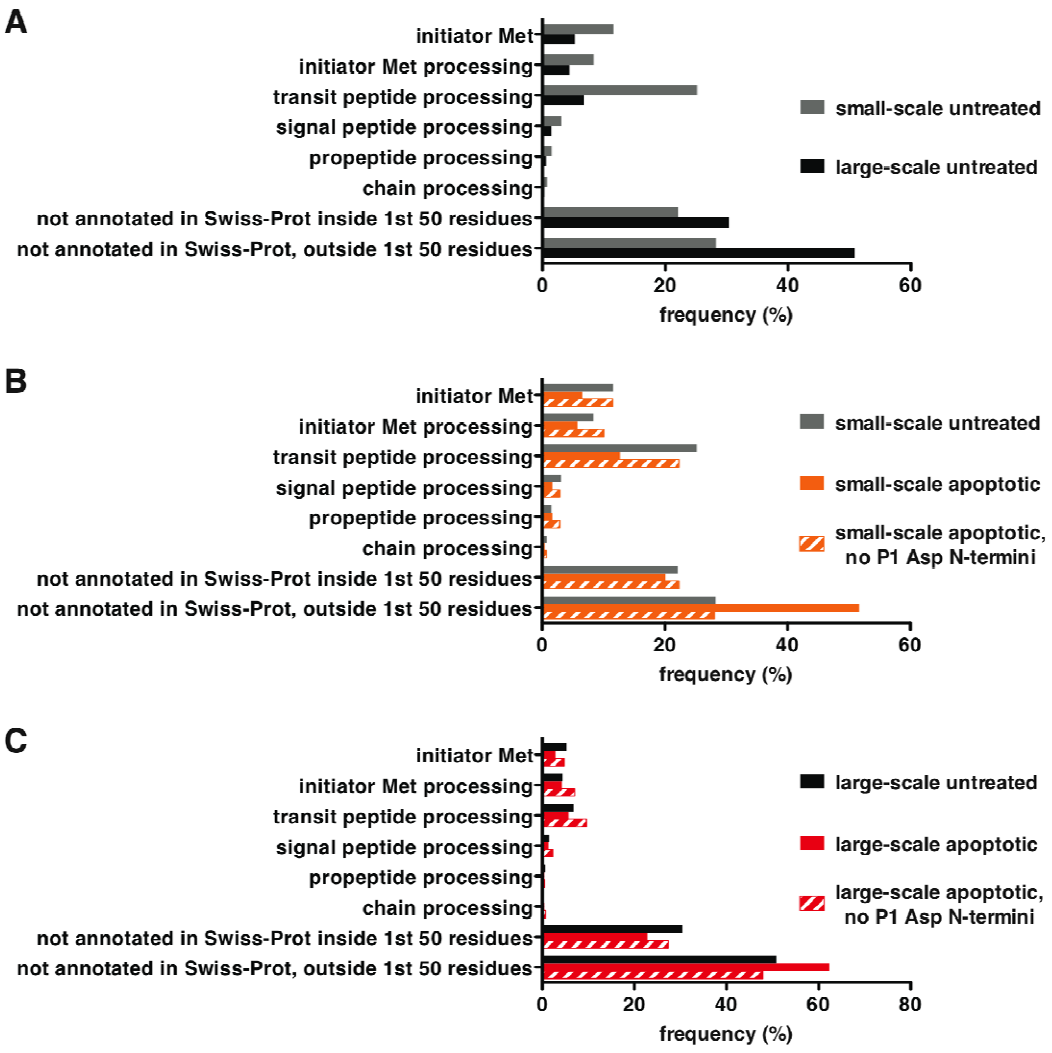

Figure S1. Classification of unique N-termini identified in untreated and apoptotic Jurkat cells according to Swiss-Prot annotation. (A) Classification of N-termini identified in small-scale and large-scale experiments with untreated cells (131 and 661 unique N-termini, respectively, combined from two experiments in both cases). (B) Classification N-termini identified in small-scale experiments with untreated cells (131 unique N-termini combined from two experiments) and apoptotic cells (244 unique N-termini combined from four experiments). (C) Classification N-termini identified in large-scale experiments with untreated cells (661 unique N-termini combined from two experiments) and apoptotic cells (733 unique N-termini combined from three experiments).

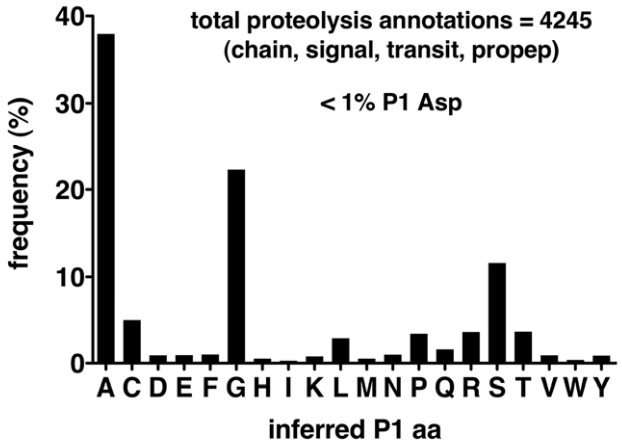

Figure S2. Inferred P1 residues for all N-termini annotated in the human Swiss-Prot database originating from chain, signal peptide, transit peptide, or propeptide processing.

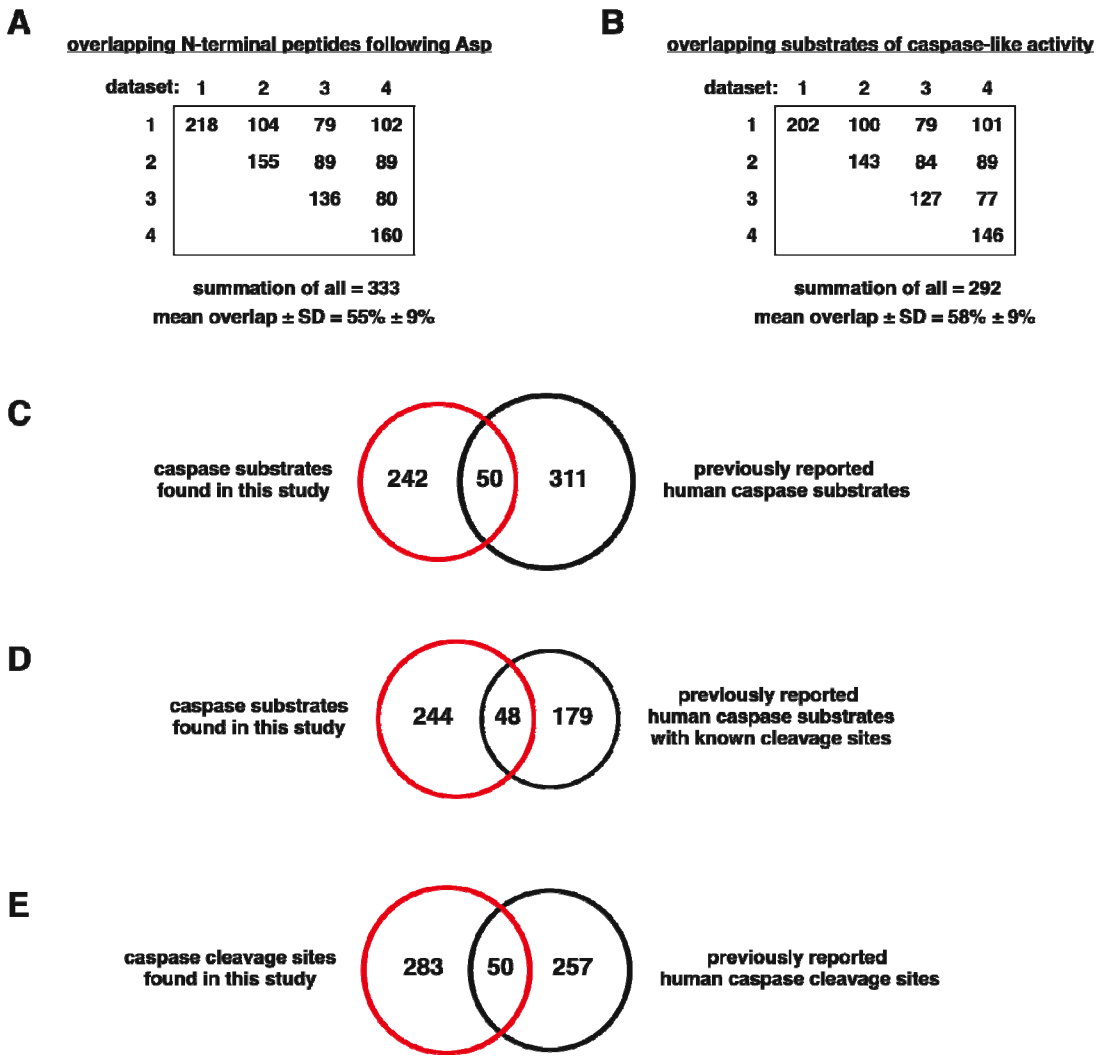

Figure S3. Overlaps between four apoptotic Jurkat cell datasets and previously reported caspase substrates and cleavage sites. Datasets 1-4 are described in **Table S4**. Only previously reported human substrates and human orthologs of previously reported rodent substrates were considered (Lüthi and Martin, 2007). (A) Peptide-level overlaps between apoptotic Jurkat cell datasets 1 through 4. (B) Protein-level overlaps betweeen apoptotic Jurkat cell datasets 1 through 4. (C) Overlap between substrates found in this study and previously reported human caspase substrates. (D) Overlap between substrates found in this study and previously reported human caspase substrates with known clevage sites. (E) Overlap between caspase-like cleavage sites found in this study and previously reported human caspase cleavage sites.

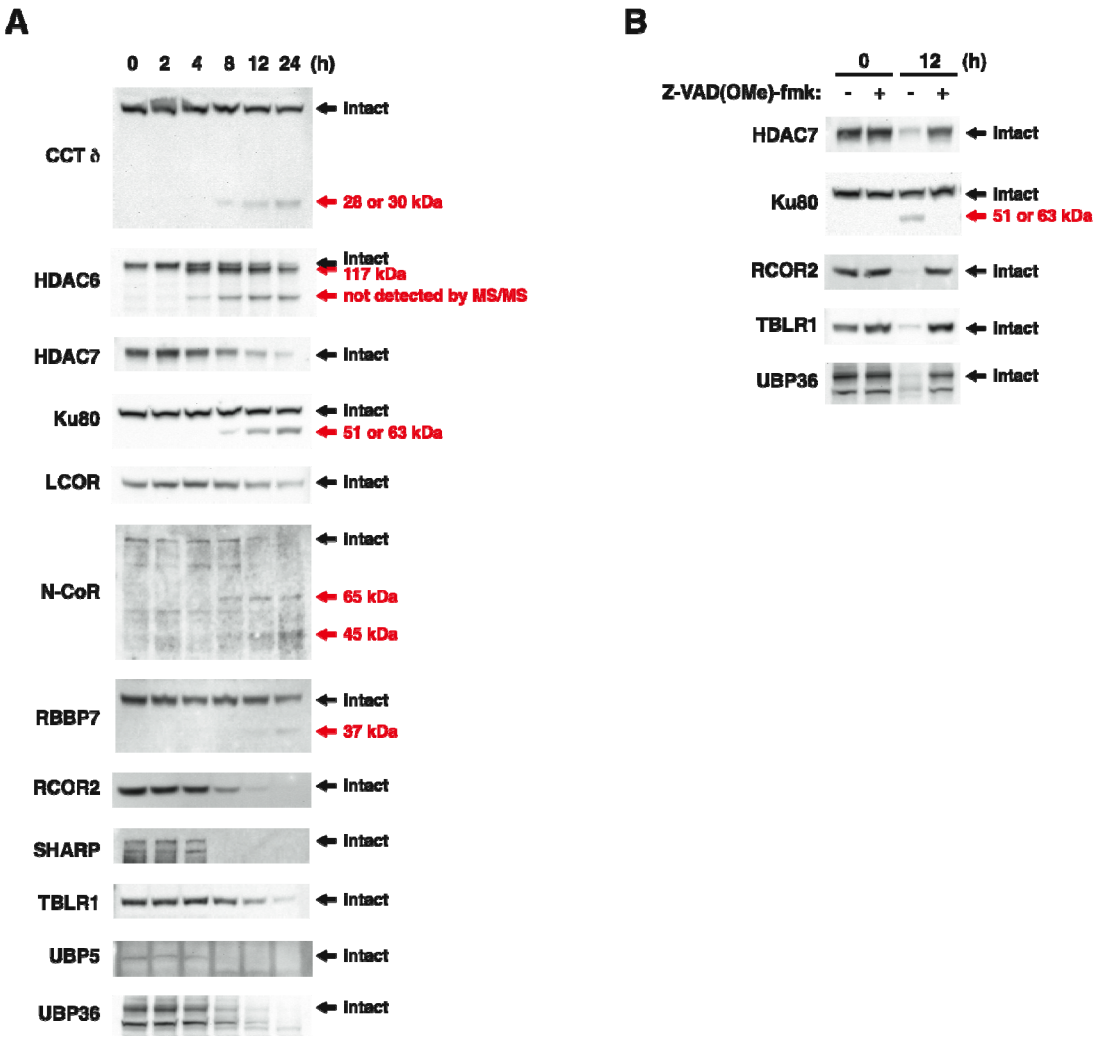

Figure S4. Analysis of proteolysis of selected proteins, all identified as caspase substrates in our proteomic studies, during apoptosis in Jurkat cells following treatment with 50 μM etoposide. Black arrows indicate full-length proteins. Red arrows indicate expected cleavage products for cleavage at the sites identified in our studies. Cleavage products were not detected in all cases. A) Time courses for the proteolysis of CCTδ, HDAC6, HDAC7, Ku80, LCOR, N-CoR, RBBP7, RCOR2, SHARP, TBLR1, UBP5, and UBP36 indicates full cleavage of HDAC6, HDAC7, N-CoR, RCOR2, SHARP, TBLR1, UBP5, and UBP36, and partial cleavage of CCTδ, Ku80, LCOR, and RBBP7. (B) The cleavage of a representative set of substrates identified in our studies, HDAC7, Ku80, RCOR2, TBLR1, and UBP36, is blocked by the broad-spectrum caspase inhibitor Z-VAD(OMe)-fmk and is thus dependent on caspase activity.

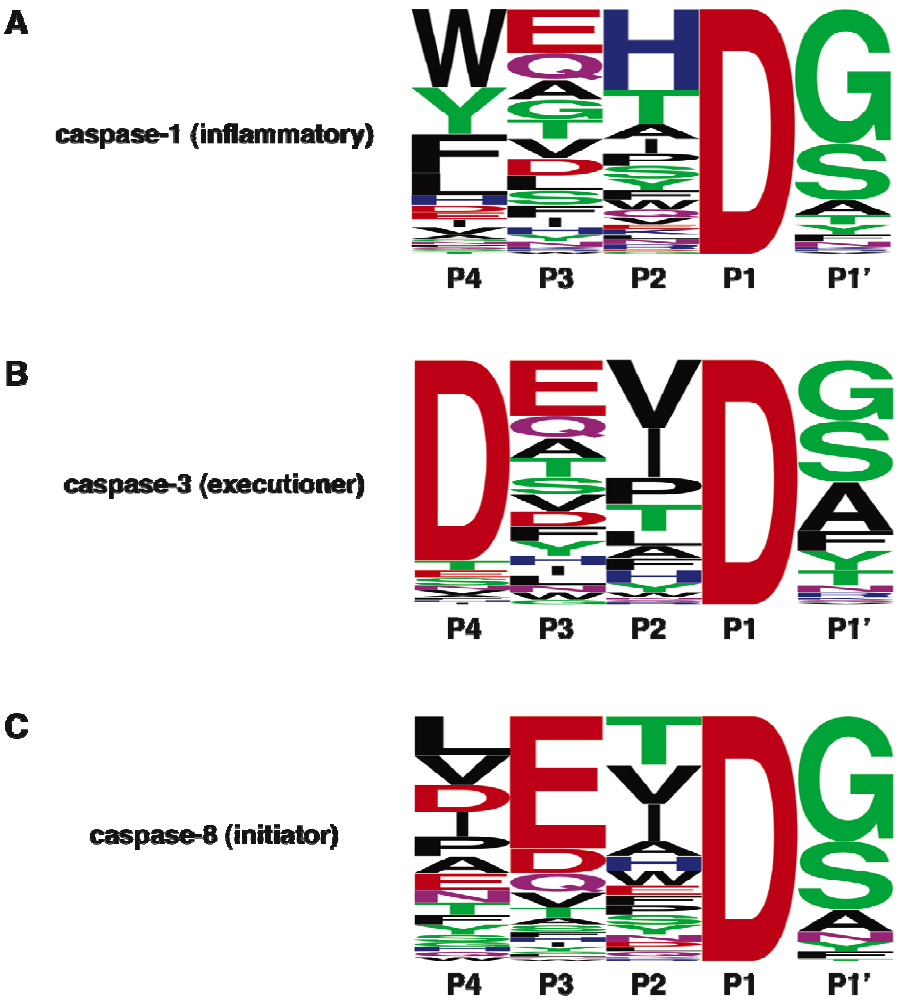

Figure S5. Sequence logo representations of prototypical inflammatory, executioner, and initiator caspase substrate specificities. These are exemplified by (A) caspase-1, (B) caspase-3, and (C) caspase-8, based on P4-P1 data adapted from Thornberry et al. (Thornberry et al., 1997) and P1′ data adapted from Stennicke et al. (Stennicke et al., 2000).

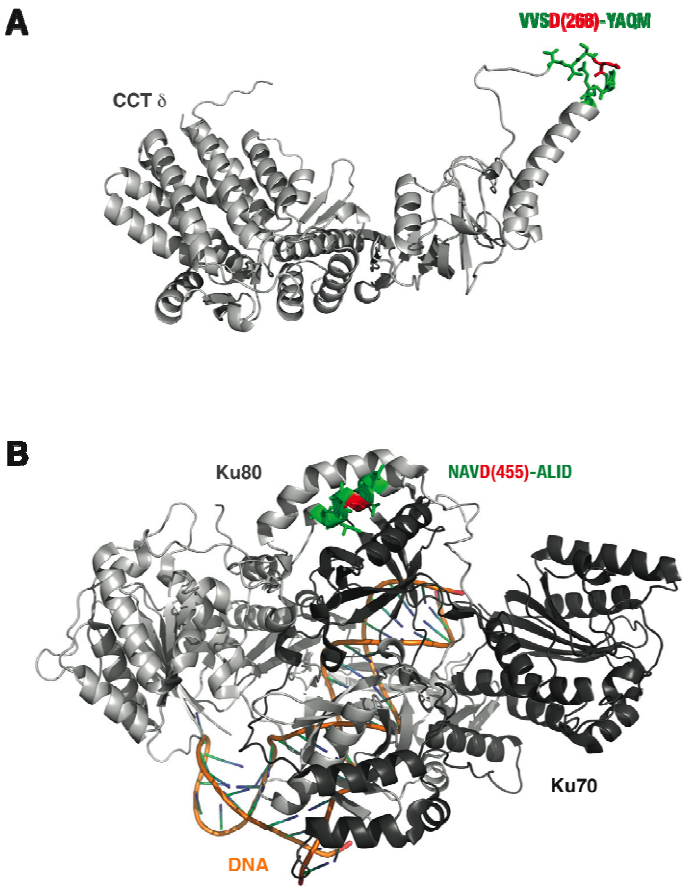

Figure S6. Examples of structural determinants of caspase substrate specificity. (A) The model structure for the type II charperonin CCTδ (TCPD_HUMAN), based on the crystal structure of an archaeal homolog with 39% sequence identity (Ditzel et al., 1998), illustrates an example of caspase cleavage in a loop region. The P1 aspartate is represented in red and all other residues in the P4-P4′ cleavage site are represented in green. (B) The crystal structure of the Ku70/80 heterodimer illustrates caspase cleavage in a helical region (Walker et al., 2001). Color representation of P4-P4′ residues in Ku80 (Ku86_HUMAN) is as for CCTδ. Immunoblot analysis indicates that both of these substrates are cleaved during apoptosis, although not to completion (**Figure S4A**).

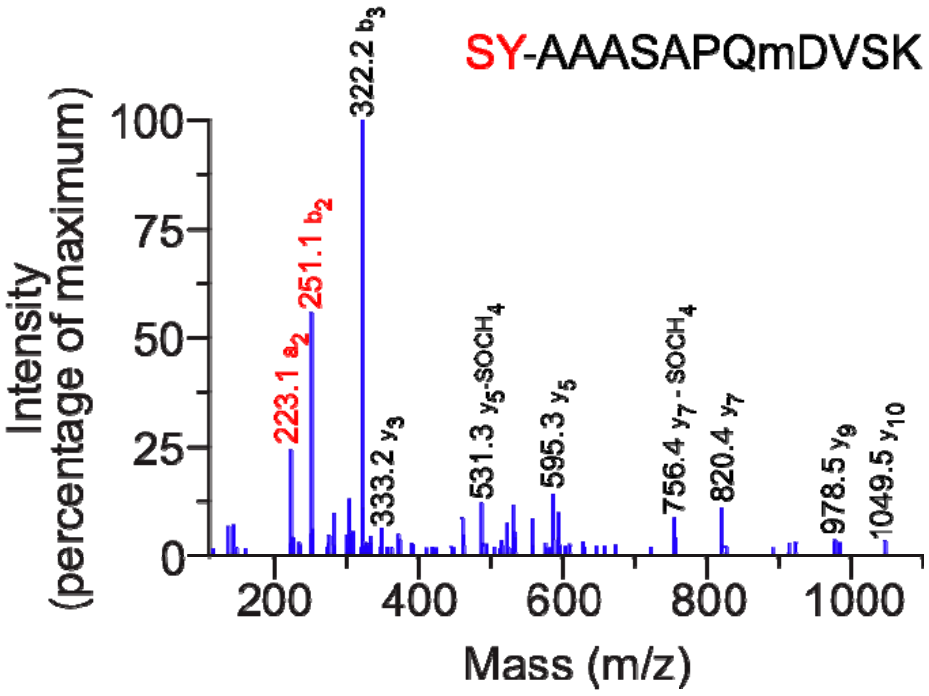

Figure S7. CID spectrum of the SY-labeled N-terminal peptide AAASAPQM(Oxidation)DVSK from N-CoR (NCOR1_HUMAN) corresponding to the P4-P4′ cleavage site ALVD(1826)-AAAS.

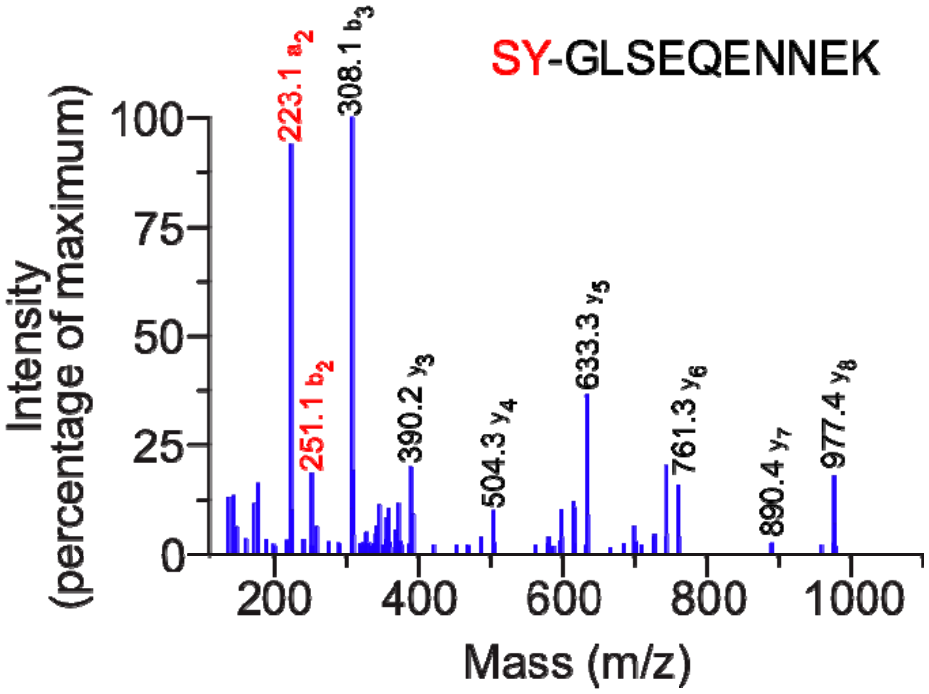

Figure S8. CID spectrum of the SY-labeled N-terminal peptide GLSEQENNEK from N-CoR (NCOR1_HUMAN) corresponding to the P4-P4′ cleavage site EIID(385)-GLSE.

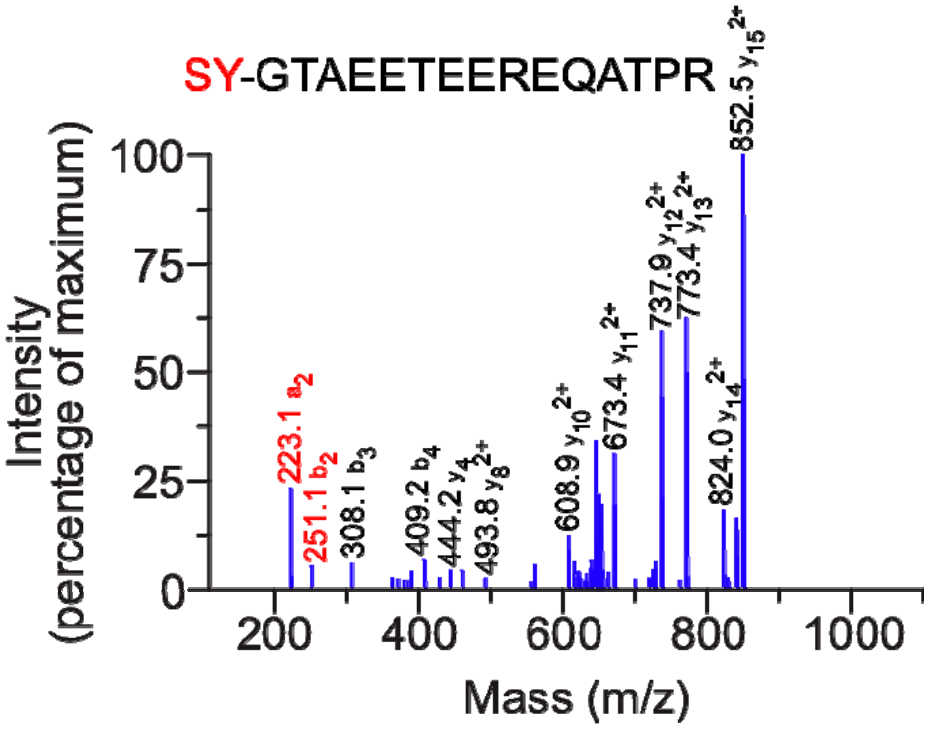

Figure S9. CID spectrum of the SY-labeled N-terminal peptide GTAEETEEREQATPR from N-CoR (NCOR1_HUMAN) corresponding to the P4-P4′ cleavage site DKID(555)-GTAE.

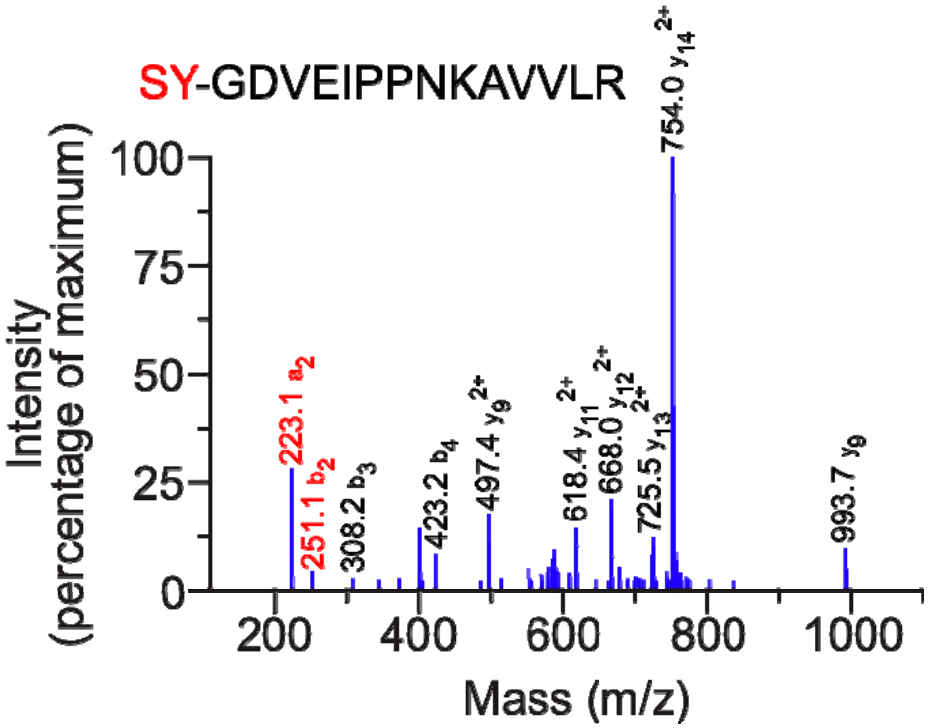

Figure S10. CID spectrum of the SY-labeled N-terminal peptide GDVEIPPNKAVVLR from TBLR1 (TBL1R_HUMAN) corresponding to the P4-P4′ cleavage site MEVD(152)-GDVE.

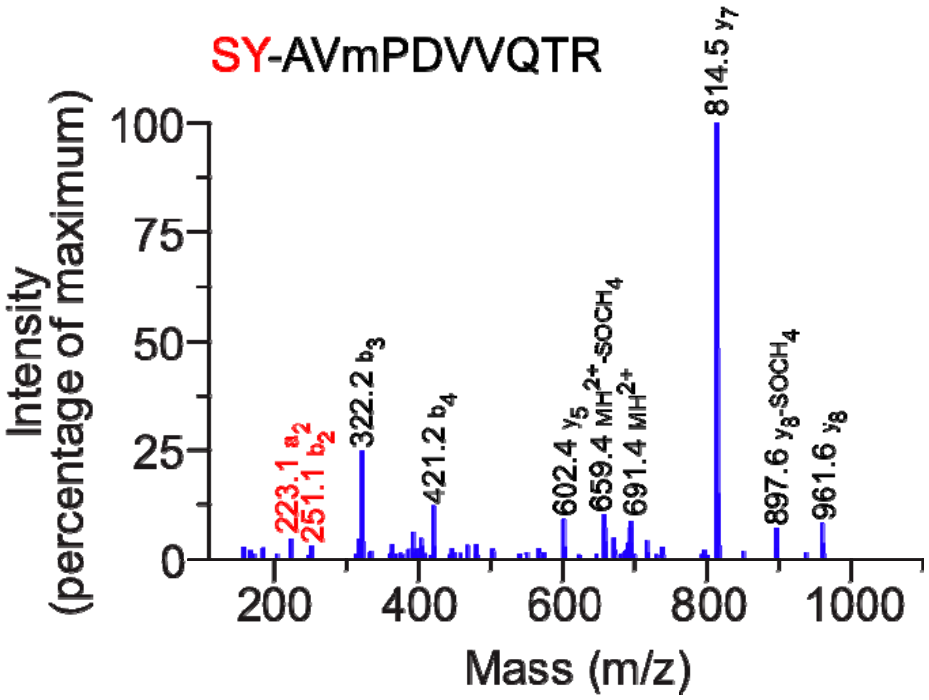

Figure S11. CID spectrum of the SY-labeled homologous N-terminal peptide AVM(Oxidized)PDVVQTR from either TBLR1 (TBL1R_HUMAN) or TBL1X (TBL1X_HUMAN) corresponding to the P4-P4′ cleavage site SLID(86)-AVMP.

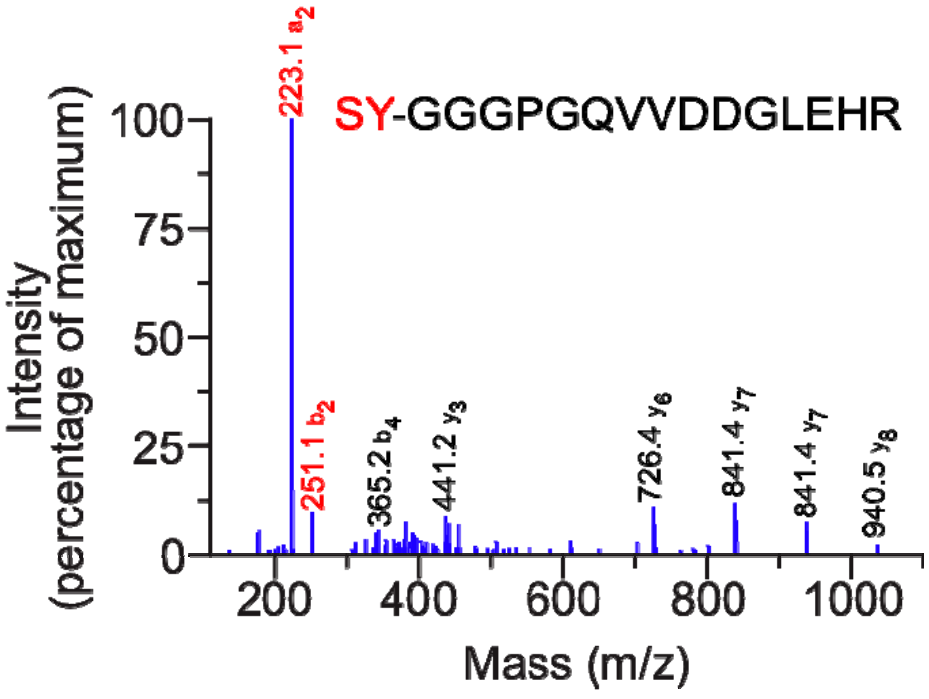

Figure S12. CID spectrum of the SY-labeled N-terminal peptide GGGPGQVVDDGLEHR from HDAC7 (HDAC7_HUMAN) corresponding to the P4-P4′ cleavage site LETD(412)-GGGP.

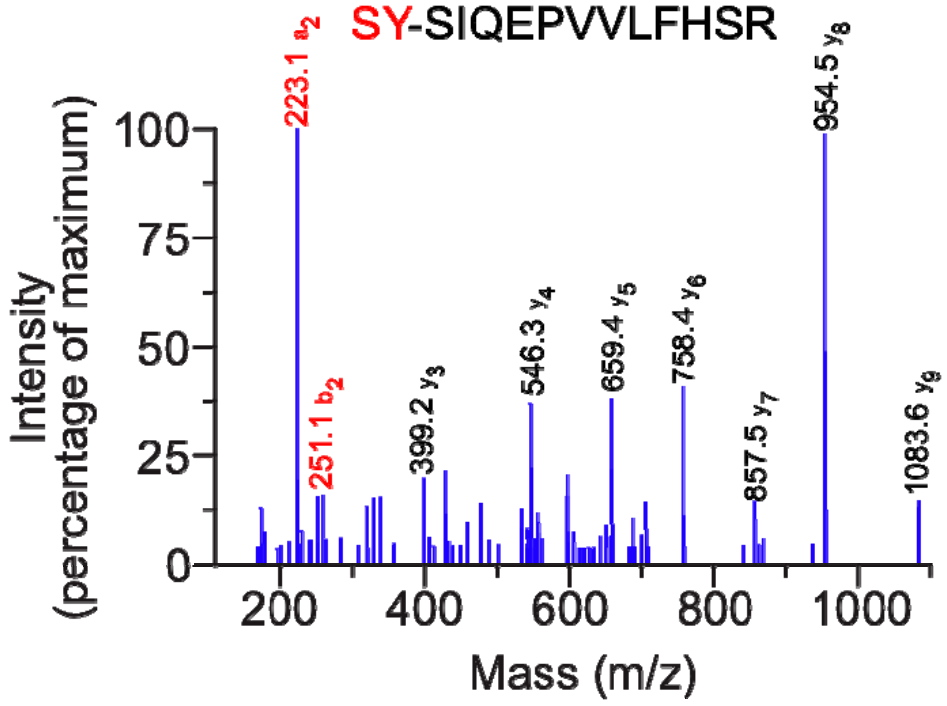

Figure S13. CID spectrum of the SY-labeled N-terminal peptide SIQEPVVLFHSR from SHARP (MINT_HUMAN) corresponding to P4-P4′ caspase-like cleavage site STTD(1574)-SIQE.

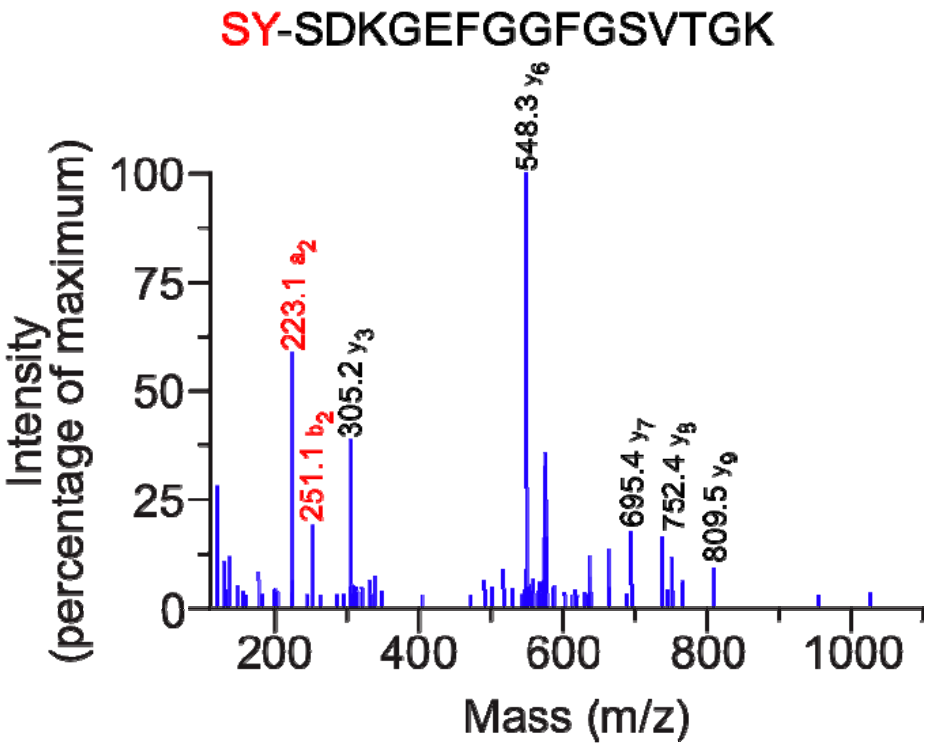

Figure S14. CID spectrum of the SY-labeled N-terminal peptide SDKGEFGGFGSVTGK from RBBP7 (RBBP7_HUMAN) corresponding to P4-P4′ caspase-like cleavage site SHCD(98)-SDKG.

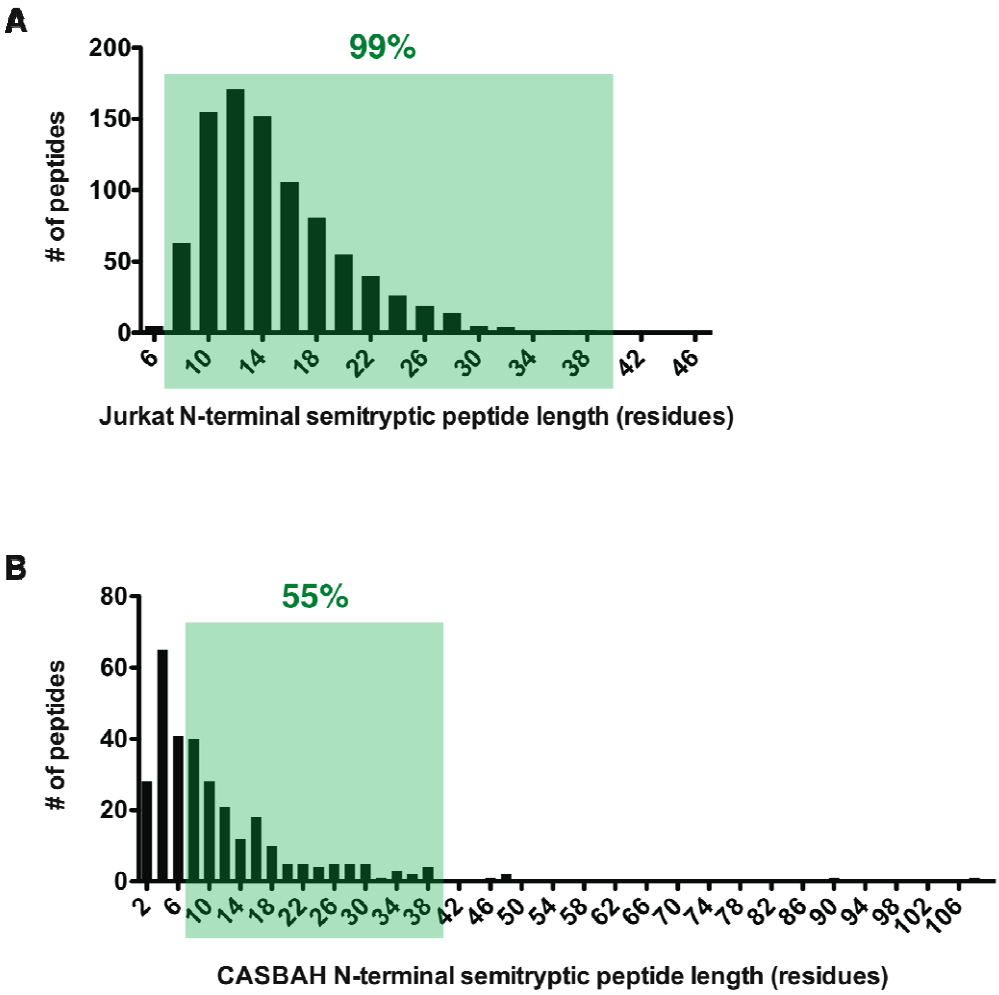

Figure S15. Subtiligase-based degradomics is limited to the identification of cleavages corresponding to N-terminal semitryptic peptides between approximately 7 and 40 residues in length. (A) Binning of all N-terminal semitryptic peptides identified in etoposide-treated Jurkat cells based on length. (B) Binning of N-terminal semitryptic peptides corresponding to 307 previously reported human caspase cleavage sites and human orthologs of previously reported rodent caspase cleavage sites (Lüthi and Martin, 2007). Boundaries and percentages of readily identified peptides are illustrated in both cases.

SUPLLEMENTAL REFERENCES

Abrahmsén, L., Tom, J., Burnier, J., Butcher, K.A., Kossiakoff, A., and Wells, J.A. (1991). Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. Biochemistry *30*, 4151-4159.

Atwell, S., and Wells, J.A. (1999). Selection for improved subtiligases by phage display. Proc Natl Acad Sci U S A *96*, 9497-9502.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S.*, et al.* (2002). The Protein Data Bank. Acta Crystallogr D Biol Crystallogr *58*, 899-907.

Braisted, A.C., Judice, J.K., and Wells, J.A. (1997). Synthesis of proteins by subtiligase. Methods Enzymol *289*, 298-313.
Chang, T.K., Jackson, D.Y., Burnier, J.P., and Wells, J.A. (1994). Subtiligase: a tool for semisynthesis of proteins. Proc Natl Acad Sci U S A *91*, 12544-12548.
Chatr-aryamontri, A., Ceol, A., Palazzi, L.M., Nardelli, G., Schneider, M.V., Castagnoli, L., and Cesareni, G. (2007). MINT: the Molecular INTeraction database. Nucleic Acids Res *35*, D572-574.
Ditzel, L., Löwe, J., Stock, D., Stetter, K.O., Huber, H., Huber, R., and Steinbacher, S. (1998). Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. Cell *93*, 125-138.
Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755-763.
Eswar, N., and Sali, A. (2007). Comparative modeling of drug target proteins. In Comprehensive Medicinal Chemistry II, J.B. Taylor, and J.D. Triggle, eds. (Oxford, UK, Elsevier Ltd), pp. 215-236.
Finn, R.D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R.*, et al.* (2006). Pfam: clans, web tools and services. Nucleic Acids Res *34*, D247-251.
Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng *14*, 993-1000.
Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R.*, et al.* (2007). IntAct--open source resource for molecular interaction data. Nucleic Acids Res *35*, D561-565.
Lüthi, A.U., and Martin, S.J. (2007). The CASBAH: a searchable database of caspase substrates. Cell Death Differ *14*, 641-650.
Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448-3449.
Melo, F., and Sali, A. (2007). Fold assessment for comparative protein structure modeling. Protein Sci *16*, 2412-2426.
Mishra, G.R., Suresh, M., Kumaran, K., Kannabiran, N., Suresh, S., Bala, P., Shivakumar, K., Anuradha, N., Reddy, R., Raghavan, T.M.*, et al.* (2006). Human protein reference database--2006 update. Nucleic Acids Res *34*, D411-414.
Pieper, U., Eswar, N., Davis, F.P., Braberg, H., Madhusudhan, M.S., Rossi, A., Marti-Renom, M., Karchin, R., Webb, B.M., Eramian, D.*, et al.* (2006). MODBASE: a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res *34*, D291-295.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.
Shen, M.Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. Protein Sci *15*, 2507-2524.
Stennicke, H.R., Renatus, M., Meldal, M., and Salvesen, G.S. (2000). Internally quenched fluorescent peptide substrates disclose the subsite preferences of human caspases 1, 3, 6, 7 and 8. Biochem J *350 Pt 2*, 563-568.
Thornberry, N.A., Rano, T.A., Peterson, E.P., Rasper, D.M., Timkey, T., Garcia-Calvo, M., Houtzager, V.M., Nordstrom, P.A., Roy, S., Vaillancourt, J.P.*, et al.* (1997). A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. J Biol Chem *272*, 17907-17911.
Walker, J.R., Corpina, R.A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature *412*, 607-614.
Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A., and Chen, E.Y. (1983). Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin in Bacillus subtilis. Nucleic Acids Res *11*, 7911-7925.