Complementary positional proteomics for screening substrates of endo- and exoproteases

Petra Van Damme[1,2], An Staes[1,2], Silvia Bronsoms[3], Kenny Helsens[1,2], Niklaas Colaert[1,2], Evy Timmerman[1,2], Francesc X Aviles[3], Joël Vandekerckhove[1,2] & Kris Gevaert[1,2]

We describe a positional proteomics approach to simultaneously analyze N- and C-terminal peptides and used it to screen for human protein substrates of granzyme B and carboxypeptidase A4 in human cell lysates. This approach allowed comprehensive proteome studies, and we report the identification of 965 database-annotated protein C termini, 334 neo–C termini resulting from granzyme B processing and 16 neo–C termini resulting from carboxypeptidase A4 processing.

C-terminal protein modifications and protein C-terminal processing in general can trigger diverse biological functions and processes such as protein sorting, membrane integration, protein activity and complex formation[1]. Although there are several available methods for N-terminal protein sequencing, deducing C-terminal sequence information is difficult; current methods for C-terminal sequencing include ladder sequencing by chemical degradation[2] and enzymatic degradation by a mixture of carboxypeptidases[3]. These technologies are, however, restricted to the analysis of individual proteins or simple protein mixtures because of incomplete reactions, harsh reaction conditions or lack of specificity, and thus far none of them have been applied to comprehensively analyze whole proteomes. Access to C-terminal peptides in shotgun proteome studies is also limited because C-terminal peptides represent only a small fraction of all peptides in a proteome digest, indicating that enrichment technologies are needed for comprehensive C-terminal proteomic analysis.

We and others have previously reported the use of strong cation exchange (SCX) to enrich protein N- and C-terminal peptides from whole-proteome digests[4,5]. Briefly, enrichment of these peptides starts from protein mixtures that are S-alkylated and in which all primary amines are blocked by (trideutero-)acetylation. Trypsin cleavage then generates α-amino–blocked N-terminal peptides, α-amino–free internal peptides and α-amino–free C-terminal peptides, all of which, except the C-terminal peptides, end with arginine. Passing this peptide mixture over a strong cation exchanger at low pH hence results in the capture of the majority of internal peptides (~98%[5]), whereas blocked N-terminal and α-amine–free C-terminal peptides are not retained because they lack a net positive charge at low pH[5]. This allows the N- and C-terminal peptides to be enriched. In addition, before SCX, the peptide mixture can be treated with glutamyl cyclase and pyrolidone–carboxylic acid peptidase, which in concert remove pyroglutamate, exposing the second residue with a free α-amino group, thus also retaining the former pyroglutamate-containing peptides on the SCX resin[5].

SCX fractionation should in principle suffice to enrich N and C termini for subsequent analysis. However, liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis of reverse-phase high-performance liquid chromatography (RP-HPLC)-separated, SCX-nonretained peptides leads to much fewer identified peptides and proteins than expected, and as described previously[6], this is due to the analysis of crowded analyte fractions, which in our work, in the case of SCX-nonretained fractions, contain both N- and C-terminal peptides. We therefore reasoned that separating C-terminal peptides from N-terminal peptides before LC-MS/MS analysis should increase the number of identifications. As C-terminal peptides differ from N-terminal peptides by the presence of a primary α-amine, we hypothesized that peptide separation could be accomplished by butyrylating these α-amines, thereby increasing the hydrophobicity and thus the reverse-phase column retention of C-terminal peptides (and of remaining internal peptides), while leaving N-terminal peptides unaffected (**Fig. 1**).

We thus report a combined fractional diagonal chromatography (COFRADIC) method in which we first fractionate the terminal peptides from the enriched SCX fraction by RP-HPLC. Next, in each RP-HPLC fraction, peptides are reacted with an *N*-hydroxysuccinimide ester of butyric acid, butyrylating the C-terminal peptides. Upon refractionation of individual RP-HPLC fractions under identical RP-HPLC conditions, the N-terminal peptides elute in the same time interval, whereas butyrylated C-terminal peptides are captured in an interval 4–12 min after the N-terminal peptides. The N- and C-terminal peptides are thus collected in distinct fractions and can be separately analyzed by LC-MS/MS. This COFRADIC strategy with the butyrylation step increased the number of identified MS/MS spectra about sevenfold compared to direct analysis of RP-HPLC–fractionated, SCX-enriched peptides, leading to an about threefold increase in the total number of identified peptides and proteins (data not shown).

[1]Department of Medical Protein Research, Vlaams Instituut voor Biotechnologie, Ghent, Belgium. [2]Department of Biochemistry, Ghent University, Ghent, Belgium. [3]Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Spain. Correspondence should be addressed to P.V.D. (petra.vandamme@vib-ugent.be) or K.G. (kris.gevaert@vib-ugent.be).

RECEIVED 28 JANUARY; ACCEPTED 28 APRIL; PUBLISHED ONLINE 6 JUNE 2010; DOI:10.1038/NMETH.1469

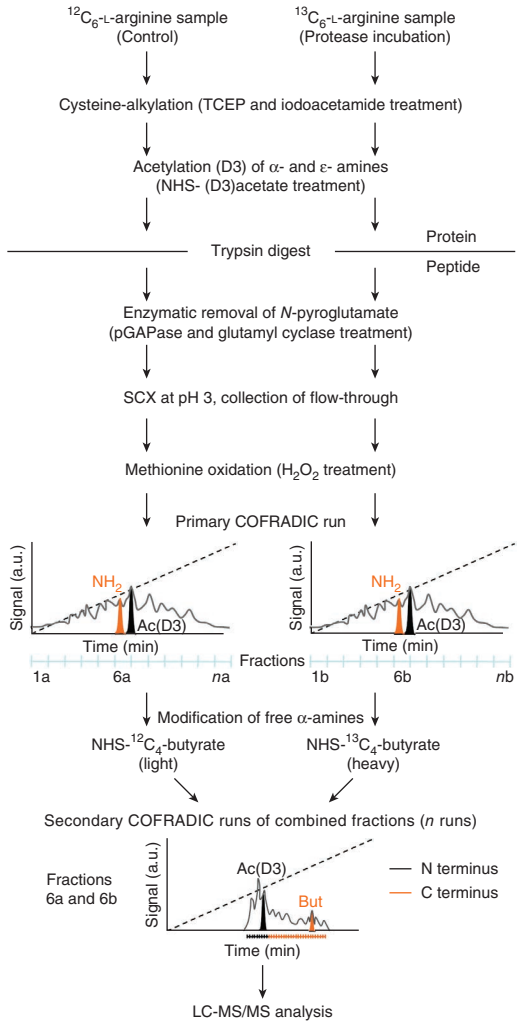

Figure 1 | Outline of the COFRADIC-based complementary positional proteomics procedure. SILAC-labeled cell lysates are incubated with a protease of interest ($^{13}C_6$-L-arginine sample) or left untreated ($^{12}C_6$-L-arginine sample). After protein S-alkylation, reduction with tris (2-carboxyethyl)phosphine (TCEP) and trideutero-acetylation of primary amines, the proteome is digested with trypsin, which now only cleaves after arginine residues. Pyroglutamate residues are enzymatically removed by the combined action of glutamyl cyclase and pyrolidone-carboxylic acid peptidase (pGAPase). Then, SCX enrichment for amino-blocked and C-terminal peptides is performed. Two parallel RP-HPLC runs with equal amounts of peptide material from the control (fractions labeled 1a–*n*a) and the protease-treated proteome (fractions labeled 1b–*n*b) are performed. ("Ac" indicates *in vivo* acetylated N-terminal peptides, and Ac(D3) indicates *in vivo* free, thus *in vitro* trideutero-acetylated peptides. The dashed lines indicate the applied gradient of organic solvent to elute peptides.) Fractionated peptides containing α-amines (C-terminal peptides) are labeled with different isotopic variants of the amine-reactive NHS-butyrate. Corresponding fractions in time (for example, fractions labeled 6a and 6b) are pooled and reseparated by RP-HPLC, upon which butyrylated (But) C-terminal peptides segregate from the N-terminal peptides. In this way, N-terminal peptides, C-terminal peptides or both are selected for LC-MS/MS analysis.

We used this COFRADIC-based complementary positional proteomics approach to study processing by the human endoprotease granzyme B (hGrB) in K-562 cell lysates. A stable isotope labeling by amino acids in cell culture (SILAC; ref. 7) $^{13}C_6$-L-Arg–labeled cell lysate served as a hGrB substrate pool, and a $^{12}C_6$-labeled proteome served as control, allowing quantification of N-terminal peptides and of peptides with new N termini generated by protease activity (neo–N-terminal peptides) that all end in arginine. We also differentially tagged protein C-terminal peptides using *N*-hydroxysuccinimide (NHS) esters of $^{12}C_4$ or $^{13}C_4$ butyric acid (**Figs. 1** and **2a**). After LC-MS/MS analysis, we identified 1,621 database-annotated N termini and 760 database-annotated C termini (**Supplementary Table 1a**,**b**). In addition, we identified 248 neo–C-terminal peptides after processing at aspartic acid, 38 neo–C termini generated by cleavage after a P1 glutamic acid and 48 neo–C termini generated upon processing at other residues. Together, these 334 neo–C termini were found in 294 proteins (**Supplementary Fig. 1** and **Supplementary Table 1c**). Furthermore, we identified 622 hGrB cleavage–generated neo–N termini in 498 proteins (**Supplementary Fig. 1** and **Supplementary Table 1d**). The number of proteins identified by both their database-annotated terminal peptides was only 202 (we identified 2,516 proteins in total based on their (neo-)terminal peptides; **Supplementary Fig. 1**), whereas we identified only 45 cleavage sites (911 different cleavage sites identified) by both of their neo-terminal peptides (**Supplementary Table 2**), illustrating that higher proteome coverage resulted from the simultaneous analyses of both N- and C-terminal peptides.

The zinc-finger HIT domain-containing protein 2, cleaved at Asp150 was an example of a protein for which we obtained complementary processing information. We identified the database-annotated N and C termini of this protein as isotopic doublets, in addition to the neo–C terminus and neo–N terminus proteolytic signature peptides that we identified as singletons (**Fig. 2b**–**f**). The potential of using neo–C termini to identify protease substrates was exemplified by caspase-3: owing to its size, high hydrophobicity and the presence of a SCX-nontolerated histidine, processing of this well-established hGrB substrate at its activation site was not identified in a previous N-terminal proteomics screen[8], but in this work, we documented it by identifying its complementary neo–C terminus, GTELDCGIETD (amino acids 165–175) (**Supplementary Table 1c**). Additionally, we identified 176 cleavage sites (53% of all sites identified by neo-C-terminal peptides) that were technically not assessable with the N terminus–only approach, highlighting the benefit of analyzing both termini (**Fig. 2g** and **Supplementary Table 1c**). iceLogo[9] analysis identified nearly identical hGrB recognition motifs, independent of whether they originated from neo–N or neo–C termini (**Fig. 3**), indicating that no notable biases were introduced when delineating the hGrB consensus cleavage motif by analyzing any of the two types of peptides.

We used a similar setup to identify the substrate repertoire of the carboxypeptidase, human metallocarboxypeptidase CPA4, which is upregulated by histone deacetylase inhibitors during differentiation of prostate epithelial cancer cells[10]. We subjected a PC3 cell lysate to proteolysis by recombinant human carboxypeptidase A4 (hCPA4). By analyzing protein terminal peptides, we identified database-annotated N and C termini of 1,750 and 687 human proteins, respectively (**Supplementary Table 1e**,**f**), leading to the identification of 2,238 proteins, for which we identified both termini of only 199 proteins. In addition, we identified 16 neo–C termini exclusively in the hCPA4-treated setup (**Supplementary Table 3**). For all neo-C termini identified, the

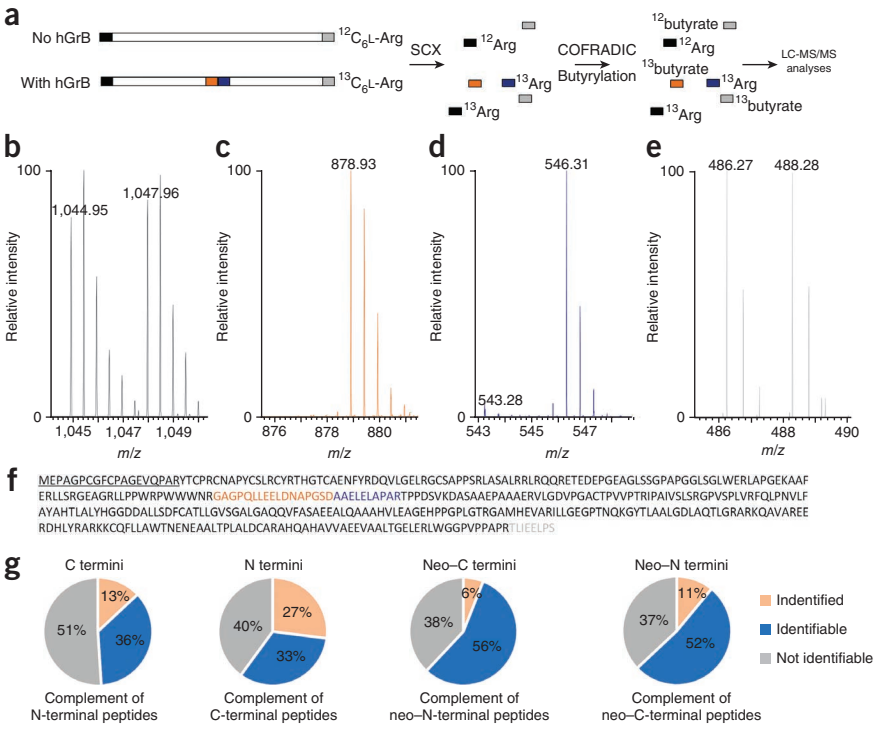

Figure 2 | N- and C-terminal peptides provide complementary data. (**a**) Protein processing with hGrB yielded protein neo–N and neo–C termini (versus only N and C termini in the unprocessed sample), which we isolated by SCX and COFRADIC, and identified by LC-MS/MS. N-terminal, C-terminal, neo–N-terminal and neo–C-terminal peptides are colored in black, gray, blue and orange, respectively. Proteolytic processing events were assigned after differential-mass tagging. (**b**) Mass spectrum of the database-annotated protein N terminus of the zinc-finger HIT domain–containing protein 2, identified as Ac-MEPAGPCGFCPAGEVQPAR (residues 1–19; Ac, *in vivo* α-acetylation), and the ion intensities of the differentially labeled forms were about equal (ratio = 1.05) indicating that this peptide was not affected by hGrB treatment. (**c**) Mass spectrum of the neo–C terminus, But$^{13}C_4$-AGPQLLEELDNAPGSD (residues 134–150; But$^{13}C_4$, α-amino group modified by $^{13}C_4$ butyric acid), which was only generated in the hGrB-treated sample. (**d**) Mass spectrum of the neo-N terminus Ac(D3)-AAELELAPAR (residues 151–160; Ac(D3), α-trideutero-acetylated amino group), which was only generated in the hGrB-treated sample. (**e**) Mass spectrum of the C terminus of the zinc finger HIT domain-containing protein 2, identified as differentially butyrylated TLIEELPS (residues 396–403) peptides with about equal intensities (ratio = 0.96), indicating that this peptide was not affected by hGrB treatment. (**f**) The amino acid sequence of human zinc finger HIT domain–containing protein 2 (Swiss-Prot identifier Q9UHR6) with the identified peptides are indicated. (**g**) Summary of the contribution of the 'identified', 'identifiable but unidentified' and 'unidentifiable' complements of the identified N termini, C termini and hGrB treatment–specific neo–N termini and neo–C termini.

ultimate residue corresponded to the penultimate residue of the protein sequence, indicating that hCPA4 acts as a monocarboxypeptidase. The fact that these 16 hCPA4 sites were picked up in a background of over 2,000 database-annotated protein termini highlights the sensitivity of our method, enabling the detection of rare processing events. In general, we found hydrophobic amino acids such as phenylalanine, isoleucine and leucine at the amino acid C-terminal to the scissile bond (P1′ site), in agreement with the known specificity of the hydrophobic S1′ pocket of hCPA4, which accommodates the substrate's P1′ site (ref. 11). Furthermore, 11 neo-C termini had a lysine residue at one of the two positions N-terminal to the scissile bond (at P1 or P2 sites but never at both sites), and no lysine residue occupied the P3 site, indicating that this carboxypeptidase most likely displays an extended specificity profile as observed in the case of peptidic substrates for metallocarboxypeptidase enzymes[12] and is thereby likely to be highly specific in substrate processing. We also identified database-annotated C termini of 13 CPA4 substrates; for 11 of these, we could calculate the extent of CPA4 proteolysis (**Supplementary Table 3**), information that is often lacking in protease studies. We observed efficient processing, indicated by ratio values of C termini differing significantly ($P \leq 0.01$) from the mean, for the substrates glutaminyl- and valyl-tRNA synthetase

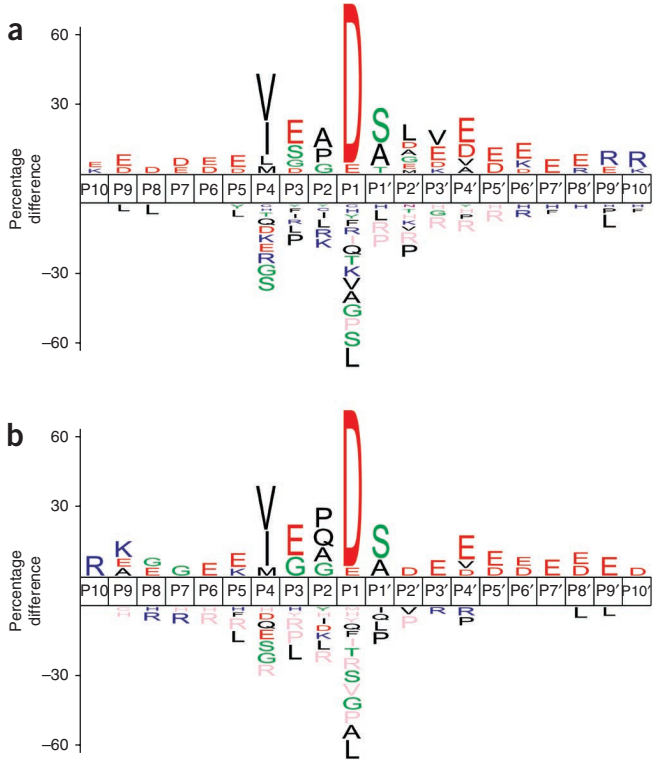

Figure 3 | Sequence analysis of identified hGrB cleavage sites. (**a**,**b**) IceLogo representations of the 334 hGrB cleavage sites identified by analysis of neo–C termini (**a**) and the 622 hGrB cleavage sites identified by analysis of neo–N termini (**b**). Peptide sequences were aligned such that cleavage occurred between position P1 and P1′. Statistically significant residues ($P < 0.01$) were plotted, and amino acids heights are indicative for the degree of conservation at the indicated position. The frequency of the amino acid occurrence at each position in the sequence set was compared with that of the human protein sequences stored in the Swiss-Prot (version 56.0) database. Residues that were completely absent in the identified terminal peptides are shown in pink.

and for the nascent polypeptide–associated complex subunit alpha: for each, we found ~80% of the precursors to be processed (**Supplementary Table 3**).

In both of the above experiments we identified more N-terminal than C-terminal peptides, though *in silico* simulation predicts a similar number of identifiable features. We predicted the theoretical overlap of identifying proteins by their terminal peptides by *in silico* digesting the human proteome (Swiss-Prot database version 57.6) C-terminal to arginines and allowing for digestion to occur between the arginine-proline bonds (**Supplementary Fig. 2**). In this model, database-annotated N- and C-terminal peptides with a calculated mass of 600–4,000 Da we considered identifiable. Of 20,332 protein entries, we found 12,904 (63.4%) to be identifiable using N-terminal peptides and 12,792 (62.9%) by C-terminal peptides. Of these, only 8,141 (40%) can be identified by both types of termini. Up to 86.3% of all human proteins were found identifiable by sampling N- and C-terminal peptides. In general, we found higher Mascot scores for N-terminal peptides compared to C-terminal peptides (**Supplementary Fig. 3**), suggesting that alternative tagging strategies that improve fragmentation might be valuable alternatives to butyrylation.

In conclusion, we developed a COFRADIC-based method to enrich both N- and C-terminal peptides from whole-proteome digests and found that both types of peptides were highly complementary for positional proteomics studies by increasing the overall fraction of a proteome that was identifiable. We also found that complementary positional proteomics is ideally suited to comprehensively analyze the substrates of endoproteases and carboxypeptidases. In this respect, whenever complementary neo-termini are identified (that is, a neo–C-terminal and a neo–N-terminal peptide following each other in a protein sequence), downstream trimming of proteolytic protein fragments by exopeptidases can be ruled out, thus pointing to primary processing events and refining the quality of reported protease substrate specificities.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Accession codes. Proteomics Identifications database (PRIDE)[13]: 12151, 12152, 12153, 12154, 12155, 12156 and 12157 (identified MS/MS spectra).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

P. Bird (Monash University) kindly provided recombinant hGrB. P.V.D. was supported by a postdoctoral fellow grant of the Research Foundation, Flanders (FWO-Vlaanderen). K.H. is supported by a PhD grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This work was supported by research grants from the Fund for Scientific Research, Flanders (Belgium) (project numbers G.0156.05, G.0077.06 and G.0042.07), the Concerted Research Actions (project BOF07/GOA/012) from Ghent University and the Interuniversity Attraction Poles (IUAP06). F.X.A. and S.B. acknowledge support from grants LSH6-2006-018830-CAMP (EC) and BIO2007-68046 (Ministerio de Ciencia y Innovación, Spain).

AUTHOR CONTRIBUTIONS

P.V.D., S.B., F.X.A., J.V. and K.G. designed the experiments. P.V.D., A.S. and E.T. generated the data. P.V.D., K.H. and N.C. analyzed the data. P.V.D., F.X.A., J.V. and K.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Chung, J.J., Shikano, S., Hanyu, Y. & Li, M. *Trends Cell Biol.* **12**, 146–150 (2002).
2. Boyd, V.L., Bozzini, M., Zon, G., Noble, R.L. & Mattaliano, R.J. *Anal. Biochem.* **206**, 344–352 (1992).
3. Samyn, B., Sergeant, K., Castanheira, P., Faro, C. & Van Beeumen, J. *Nat. Methods* **2**, 193–200 (2005).
4. Dormeyer, W., Mohammed, S., Breukelen, B., Krijgsveld, J. & Heck, A.J. *J. Proteome Res.* **6**, 4634–4645 (2007).
5. Staes, A. *et al. Proteomics* **8**, 1362–1370 (2008).
6. Washburn, M.P., Wolters, D. & Yates, J.R. III. *Nat. Biotechnol.* **19**, 242–247 (2001).
7. Ong, S.E. *et al. Mol. Cell. Proteomics* **1**, 376–386 (2002).
8. Van Damme, P. *et al. Mol. Cell. Proteomics* **8**, 258–272 (2009).
9. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. *Nat. Methods* **6**, 786–787 (2009).
10. Huang, H. *et al. Cancer Res.* **59**, 2981–2988 (1999).
11. Pallares, I. *et al. Proc. Natl. Acad. Sci. USA* **102**, 3978–3983 (2005).
12. Bayes, A. *et al. Biochemistry* **46**, 6921–6930 (2007).
13. Martens, L. *et al. Proteomics* **5**, 3537–3545 (2005).

ONLINE METHODS

Cell culture. Human K-562 cells (American Type Culture Collection (ATCC)) were grown in RPMI 1640 glutamax medium (Invitrogen) containing either 57.5 μM of natural or $^{13}C_6$ L-arginine (Cambridge Isotope Labs). Media were supplemented with 10% dialyzed FBS (Invitrogen), 100 units ml^{-1} of penicillin and 100 μg ml^{-1} of streptomycin. Human PC3 cells (ATCC) were grown in DMEM/F12 glutamax medium (Invitrogen) containing 10% FBS, 100 units ml^{-1} of penicillin and 100 μg ml^{-1} of streptomycin. All cells were grown at 37 °C and 5% CO_2. Human K-562 cells were passaged for at least six population doublings for complete incorporation of the labeled arginine.

Sample preparation and recombinant hGrB or hCPA4 treatment. K-562 cells were washed in PBS and resuspended at 7×10^6 cells ml^{-1} in 50 mM Tris-HCl (pH 7.9) and 100 mM NaCl containing the caspase-inhibitors z-DEVD-fmk and z-VAD-fmk (each at 100 μM). Cells were incubated with these inhibitors for 15 min at 37 °C and then subjected to three rounds of freeze-thaw lysis. Lysates were cleared by centrifugation and equal amounts of protein were either left untreated ($^{12}C_6$ L-arginine–labeled sample) or treated with 200 nM recombinant hGrB ($^{13}C_6$ L-arginine–labeled sample) for 1 h at 37 °C. Recombinant hGrB was expressed and purified from *Pichia pastoris*, and its activity was determined as described previously[14,15].

PC3 cells were washed in PBS and resuspended at 9×10^6 cells ml^{-1} in 50 mM Tris-HCl (pH 7.5) and 100 mM NaCl, and then subjected to three rounds of freeze-thaw lysis. The lysate was cleared by centrifugation, and equal amounts of proteome material were either left untreated or treated with 25 nM recombinant human carboxypeptidase A4 (hCPA4) for 1.5 h at 25 °C. hCPA4 was produced and purified as described previously[11].

Isolation of terminal peptides. After protease incubation, guanidinium hydrochloride was added to the cell lysates to a final concentration of 4 M to denature and inactivate proteases. The protein samples were then reduced, S-alkylated and subjected to tri-deuteroacetylation of primary amines and trypsin digestion, as described previously[5,16]. The so-generated peptide mixtures were vacuum dried. Peptides starting with pyroglutamate were unblocked before SCX prefractionation: 25 μl of pGAPase (pyroglutamyl aminopeptidase, 25 U ml^{-1}) (TAGZyme kit; Qiagen) was activated for 10 min at 37 °C by addition of 1 μl of 50 mM EDTA (pH 8.0), 1.25 μl of 800 mM NaCl and 11 μl of freshly prepared 50 mM cysteamine-HCl. To the pGAPase solution, 25 μl of glutamine cyclotransferase (Qcyclase; 50 units ml^{-1}; TAGZyme) was added. The dried peptides were redissolved in 212 μl of buffer containing 16 mM NaCl, 0.5 mM EDTA, 3 mM cysteamine and 50 μM aprotinin. The activated pGAPase and Q-cyclase mixture was added to the sample, and the mixture was incubated for 60 min at 37 °C (total volume of 275 μl). Then, 275 μl acetonitrile was added, and the sample was acidified to pH 3 using a stock solution of 1% trifluoroacetic acid (TFA) in 50% acetonitrile. The sample was diluted with 10 mM sodium phosphate in 50% acetonitrile to a final volume of 1 ml. SCX enrichment of α-amino–blocked N-terminal peptides and C-terminal peptides was performed as described previously[5].

The SCX fraction containing the N- and C-terminal peptides was dried *in vacuo* and re-dissolved in 100 μl of HPLC solvent A (10 mM ammonium acetate; pH 5.5). To prevent oxidation of methionine between the primary and secondary RP-HPLC separations (and thus unwanted shifting of methionyl peptides[17]), methionines were oxidized before the primary run by adding 2 μl of 30% (wt/vol) H_2O_2 (final concentration of 0.06%) for 30 min at 30 °C. We injected 100 μl of this peptide mixture (an equivalent of about 300 μg of digested proteins) onto an RP column (Zorbax 300SB-C18 Narrowbore, 2.1 mm (internal diameter) × 150 mm length, 5 μm particles; Agilent Technologies) and the RP-HPLC separation was performed as described previously[5]. Fractions of 4 min were collected from 20 to 80 min after sample injection. These primary fractions were dried and redissolved in 20 μl of 100 mM sodium phosphate (pH 8.0). Then 20 μl of a freshly made solution of NHS-butyric acid (60 mM in 100 mM sodium phosphate; pH 8.0) was added ($^{12}C_4$-butyrylation was used for the control sample and $^{13}C_4$-butyrylation for the hCPA4- or hGrB-treated sample) and acylation of primary amines was allowed for 2 h at 30 °C. The reaction was stopped by adding 12 μl of 1 M glycine and O-acetylation was reversed by heating the peptide mixture at 95 °C for 60 min[18]. Corresponding primary fractions from the two parallel primary RP-HPLC runs (the control and hGrB-treated sample) were pooled and acidified with 4 μl of acetic acid. The butyrylated fractions were reseparated on the same RP column using the same gradient as during the primary run. Secondary fractions, highly enriched for N-terminal peptides, were collected during the same time interval as that used for the primary fraction but by using eight 30-s collection intervals, and the secondary fractions highly enriched for C-terminal peptides were collected up to 12 min after the primary collection interval, again using 30-s collection intervals. This resulted in 32 collected fractions per secondary run and 480 fractions in total. To reduce LC-MS/MS analysis time, secondary fractions eluting 12 min apart were pooled, dried and redissolved in 40 μl of 2% acetonitrile. In total, 120 samples were analyzed by LC-MS/MS.

LC-MS/MS analysis. LC-MS/MS analysis was performed using an Ultimate 3000 HPLC system (Dionex) in-line connected to a LTQ Orbitrap XL mass spectrometer (Thermo Electron). Per LC-MS/MS analysis, 5 μl of sample was consumed. LC-MS/MS analysis and generation of MS/MS peak lists were performed as described previously[19]. These MS/MS peak lists were then searched with Mascot using the Mascot Daemon interface (version 2.2.0; Matrix Science). The Mascot search parameters were set as follows. Searches were performed in the Swiss-Prot database with taxonomy set to human (database version 56.4; 20,328 human protein sequences). Trideutero-acetylation at lysines, carbamidomethylation of cysteine and methionine oxidation to methionine-sulfoxide were set as fixed modifications. Variable modifications were trideutero-acetylation and acetylation of protein N termini. Arg-C (for identification of database-annotated termini) or semi-ArgC (for identification of neo-termini) were set as the used protease (one missed cleavage was allowed) and the mass tolerance on the precursor ion was set to 10 p.p.m. and on fragment ions to 0.5 Da. In addition, Mascot's C13 setting was set to 1. All quantifications (SILAC ($^{12}C_6$ L-arginine versus $^{13}C_6$ L-arginine) and butyrylation ($^{12}C_4$-butyrylated versus $^{13}C_4$-butyrylated) (hGrB proteome) or butyrylation alone (hCPA4 proteome)) were carried out using the Mascot Distiller Quantitation tool (version 2.2.1). The quantification method

doi:10.1038/nmeth.1469

details were as follows: constrain search, yes; protein ratio type, average; report detail, yes; minimum peptides, 1; protocol, precursor; allow mass time match, yes; allow elution shift, no; all charge states, yes; and fixed modifications, mass values. Ratios for the proteins were calculated by comparing the extracted ion chromatogram peak areas of all matched light peptides with those of the heavy peptides, and the results were all verified by visual inspection of mass spectrometry spectra.

Synthesis of N-hydroxysuccinimide esters. An NHS ester of trideutero-acetate (Sigma-Aldrich) was synthesized as described previously[20]. An NHS ester of ($^{13}C_4$)-butyrate (Sigma-Aldrich) was synthesized as follows: 5 mmol of NHS (Sigma-Aldrich) was mixed with 5 mmol of sodium$^{13}C_4$-butyrate dissolved in 15 ml of acetone. Then 20 mmol of sodium bicarbonate was added to the mixture, and the mixture was vortexed thoroughly. To the reaction mixture, we added drop-wise 6 mmol of diphenyl phosphoryl chloride dissolved in 5 ml of acetone and mixed for 24 h at 25 °C. To this reaction mixture, 25 ml of acetone was added, and the mixture was stirred at 50 °C for 3 h. The resulting white suspension was vacuum dried. Thirty milliliters of dichloromethane was added to the precipitate to extract the NHS-ester. The dichloromethane fraction was then taken away and pooled with the second extraction using 5 ml of dichloromethane. By the addition of 40 ml of water and multiple up-and-down pipetting, the pooled dichloromethane fractions were subjected to a water-purging step to dissolve and remove any remaining sodium bicarbonate. The watery phase from the biphasic water-dichloromethane system was subsequently discarded, and the remaining dichloromethane fraction was vacuum-dried. About 200 mg of NHS–$^{13}C_4$-butyrate was obtained.

14. Sun, J. *et al. J. Biol. Chem.* **279**, 16907–16911 (2004).
15. Van Damme, P. *et al. Nat. Methods* **2**, 771–777 (2005).
16. Sun, J. *et al. Biochem. Biophys. Res. Commun.* **261**, 251–255 (1999).
17. Gevaert, K. *et al. Mol. Cell. Proteomics* **1**, 896–903 (2002).
18. Abello, N., Kerstjens, H.A., Postma, D.S. & Bischoff, R. *J. Proteome Res.* **6**, 4770–4776 (2007).
19. Ghesquiere, B. *et al. Mol. Cell. Proteomics* **8**, 2642–2652 (2009).
20. Van Damme, P. *et al. BMC Proc.* **3** (Suppl. 6), S6 (2009).