Methods 54 (2011) 351–360

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

The importance of the digest: Proteolysis and absolute quantification in proteomics

Philip Brownridge, Robert J. Beynon *

Protein Function Group, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK

ARTICLE INFO

Article history:
Available online 6 June 2011

Keywords:
Quantitative proteomics
QconCAT
Proteolysis
Trypsin
Selected reaction monitoring
Quantotypic peptides
Endopeptidase LysC
Label-free quantification
Stable isotope labelling

ABSTRACT

Virtually all mass spectrometric-based methods for quantitative proteomics are at the peptide level, whether label-mediated or label-free. Absolute quantification in particular is based on the measurement of limit peptides, defined as those peptides that cannot be further fragmented by the protease in use. Complete release of analyte and (stable isotope labelled) standard ensures that the most reliable quantification data are recovered, especially when the standard peptides are in a different primary sequence context, such as sometimes occurs in the QconCAT methodology. Moreover, in label-free methods, incomplete digestion would diminish the ion current attributable to limit peptides and lead to artifactually low quantification data. It follows that an essential requirement for peptide-based absolute quantification in proteomics is complete and consistent proteolysis to limit peptides. In this paper we describe strategies to assess completeness of proteolysis and discuss the potential for variance in digestion efficiency to compromise the ensuing quantification data. We examine the potential for kinetically favoured routes of proteolysis, particularly at the last stages of the digestion, to direct products into 'dead-end' mis-cleaved products.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

In the 1940's, Linderstrøm-Lang studied the action of proteases on proteins. He proposed two different mechanisms; an 'all or none' process, whereby a protease bound to a substrate molecule and remained associated until the protein was fully digested, and a 'zipper' process, whereby a protease interacted with intact substrate and partially degraded fragments, until proteolysis was complete [1]. The primary difference between the two processes was that in the former, there could be no degradation intermediates free in the digestion reaction whereas the zipper mechanisms could release partially cleaved products. We now know that the all-or-none mechanism does not operate (other than in the confines of the 20S proteasomal core), and that degradation intermediates are therefore not only likely but obligatory for simple endopeptidases.

When a protein undergoes an initial proteolytic event, the products can become more or less susceptible to further proteolysis. For example, the proteolytic action of enteropeptidase on trypsinogen produces active trypsin by virtue of the loss of an N-terminal hexapeptide. The activated enzyme is less likely to undergo further proteolysis by enteropeptidase – if this were not the case; the active enzyme would be degraded more rapidly and would not persist. Alternatively, a protein can be destabilised by the initial proteolytic cleavage, such that the products are more rapidly cleaved into multiple further products. A feature of the latter behaviour is that the route of digestion might not follow the same pathway for each protein molecule, and thus, a large number of discrete, partially digested species are generated. It is only as the sequential proteolytic reactions reach completion that the different pathways converge to the same products (Fig. 1). When all peptide bonds that **can** be cleaved **have** been cleaved, the resultant set of peptides are referred to as 'limit peptides'; peptides that lack any further endoproteolytic sites compatible with the endopeptidase being used.

Despite the development of top-down analytical approaches, most proteomics workflows require a proteolytic step prior to mass spectrometric analysis of the peptides generated by the hydrolytic reaction. In most instances, the endopeptidase that is used is trypsin, reflecting the very restricted specificity of this enzyme (Arg-X, Lys-X, and under normal circumstances, zero or low frequency cleavage at Arg-Pro, Lys-Pro) and the fact that most products from a tryptic digest have a minimum of two protonatable sites (the N-α amino group and the C-terminal basic residue) and thus generate $[M+2H]^{2+}$ ions, enhancing the generation and enhancement of gas phase fragmentation products. In many proteomes the residues arginine and lysine are each present at about 5% of the amino acids, making a tryptic fragment approximately 10–15 amino acids residues long. Assuming complete fragmentation, the limit peptides that are detectable are usually between 1000 (about eight amino acids) and 3000 Da (about 25 amino acids), optimally aligned to the m/z range of the mass analysers used in mass spectrometers that feature in proteomics studies.

* Corresponding author.
E-mail address: r.beynon@liv.ac.uk (R.J. Beynon).

1046-2023/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2011.05.005

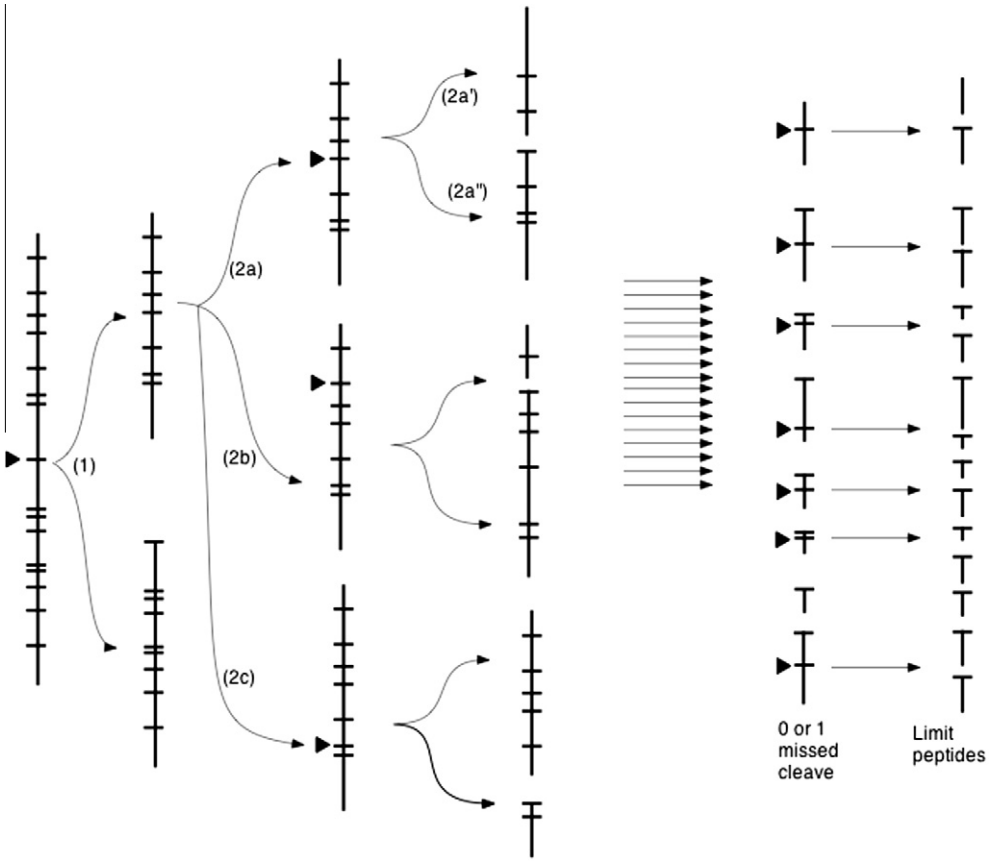

Fig. 1. Routes of proteolysis of a protein. In the absence of higher order structural factors that can modify the propensity of sites to be digested, the conversion of an intact protein to limit peptides can take many different routes; the relative occupancy of such routes is a consequence of the intrinsic digestibility of each scissile bond. Eventually, multiple pathways converge to oligopeptides that are defined as 'mis-cleaved peptides'. Completeness of digestion can be assessed by monitoring the ratio of mis-cleaved peptides and limit peptides.

Although proteolysis of an entire proteome is often predicated on the complete hydrolysis of all proteins to limit peptides, many protein identification strategies are tolerant to a small number of mis-cleavages (typically one or two), which might even enhance the strength of the identification, since a mis-cleaved product restores some of the lost connectivity that is inherent in a set of limit peptides – in a fully proteolysed proteome we do not know which peptides are 'adjacent' to each other. The gain in identifiability is more critical in peptide mass fingerprinting, because the only piece of information obtained from the peptide is the mass whereas in tandem mass spectrometry further information is gained from each peptide according to sequence specific fragmentation.

Although it may be possible to optimise complete proteolysis for a single protein, a proteome offers a large and complex reaction space. Residues at least three positions distal to the cleavage site can affect proteolysis, predominantly through changes in the affinity of the endopeptidase for the substrate [2]. This creates a large number of possible (approximately $20^6 = 64$ million) different cleavage sites, although in practice only a subset of these are evident in any proteome; for example, the *Saccharomyces cerevisiae* proteome has approximately 250,000 tryptic sites. Many of the tryptic sites will be efficiently and completely cleaved, but some, for example, those with acidic residues C-terminal to the scissile bond in positions P1′ and P2′ (nomenclature of Schechter and Berger [3]), will be slow to hydrolyse and therefore difficult to digest to completion [4].

Although mis-cleaved products can sometimes enhance the quality of an identification workflow, there are circumstances in which they can compromise quantitative proteomics. Peptide-level quantification can be conducted by label-mediated methods or label-free approaches. In label-mediated methods, a differentially stable isotope labelled standard peptide [5] of known amount is co-analysed with the analyte, and the ratio of the analyte to standard reveals the abundance of the analyte. Relative quantification (whether label free or isotope coded, such as is obtained with metabolically labelled samples) may be more tolerant to incomplete proteolysis, provided that it can be assumed that the labelled and unlabelled proteins undergo the same extent of proteolysis. For absolute quantification, whether using stable isotope labelled chemically synthesized peptides (AQUA peptides) or peptides derived from hydrolysis of a protein standard (QconCAT or PSAQ) it is necessary to compare the analyte with a standard, usually at the peptide level, in assays wherein the quantities of one or more (tryptic) peptides are considered to be formally representative of the quantity of the parent protein. Complete proteolysis of analyte (AQUA) or analyte and standard (PSAQ, QconCAT) is thus far more critical in quantification workflows than it is in discovery workflows. Completeness of proteolysis is also important in label-free methods that make use of the number of tryptic fragments observable (spectral counting) or the inherent intensity of the mass spectrometric signal for one or more peptides [6,7]. Implicit in either of these approaches is that the optimal data will be obtained if the analyte (and in some instances, standard) signal is delivered by limit peptides.

The requirement for complete proteolysis is never more important than in QconCAT quantification workflows [8–11]. QconCATs are artificial proteins that are concatenated tryptic peptides from a large number of different analyte proteins, typically two peptides for each protein. The gene that would direct the synthesis of the QconCAT is synthesized *de novo*, and expressed heterologously in bacteria, in stable isotope labelled media. Once purified, a known

Download English Version:

https://daneshyari.com/en/article/10826077

Download Persian Version:

https://daneshyari.com/article/10826077

Daneshyari.com