

Taking a bite: proteasomal protein processing

Michael Rape and Stefan Jentsch

The proteasome is a hollow cylindrical protease that contains active sites concealed within its central cavity. Proteasomes usually completely degrade substrates into small peptides, but in a few cases, degradation can yield biologically active protein fragments. Examples of this are the transcription factors NF- κ B, Spt23p and Mga2p, which are generated from precursors by proteasomal processing. How distinct protein domains are spared from degradation remains a matter of debate. Here, we discuss several models and suggest a novel mechanism for proteasomal processing.

Being choosy and having a restricted appetite is a necessity for a cellular protease; otherwise, proteases would kill cells instantly by chewing up all cellular proteins and polypeptides. One elegant solution to this problem is the so-called 'self-compartmentalizing' proteases¹. These are usually hollow cylindrical particles that have their active sites concealed within their central cavities. Substrates can only reach the catalytic sites through narrow ports that control substrate entry. Proteases of this type are found in all forms of life, the proteasome being the best-known member^{1,2}.

In eukaryotes, the proteasome is composed of a 20S cylindrical core particle, which has six active sites and two 19S caps, which sit above the openings of the 20S cylinder, on each of its two ends (for a review see ref. 2). Substrate entry into the 20S proteasome proceeds by a gating mechanism, which seems to be controlled by the 19S particle^{3,4}. To reach the active sites, the substrates are unfolded and threaded through the narrow openings. Concomitant with the translocation of the polypeptide into the proteolytic chamber, the substrates are cleaved into peptides and proteolysis seems to continue until the fragments are small enough to leave the particle by diffusion.

The menu: NF- κ B & Co.

The observed processivity of the reaction suggested that proteasomes completely degrade proteins into small peptides. However, in the early 1990s it was discovered that proteasomes are also responsible for the processing of p105, the precursor of the p50 subunit of the mammalian transcription factor NF- κ B^{5,6}. This exciting finding identified a novel role of the ubiquitin/proteasome pathway in activating dormant signalling molecules. Notably, the authors showed that the carboxy-terminal domain of p105 is completely degraded by a ubiquitin/proteasome-dependent mechanism, whereas the

amino-terminal transcription factor domain, p50, is left intact. Proteasomal processing occurs after ubiquitination of p105, possibly at a site within the C-terminal domain that gets degraded⁷. When expressed in yeast, p105 is also processed, generating an almost identically sized p50 polypeptide⁸. Notably, processing in mammalian cells and yeast was shown to be dependent on the proteasome, as proteasome inhibitors and yeast proteasome mutants, respectively, blocked the reaction. The most logical explanation at this time was that proteasomes degrade p105 processively, starting from the C-terminal end of the precursor (see Fig. 1a). A hypothetical 'stop-transfer signal' in the centre of p105 was proposed to block further degradation. It has been suggested that the glycine-rich repeat (GRR) in the centre of p105, which seems necessary for processing, may function in this manner^{7,9}. But how a supposedly unstructured domain could prevent further translocation of the substrate into the proteasome is unclear.

A different model of NF- κ B processing was suggested by the finding of other studies, which showed that apparently a substantial pool of p50 is generated cotranslationally on polysomes^{10,11}. It was suggested that the N-terminal p50 segment emerging from the leading ribosome folds cotranslationally and dimerizes with another folded p50 domain of a second emerging p105 molecule synthesized by the trailing ribosome^{10,11}. Subsequently, the nascent polypeptide is processed by the proteasome. Because nascent polypeptide chains have no free C-terminal ends, cotranslational proteasomal processing is thought to be initiated by an internal ('endoproteolytic') cleavage. The proposed product of this reaction is a processed p50 molecule bound to a full-length p105 partner (see Fig. 1b). This model is appealing, as it elegantly explains how the p50 domain of NF- κ B escapes proteolysis. However, it cannot account for all the p50 generated in cells,

because stimulated processing from presynthesized precursors has also been reported¹².

Possible new explanations as to how proteasomes mediate post-translational protein processing came from an unexpected direction. Recently, two distant homologues of NF- κ B have been discovered from the budding yeast *Saccharomyces cerevisiae*: Spt23p and Mga2p^{13,14}. These two related proteins have overlapping functions and drive the transcription of *OLE1*, encoding fatty acid desaturase¹³. In common with NF- κ B, they are initially synthesized as inactive precursors (p120) and become activated by regulated ubiquitin/proteasome-dependent processing (RUP)¹⁴. However, in striking contrast to NF- κ B p105, Spt23p/Mga2p p120 proteins are integral membrane proteins that are anchored to the endoplasmic reticulum (ER) through their C-terminal tails; the bulk of the proteins facing the cytosol¹⁴. Also similar to NF- κ B activation, processing of p120 completely eliminates the C-terminal tails (including the membrane spanning regions), whereas the N-terminal transcription factor domains are left intact. Notably, processing of Spt23p/Mga2p occurs post-translationally, directly at the ER membrane^{14,15}. This is of particular importance for Spt23p because the full-length molecule may function as a sensor for membrane composition; its processing is negatively feed-back regulated by unsaturated fatty acids, the products of the ER-bound Ole1p enzyme^{14,16}. Clear evidence that p120 processing is also mediated by the ubiquitin/proteasome system comes from biochemical and genetic data. The Rsp5p ubiquitin ligase was shown to physically interact with Spt23p p120 through one of the WW protein-protein interaction domains of Rsp5p, resulting in ubiquitination of p120¹⁴. Furthermore, and importantly, processing of p120 is blocked in *rsp5* and proteasome mutants¹⁴. As the C-terminal tails of Spt23p/Mga2p p120 are not accessible to the proteasome (they are

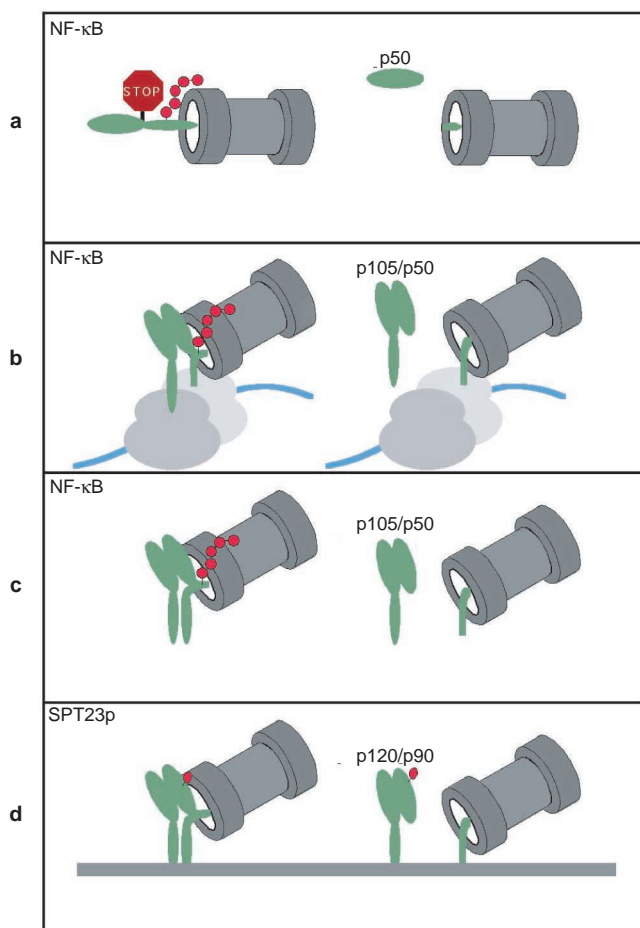


Figure 1 Hypothetical models for proteasomal protein processing of NF- κ B p105 and Spt23p p120. Processing of substrates (green) by proteasomes (grey barrel) is initiated by ubiquitination (red balls). **a**, The 'Stop-transfer model' for post-translational processing of NF- κ B. p105 is proposed to be degraded processively from its C-terminal end until the proteasome encounters a stop-transfer signal (STOP sign). The N-terminal domain (p50) is spared from degradation and released, whereas the C-terminal half of the molecule is completely eliminated. **b**, A cotranslational model for NF- κ B p105 processing. The N-terminal p50 segment emerging from the ribosome (light grey; front-left shows the leading ribosome, back-right shows the trailing ribosome) folds cotranslationally and dimerizes with another folded p50 domain of a second p105 molecule synthesized by another ribosome. The nascent polypeptide chain (still attached to the ribosome through a tRNA) is cleaved 'endoproteolytically' (probably involving loop formation) and a dimer composed of p105 and a cleaved p50 molecule is released. **c, d**, Alternative 'loop models' for post-translational processing involving hairpin loop formation and initiation of proteolysis by an internal cleavage. Tight association of a processing template (left molecule of the pair) and a processing substrate (right molecule of the pair) protects the N-terminal p50 domain from unfolding and degradation; however, the free C-terminal domain of the processed substrate is completely degraded. The processed substrate (NF- κ B p50 or Spt23p p90) is released from the proteasome in a complex with the processing template (NF- κ B p105 or Spt23p p120). Note that the ubiquitination sites for NF- κ B and Spt23p may be located at the C- or N-terminal domains of the precursors, respectively. Spt23p p90 is released as a mono-ubiquitinated protein and segregated from its p120 partner by the chaperone-like CDC48^{UFD1/NPL4} complex; NF- κ B may be released from its partner by the activity of the chaperone BCL-3 (not shown). In contrast to NF- κ B p105, Spt23p p120 is a tail-anchored membrane protein of the ER (grey line).

embedded into the membrane), proteasomal processing yielding free N-terminal

protein domains was proposed to be initiated by an internal cleavage event^{14–16}.

Internal cleavage: polypeptide loops as substrates

A key feature of the models for cotranslational NF- κ B p105 processing (see above) and post-translational processing of Spt23p/Mga2p at the ER is that substrate processing is initiated by an internal cleavage. However, this aspect of the two models raises several questions. First: how can proteasomes generate internal cuts? A possible answer is that a segment of the substrate polypeptide chain unfolds into a hairpin loop, which translocates through the narrow entry channel of the proteasome, thereby reaching the active sites of the 20S barrel.

However, this leads to a second question: does the architecture of the proteasome allow the passage of hairpin loops? The active sites of the proteasome are concealed in its central chamber, which is composed of the two β -rings from the 20S core particle¹⁷. This proteolytic chamber is flanked symmetrically by two antechambers. Substrate entry into the antechambers is restricted by gated channels, which are controlled by Rpt2p, an AAA-ATPase of the proteasome 19S cap^{3,4}. Opening of the gate through Rpt2p generates a channel with a diameter of approximately 13 Å^{3,4}. This is wide enough to accommodate two juxtaposed unfolded polypeptide chains, or alternatively, a hairpin loop (M. Groll, personal communication; see Fig. 2). The distance from the outside of the 20S barrel to the active sites is equivalent to approximately 20 amino acids of an unfolded polypeptide¹⁷, indicating that a polypeptide loop that stretches down to the active sites should be at least 40 amino acids in length.

Other questions remain: for example, how is the process initiated and what promotes loop formation? In all likelihood, ubiquitination of the substrates is a prerequisite for proteasome association. We speculate that flexible domains of substrate proteins are usually the first segments that pass the gate of the proteasome. These could be either the free N- or C-terminal ends of proteins, or flexible regions within the centre of a polypeptide chain. A highly attractive candidate for a centrally positioned flexible domain is the GRR element in NF- κ B p105. Studies have shown that this domain is essential for p105 processing and that if placed between two stably folded domains, proteasomes seem to cut in the middle of the fusion protein, yielding two protein products¹⁸. Unfolding of the GRR probably does not require much energy and the lack of bulky side chains facilitates passage through the gate. Spt23p and Mga2p do not possess GRRs, but in common with Spt23p homologues from *Schizosaccharomyces pombe* and *Candida albicans*, they have other regions of low sequence complexity at related positions of their polypeptide chains^{14,15}. Whether unstructured domains are sufficient for promoting loop formation

or whether ubiquitination in close proximity to the flexible domain is additionally required, is not known. Notably, however, whereas the ubiquitination sites of p105 may localize in the C-terminal domain that gets lost by the processing reaction⁷, Spt23p is ubiquitinated within the N-terminal transcription factor domain¹⁵.

Limited proteolysis

In the previous section, we discussed how we envision endoproteolytic attacks by proteasomes and we argued that both co- and post-translational processing events might be initiated by internal cleavages. But if a loop enters the proteasome, why do proteasomes not nibble in both directions, starting from the first cut and eliminating both halves of the protein completely? In fact, we believe that this often happens and that possibly many cellular proteins are completely degraded in this manner. In other words, initiating proteasomal degradation by internal cleavage may not be restricted to substrate processing. But how does it work for protein processing? How are certain regions spared from degradation? Several observations suggest that specific protein domains may not be accessible, or are resistant to, the unfolding forces of the 19S cap ATPases of the proteasome, and are therefore left behind. The first evidence came from an elegant study demonstrating that tightly folded domains can prevent proteasomal degradation¹⁹. The substrate used in this study was a short-lived variant of dihydrofolate reductase (DHFR). Addition of the DHFR-ligand methotrexate induced a stably folded enzyme conformation, which was resistant to proteasomal processing, although it was ubiquitinated¹⁹. These findings were recently bolstered by systematic and refined studies showing that substrates are unravelled for proteasomal degradation, starting from their initial sites of proteasomal action (for example, their degradation signals)²⁰. By experimentally placing tightly folded domains at various locations along a polypeptide chain, the authors could show that proteasomes stop when they reach these domains because they are unable to unfold them²⁰.

Another strategy of protection against complete proteasomal degradation seems to involve dimerization with a 'life-saving' partner. We propose that this is the main strategy for NF- κ B and its relatives, and that it works cotranslationally, as well as post-translationally. Indeed, processing of the NF- κ B (NF- κ B1 and NF- κ B2) precursors and Spt23p/Mga2p seems to be dependent on homodimerization^{11,15}. This is mediated by the immunoglobulin-like IPT domain of the transcription factors^{11,15}, and, notably, the majority of the critical residues at the dimerization interface are well conserved between the mammalian

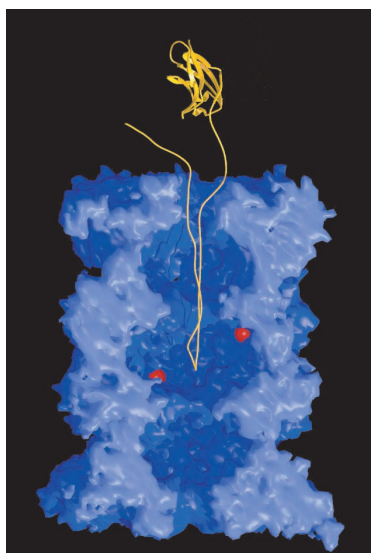


Figure 2 A polypeptide loop as a possible substrate for the proteasome. The model shows a vertically sliced and opened 20S proteasome, as determined by X-ray crystallography studies^{3,17}. A hairpin loop of a substrate composed of an unstructured polypeptide chain is modelled into the structure. A folded NF- κ B (p50) domain of a hypothetical substrate is shown for comparison. The dimensions of the gate are ~ 13 Å and the distance from the outside of the 20S proteasome to the active sites (shown in red) are ~ 70 Å (equivalent to ~ 20 amino acids of an extended polypeptide chain).

and yeast transcription factors²¹. Interestingly, deletions or mutations within the IPT domain prevent proteasomal processing^{15,20}. Thus, we suggest that the NF- κ B p50 domain and the Spt23p/Mga2p p90 domains are left intact because they tightly associate with homologous partner molecules outside of the proteasome (Fig. 1c,d). In this model, degradation is initiated by an internal cut within a polypeptide loop of the substrate and continues by processive degradation of the two chains. Degradation towards the N-terminal end of the original substrate will come to a halt when the proteasome reaches the tightly folded dimerization domain, whereas processive degradation in the direction of the C-terminal end of the original molecule is not restricted by barriers. Interestingly, in the case of Spt23p/Mga2p, degradation of the C-terminal domain includes the complete elimination of the ER-embedded transmembrane span, suggesting that this process may be related to ER-associated degradation (ERAD)^{14,15,22}.

A significant aspect of our model is that only one molecule is processed (processing

substrate) within one homodimer, whereas its homologous binding partner (processing template) remains intact. Indeed, in all known processing events, that is, for NF- κ B and Spt23p/Mga2p, heterodimeric complexes composed of one full-length molecule and one processed molecule have been recovered^{11,15,23}. What determines which molecule will be the substrate and which the template is not clear, but a model in which the substrate is chosen by stochastic ubiquitination of just one subunit of a dimer is attractive. Complexes composed of the processed substrate and the template can be regarded as intermediates of processing, and it seems that specific chaperones or 'segregases' are needed to separate the two homologous proteins¹⁵. In the case of Spt23p, the enzyme is the ubiquitin-selective segregase CDC48^{UFD1/NPL4} (ref. 15), whereas NF- κ B p50 may be liberated from its p105 binding partner by the ankyrin-repeat-containing chaperone BCL-3 (ref. 24).

Novel proteasome function

The functional repertoire of the proteasome is greatly expanded by its ability to function as a processing enzyme. Activation of presynthesized dormant signalling molecules, such as transcription factors, seems to be the major purpose, as it allows a rapid response to cellular cues¹⁶. Until now, only a handful of proteins undergoing proteasomal processing have been identified. In addition to the examples described here, the *Drosophila melanogaster* protein Cubitus interruptus (Ci), which is involved in Hedgehog (Hh) signalling, may be processed by a similar ubiquitin/proteasome-dependent mechanism^{25,26}. In the absence of Hh signalling, full-length Ci gets proteolytically processed and the 75K N-terminal segment remains intact and functions as a transcriptional repressor. The full-length molecule, however, functions as an activator²⁷. Although Ci does not share detectable similarity with NF- κ B or Spt23p, it is interesting to note that Ci has to be assembled in a multiprotein complex to be processed²⁸. This finding suggests that not only homodimerization, but also an interaction with heterologous partners, may prevent complete degradation.

Although our model for post-translational proteasomal processing must be validated with further experiments, it is consistent with most findings obtained from mammalian and yeast systems. Several aspects of our model can now be thoroughly tested and new important details are expected to emerge. In particular, the proposed mechanism that relies on processing substrates and templates offers several interesting potential levels of regulation – both during and after the processing step. First, processed mole-

cules may be stored in inactive intermediary complexes with processing templates, and a regulated chaperone-mediated release may result in instant activation of these proteins. Second, mechanisms that control the affinity between processing substrates and templates may function as powerful molecular switches. If the affinity is high, protein processing will occur, and signalling will be turned on. Conversely, if the affinity is weak, the molecules will be completely degraded and thus the signalling pathway will be turned off. It will be interesting to learn whether such potential regulatory mechanisms are indeed relevant for cellular signalling pathways that involve proteasomal processing. Furthermore, the concept of 'life-saving' associations that prevent translocation of a protein into the proteasome should be viewed in a larger context. The glycine-alanine repeats of EBNA1, an Epstein-Barr virus protein that is resistant to proteasomal degradation²⁹, and the polyglutamine stretches of several proteins (for example, of the huntingtin

protein, associated with Huntington's disease³⁰) may promote micro-aggregate formation, which may withstand the appetite of the proteasome. □

Michael Rape and Stefan Jentsch are in Department of Molecular Cell Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18a, 82152 Martinsried, Germany
e-mail: jentsch@biochem.mpg.de

- Lupas, A., Flanagan, J. M., Tamura, T. & Baumeister, W. *Trends Biochem. Sci.* **22**, 399–404 (1997).
- Baumeister, W., Walz, J., Zuhl, F. & Seemuller, E. *Cell* **92**, 367–380 (1998).
- Groll, M. *et al. Nature Struct. Biol.* **7**, 1062–1067 (2000).
- Kohler, A. *et al. Biochimie* **83**, 325–332 (2001).
- Fan, C. M. & Maniatis, T. *Nature* **354**, 395–398 (1991).
- Palombella, V. J., Rando, O. J., Goldberg, A. L. & Maniatis, T. *Cell* **78**, 773–785 (1994).
- Ciechanover, A. *et al. Biochimie* **83**, 341–349 (2001).
- Sears, C., Olesen, J., Rubin, D., Finley, D. & Maniatis, T. *J. Biol. Chem.* **273**, 1409–1419 (1998).
- Orian, A. *et al. Mol. Cell Biol.* **19**, 3664–3673 (1999).
- Lin, L., DeMartino, G. N. & Greene, W. C. *Cell* **92**, 819–828 (1998).
- Lin, L., DeMartino, G. N. & Greene, W. C. *EMBO J.* **19**, 4712–4722 (2000).
- Xiao, G., Harhaj, E. W. & Sun, S. C. *Mol. Cell* **7**, 401–409 (2001).
- Zhang, S., Skalsky, Y. & Garfinkel, D. *J. Genetics* **151**, 473–483 (1999).
- Hoppe, T. *et al. Cell* **102**, 557–586 (2000).
- Rape, M. *et al. Cell*, **107**, 667–677 (2001).
- Hoppe, T., Rape, M. & Jentsch, S. *Curr. Opin. Cell Biol.* **13**, 344–348 (2000).
- Groll, M. *et al. Nature* **386**, 463–471 (1997).
- Lin, L. & Ghosh, S. *Mol. Cell Biol.* **16**, 2248–2254 (1996).
- Johnston, J. A., Johnson, E. S., Waller, P. R. & Varshavsky, A. *J. Biol. Chem.* **270**, 8172–8178 (1995).
- Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M. & Matouschek, A. *Mol. Cell* **7**, 627–637 (2001).
- Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. *Nature* **373**, 311–317 (1995).
- Braun, S., Matuschewski, K., Rape, M., Thoms, S. & Jentsch, S. *EMBO J.* **21**, (2002).
- Betts, J. C. & Nabel, G. J. *Mol. Cell Biol.* **16**, 6363–6371 (1996).
- Watanabe, N., Iwamura, T., Shinoda, T. & Fujita, T. *EMBO J.* **16**, 3609–3620 (1997).
- Chen, C. H. *et al. Cell* **98**, 305–316 (1999).
- Jiang, J. & Struhl, G. *Nature* **391**, 493–496 (1998).
- Aza-Blanc, P., Ramirez-Weber, F. A., Laget, M. P., Schwartz, C. & Kornberg, T. B. *Cell* **89**, 1043–1053 (1997).
- Ingham, P. W. *Nature* **394**, 16–17 (1998).
- Levitkaya, J. *et al. Nature* **375**, 685–688 (1995).
- Orr, H. T. *Genes Dev.* **15**, 925–932 (2001).

ACKNOWLEDGEMENTS

We thank M. Groll for preparing Fig. 2, and W. Baumeister, G. DeMartino, M. Groll, T. Hoppe, R. Huber, L. Lin and A. Matouschek for discussions.