COMMENTARY

Finding the needles in the haystack: mapping constitutive proteolytic events *in vivo*

Matthew BOGYO¹

Department of Pathology and Department of Microbiology and Immunology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305-5324, U.S.A.

Our quest to understand the complex inner workings of the cell depends on the development of new technologies that allow the study of global regulatory events as they happen within their native cellular environment. Post-translational processing of proteins by proteases is one such regulatory process that can control many aspects of basic cell biology. In this issue of the *Biochemical Journal*, Timmer et al. describe a new proteomic approach that can be used to globally monitor constitutive proteolytic events *in vivo*. Using bacterial, human, yeast and mouse

cells, the authors show that this methodology provides a comprehensive map of constitutive trimming events mediated by regulatory proteases such as methionine aminopeptidase. This study also identifies previously uncharacterized processing events that highlight potential novel regulatory mechanisms mediated by proteolysis.

Key words: mass spectrometry, methionine aminopeptidase, protease, proteolysis, substrate identification.

WHY MONITOR PROTEOLYSIS?

At the time of their discovery over a century ago, proteases were considered to be simple degradative enzymes that marked the ‘end of the road’ for proteins. However, research undertaken over the last few decades has shown that proteases play key roles in regulating diverse cellular processes, ranging from the initiation of cell death to controlling the rates of cell division, with exquisite precision. Furthermore, many proteases carry out processing events that produce products with specific regulatory functions. Such limited proteolysis is perhaps one of the most interesting jobs of a protease, but also one of the most difficult to monitor within a complex cellular environment. In addition, the completion of a number of genome sequences has now demonstrated that 1–2% of most genes in an organism code for proteases. This predicts a total of more than 500 proteases in the human genome, of which very few have been thoroughly functionally characterized. Thus proteolysis must be considered to be one of the most important, but also most difficult to study, post-translational regulatory mechanisms used by any cell. Such a difficult problem therefore requires development of technologies that will allow proteolytic events to be monitored on a system-wide level. In this issue of the *Biochemical Journal*, Timmer et al. [1] describe an important new biochemical method for enrichment of the products of proteolysis. This work represents yet another step forward in our quest to understand protease function.

DEGRADOMICS: THE GRAND CHALLENGE

The term ‘degradomics’ has been coined to describe the system-wide study of proteases and their regulation of downstream substrates. Like most other ‘omics’, this field has its share of challenges that, for the most part, will be met through hard work and innovation. Although it is now possible to use simple computational methods to define a relatively complete list of all genes that code for proteases within a given genome, the assignment of function to these proteases remains far more challenging.

The key to understanding the function of a given protease is to define how, when and what substrates it processes. Yet the identification of protease substrates remains a daunting task. On the positive side, the last decade has seen a proliferation of creative methods that allow the primary amino acid specificity of a given protease to be measured (for a review, see [2]). These methods provide valuable information that can be used to help guide the selection of possible substrates based on predicted recognition sequences. However, the list of substrates that a protease can process when mixed with a large set of proteins or peptides in a test tube and what it does process in the cell are often non-overlapping. Thus attention has recently turned to methods that allow native protease substrates to be identified directly. Much like looking for a needle in a haystack, these methods must separate fragments of proteins that are produced by the action of a protease from the myriad other proteins and peptides that exist within cells. Thus the key to solving this problem is finding ways to enrich for these elusive protein fragments.

FINDING THE PRODUCTS OF PROTEOLYSIS

Proteases catalyse one of the most basic of chemical reactions: hydrolysis. In this process, water is added across the amide bond, resulting in the formation of two new products: a carboxy group and an amine. Since relatively few chemical methods exist to react specifically with carboxy groups, the amine has become the chemical ‘handle’ of choice for the isolation of newly formed peptide fragments. Amines can easily be chemically modified with a range of diverse tags that facilitate their subsequent biochemical enrichment. However, use of the N-terminal amine is complicated by the presence of other free amino groups on the side chains of lysine residues. These amines are virtually identical with the backbone free amine, except that they are slightly more basic. Thus, to specifically enrich for a newly formed protein fragment, one must devise methods to specifically block the unwanted lysine amines.

¹ email mbogyo@stanford.edu

Figure 1 Comparison of proteomic methods for the global mapping of proteolytic events

(A) The method of Timmer et al. [1] from this issue of the *Biochemical Journal*. Highlights include the specific modification of lysine side chains, leaving the free N-terminus available for modification with a biotin-affinity tag. This results in positive selection of the desired peptide using an avidin-affinity column. (B) The method of Gevaert et al. [3] and McDonald et al. [4] which uses a general acylation method to block all free amines. After blocking and digestion with trypsin, the unwanted internal amines are modified with biotin for removal with avidin beads or with a charged capping group that shifts their mobility in the liquid chromatography portion of the proteomic analysis. Thus the desired peptide is obtained by negative selection.

In the method described in this issue of the *Biochemical Journal*, Timmer et al. [1] have developed a selective modification strategy that allows the amines on the side chains of lysine to be guanidylated while leaving the free N-terminus of the protein intact (Figure 1A). This allows the desired N-terminal amine to be labelled with an affinity tag for specific enrichment by affinity chromatography. After selection of the N-terminal peptide and subsequent LC-MS/MS (liquid chromatography-tandem MS) analysis, primary sequence information can be used to identify target proteins by searching protein databases. By comparison of the sequences of the isolated peptides with genome sequences, it is possible to determine which peptides represent native N-termini and which have been trimmed by the action of a protease. The results from the study by Timmer et al. [1] show that this method provides a relatively comprehensive list of the N-terminal peptides from a large number of proteins. Of particular note, the analysis of *Escherichia coli*, human, mouse and yeast cells provides a first comprehensive look at the specificity of methionine aminopeptidase, a protease that plays the important role of removing the initial N-terminal methionine found on most newly synthesized proteins. This protease, like many others, has been extensively studied biochemically *in vitro*, and thus general information about its substrate specificity is available. These latest results provide direct experimental validation of the specificity of this important protease target. In addition, by using an unbiased method to identify products of proteolysis, the authors have been able to identify new N-termini that were not predicted on the basis of known constitutive processing events. In particular, the identification of alternative processing events in the sequences of mitochondrial proteins suggests that additional trimming by aminopeptidases may play an important role in their trafficking and/or stability. Such findings may lead to a better understanding of the basic proteolytic regulatory mechanisms that govern protein trafficking.

DIFFICULT PROBLEMS REQUIRE MANY SOLUTIONS

Although the work presented by Timmer et al. [1] provides a new method to approach the difficult task of globally mapping proteolytic events, a number of equally elegant methods have also been devised that serve to complement this approach. In particular, a number of research groups have made use of acylation to block free amino groups both on the N-terminal backbone nitrogen and also on lysine side chains [3,4]. This approach, although potentially technically more simple than the specific guanidinylation approach, has the disadvantage of relying on a negative selection process to remove unwanted N-termini that are produced during digestion of proteins to peptides (Figure 1B). It is likely that both methods will be valuable, and perhaps each will prove to be optimal for specific applications. It is also important to note that both methods make use of a biochemical purification strategy that ultimately results in the purification of only a single peptide for each target protein. Thus, unlike the usual procedure for identification of proteins by MS-based sequencing, which relies on the determination of multiple peptide sequences for a definitive identification, this method must make confident predictions of protein identity based only on a single peptide. Thus one must use much more stringent criteria for assessment of MS/MS data and also use the most accurate MS methods possible. The development of commercially available Fourier-transform and other ultra-high-resolution mass spectrometers will undoubtedly make this limitation insignificant in the future.

FUTURE DIRECTIONS

Although the global proteomic method presented here has already provided some valuable new information about constitutive proteolytic events, it clearly has the potential to be translated to the study of more dynamic processes. In fact, a number of new

studies have recently been reported that make use of system-wide proteomic methods to carry out unbiased searches for substrates of specific proteases [5–7]. These methods take advantage of isotopic labels to compare pools of proteins that either have or have not been exposed to a protease of interest. For example, by overexpressing a protease on the surface of a cell, it is possible to monitor changes in the levels of specific proteins that are released as the result of proteolytic shedding. These approaches have the advantage that they allow specific protein substrates to be linked to a defined protease of interest, but they generally do not allow assignment of the exact site of proteolysis. An obvious next step in the use of the tagging and enrichment strategy described by Timmer et al. [1] would be to use isotopically labelled versions of the biotin tag that would allow direct quantitative monitoring of dynamic changes in proteolytic processing. Significantly, this method would allow the exact site of proteolysis to be mapped at the same time that the substrate is identified. It is clear that the authors have begun to address this future application, and its development is likely to greatly enhance the value of this exciting new proteomic approach.

REFERENCES

- 1 Timmer, J. C., Enoksson, M., Wildfang, E., Wenhong, Z., Igarashi, Y., Denault, J. B., Ma, Y., Dummitt, B., Chang, Y. H., Mast, A. E. et al. (2007) Profiling constitutive proteolytic events *in vivo*. *Biochem. J.* **407**, 41–48
- 2 Diamond, S. L. (2007) Methods for mapping protease specificity. *Curr. Opin. Chem. Biol.* **11**, 46–51
- 3 Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R. and Vandekerckhove, J. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.* **21**, 566–569
- 4 McDonald, L., Robertson, D. H., Hurst, J. L. and Beynon, R. J. (2005) Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. *Nat. Methods* **2**, 955–957
- 5 Dean, R. A. and Overall, C. M. (2007) Proteomics discovery of metalloproteinase substrates in the cellular context by iTRAQ labeling reveals a diverse MMP-2 substrate degradome. *Mol. Cell. Proteomics* **6**, 611–623
- 6 Overall, C. M. and Blobel, C. P. (2007) In search of partners: linking extracellular proteases to substrates. *Nat. Rev.* **8**, 245–257
- 7 Tam, E. M., Morrison, C. J., Wu, Y. I., Stack, M. S. and Overall, C. M. (2004) Membrane protease proteomics: isotope-coded affinity tag MS identification of undescribed MT1-matrix metalloproteinase substrates. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6917–6922

Received 13 August 2007; accepted 14 August 2007

Published on the Internet 12 September 2007, doi:10.1042/BJ20071096