Identification of protease cleavage-sites in *E. coli* lysate by differential N-terminomic analysis

CURRENT STATUS: POSTED

John C. Timmer
Biomedical Sciences Graduate Program at University of California San Diego. Burnham Institute for Medical Research.

Guy S. Salvesen
Burnham Institute for Medical Research.

DOI:
10.1038/nprot.2009.169

SUBJECT AREAS
Computational biology and bioinformatics *Biotechnology*

KEYWORDS
N-terminomics, protease, substrate, cleavage, cleavage-site, kinetics

Introduction

N-terminomics is an approach that identifies N-terminal peptides originating from protease cleavage-sites [1]. This technique has been instrumental for the investigation of the substrate structures cleaved by proteases. The protocol presented here describes how purified proteases can be applied to the *E. coli* proteome and screened for substrate cleavage-sites using N-terminomics and the companion software N-TerProt. This analysis can yield important amino acid specificity information about proteases, as well as their structural tolerance.

Reagents

E. coli K12 MG1655

Difco 2x yeast extract tryptone medium (BD)

Stericup & Steritop 0.45 micro m HV Durapore Membrane (Millipore)

SealRite microcentrifuge tubes 2.0 mL (USA Scientific Inc.)

SealRite microcentrifuge tubes 1.5 mL (USA Scientific Inc.)

sucrose (MP Biomedicals LLC)

guanidine HCl 98% (Acros Organics)

dithiothreitol (DTT) (Bio Vectra)

iodoacetamide (Sigma)

o-methylisourea hemisulfate (Acros Organics)

sodium hydroxide (NaOH) 97% (Aldrich)

hydrochloric acid (HCl) (Fisher)

urea (Mallinckrodt)

AG 501-X8 resin (BioRad)

0.22 micro m filter (GE Healthcare)

4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Fisher)

sodium chloride (Fisher)

PD-10 columns G-25M (GE Healthcare)

Ammonium Bicarbonate (AmmBic) (Acros Organics)

C18 Sep-Pak Vac 6cc cartridges (Waters)

tris(2-carboxyethyl)phosphine (TCEP) (Invitrogen)

trifluoracetic acid (TFA) (Sigma)

sequence grade modified trypsin (Promega)

high capacity neutravidin agarose resin (Thermo)

dimethylformamide (DMF) (Fisher)

sulfo NHS-SS-biotin (Pierce)

hydroxylamine (Sigma)

Disposable 2 mL polystryene columns (Pierce)

isopropanol (Fisher)

SafeSeal microcentrifuge tubes 1.7 mL (Sorenson BioScience, Inc.)

acetonitrile (MeCN) (Alfa Aesar)

Costar 96-well flat bottom microtiter plate (Corning)

GelCode Blue (Thermo)

Equipment

orbital shaking incubator

1 L baffled glass Erlenmeyer flasks

RC 5B superspeed centrifuge (Sorvall)

SLA-3000 rotor (Sorvall)

SS-34 rotor (Sorvall)

sonifier (Branson)

pH meter with small electrode (Beckman)

heat block (VWR)

tabletop microcentrifuge (Eppendorf)

LC-MS/MS system

Macintosh computer with OS 10.4 or 10.5

Microsoft Excel for Mac 2004 or 2006

Script Editor version 2.1.2 or 2.2.1

Gel imaging station with densitometry capability

Procedure

A. Prepare *E. coli* lysate

1. Streak out E. coli K12 MG1655 on 2xYT plates with no selection for isolation, and grow overnight at 37^{o}C.
2. Pick a well-isolated colony into 50 mL of 2xYT medium. Grow at 37^{o}C with shaking at about 225 revolutions per minute overnight.
3. Innoculate two 500 mL cultures with 10 mL of the overnight culture. Grow at 37^{o}C with shaking, and monitor the optical density at 600 nm.
4. When the optical density at 600 nm equals 1.0 then harvest the cultures in an RC 5B centrifuge at 4^{o}C with the SLA-3000 rotor at 5000 rotations per minute (rpm) (about 2,000 relative centrifugal force (rcf)) for 10 minutes to pellet the bacteria.
5. Decant the media, and scrape the pellets into a 50 mL conical tube.
6. Add 30 mL of ice cold buffer that is appropriate for the protease cleavage reaction (such as 50 mM HEPES including 10% sucrose to maintain natively folded proteins.
7. Resuspend the pellet by vortexing.
8. Sonicate with a large probe with 50% intensity and 50% duty cycle for 2 minutes. Keep resuspended bacteria on ice at all times.
9. Transfer sonicated bacterial lysate into a 35 mL polypropylene tube. Pellet the insoluble material in an RC 5B centrifuge at 4^{o}C with an SS-34 rotor at 12,000 rpm (~10,000 rcf) for 30 minutes.
10. Clarify the supernatant with a Stericup 0.45 micro m vacuum filter (Millipore).
11. Measure the protein concentration (Bradford for example). The total protein should be between 5 to 20 mg/mL for a good prep.

12. Aliquot 1 mL of lysate into 1.5 mL microcentrifuge tubes, and store at -80^{o}C.

B. Prepare protease treated and control samples

1. Determine appropriate conditions for the protease cleavage reaction that are appropriate for the question being asked. Studies of limited proteolysis should seek conditions that capture the most susceptible cleavage events (do not to over-digest proteins in a degrading fashion). These conditions will likely need to be determined empirically. A general approach is to prepare several cleavage reactions with 10-fold dilutions of the protease of interest (I found that 1 microM protease is a nice 'high' concentration for a 1 hour cleavage reaction using caspase-3 and Staphylococcal GluC). Control samples should be run at the same time and may include some or all of the following; lysate alone, a catalytically inactive mutant of the protease, or a pre-inhibited protease using a chemical or protein inhibitor.
2. Setup the cleavage reaction such that the final volume is about 1 mL final volume, having a concentrated protease in 1x cleavage buffer, and the *E. coli* lysate in 1x cleavage buffer.
3. Prepare samples with 1 to 10 mg of *E. coli* lysate in 1x cleavage buffer, and preincubate at the temperature of the cleavage reaction (typically 37^{o}C) for ~15 minutes.
4. Preincubate protease in 1x cleavage reaction buffer at the temperature used for the cleavage reaction for ~15 minutes.
5. Start cleavage reaction by adding the protease to the *E. coli* lysate, mix gently by inverting several times, and maintain at the desired temperature for a predetermined time.
6. Weigh out dry guanidine HCl for a 6M solution, assuming the final volume will be ~1.65 times the cleavage reaction volume (a 1 mL cleavage reaction will turn into

1.65 mL after the addition of 0.946 g of guanidine HCl).

7. Stop the reaction by adding the dry guanidine HCl to the cleavage reaction, add DTT to 10 mM final concentration, vortex well, and incubate on a 95^{o}C heat block for 10 minutes to denature proteins and inactivate proteases.

C. N-terminomic sample preparation

1. Cool denatured sample to room temperature.
2. Alkylate cysteins with 30 mM iodoacetamide in the dark at room temperature for 30 minutes.
3. Guanidinate lysine side chains by adding dry o-methylisourea hemisulfate to 0.5 M final concentration.
4. Adjust the pH to greater or equal to 10.3 with 5 N NaOH, monitor with a small electrode probe.
5. Incubate at 4^{o}C for ~16 hours.
6. Prepare an appropriate volume of 9.1 M urea (~500 mL).
7. Add 2 or 3 large scoops of AG 501-X8 resin to the urea solution, and stir for >=30 minutes.
8. Clarify urea with a 0.22 micro m filter.
9. Add 1M stock HEPES pH 7.8 to the urea solution making it 50 mM HEPES, and 8 M urea (8 M urea buffer).
10. Equilibrate PD-10 columns with 25 mL of 8M urea buffer.
11. Load sample onto PD-10 column (a smaller volume is better, try to load <=1.5 mL of sample).
12. Add 8 M urea buffer to 2.5 mL total volume including the sample.
13. Collect the next 2.5 mL coming off the PD-10 column (load 2.5 mL of 8 M urea buffer).

14. Label N-terminal amines with 5 mM sulfo NHS-SS-biotin (dissolve 100 mg into 1 mL of DMF making a 165 mM stock, and store at -20^{o}C) for 1 hour at 37 ºC.
15. Stop reaction by adding 1 M stock AmmBic pH 7.8 to 50 mM final concentration.
16. Reverse ester side reactions with Ser and Thr residues by adding 1 M stock hydroxylamine to 40 mM final concentration, and incubate at 37^{o}C for 15 minutes.
17. Buffer exchange to 8 M urea buffer with PD-10 columns as before. The volume of the sample will be greater than 1.5 mL, so split it into two separate buffer exchange runs on the PD-10 column. Load the sample, add buffer to 2.5 mL. Collect the next 2.5 mL.
18. Dilute the sample to 2 M urea with 50 mM AmmBic buffer (Although this results in a large sample volume, it minimizes the precipitation of proteins in the sample as they transition back to a more 'non-denaturing' condition.
19. Add 5 to 10 micro g of modified sequence grade trypsin to digest proteins into peptides. Incubate at 37^{o}C overnight.
20. If any flocculant is visible, spin down the insoluble material at high speed (10,000 rcf) for 5 minutes at room temperature.
21. Calculate the amount of high capacity neutravidin resin needed for all samples. Generally 500 micro L is sufficient for each sample.
22. Wash the high capacity neutravidin resin with 2 M urea buffer.
23. Add high capacity neutravidin resin to each sample. Neutate or rock on a rocker for 30 minutes at 4^{o}C to bind biotinylated peptides.
24. Load samples on disposable 2 mL polystyrene columns (Pierce).
25. Wash extensively with 2 M urea buffer 10 times.
26. Wash with 50 mM AmmBic pH 7.8 buffer 3 times.
27. Transfer resin to a 2 mL microcentrifuge tube in 50 mM AmmBic buffer.

28. Spin down resin quickly, and remove buffer to the 1 mL marking.

29. Add 10x TCEP to a final concentration of 5 mM. Incubate at 37^{o}C with rocking for 30 minutes.

30. Collect eluted peptides by loading the sample on to MicroBio spin columns and taking the flow through. Flush the resin with 500 µL of 50 mM AmmBic buffer.

31. Cleanup peptides using C_{18} Sep-Pak Vac 6cc cartridges. Condition the columns with 6 mL of isopropanol, equilibrate with 6 mL of 0.1% TFA, and load samples.

32. Wash 5 times with 2 mL of 4.25% MeCN, 0.1% TFA.

33. Elute 5 times with 1 mL of 50% MeCN, 0.1% TFA. Samples are usually only in the third elution.

34. Dry samples in a vacuum centrifuge at room temperature.

35. Solubilize peptides in 40 micro L of 2% MeCN, 0.1% TFA, and store at 4^{o}C until analysis by LC-MS/MS.

D. Sample analysis by nano-LC MS/MS

1. First separate peptides by liquid chromatography prior to MS/MS. Reverse phase chromatography has performed very well for this purpose. Although multi-dimensional chromatography approaches may improve the number of non-redundant peptide identifications, preliminary studies utilizing an upfront ion exchange and subsequent reverse phase separation have not improved the overall number of peptides identified.

2. Load 5-8 micro L of sample onto the reverse phase column at 10 micro L per minute, then wash with solvent A (2% MeCN, 0.1% TFA), and eluted over a two hour linear gradient from 10-60% solvent B (80% MeCN, 0.1% TFA).

3. Elute directly into the mass spectrometer. Both the LTQ and the Orbitrap LTQ MS/MS instruments perform well for peptide identification, although other MS/MS

instruments will almost certainly work as well. Run the instrument with a duty cycle consisting of a single MS scan followed by 4 MS/MS scans. Enable dynamic exclusion to manage highly abundant peptides.

4. Run each sample 3 or more times to increase the number of non-redundant peptide identifications and improve the spectral counts, which serve as a semi-quantitative reference when comparing samples with different concentrations of protease.

E. Database searching

1. Combine the MS/MS spectra from repeat runs of the same sample. Analyze these spectra with Sorcerer SEQUEST using the following modifications. N-terminal cleaved biotin stub = 88 Da (variable)

Cys carboxyamidomethylation = 57 Da (fixed)

Lys guanidination = 42 Da (variable)

Met oxidation = 16 Da (variable)

2. Use a concatenated forward and reverse database to estimate the false discovery rate (reference 2), and make sure to derive the database from the SwissProt proteome database of the pertinent organism. Use a semi-enzyme (usually semi-tryptic) database so as to identify protease cleavage-sites that are not Lys/Arg cleavages.

F. Annotate N-terminal peptides with N-TerProt software

1. Export the Peptide Prophet results into Excel spreadsheets.

2. Add 3 additional columns to each sample dataset with the following headers and content: 'sorcerer', and fill down the unique numerical sample identifier number.

'sample', and fill down the unique sample description (shorter is better).

'enzyme', fill down the enzyme used to generate peptides for MS/MS analysis (usually trypsin).

3. Download 'N-TerProt' from "http://www.burnham.org/labs/Salvesen/Salvesen%20Lab%20-%20Links.html":http://www.burnham.org/labs/Salvesen/Salvesen%20Lab%20-

%20Links.html, or contact "gsalvesen@burnham.org":gsalvesen@burnham.org Save the folder 'N-TerProt' to: 'Macintosh HD/users/(username)/library/scripts'.

Includes a test dataset 'Sample data'.

4. Open the file 'N-TerProt V3.0'
5. Run the N-TerProt script, and click 'YES' when prompted to run 'Curate Probability & Xcorr'. This will condense the dataset to peptides with a probability score of 0.8 or more and cross correlation of 2.0 or less and copy the resulting information into a new spreadsheet. These parameters can be modified in the script by going to the N-TerProt folder and opening the file '2009.07.19 curate prob & xcorr'. Simply change the value for the probability score cutoff variable 'probCutOff', or the cross correlation cutoff variable 'xcorrCutoff'.
6. Run the N-TerProt script, and click 'YES' when prompted to run 'Fix unlabelled N-terminal artifacts'. This corrects an error resulting from lower resolution LTQ data and not seen with Orbitrap LTQ data, whereby peptides from protein N-termini that have their initiator Met removed and are labeled with the cleavable biotin tag are identified by SEQUEST as a Met containing peptide that is unlabelled. This is because the mass difference is not distinguishable within the accuracy of the instrument.
7. Run the N-TerProt script, and click 'YES' when prompted to run 'Peptide Stuff'. This script creates some new columns with more peptide information that the script will use later.
8. Run the N-TerProt script, and click 'YES' when prompted to run 'Weird NT & CT Analysis'. This part looks for non-enzyme (usually non-tryptic) ends of peptides.
9. Run the N-TerProt script, and click 'YES' when prompted to run 'NT Index'. This script makes an index of the non-redundant N-termini, the non-redundant peptides, and the spectral counts per non-redundant N-terminus.

10. Run the N-TerProt script, and click 'YES' when prompted to run 'New ES'. This part returns columns displaying the samples that each peptide are identified in, as well as the enzyme(s) (trypsin, GluC, etc), and the number of spectral counts for each respective sample and enzyme.
11. Optional: run the N-TerProt script, and click 'YES' when prompted to run 'Enzyme & Sample column separation'. This will separate the columns derived in the last script into individual columns for each sample/enzyme with the respective spectral counts.
12. Run the N-TerProt script, and click 'YES' when prompted to run 'Swiss-Prot Stuff'. This script gathers information about the substrate and cleavage-site from SwissProt, and deposits it in several new columns. The 'protein name' is hyperlinked to each SwissProt entry, 'FT & P1 list' is a list of all the potential annotated proteolytic events for that protein, 'protein sequence' is the amino acid sequence of that protein, 'obs P1' is the amino acid number of the P1 residue, 'obs FT' is the proteolytic feature that matches the 'obs P1' residue (if any), 'cut site' is the P4-P4^{I} amino acid sequence of the cleavage-site with a hyphen at the scissile bond, 'group' shows the annotated cleavage features as they are or as 'Unannotated', 'annotation' is the confidence level that SwissProt has designated for the annotated proteolytic event observed (if any).
13. Run the N-TerProt script, and click 'YES' when prompted to run 'Get Category'. This will further categorize the 'Unannotated' cleavage-sites. For example, a sample that has been treated with trypsin and searched against a semi-tryptic database will produce 3 types of unascribed peptides. 'Bonus-1' peptides have a non-tryptic N-terminus and a tryptic C-terminus, 'Bonus-2' peptides are fully tryptic, and 'Mystery-1' peptides are tryptic at the N-terminus, and non-tryptic at the C-terminus. When searching a non-enzyme database an additional category emerges, 'Mystery-2' that is

non-tryptic at both ends. When analyzing samples with two different proteases, such as tyrpsin and GluC, then a small fraction of peptides will be derived from the same cleavage-site at the N-terminus, but have a different C-terminus. These N-termini are called 'Bonus-0'. These sub-categories help to assess the confidence in unascribed cleavage-sites.

14. Run the N-TerProt script, and click 'YES' when prompted to run 'Find unascribed sites'. This part simplifies the annotation to 'Unascribed' or 'Annotated'.
15. Run the N-TerProt script, and click 'YES' when prompted to run 'Negative Control'. This shows which cleavage-sites are also found in the negative control samples, and which ones are specific to the protease treated samples. The default controls are 'L' for lysate, 'M' for catalytic mutant, and 'I' for inhibited protease; however, customized control samples can be added to the script. Go to the N-TerProt folder and open the file named '2008.09.02 No Neg Control'. Add new control descriptions to the variable list 'negList' keeping the text in quotes.
16. Run the N-TerProt script, and click 'YES' when prompted to run 'Unascribed Frequency Distribution'. This script creates a new worksheet with the percent frequency distribution of each amino acid for unascribed cleavage-sites in the experimental vs shared & control samples. This data can then be plotted in a bar graph format to distinguish amino acids enriched in the protease treated sample. The script default is to analyze the P1 position, but can be modified to scan any position from P4 to P4’ by going to the N-TerProt folder and open the file named '2009.07.17 UFD'. Change the variable 'myPosition' in the third line to the desired position keeping the text in quotes.
17. Optional: run the N-TerProt script, and click 'YES' when prompted to run 'Statistics'. This script makes a new spreadsheet with the number of spectra and non-redundant

N-termini identified for each category, either annotated or unascribed, and the total tally as well.

18. All of the unascribed cleavage-sites not found in control samples ('OK' in the 'neg control' column) are candidate substrate cleavage-sites for the protease used in the cleavage reaction. Additional information known about the specificity of the protease of interest may be used to select candidates for biochemical validation and kinetic measurement.

G. Validation of cleavage-sites by in vitro cleavage assay

1. Clone candidate protease substrates into an amenable expression vector for expression in *E. coli*.
2. Express and purify candidate substrates proteins aiming for a protein concentration of 5 micro M. Elute proteins in buffer compatible with the cleavage reaction buffer.
3. Perform the cleavage reaction with the maximum protease concentration 5-10 fold greater than the conditions in the original sample used for the N-terminomics analysis, while maintaining the time, temperature, and buffer conditions. Assay protease concentrations by 1/3 dilution, and including both substrate alone, and maximum protease alone. The 1/3 dilution is superior to smaller dilution factors because it allows for more points to be used for the uncleaved precursor value. Inacurate measurement of the full-length uncleaved substrate can dramatically alter the denominator when determining the enzyme concentration required for 50% cleavage.
4. Prepare the cleavage reaction in a Costar 96-well flat bottom microtiter plate (Corning) in a volume of 100 micro L.
5. Add substrate to appropriate wells, then start the cleavage reaction by adding the protease dilution series to the respective wells.
6. Stop the cleavage reaction by adding 50 micro L of 3x sample buffer with 30 mM

DTT, and denaturing on a 95^{o}C heat block for 10 minutes.

7. Visualize cleavage reaction by loading 10-15 micro L per lane on gel for SDS-PAGE.
8. Stain protein (GelCode Blue, or coomassie), and destain as directed.
9. Confirm that the anticipated cleavage products match the banding pattern of the cleavage assay. If the banding pattern does not match, then the substrate may be cleaved at an additional site. If the substrate is not cleaved under the conditions assayed, try increasing the protease concentration as the N-terminomics analysis can be much more sensitive than SDS-PAGE. It is possible that the substrate will not be cleaved in the cleavage assay due to incomplete sampling. Protein-protein interactions in the lysate may also conformationally alter the substrate such that it will no longer be susceptible to proteolysis in a purified in vitro cleavage assay.
10. Measure the full-length substrate in all lanes by densitometry to determine the enzyme concentration required for 50% cleavage ($E_{1/2}$).
11. Plot the densitometry and determine the $E(_{1/2})$ value.
12. Calculate kobs value from the following equation where ln2 is the natural log of 2, t is time in seconds, and $E_{1/2}$ is the enzyme concentration needed for half of the substrate to be cleaved: $k_{obs} = \ln 2/(t*E_{1/2})$

Timing

2 or more weeks depending on the mass spec turn around time

Troubleshooting

Insufficient peptides identifications from the N-terminomic analysis.

1. Make sure the proteins/peptides are not being lost at any stage. Save portions after steps with potential losses (buffer exchanges, peptide cleanup by SepPak). Quantitate proteins by Bradford assay, and sulfhydryl groups of the cleaved biotin tag for peptides by DTNB assay. Expect about 100 µg of eluted peptide from the C_{18}

SepPak column.

2. Run peptide standards to make sure the LC-MS/MS system are operating optimally.

Anticipated Results

The anticipated proportion and number of spectra and non-redundant N-termini can be seen on the 'statistics' worksheet after running N-TerProt on the 'Sample data'.

References

1 J. C. Timmer, M. Enoksson, E. Wildfang et al., Biochem J 407, 41 (2007).

2 J. E. Elias, W. Haas, B. K. Faherty et al., Nat Methods 2 (9), 667 (2005).

Acknowledgements

This work was supported by NIH Roadmap Initiative National Biotechnology Resource Center grant RR20843 for the Center on Proteolytic Pathways, CA69381 from the NCI, and by Training Grant 5T32CA77109-9 from the NCI. We thank Mari Enoksson, Wenhong Zhu, and Eric Wildfang for their valuable contributions to developing these methods.

Structural and kinetic determinants of protease substrates

by John C Timmer, Wenhong Zhu, Cristina Pop, +4

Nature Structural & Molecular Biology (19 July, 2009)