

Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini

Guoqiang Xu, Sung Bin Y. Shin, and Samie R. Jaffrey¹

Department of Pharmacology, Weill Medical College, Cornell University, New York, NY 10065

Edited by Solomon H. Snyder, Johns Hopkins University School of Medicine, Baltimore, MD, and approved September 25, 2009 (received for review August 6, 2009)

Proteolysis has major roles in diverse biologic processes and regulates the activity, localization, and intracellular levels of proteins. Linking signaling pathways and physiologic processes to specific proteolytic processing events is a major challenge in signal transduction research. Here, we describe N-CLAP (N-terminalomics by chemical labeling of the α -amine of proteins), a general approach for profiling protein N-termini and identifying protein cleavage sites during cellular signaling. In N-CLAP, simple and readily available reagents are used to selectively affinity label the α -amine that characterizes the protein N terminus over the more highly abundant ϵ -amine on lysine residues. Protein cleavage sites are deduced by identifying the corresponding N-CLAP peptides, which are derived from the N-termini of proteins, including the N-termini of the newly formed polypeptide products of proteolytic cleavage. Through selective affinity purification and tandem mass spectrometry analysis of 278 N-CLAP peptides, we characterized proteolytic cleavage events associated with methionine aminopeptidases and signal peptide peptidases, as well as proteins that are proteolytically cleaved after cisplatin-induced apoptosis. Many of the protein cleavage sites that are elicited during apoptotic signaling are consistent with caspase-dependent cleavage. These data demonstrate the utility of N-CLAP for proteomic profiling of protein cleavage sites that are generated during cellular signaling.

caspase | N-CLAP | N-terminalomics | tandem mass spectrometry

Proteolytic processing of proteins is a major regulatory mechanism in signal transduction (1). Proteases regulate signaling pathways by altering protein function. For example, proteases can either convert an inactive proenzyme to an active enzyme or inactivate specific proteins (1). The human genome is predicted to contain 561 proteases or protease-like enzymes, suggesting a potentially broad role for proteases (2). However, in the vast majority of cases, the specific proteins that are targeted by a protease are unknown. Similarly, the proteins and the sites that are cleaved during a signaling event are difficult to identify.

To characterize protein processing on a proteomic scale, one strategy is to systematically profile protein N-termini in tissue or cellular samples. "N-terminalomics" is motivated by the idea that the population of N-termini changes after protease activation. This is because any single protein cleavage event results in the formation of an "internal" N terminus. Proteomic approaches typically use protein samples that have been digested with proteases, such as trypsin, to generate peptides suitable for tandem mass spectrometry (MS/MS) analysis. Selective enrichment of N-terminal peptides (i.e., the peptides derived from the N-termini of proteins) would be required to identify protein N-termini and proteolytic cleavage sites. Thus, there is considerable interest in the development of strategies that selectively enrich for N-terminal peptides before MS analysis.

Several techniques to enrich for N-terminal peptides rely on "negative selection," in which non-N-terminal peptides are removed from a sample, leaving N-terminal peptides (3–6). However, because most peptides are not N-terminal peptides, the efficiency needs to be very high to obtain a preparation enriched in N-terminal peptides. In contrast, "positive selection" involves the purification of N-terminal peptides from a mixture of peptides.

Selective labeling of the N-termini of proteins would allow N-terminal peptides to be readily purified after protease digestion (7, 8). The challenge of positive selection is to selectively recognize the α -amino group, which is present exclusively on the N-termini of proteins, but not the ϵ -amino group, which is present at high levels in proteins on lysine side chains.

Here we describe a proteomic strategy, N-terminalomics by chemical labeling of the α -amine of proteins (N-CLAP). N-CLAP utilizes features of Edman chemistry to enable the selective labeling and capture of N-terminal peptides from proteins in complex biological samples. We demonstrate that N-CLAP can be used to characterize the diversity of N-terminal processing as well as signal peptidase specificity. Using this method, we also identify protein processing that occurs during apoptosis and identify several unique potential targets of caspase processing. Together, these data demonstrate that N-CLAP can be used to identify cleavage sites in proteins and identify proteolytic cleavage events that occur during cell signaling.

Results

A Chemoselective Strategy for Labeling and Enriching N-Terminal Peptides. The N terminus of proteins is characterized by an α -amine, as opposed to the ϵ -amines that are on lysine side chains. The α -amine (the pK_a of its ammonium ion ≈ 8.95) is notable for its slightly reduced pK_a relative to that of ϵ -amines (the pK_a of its ammonium ion ≈ 10.53) (9). However, given this narrow pK_a window, there is no pH at which the α -amine can be selectively labeled with an amine-labeling reagent, because ϵ -amines, albeit less reactive, are typically much more abundant (10). The circumstances are further complicated by the fact that the pK_a of the ammonium ion for the α -amine and the ϵ -amine can be affected by the side chain of the adjacent amino acids, hydrogen bonding, and other intramolecular interactions (11, 12).

To address these issues, we have developed a chemical labeling strategy, N-CLAP, that chemoselectively labels the N terminus. This labeling strategy is based on the chemistry of Edman degradation, which is used for N-terminal amino acid sequencing (13) [SI Appendix, Fig. S14]. In N-CLAP, the amine-reactive Edman reagent, phenyl isothiocyanate (PITC), is used to block all of the amines in proteins (Fig. 1). TFA is then used to trigger an intramolecular cyclization of the PITC-modified α -amine, which results in cleavage of the peptide bond between the first and second amino acid. This intramolecular cyclization does not occur for the PITC adduct on the ϵ -amine because an energetically favorable five-membered ring cannot form at this position (14) and leaves the PITC-modified ϵ -amine on the lysines intact (15). Therefore, the

Author contributions: G.X. and S.R.J. designed research; G.X. and S.Y.S. performed research; G.X. analyzed data; and G.X. and S.R.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed at: Department of Pharmacology, Weill Medical College, Cornell University, 1300 York Avenue, Box 70, New York, NY 10065. E-mail: srij2003@med.cornell.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0908958106/DCSupplemental.

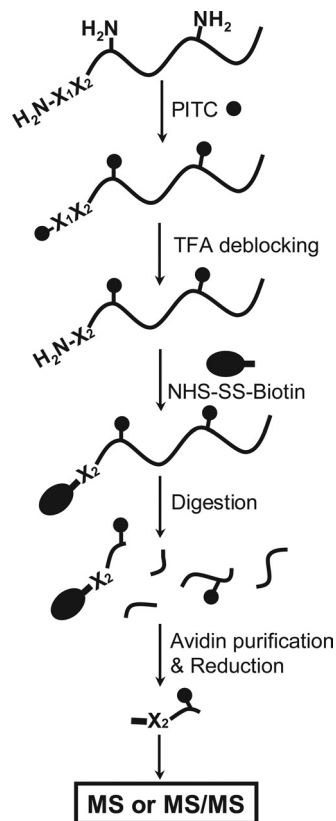


Fig. 1. Strategy for N-CLAP. After PITC treatment, all amines in a protein are blocked (filled circle). Reaction with TFA selectively deblocks the N terminus but not PITC-modified lysines. The newly generated amino groups at the N-termini are then reacted with an amine-specific labeling reagent, such as EZ-Link Sulfo-NHS-SS-biotin (oval). After protein digestion, the N-terminal peptides are recovered using avidin-based resins, eluted with a reducing agent, and then identified by MS.

PITC adducts on the N-termini are selectively “deblocked.” After TFA treatment, the protein is shortened by one amino acid, and importantly, contains only a single α -amine, whereas all lysines remain blocked. The α -amine is then labeled using an amine-reactive tagging reagent, such as a biotinylating agent, resulting in selective tagging of the N terminus of each protein. Proteins are digested, and tagged peptides are purified, for example using avidin agarose. N-terminal peptides are eluted and identified by MS (Fig. 1).

We first addressed whether N-terminal peptides generated by N-CLAP would be sufficiently unique to identify parent proteins. The majority (74%) of N-CLAP peptides prepared by virtual trypsin digestion of proteins listed in the Swiss-Prot human protein database (v55.6, July 1, 2008) (16) are at least 5 aa long, a length that may provide enough sequence information for protein identification. Of these peptides, 68% identify a single protein in the database (SI Appendix, Fig. S1 B–D). Most of the peptides that match more than one protein match a family of highly similar isoforms. Thus, for N-CLAP peptides prepared with trypsin, 54% of the human proteome could be unambiguously detected. To increase the coverage, the results from parallel experiments, each using a separate protease, could be merged. For example, using trypsin and Glu-C would increase coverage of the proteome to 96.6% (SI Appendix, Fig. S1E). Thus, a significant fraction of potential N-CLAP peptides would provide unambiguous identification of parent proteins. In some cases, the masses of N-CLAP peptides alone could be used to identify many proteins. This would be most useful when analyzing

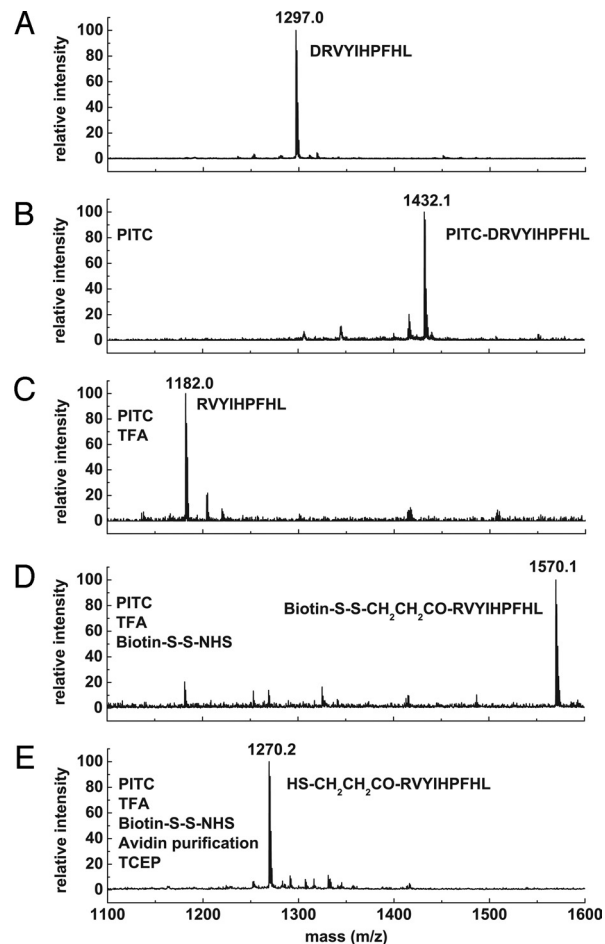


Fig. 2. Selective and efficient labeling of angiotensin I using N-CLAP. (A) MALDI-TOF-MS of angiotensin I results in a major peptide ion at 1297.0. (B) PITC-treated angiotensin I results in a major peptide ion reflecting the expected thiocarbamoyl PITC modification. (C) TFA treatment of PITC-modified angiotensin I results in a single prominent peptide ion with a mass corresponding to angiotensin I missing the first N-terminal residue. (D) The deblocked angiotensin I was reacted with EZ-Link Sulfo-NHS-SS-biotin. MALDI-TOF-MS revealed a single major peak corresponding to the modified peptide. (E) Avidin-purified and TCEP-eluted sample results in a 1,270.2-Da peptide ion, corresponding to angiotensin I peptide without N-terminal aspartic acid and with a remnant from the cleavable biotin tag. The treatments, peptide sequence, and the chemical modification for each sample are indicated.

small proteomes with instrumentation capable of providing highly accurate masses (SI Appendix, Fig. S1F).

Selective Labeling of Peptide N-Termini. To evaluate N-CLAP, we first characterized the efficiency of each chemical reaction using a model peptide, angiotensin I, which contains a single amine at its N terminus. After reaction of angiotensin I with PITC, the apparent mass increased by the expected 135 Da (Fig. 2A and B). After TFA treatment, the peptide mass is reduced by 250 Da, which corresponds to the mass of the N-terminal residue, aspartic acid, modified with PITC (Fig. 2C). A mass increase of 388 Da was detected after labeling the peptide with the amine-reactive cleavable labeling reagent EZ-Link Sulfo-NHS-SS-biotin (Fig. 2D). The final step, removal of the biotin by Tris(2-carboxyethyl)phosphine (TCEP)-mediated reduction of the disulfide linker, resulted in the expected loss of 300 Da, corresponding to the removal of the thiol-containing biotin moiety (Fig. 2E). Similar results were obtained with another peptide, derived from β -catenin (SI Appendix, Fig. S2 A–F).

We next performed a similar analysis using the c-Myc peptide, which contains both an α -amine and an ϵ -amine from an internal lysine (*SI Appendix*, Fig. S2 G–K). As with angiotensin I, the c-Myc peptide was readily modified by PITC, and the first amino acid was readily removed with TFA. Biotinylation resulted in selective labeling of the α -amine, whereas the PITC-blocked ϵ -amine was not labeled (*SI Appendix*, Fig. S2J). Notably, the PITC modification did not impair peptide fragmentation, indicating that MS/MS spectra obtained from PITC-modified peptides could be used for peptide identification. Together, these data suggested that each step of N-CLAP can occur with high efficiency and with negligible side reactions.

Selective Enrichment of N-Terminal Peptides from Proteins. We next applied N-CLAP to individual proteins. To monitor the completeness of each step, an assay for amines in proteins was developed. Amines were labeled with biotin *N*-hydroxysuccinimide ester (biotin-NHS). Signals obtained after Western blotting with a biotin antibody indicate the presence of unmodified amines. Incubation of insulin, casein, or RNase A with PITC resulted in complete loss of amine reactivity (Fig. 3A). After TFA treatment, amine reactivity was recovered, consistent with selective deblocking of the N terminus. In the case of insulin, the recovered signal is approximately half as intense as the signal for the unmodified protein (Fig. 3A, *Top*), consistent with the presence of two amines in the heavy chain of insulin, one α -amine and one ϵ -amine at its sole lysine residue. However, the recovered signal is much weaker for casein and RNase A (Fig. 3A, *Middle and Bottom*) than that of the original samples, which is expected because these two proteins have many lysines (12 and 7 for casein and RNase A, respectively).

We next addressed whether the N-terminal tryptic peptide could be enriched using N-CLAP. After PITC and TFA treatment, RNase A was labeled with EZ-Link Sulfo-NHS-SS-biotin. MALDI-TOF-MS of the tryptic digest reduced by TCEP revealed numerous peptides, including a relatively low abundance peak with a mass corresponding to the expected mass of the N-terminal peptide containing a PITC-modified lysine and a mercaptoethanoate (HSCH₂CH₂CO⁻) amine adduct that remains after the cleavage of the biotin tag (Fig. 3B). However, after avidin affinity purification, this peak was the predominant peak (Fig. 3C and *SI Appendix*, Fig. S3). Additionally, a peak representing a peptide in which the PITC modification on lysine is hydrolyzed, which would appear as a 1,196-Da peak, was not detected, confirming the stability of the PITC adduct. Together, these data demonstrate the usefulness of N-CLAP for positive purification of N-terminal peptides.

Characterization of the Diversity of N-Terminal Methionine Processing. We next examined the ability of N-CLAP to generate N-terminal peptides from cellular lysates. Cell lysate was prepared from human Jurkat T cells and subjected to N-CLAP. As with the individual proteins, PITC effectively blocked the amines in the lysate (Fig. 4A), and TFA resulted in the reappearance of a fraction of the amine reactivity. Amines were biotinylated, proteins were digested, and N-terminal peptides were affinity isolated. A total of 80 N-terminal peptides were identified using liquid chromatography (LC) quadrupole (Q)-TOF MS/MS (*SI Appendix*, Fig. S4 and *Table S1*). As an example, the MS/MS spectrum of the N-terminal peptide from 26S proteasome non-ATPase regulatory subunit 6 is shown (Fig. 4B). As with many N-CLAP peptides, it starts at the third amino acid predicted from the coding sequence. This reflects the removal of the original N-terminal methionine by endogenous methionine aminopeptidases (MetAPs) and the loss of the second amino acid during the TFA deblocking step in the N-CLAP procedure. The identity of the residue lost during N-CLAP can be extrapolated from protein sequence databases.

Many of the N-CLAP peptides detected in Jurkat T cell lysate were derived from the N-termini of proteins lacking the N-terminal methionine. However, 27.2% of peptides were derived from pro-

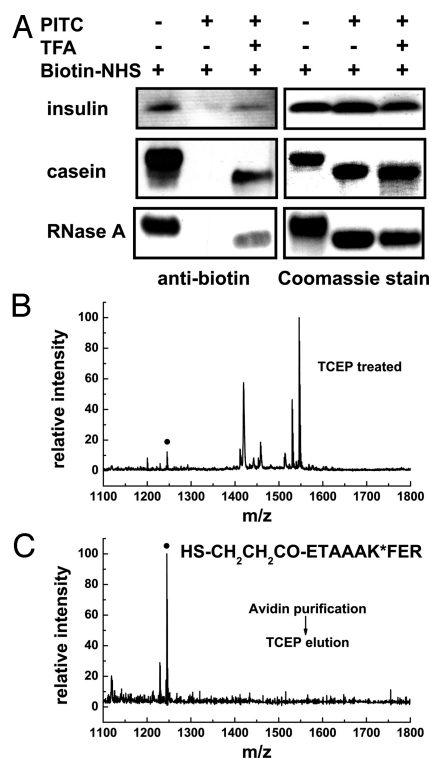


Fig. 3. Assessment of the N-CLAP for enrichment of the N-terminal peptide from individual proteins. (A) PITC completely modifies amines in proteins, and TFA generates new N-terminal amines. The presence of amines in the unreacted, PITC-treated, and both PITC- and TFA-treated samples (insulin, α -casein, and RNase A) was measured by reaction with biotin-NHS and visualized by Western blotting with a biotin antibody. The band intensity reflects the relative amount of free amines. Coomassie staining confirms comparable loading in each lane. A slight increase in the mobility of PITC-modified proteins is typically detected and likely reflects the reduced positive charge on the protein due to lysine modification. (B) MALDI-TOF-MS spectrum of the tryptic peptides generated after RNase A was subjected to N-CLAP, digested, and reduced (without avidin purification). MALDI-TOF reveals multiple peptides, including a peptide consistent with the N-terminal RNase A N-CLAP peptide (filled circle). (C) N-CLAP enriches for N-terminal peptides. After RNase A is subjected to N-CLAP, the tryptic peptides are purified on avidin agarose and eluted with TCEP. MALDI-TOF-MS spectrum of the TCEP eluate reveals a predominant peak. This peptide corresponds to the N-CLAP peptide from the RNase A N terminus, containing a PITC adduct on an internal lysine (asterisk) and an N-terminal mercaptoethanoate amine adduct (shown in sequence) derived from the cleaved biotinylation reagent.

teins that retained the N-terminal methionine. This likely reflects inefficient removal of methionine by MetAPs, as has been described for methionine residues followed by charged or bulky amino acids (17). MS/MS analysis of N-CLAP peptides also indicated that numerous proteins exist in the cytosol as a heterogeneous population, with a fraction containing the N-terminal methionine and another fraction missing the N-terminal methionine. Some proteins (16.2%) were the source of three N-CLAP peptides, reflecting variable trimming of N-terminal residues. These data indicate that the identity of the N-terminal residue can be considerably different from what is predicted by sequence databases.

Determination of Organelle-Targeting Peptide Cleavage Sites Using N-CLAP. Many N-CLAP peptides were derived from proteins that were processed to remove the first 15 or more N-terminal amino acids (*Table S2*). Some N-CLAP peptides were derived from transmembrane or secreted proteins, in which signal peptides are removed by signal peptide peptidases (SPPs) (18). Other N-CLAP peptides were derived from mitochondrial proteins, which are processed by mitochondrial processing peptidases (MPPs) to re-

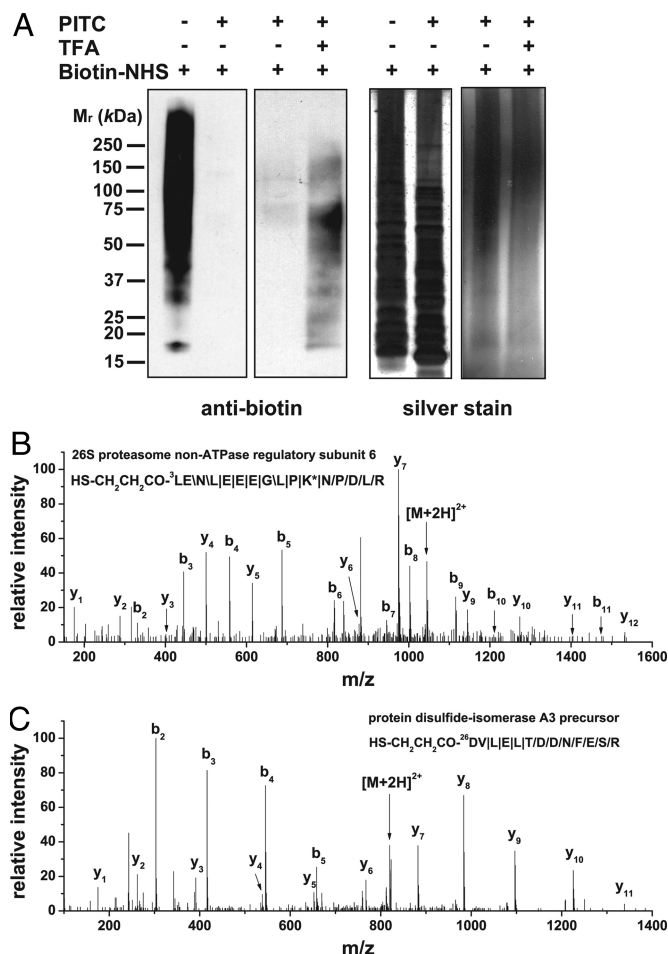


Fig. 4. Identification of protein N-termini from a complex mixture using N-CLAP. (A) Evaluation of the completeness of PITC and TFA treatment on Jurkat T cell lysate. The amine content of a Jurkat T cell lysate is monitored by Western blotting with a biotin antibody at different stages in the N-CLAP protocol. Note that the exposure time for the right panel is much longer than for the left panel to detect the recovered signal after TFA treatment. Silver staining showed similar loading. (B and C) Representative MS/MS spectra of 2 N-CLAP peptides from Jurkat T cell lysate. The sequence map of the peptides, the N-terminal modification, the position of the first amino acid in the coding sequences, and the protein names are indicated in the spectra. PITC modification on lysine is indicated by an asterisk. In B, the N-CLAP peptide starts from the third amino acid from the coding sequence, which indicates that the initial methionine is removed from the nascent protein. However, in C the N-CLAP peptide starts from the 26th amino acid from the coding sequence, which indicates that the first 24 aa are removed in the nascent protein because the N-CLAP removes one additional N-terminal amino acid. The symbols b , y , and l represent b -ions, y -ions, and both b -ions and y -ions, respectively.

move the transit peptides (19). As an example, the N-CLAP peptide derived from protein disulfide-isomerase A3 indicates that the first 24 aa of the nascent protein is removed (Fig. 4C), consistent with previous MS results (3).

Signal peptides lack simple consensus sequences can possibly be predicted using computational approaches (20). Because different algorithms predict different cleavage sites, the precise cleavage sites have to be experimentally verified. Signal peptides typically show 3 distinct regions: a positively charged N-terminal region, a hydrophobic region, and a C-terminal uncharged polar region. Examination of N-CLAP peptides (*SI Appendix*, Fig. S4 and *Table S2*) indicates that only 57%, 93%, and 28% of signal peptides contain a positively charged N terminus, a hydrophobic region, and a C-terminal uncharged polar region, respectively, and only 14% of

signal peptides have all three regions. These results indicate considerable diversity in signal peptides and confirm previous studies reporting that signal peptides do not conform to a simple consensus sequence (20, 21).

All of the N-CLAP-identified putative mitochondrial transit peptides, except one, were enriched in Arg, Ser, and Ala and deficient in negatively charged residues, consistent with proposed characteristics for these sequences (22, 23). Most (68%) have a conserved Arg close to the cleavage site. However, there was also heterogeneity in cleavage sites, even in the same protein. For example, several distinct N-CLAP peptides were identified from ATPase inhibitor, reflecting distinct pools of protein cleaved after the 24th, 25th, and 26th amino acid (*Table S2*). These noncanonical cleavage events may reflect broad cleavage specificity of MPPs or may reflect additional N-terminal processing by distinct peptidases. More than half (52.6%) of the signal or transit sequences and the cleavage sites identified in our analysis had not been previously experimentally validated, according to the annotation in the Swiss-Prot database (16).

Identification of Caspase Cleavage Sites Using N-CLAP. We next examined proteolytic processing events elicited by cisplatin, a chemotherapeutic agent. On the basis of a time course analysis of cisplatin-induced apoptosis in Jurkat T cells (*SI Appendix*, Fig. S5), cells were harvested after treatment with vehicle or cisplatin for 8 h and then processed using the N-CLAP procedure. The majority of N-CLAP peptides from both vehicle and cisplatin-treated Jurkat T cells were derived from the N terminus of proteins, from internal cleavages, or from the N terminus of proteins that were generated after the removal of signal sequences (Fig. 5A). However, after cisplatin treatment, many additional internal N-terminal peptides were also identified (Fig. 5B and C), some of which were derived from proteins cleaved after well-established consensus caspase cleavage sites, such as DEXD and (I/L/V)EXD (24). Other proteins were cleaved at sites that did not resemble canonical caspase cleavage sites but occurred after aspartate, the minimal feature for caspase cleavage (*SI Appendix*, Fig. S4 and *Table S3*) (7, 24). Among the putative caspase substrates, only 17% had canonical caspase cleavage sites, whereas nearly half (48%) had the consensus sequence of DXXD. These data suggest physiologic utilization of heterogeneous cleavage sites by caspases.

Most of the proteins cleaved after cisplatin treatment were previously identified as caspase targets using various approaches (7, 25, 26). However, some of them, such as protein phosphatase 1G, myosin-10, HERV-H.3q26, and pyridoxal-dependent decarboxylase domain-containing protein 1, have not been previously reported as being cleaved during apoptosis. Overall, 56% of the cleavage sites have not been previously reported. To confirm that these cleavage events were caspase dependent, Jurkat T cells were treated either with vehicle, cisplatin, or cisplatin and z-VAD-fmk, a pan-caspase inhibitor. We selected proteins that included previously reported and newly discovered caspase substrates for validation, in part on the basis of commercial availability of antibodies. In each case, the selected proteins were either completely or partially degraded after the addition of cisplatin, and the cleavage was blocked by the caspase inhibitor, confirming that the cleavage events were indeed caspase dependent (Fig. 5D).

In addition to caspase cleavage sites, additional classes of cleavage sites were identified (*SI Appendix*, Fig. S4 and *Table S4*), consistent with findings that cisplatin induces the activation of proteases other than caspases (27). In some cases, N-CLAP identified proteins that had previously been shown to be cleaved during apoptosis, but the specific cleavage sites was not identified (28, 29). Together, these results demonstrate the ability of N-CLAP to identify numerous caspase-dependent and caspase-independent cleavage events after cisplatin treatment.

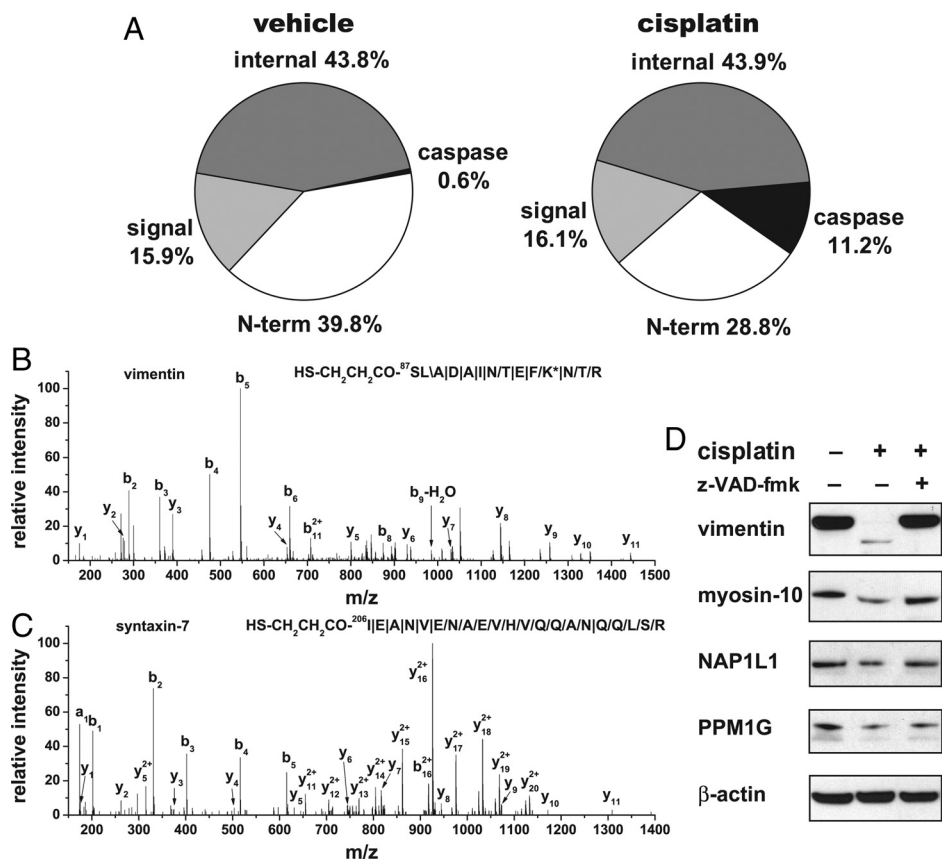


Fig. 5. Identification of caspase cleavage sites induced by apoptosis by N-CLAP. (A) Distribution of N-CLAP peptides prepared from vehicle- and cisplatin-treated Jurkat T cells. The N-CLAP peptides are classified into 4 categories: (i) N-terminal peptides (N-term), which comprise N-CLAP peptides starting from the 2nd, 3rd, and 4th amino acid from the coding sequence; (ii) N-terminal peptides from proteins with signal or transit peptides removed (signal); (iii) N-terminal peptides derived from internal N-termini of proteins cleaved in a manner consistent with caspase cleavage (caspace); and (iv) N-terminal peptides derived from internal N-termini of proteins cleaved at sites that do not resemble caspase cleavage sites (internal). (B and C) Representative MS/MS spectra of the N-CLAP peptides derived from the internal N-termini of vimentin and syntaxin-7. Jurkat T cells were treated with cisplatin (200 μM for 8 h). (D) Biochemical validation of caspase substrates. Western blot analysis of vimentin, myosin-10, NAP1L1 (nucleosome assembly protein 1-like 1), PPM1G (protein phosphatase 1G), and β -actin, a negative control, was used to assess the caspase dependence of their cleavage after cisplatin treatment. Jurkat T cells were treated with either DMSO, 200 μM cisplatin, or 200 μM cisplatin after 2 h pretreatment with z-VAD-fmk (20 μM). The levels of each of these proteins were reduced after cisplatin treatment, and the effect was blocked by inclusion of the caspase inhibitor. The size of the vimentin cleavage product (43 kDa, asterisk) is consistent with cleavage after Asp-85, which is indicated by the vimentin N-CLAP peptide.

Discussion

Profiling the N-termini of proteins in complex mixtures is a powerful approach for elucidating proteolysis signaling pathways. Obtaining N-terminal peptides for N-terminalomics is challenging because the α -amine at the protein N terminus is nearly chemically indistinguishable from the ϵ -amines on lysines. We have developed a strategy for chemoselective labeling of the N terminus, N-CLAP, which has resulted in the identification of 278 N-CLAP peptides in normal and apoptotic Jurkat T cells. N-CLAP provides a straightforward chemical strategy for selective N-terminal labeling of proteins and for enrichment of N-terminal peptides.

Using N-CLAP to characterize protein N-termini in cellular lysates reveals a surprising degree of trimming of the protein N terminus, considering that MetAPs are the major enzymes that have been linked to N-terminal processing. It is not clear whether amino acid cleavage after the removal of the initial methionine is also mediated by MetAPs or whether other aminopeptidases are involved. Conceivably, the differentially processed N-terminal forms of these proteins have different stabilities, due to effects of N-terminal residues on protein stability (30), or may endow the protein with different functions. Similarly, the N-termini of membrane proteins revealed considerable variability, suggesting heterogeneity in the specificity of SPPs and MPPs or secondary protein processing events after initial cleavage.

Characterization of protease processing induced by cisplatin reveals cleavage at numerous different types of sites, including consensus caspase cleavage sites. Many of these sites have been previously described, validating the accuracy of N-CLAP. Additionally, we biochemically validated the caspase dependence of the cleavage of several proteins. In numerous cases, N-CLAP identified proteins that were already known to be cleaved during apoptosis, but the specific sites were not known. Thus, N-CLAP provides a strategy to readily identify cleavage sites in proteins, thereby

permitting mutagenesis and the generation of cleavage-resistant proteins for mechanistic studies. Of note, we identified fewer cleavage sites than other studies using N-terminal labeling approaches (7). This likely reflects the limited number of N-CLAP samples subjected to MS and differences in MS instrumentation, peptide identification algorithms, and settings and parameters used for positive identification, as well as different apoptotic stimuli. These differences are unlikely to reflect an intrinsic inability to detect certain peptides that have been processed in the N-CLAP procedure by MS/MS. Indeed, a β -catenin peptide that was not detected in Jurkat T cell lysates in our initial N-CLAP experiments, but was detected in earlier studies (7), is detectable when a synthetic peptide is used (SI Appendix, Fig. S2 A–F).

Another potential application of N-CLAP is to identify the substrates of a specific protease. Comparison of N-CLAP peptides derived from cells expressing different levels of specific proteases or treated with an inhibitor of a protease of interest would reveal protease substrates. These analyses would be facilitated by tagging N-CLAP peptides using standard quantitative proteomic approaches (31).

In addition to monitoring protein processing, N-terminalomics can be used to enhance the sensitivity of quantitative proteomics. The large number of peptides obtained after tryptic digestion can prevent thorough identification of a significant fraction of the proteins in the sample owing to instrument duty cycle, ion suppression, and dynamic range. Although this is partially alleviated using techniques such as isotope-coded affinity tags, in which only cysteine-containing peptides are analyzed (32), N-CLAP can provide further simplification of a peptide mixture by providing a single peptide per protein. Additionally, using this approach, peptide identification can be enhanced by searching databases solely comprising N-terminal peptides (6) against MS/MS spectra, or highly accurate MS spectra when examining small proteomes (Fig. S1).

