nature structural & molecular biology

A conserved processing mechanism regulates the activity of transcription factors Cubitus interruptus and NF-κB

Lin Tian, Robert A Holmgren & Andreas Matouschek

The proteasome degrades some proteins, such as transcription factors Cubitus interruptus (Ci) and NF-κB, to generate biologically active protein fragments. Here we have identified and characterized the signals in the substrate proteins that cause this processing. The minimum signal consists of a simple sequence preceding a tightly folded domain in the direction of proteasome movement. The strength of the processing signal depends primarily on the complexity of the simple sequence rather than on amino acid identity, the resistance of the folded domain to unraveling by the proteasome and the spacing between the simple sequence and folded domain. We show that two unrelated transcription factors, Ci and NF-κB, use this mechanism to undergo partial degradation by the proteasome *in vivo*. These findings suggest that the mechanism is conserved evolutionarily and that processing signals may be widespread in regulatory proteins.

ATP-dependent proteolysis participates in the regulation of cellular processes such as the cell cycle, signal transduction and the immune response. In eukaryotes, this activity is carried out by the ubiquitin-proteasome system. Proteins that are to be degraded are first tagged with several ubiquitin chains and then hydrolyzed by the large multi-component protease called the 26S proteasome[1,2].

The proteasome is a barrel-shaped particle with its active sites of proteolysis buried deep inside the structure and accessible only through a channel that is too small to allow passage of folded proteins[3]. The proteasome unfolds substrate proteins by unraveling them from the degradation initiation site. Unfolding and proteolysis can proceed sequentially in either direction along a substrate's polypeptide chain, from the N terminus toward the C terminus or from the C terminus toward the N terminus[4,5]. The proteasome is highly processive and produces small peptides of about 8 residues in length[6]. The complete degradation of proteins avoids the creation of protein fragments whose activity is not held in check by regulatory domains.

However, exceptions do occur. The proteasome degrades a ubiquitin–β-galactosidase fusion protein only partially[7] and has since been implicated in several protein processing reactions *in vivo*[8–12]. The two examples discussed here concern the transcription factors Ci[10,11] and NF-κB[12]. Ci is the key regulator of Hedgehog (Hh) target genes and controls several aspects of body-plan formation in *Drosophila*. Hh signaling stabilizes and activates the full-length form of Ci, Ci155, which leads to the transcription of downstream target genes[13]. In the absence of Hh, Ci155 is processed to the truncated N-terminal fragment Ci75, which lacks the C-terminal transactivation domain and acts as a repressor of Ci target gene expression[14]. Several lines of evidence indicate that the proteasome is involved in the processing reaction of Ci. The formation of Ci75 requires the protein ubiquitin ligase Slimb[11], is prevented by the proteasome inhibitor MG132 (refs. 9,10) and is regulated by phosphorylation[15,16].

In mammalian cells, the p50 subunit of NF-κB is generated from a larger precursor, p105, in a process that also involves the ubiquitin-proteasome pathway[12,17]. The NF-κB family of transcription factors functions in regulating immune and inflammatory responses. p105 functions as an inhibitor of NF-κB by trapping several of its subunits in the cytosol in a manner that is similar to IκB proteins[18,19]. The C-terminal half of p105 contains seven ankyrin repeats, which bind and mask a nuclear localization signal that is located next to the DNA-binding domain in the N-terminal half of the protein. Upon stimulation, p105 is ubiquitinated by the SCF β–TrCP ubiquitin ligase[20] and its C-terminal part is degraded to yield the N-terminal fragment, p50 (ref. 12). The processing allows p50 to move from the cytosol into the nucleus where it regulates gene transcription as a dimer with other Rel subunits[21]. Proteasome inhibitors markedly block the formation of p50 and the subsequent NF-κB activation in mammalian cells[12].

Several models of p105 processing by the proteasome have been discussed[22], but none of them has been extended to Ci processing and the signal that controls the processing event is not understood. Earlier research found that a glycine-rich region (GRR) adjacent to the processing point in p105 is required for p50 generation in mammalian cells[23,24]. It has been suggested that this sequence protects p50 during proteasome degradation of p105 (refs. 22,24) or that it serves as a transferable degradation signal[23]. Another model proposed that p50 is produced cotranslationally by proteasome processing of the nascent polypeptide chain of p105 (ref. 25). Notably, the GRR is not required for p105 processing in yeast[26]. Ci processing also depends on sequences flanking the cleavage site[27], but this region is not rich in

Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2205 Tech Drive, Evanston, Illinois 60208, USA and Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Chicago, Illinois 60611, USA. Correspondence should be addressed to A.M. (matouschek@northwestern.edu).

Received 1 August; accepted 14 October; published online 20 November 2005; doi:10.1038/nsmb1018

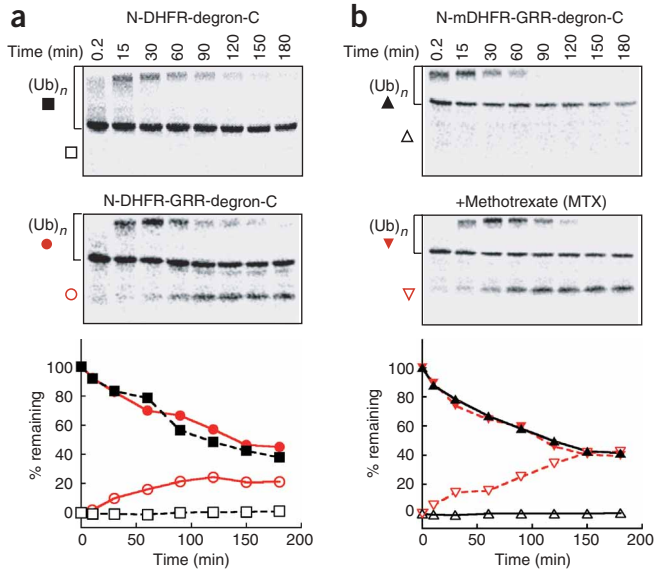

Figure 1 A GRR in combination with a folded domain inhibits progression of proteasomal degradation. ^{35}S-labeled fusion proteins containing a C degron were degraded by the proteasome at 30 °C. (**a**) Degradation of DHFR and DHFR-GRR fusion protein from the C degron are shown as the autoradiogram of an SDS-PAGE gel separating the reaction products. Positions of the full-length protein and its ubiquitinated forms are indicated by square brackets (DHFR, solid black square; DHFR-GRR, solid red circle). Partially degraded product (open red circle) accumulated only when the GRR was inserted between DHFR and the C degron. The extent of degradation and fragment formation was quantified as the remaining percentage of the total amount of full-length protein present at the beginning of the reaction. (**b**) Degradation of mouse DHFR-GRR fusion protein in the absence (solid black triangle) and presence (solid red triangle) of its ligand MTX. The extent of degradation and fragment formation (–MTX, open black triangle; +MTX, open red triangle) was quantified and plotted.

glycine residues. Other results suggested that the Rel-homology domain in p50 contributes to processing. Highly stabilized proteins cannot be degraded by the proteasome[28], and destabilizing the Rel-homology domain by mutagenesis abolishes p50 accumulation[4]. In addition, dimerization of the Rel-homology domain is required for p50 formation[29].

Here we have identified and characterized the factors that cause processing. Processing occurs when the progression of the proteasome along its substrate's polypeptide chain is stopped by a signal consisting of two components: a sequence of simple amino acid composition (simple sequence) followed by a tightly folded domain in the direction of proteasome movement. In Ci, the simple sequence is rich in asparagine, serine and glutamine residues and the folded domain is a C2H2 zinc finger, whereas in p105 the simple sequence is rich in glycine residues and the folded domain is the Rel-homology domain. The strength of the signal depends on the complexity of the simple sequence and the resistance of the folded domain to unraveling, and both components of the signal are transferable. Thus the processing mechanism but not the sequence of the signal is conserved between flies and humans, suggesting that other examples remain to be found.

RESULTS

Inhibition of proteasome progression

The proteasome degrades folded domains efficiently. A substrate protein consisting of an N-terminal folded domain (*Escherichia coli* dihydrofolate reductase, or DHFR) followed by a C degron (C-terminal degradation signal; see Methods) was unraveled sequentially from its C terminus toward its N terminus and degraded completely[5] (**Fig. 1a**). When a glycine-rich sequence of 37 residues was inserted between the folded domain and the ubiquitination site, degradation began normally but its progression along the substrate was blocked, leading to the production of a partially degraded substrate that consisted of the DHFR domain and a tail of ~80 residues (**Fig. 1a**). Fragment formation depended on the ubiquitin-proteasome pathway because it required ATP and was prevented by the proteasome inhibitor MG132 or an excess of mutant ubiquitin lacking lysine residues (**Supplementary Fig. 1** online). The GRR did not lead to substrate aggregation, and the fusion protein remained monomeric as judged by size exclusion chromatography (**Supplementary Fig. 2** online).

Partial degradation depended not only on the presence of the GRR but also on the presence of a stably folded domain adjacent to the simple sequence. When *E. coli* DHFR was replaced with the homologous (27% identical) but thermodynamically less stable mouse DHFR (free energies of unfolding are 6.1 kcal mol^{-1} and 4.4 kcal mol^{-1}, respectively[30,31]), the fragment no longer accumulated (**Fig. 1b**). Stabilizing mouse DHFR against unfolding with its ligand

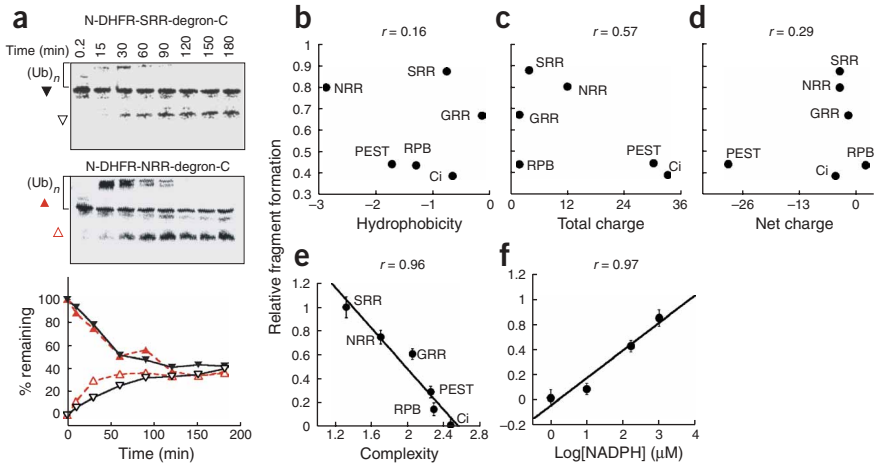

Figure 2 Progression of proteasomal degradation can be inhibited by simple sequences. (**a**) Fusion proteins containing an SRR (solid black triangle) or an NRR (solid red triangle) were degraded by the proteasome, which led to the accumulation of fragments (SRR, open black triangle; NRR, open red triangle). The extent of degradation and fragment formation was quantified and plotted as in **Figure 1**. (**b**–**f**) The amount of fragment formation is correlated with both the complexity of the simple sequence insert and the stability of the folded domain. Relative fragment formation represents the amount of fragment formed in the presence of 10 μM NADPH relative to the amount of full-length protein degraded. Relative fragment formation for substrates with six different simple sequence inserts was plotted according to the hydrophobicity (**b**), total charge (**c**), net charge (**d**) and amino acid–sequence complexity (**e**) of the simple sequence and according to the stability of the folded domain (**f**). The simple sequences used were SRR, NRR, GRR, the repeat region from the C terminus of RNA polymerase II (RPB), the PEST domain from IκBα (PEST) and the Ci cleavage region (Ci).

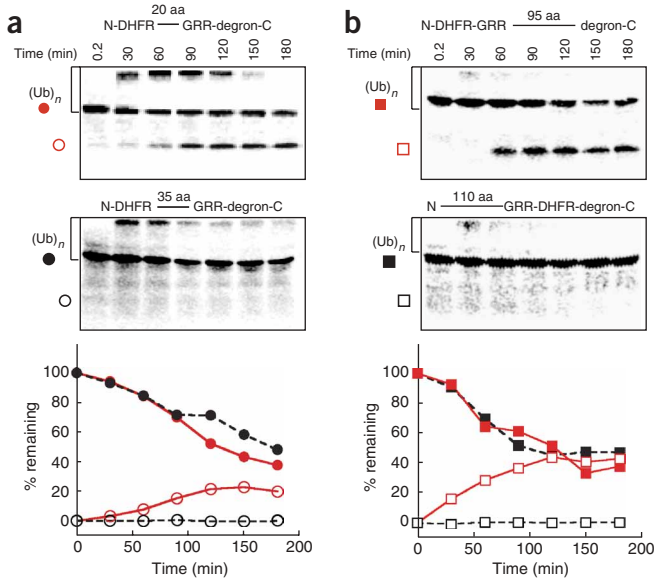

Figure 3 The spacing of the simple sequence inserts relative to the folded domain determines fragment formation. ^{35}S-labeled fusion proteins were degraded by the proteasome from a C degron at 30 °C. The extent of degradation and fragment formation was quantified as in **Figure 1**. (**a**) Inserting a 20-residue-long (20 aa) spacer region between DHFR and GRR (full-length protein, solid red circle) did not prevent fragment (open red circle) formation, whereas inserting a 35-residue spacer did prevent it (full-length protein, solid black circle). (**b**) Inserting a 95-residue spacer between the GRR and C degron (full-length protein, solid red square) did not affect the amount of fragment (open red square) accumulating. However, when the GRR insert was placed on the opposite side of DHFR to the C degron, the full-length protein (solid black square) was degraded completely and no partially degraded product (open black square) was detected.

methotrexate (MTX)[4,28] restored fragment accumulation (**Fig. 1b**). Thus progression of degradation is blocked by a signal consisting of two components: a glycine-rich sequence adjacent to a stably folded domain. The glycine-rich sequence by itself has no effect on proteasomal degradation.

Effects of simple sequences

The inhibition of proteasomal degradation progression is not specific to the identity of the amino acids in the simple-sequence region. For example, replacing the GRR with a serine-rich region (SRR) or an asparagine-rich region (NRR) did not substantially affect fragment accumulation during proteasomal degradation (**Fig. 2a**). To compare the magnitude of the effect of different simple sequences on proteasomal degradation, we measured the amount of partially digested fragment formed during degradation of a series of *E. coli* DHFR substrates in which different sequences were inserted next to the DHFR domain. The amount of fragment formed was standardized to the amount of full-length protein degraded.

When six unrelated sequences of varying complexity were inserted between an *E. coli* DHFR domain and the C degron, proteasome degradation of the resulting proteins led to the accumulation of different amounts of partially degraded substrate (**Fig. 2**). The extent of fragment formation did not correlate with the hydrophobicity (**Fig. 2b**) or charge (**Fig. 2c,d**) of the inserted sequences but instead with their compositional complexity as defined by the SEG algorithm[32] (**Fig. 2e**). The greatest accumulation of fragments occurred with the sequence of lowest complexity. The accumulation of fragments decreased with increasing complexity of the inserted simple sequence (**Fig. 2e**). At the same time, the amount of fragments formed also depended on the stability of the folded domain. Modulating the stability of *E. coli* DHFR by varying the amount of the more weakly binding DHFR ligand NADPH in the degradation reaction affected the accumulation of fragments (**Fig. 2f**). The more stable the DHFR domain was against unfolding, the more fragments accumulated.

Spacing requirements

Partially degraded fragments accumulated when the GRR was inserted immediately next to an *E. coli* DHFR domain (**Fig. 1a**) or when a 20-residue spacer separated the folded domain and the simple sequence (**Fig. 3a**). However, fragment formation was abolished when the GRR was moved 35 residues (**Fig. 3a**) or more (55 or 95 residues, data not shown) away from the folded domain. The spacers were unstructured, because they were highly sensitive to trypsin and proteinase K digestion and migrated slower than expected in gel filtration chromatography experiments (data not shown), but they were not low-complexity regions as determined by the SEG algorithm[32]. The spacing of the simple sequence relative to the C degron did not affect fragment formation (**Fig. 3b**). However, the simple sequence had to be located between the folded domain and the C degron for fragment formation. Placing it on the N-terminal side of a folded domain, which in turn was followed by a C degron, led to complete degradation (**Fig. 3b** and **Supplementary Fig. 3** online). The same spacing requirements were also observed for other hydrophilic simple sequences such as an NRR (data not shown).

Progression of degradation can also be stopped when it proceeds from the N terminus toward the C terminus, but the two components of the stop signal have to be spaced differently relative to each other. In constructs targeted for proteasome degradation by the N-end rule[33], fragments accumulated when a GRR was placed at a distance of 55 residues before *E. coli* DHFR (**Fig. 4a**) but not when the GRR was immediately adjacent to the folded domain (**Fig. 4a**). Fragment formation was dependent on the presence of a tightly folded domain, because replacing *E. coli* DHFR with mouse DHFR prevented fragment formation unless DHFR was stabilized with MTX (**Fig. 4b**). As before, the GRR could be replaced with an NRR or SRR without affecting the extent of fragment accumulation substantially (**Fig. 4c**). Fragment formation was due to the ubiquitin-proteasome system, because it could be abolished by mutating the N-end rule residue (**Supplementary Fig. 1**) to interfere with the ubiquitination machinery[33,34].

Initiation of proteasome degradation

To obtain insights into the biochemical mechanism by which a simple sequence in combination with a folded domain leads to partial degradation, we investigated the effect of the amino acid composition of a substrate on its interaction with the proteasome. The proteasome degrades its substrates by binding to their ubiquitin modification and then initiating degradation at an unstructured region[5]. The proteasome engages its substrate at the initiation site, possibly by threading it into the degradation channel. Simple sequences also prevented the initiation step of degradation. For example, a DHFR substrate protein was inefficiently digested when a degron consisting of four ubiquitin moieties was fused directly to its N terminus[5] (**Fig. 4d**). Attachment of unstructured extensions to the substrate's C terminus led to its rapid degradation (**Fig. 4d**). However, replacement of the unstructured extension in the DHFR constructs with a GRR abolished degradation (**Fig. 4d**). Similarly, none of the other hydrophilic simple sequences tested here functioned as an initiation site (**Fig. 4d**). The simple

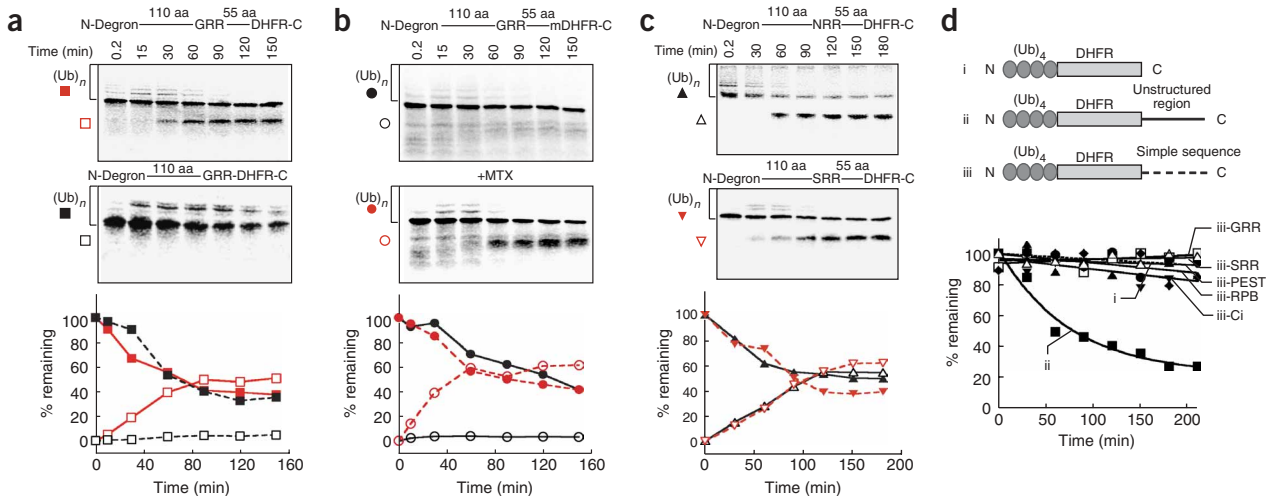

Figure 4 Simple sequence inserts can inhibit proteasome progression in both directions of degradation. Simple sequences do not function as efficient degradation initiation sites. ^{35}S-labeled fusion proteins were degraded by the proteasome from an N-terminal degron (N degron) at 30 °C. The extent of degradation and fragment formation was quantified as in **Figure 1**. (**a**) Substrate protein containing an N degron followed by a glycine-rich insert, a 55-residue spacer and an *E. coli* DHFR domain (solid red square) was degraded by the proteasome, which resulted in the accumulation of partially degraded product (open red square) corresponding in size to the DHFR domain plus ~80 residues as judged by its gel mobility (not shown). Placing the GRR insert adjacent to the N terminus of DHFR by removing the spacer region abolished fragment formation (full-length protein, solid black square). (**b**) When *E. coli* DHFR was replaced with the less stable mouse homolog (mDHFR, solid black circle), the proteasome degraded the substrate protein completely and no fragment accumulated (open black circle). Stabilizing mDHFR with 320 μM MTX restored fragment formation (full-length protein, solid red circle; fragment, open red circle). (**c**) The GRR insert can be replaced with an asparagine-rich (solid black triangle) or serine-rich (solid red triangle) sequence without reducing the accumulation of partially degraded product (NRR, open black triangle; SRR, open red triangle). (**d**) Degradation of a fusion protein consisting of four ubiquitin sequences fused in frame to *E. coli* DHFR by the proteasome at 25 °C. Construct i lacked an extended unstructured region that could serve as a degradation initiation site and was therefore hydrolyzed only inefficiently. Degradation was accelerated when an unstructured region was attached to the C terminus of the protein (ii). However, proteins with hydrophilic simple sequences (GRR, SRR, PEST, RPB or Ci, defined as in **Figure 2**) fused to their C termini as initiation sites (iii) were not degraded.

sequence tags were all highly sensitive to proteolysis by proteases such as trypsin and proteinase K, suggesting they are unstructured and accessible for initiation of degradation (data not shown). These results are very similar to the earlier observation that a stretch of glycine residues cannot serve as an N degron, even if it contains a destabilizing N-end rule residue and an appropriate lysine residue as ubiquitin acceptor[34]. They suggest that the proteasome cannot engage its substrates at a simple sequence during the initiation of degradation. Specifically, the GRR cannot serve as a degradation initiation or proteasome entry site as proposed previously[22,23].

Ci processing *in vitro*

The partial degradation of our model substrates is reminiscent of the proteasome-dependent processing of the transcription factors Ci and NF-κB. We proposed that processing of these transcription factors may be caused by a degradation stop signal consisting of a folded domain and a simple sequence.

In cells that do not receive an Hh signal, Ci155 is converted into its Ci75 fragment[14]. The Ci protein contains a DNA-binding domain consisting of five tightly folded C2H2 zinc fingers between residues 451 and 603 followed by a cleavage region from residues 615 to 760 (refs. 14,27) and an acidic C-terminal activation domain (**Fig. 5a**). The cleavage region contains three stretches of simple sequences, a stretch rich in glutamine and asparagine between residues 626 and 642, a serine-rich stretch between residues 655 and 670, and an aspartate-rich stretch between residues 689 and 712. We suspected that processing occurs when the proteasome sequentially degrades Ci from a degron near the C terminus, but its progression toward the N terminus is blocked by a stop signal consisting of the simple sequences and the fifth zinc finger of the DNA-binding domain.

To test this proposal, we constructed a protein substrate consisting of the DNA-binding and cleavage region of Ci (residues 440–760) fused to the C degron used in the model proteins described above (**Fig. 5a**). The construct was ubiquitinated in reticulocyte lysate, and its degradation yielded the accumulation of a partially degraded fragment (**Fig. 5b**). Fragment formation depended on a stably folded fifth zinc-finger domain. Removing the Zn^{2+} in the reaction mixture with the chelator tetrakis-(2-pyridylmethyl)-ethylenediamine (TPEN; **Fig. 5b**) or destabilizing this domain by mutating the two Zn^{2+}-binding cysteine residues to alanine also led to complete degradation of the protein (**Fig. 5b**). Fragment formation required the presence of the cleavage region, because deletion of residues 615–760 abolished processing (**Fig. 5c**). However, when the cleavage region was replaced by entirely unrelated glycine-rich or serine-rich sequences, fragment formation was restored (**Fig. 5c**). Indeed, these replacements led to the accumulation of larger amounts of partially degraded protein, presumably because the amino acid compositions of the glycine-rich and serine-rich sequences are less complex than that of the simple sequences in the Ci cleavage region (**Fig. 2e**).

Ci processing in cell culture

The *in vitro* assays were conducted in a mammalian cell lysate using a heterologous C degron, and thus it was possible that processing in *Drosophila* cells occurred by a different mechanism. Therefore, we monitored processing of Ci (**Fig. 6a**) in *Drosophila* Kc and Schneider 2 (S2) culture cells. Transient transfection of Kc cells with hemagglutinin

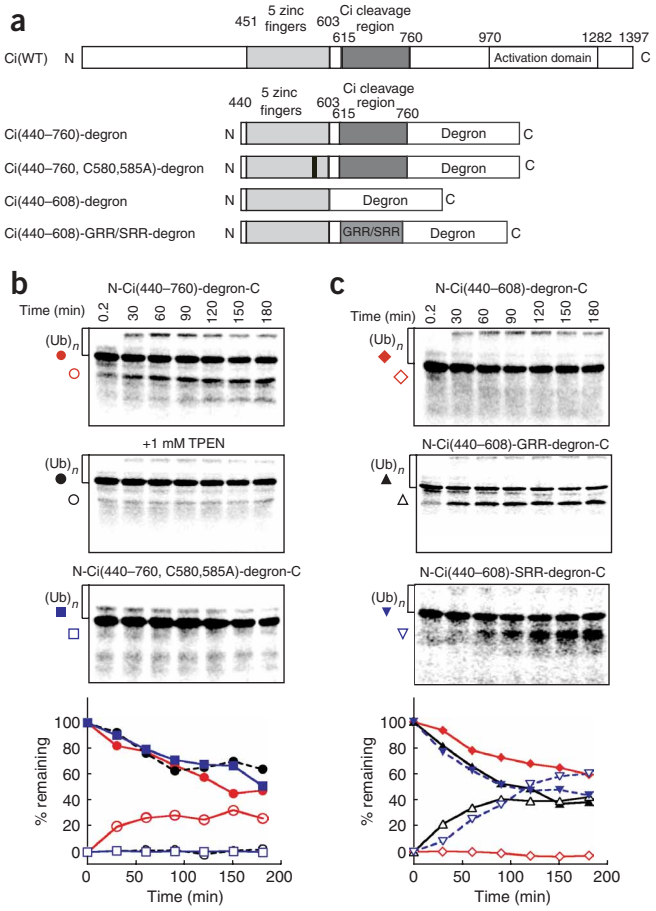

Figure 5 Processing of Ci by the proteasome *in vitro*. ^{35}S-labeled proteins containing Ci sequences were degraded from a C degron at 30 °C. The extent of degradation and fragment formation was quantified as in **Figure 1**. (**a**) Schematic representation of Ci-derived substrates for *in vitro* degradation. (**b**) Degradation of a fusion protein containing Ci(440–760) from a C degron (solid red circle) resulted in the accumulation of a partially degraded product (open red circle). Destabilizing the last zinc-finger motif by chelating the free zinc ions by 1 mM TPEN (solid black circle) or point mutations C580A and C585A (full-length protein, solid blue square) led to complete degradation. (**c**) Deletion of the Ci cleavage region (residues 615–760) (full-length protein, solid red diamond) also abolished fragment formation. Replacing the cleavage region with glycine-rich (solid black triangle) or serine-rich (solid blue triangle) inserts also led to the accumulation of partially degraded product (GRR, open black triangle; SRR, open blue triangle).

(HA)-tagged Ci constructs resulted in the expression of full-length Ci protein (**Fig. 6b**). Wild-type Ci was processed so that both full-length Ci155 and processed Ci75 were detected by western blotting. Deletion of the cleavage region (residues 615–760) abolished accumulation of the processed form (**Fig. 6b**), as was reported for Cl-8 cells[27]. Replacement of the cleavage region with the glycine-rich sequence increased the level of Ci75 substantially compared with that observed for wild-type Ci (**Fig. 6b**). This enhanced processing in Kc cells correlated with the effect of the glycine-rich sequence on processing *in vitro* (**Figs. 2** and **5**). Destabilization of the last zinc finger by mutating two Zn^{2+} ligands (cysteines 580 and 585 to alanine) abolished Ci75 formation (**Fig. 6b**). Experiments with S2 cells yielded equivalent results (data not shown).

To determine whether the processing of Ci and its derivatives described earlier affected the downstream signaling pathway, we examined the transcriptional activity of these constructs in Kc and S2 cells. For this purpose, we cotransfected Kc cells with the different Ci constructs, a reporter construct containing the firefly luciferase gene under control of the patched (*ptc*) promoter, which contains three Ci-binding sites[35], and a vector that either was or was not expressing Hh. Deletion of the cleavage region in Ci led to a ~3.5-fold increase in luciferase activity compared with wild-type Ci (**Fig. 6c**), presumably because formation of the Ci repressor form was abolished. In contrast, when the cleavage region was replaced with the GRR, Ci-regulated luciferase activity was reduced about four-fold compared with the wild-type Ci (**Fig. 6c**), presumably because the GRR is more effective at stopping proteasome progression (**Fig. 6b**). Notably, Ci in which the whole cleavage region was replaced by the GRR was still regulated by Hh signaling in the expected way. Together with the results obtained *in vitro*, these observations suggested that Ci processing is caused by a degradation stop signal consisting of the stably folded zinc-finger domain adjacent to a simple sequence that

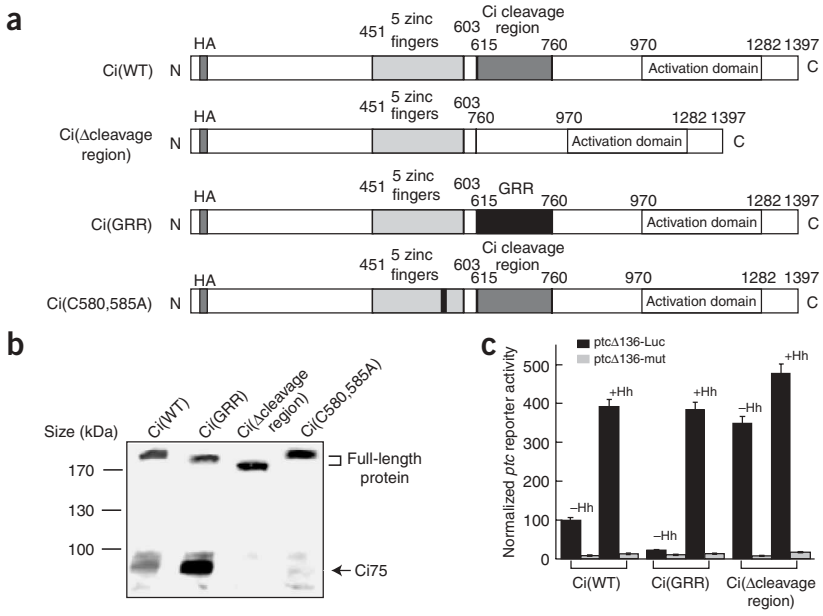

Figure 6 Processing of Ci *in vivo*. (**a**) Schematic representation of HA-tagged Ci proteins under the control of an actin 5C promoter. (**b**) Western blot analysis of Ci processing in Kc167 cells. The positions of full-length wild-type Ci protein and its partially degraded 75-kDa fragment detected with an antibody to HA are indicated. (**c**) Transcriptional activity of Ci and its mutants in Kc167 cells containing a reporter plasmid with luciferase under the control of a modified patched promoter (ptcΔ136-Luc) or a patched promoter in which the Ci-binding sites were mutated (ptcΔ136-mut). Whole-cell extracts were assayed for luciferase activity and normalized to cotransfected *Renilla* luciferase. Luciferase activities each represent the average from triplicate samples, with s.e.m. shown.

prevents further proteasomal degradation and thus spares the N terminus of Ci (Ci75) from proteolysis.

p105 processing

The other known example of proteasome-mediated regulation of transcription factor activity is the generation of the p50 subunit of NF-κB from its precursor protein p105. Starting from the N terminus, the processed p50 contains a Rel-homology domain followed by a glycine-rich cleavage region[23,24] (**Fig. 7a**).

Processing of p105 to p50 is due to a two-part signal as described earlier, with the GRR serving as the simple sequence and the dimerization domain of the Rel-homology region forming the tightly folded domain. When the processing reaction was reconstituted in reticulocyte lysate *in vitro* (**Fig. 7b**), changing the spacing between the GRR and the stably folded DNA-binding domain by insertion of a 95-residue spacer prevented the accumulation of p50 (**Fig. 7b**). Fragment formation did not depend on the amino acid composition of the GRR, and replacing it with an SRR did not abolish processing (**Fig. 7c**). Destabilization of the Rel-homology domain by a point mutation abolished formation of p50 (ref. 4) (**Fig. 7c**).

Processing of p105 constructs in mammalian culture cells followed the same rules as observed in the *in vitro* experiments. Flag-tagged p105 transiently transfected into HEK 293T cells was processed to the p50 form (**Fig. 7d**). Destabilizing the Rel-homology domain with the mutation F348A prevented the accumulation of p50 (**Fig. 7d**), as did disrupting the spacing between the Rel-homology domain and the GRR with the insertion of a 95-residue spacer. Replacing the GRR with an SRR did not affect processing substantially (**Fig. 7d**).

To investigate how processing of p105 and its mutants affects NF-κB transcriptional activity, an NF-κB–dependent luciferase reporter gene assay was done. Expression of wild-type p105 inhibited luciferase activity approximately two-fold (**Fig. 7e**), consistent with efficient production of p50. Replacing the GRR with an SRR also inhibited the reporter (**Fig. 7e**). However, inactivating the processing signal, either by disrupting the appropriate spacing between the GRR and the folded domain (p105(spacer)) or by destabilizing the folded domain (F348A), restored luciferase activity to the levels observed in the absence of p105 (**Fig. 7e**), consistent with a failure to form p50 (**Fig. 7e**). The processing signal in p105 is already strong enough to allow the complete conversion of full-length protein into fragments; thus, replacing the GRR with an SRR did not enhance the amount of p50 formation.

Therefore, our *in vivo* results confirmed that a signal consisting of two components can block progression of proteasomal degradation along the polypeptide chain and cause processing. The two components of the signal are a simple sequence followed closely by a stably folded domain in the direction of proteasome movement.

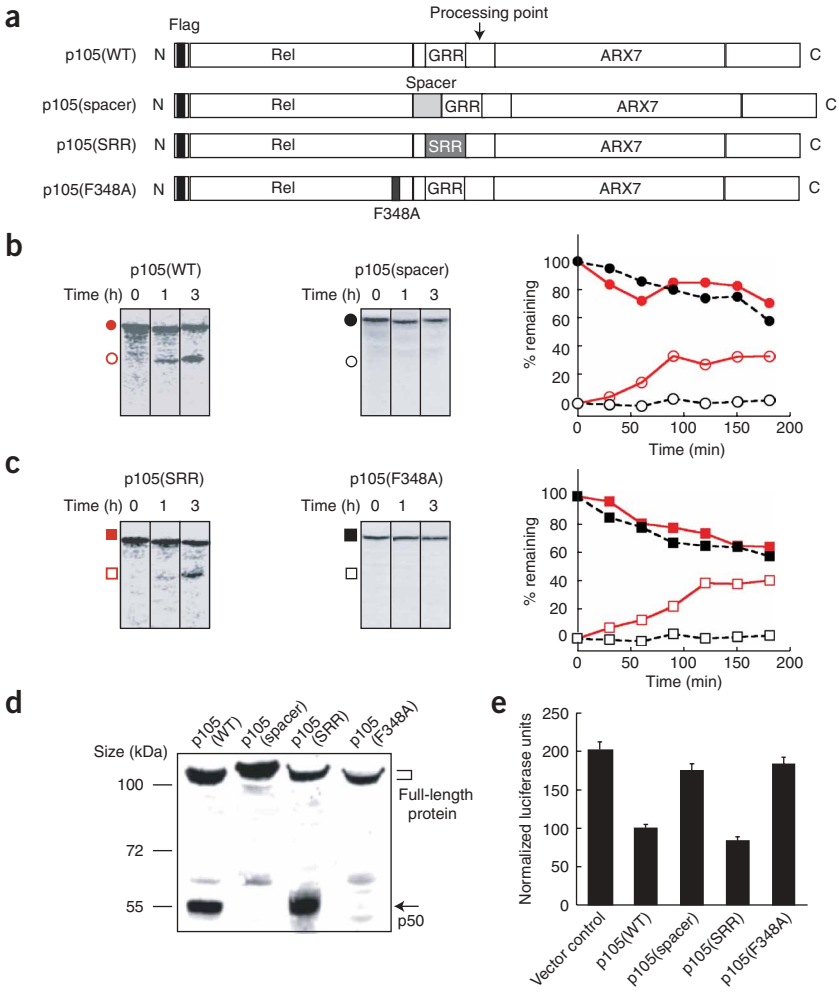

Figure 7 Processing of NF-κB precursor p105 depends on the presence of a folded Rel-homology domain and a GRR. (**a**) Schematic representation of wild-type p105 and its mutants. (**b**) ^{35}S-labeled p105 (solid red circle) was degraded by the proteasome at 30 °C, which led to the formation of processed p50 (open red circle), as shown in autoradiograms of SDS-PAGE gels. p105(spacer) protein (solid black circle) containing a 95-residue insertion between the GRR and the Rel-homology domain was degraded but did not lead to formation of p50. (**c**) When the GRR was replaced with an unrelated SRR, the resulting protein p105(SRR) (solid red square) was processed like wild-type p105 to p50 (open red square). Destabilizing the Rel-homology domain by point mutation (F348A, solid black square) completely prevented p50 formation. The extent of degradation and fragment formation in **b** and **c** was quantified as in **Figure 1**. (**d**) Western blot analysis of p105 processing *in vivo*, as in **Figure 6**. (**e**) Generation of p50 inhibited the intrinsic NF-κB activity in 293T cells. Luciferase activity was measured 24 h after transfection and normalized to cotransfected *Renilla* luciferase. The values represent the average from triplicate samples, with s.e.m. shown.

DISCUSSION

Proteasomal proteolysis controls the cellular concentrations of hundreds of regulatory proteins. Normally, the proteasome degrades its substrates completely into small peptides by sequentially running along their polypeptide chain and hydrolyzing the peptide bonds approximately every 8 residues[6]. The proteasome can also function as a processing enzyme that produces functional protein fragments from larger precursors by partial degradation[22]. Processing occurs

when the proteasome encounters a stop signal during its sequential hydrolysis of a substrate protein. The stop signal consists of two components: a sequence of low compositional complexity followed by a tightly folded domain in the direction of proteasome movement. Glycine-alanine repeat regions in the Epstein-Barr virus nuclear antigen-1 are known to protect the protein from proteasomal degradation[36] and to be transferable to other proteins[36,37]. Here, we find that many different simple sequences can cause partial degradation, but only in combination with a tightly folded domain and at the appropriate spacing.

Partial degradation can modulate the function of regulatory proteins, as shown for p105 (ref. 17) and Ci[14], and provides a simple mechanism for directly switching a signaling pathway from an active state to a repressed state and vice versa. Processing can also produce more subtle changes in activity. The amount of fragment formed during the degradation of a protein depends on the strength of the processing signal. More fragment is formed the less complex the simple sequence (**Fig. 2e**) and the more stable the folded domain (**Fig. 2f**). For example, the low-complexity region of Ci is not as simple as the GRR in p105. Therefore, less Ci repressor is formed in the absence of Hh signaling (**Fig. 6b**), and a reporter target gene is less tightly repressed (**Fig. 6c**) than possible if the low-complexity regions of Ci were replaced with the p105 GRR (**Fig. 6b,c**). The ratio of Ci activator to Ci repressor in turn determines the activity of the Ci target genes, and, therefore, processing efficiency affects the shape of the Ci activity gradient at the distal edge of Hh signaling. The second component of the processing signal is the susceptibility of folded domains to unraveling by the proteasome, which depends on the stability of the local structure first encountered by the proteasome[4]. A very stable domain, such as MTX-stabilized DHFR, can lead to fragment formation without a neighboring simple sequence[4,28]. It may be possible to modulate the strength of a processing signal in the cell by modifying the simple sequence, for example by phosphorylation, or by adjusting the stability of the folded domain, for example by ligand binding.

Processing signal function depends on the direction of degradation, because the proteasome has to encounter the simple sequence before the folded domain (**Fig. 3b**) and because the susceptibilities of folded domains to unraveling from the N or C terminus can differ[5]. This polarity provides a simple mechanism for the degradation of the protein fragments when they are no longer needed. In the eye, the protein-ubiquitin ligase Cul3 targets Ci for complete degradation using unknown ubiquitination sites[38]. We predict that Cul3 ubiquitinates in the N-terminal region of Ci, which would lead to its complete degradation by the proteasome unimpaired by the processing signal.

Other examples of protein processing by the proteasome probably exist. The NF-κB subunit p52 is synthesized as the larger precursor p100 (ref. 39), which is homologous to p105 and presumably processed by the same mechanism. Vertebrates have three Ci homologs, Gli1, Gli2 and Gli3, but only the latter two are processed to a smaller form, probably in a proteasome-dependent manner[40]. Consistent with this observation, only Gli2 and Gli3 seem to have a processing signal and Gli1 may not be ubiquitinated. Processing could also exist in proteins unrelated to Ci or NF-κB and does not have to be limited to transcription factors. Regions of low compositional complexity are common and found in half of all predicted eukaryotic proteins[32], but to form a processing signal, a simple sequence must be positioned adjacent to a tightly folded domain at the appropriate spacing. Standard sequence alignments cannot detect the processing signals, because folded domains can be formed by unrelated sequences and the function of the simple sequence does not depend on the identity of the repeated amino acids.

The ubiquitin-proteasome system is also involved in the activation of the two membrane-bound yeast transcription factors Spt23 and Mga2 by the release of N-terminal fragments of these proteins from the membrane[8,41]. Processing requires folded domains that are homologous to the Rel-homology domain of p105 (refs. 8,22,41), but the proteins do not contain simple sequences at the expected places. Notably, p105 processing in yeast also does not depend on the presence of a simple sequence[26]. However, because Spt23 and Mga2 are membrane anchored and degradation seems to proceed from an internal loop[8,22,41], it is also possible that the processing is more complicated than the simple mechanism described here.

The biochemical mechanism by which the processing signal causes partial degradation is not known. Prokaryotic ATP-dependent proteases can release their substrate when they reach protein domains that are hard to unfold[42]. In all the substrates described here, the ubiquitination sites had been degraded by the time the proteasome reached the folded domain in their substrate. Thus, the protease was associated with its substrate only through the part of the substrate that was about to be degraded, which contained the simple sequence. The spacing requirement between simple sequence and folded domain for processing differs with the direction of proteasome movement, and this disparity may indicate that the proteasome interacts with its substrates differently depending on the direction of degradation. The simple sequences could then lead to processing if they reduced the affinity of the substrate for the proteasome. In apparent agreement with this proposal, simple sequences cannot serve as efficient degradation initiation sites (**Fig. 4d**). Once the substrate is released from the proteasome, it has escaped proteasomal degradation because the ubiquitination site has already been removed. Thus, according to this model, protein processing occurs when a tightly folded domain located at the entrance to the degradation channel stalls the progression of the proteasome along the polypeptide chain and a simple sequence in the channel accelerates the release of the folded domain and the remaining protein from the proteasome.

In summary, we showed that a combination of a simple sequence followed by a folded domain in the direction of proteasome movement inhibits proteasome progression, which results in the accumulation of a partially degraded fragment. Proteasomal processing by partial degradation has an essential function in at least two unrelated cellular signaling pathways. The processing mechanism is conserved between flies and humans, and more examples of this process probably exist.

METHODS

Substrate proteins. Protein substrates were derived from *E. coli* or mouse DHFR, the transcription factor Ci and the NF-κB precursor protein p105. Model substrates were targeted for ubiquitination and degradation by the proteasome from their C terminus by a C degron corresponding to residues 430–646 of p105 or from their N terminus by the N-end rule N degron[34]. The spacer region in protein substrates was derived from residues 1–20, 1–35, 1–55 or 1–95 of cytochrome b_2 in which all the lysine residues were replaced by arginine. Human NF-κB precursor p105 was expressed from pT7β105 provided by A. Ciechanover (Technion-Israel Institute of Technology, Haifa, Israel). The linker region used was protein barnase. In the p105(spacer) construct, residues 1–95 of cytochrome b_2 but with all lysines replaced by arginines were placed between residues 368 and 369 of p105; in the p105(SRR) construct, the GRR (residues 367–403) of p105 was replaced with the SRR described later. Radioactive proteins were expressed from the T7 promoter by *in vitro* transcription and translation in rabbit reticulocyte lysate (Promega) supplemented with [^{35}S]methionine.

Simple sequences. The glycine-rich insert corresponded to residues 367–403 of p105; the serine-rich insert contained two tandem repeats of residues 178–196 of transcription factor ICP4; the asparagine-rich insert contained two tandem repeats of residues 373–386 from transcription factor SPT23; the RPB repeat insert contained three tandem repeats of residues 1689–1702 of RNA polymerase II subunit 1 (RPB1); the PEST insert was composed of two tandem repeats of residues 284–307 of IκBα; the Ci cleavage region insert consisted of residues 615–760 of Ci. The complexity values of the simple sequence inserts were calculated by SEG[32]. Hydrophobicity per residue of the simple sequence inserts was calculated by applying the Kyte-Doolittle algorithm[43], and total and net charge were calculated using the ProtParam tool in the ExPASy suite[44].

Gel filtration chromatography. ^{35}S-labeled proteins in degradation buffer (5 mM Tris-HCl, pH 7.5, 25 mM KCl, 5% (v/v) glycerol, 2 mM $MgCl_2$) were chromatographed over a 20-ml Sephacryl S-200 column (Amersham Biosciences). Fractions of 250 µl were collected and precipitated with 10% (w/v) trichloroacetic acid. The pellet was dissolved in SDS gel sample buffer and resolved by 10% SDS-PAGE, and the radioactive signal was quantified using electronic autoradiography (Instant Imager; Packard).

Proteasome degradation assay *in vitro*. Proteasome degradation assays were conducted in ATP-depleted rabbit reticulocyte lysate (Green Hectare) essentially as described in ref. 33. *In vitro*–translated radiolabeled protein was precipitated with ammonium sulfate and then resuspended in 10 µl resuspension buffer (240 mM Tris-HCl (pH 7.5), 1.2 M KCl, 25% (v/v) glycerol, 100 mM $MgCl_2$). Typically, 3 µl resuspended protein was added to 141 µl reaction mixture (2 mM DTT, 25 µM added ubiquitin and 85% (v/v) reticulocyte lysate, final concentrations) and prewarmed at 30 °C for 3 min. Degradation was initiated by the addition of 6 µl ATP and ATP regeneration system (1 mM ATP, 10 mM creatine phosphate, 0.1 mg ml^{-1} creatine phosphokinase, final concentrations). Incubation was continued at 30 °C, and at the indicated times samples were withdrawn from the reaction mix and added to SDS-PAGE sample buffer to stop the reaction. Samples were analyzed by SDS-PAGE and electronic autoradiography (Instant Imager, Packard).

Cell culture experiments. HA-tagged wild-type Ci (pPac5C-HA-Ci) was a gift from S. Smolik[15] (Oregon Health & Science University, Portland, Oregon, USA). Ci(GRR) was constructed by replacing the cleavage region (residues 615–760) of Ci in pPac5C-HA-Ci with two tandem repeats of the GRR. Ci(C580,585A) and Ci(Δ cleavage region) were generated by QuikChange mutagenesis (Stratagene). Flag-tagged NF-κB precursor protein p105 and its mutants were subcloned from the corresponding pT7β105 plasmids into the pcDNA3 expression vector (Invitrogen) between the HindIII and NotI restriction sites. Where indicated, 1 µg of pPac5C-Hh or pPac5C without insert was cotransfected to allow the expression of the Hh.

Ci activity reporter plasmids ptcΔ136-Luc and ptcΔ136-mut containing the firefly luciferase gene under the control of wild-type and mutant *ptc* promoter, respectively, were gifts from R. Carthew (Northwestern University, Evanston, Illinois, USA). The NF-κB activity reporter plasmid (4XPRDII-luciferase) contained NF-κB–binding sequences from the interferon-γ promoter in its control region[45] and was provided by C.M. Horvath (Northwestern University).

Drosophila Kc 167 cells and S2 cells[35] were grown and maintained as described in ref. 15. HEK 293T cells were maintained in Dulbecco's modified Eagle's medium (Gibco-BRL) supplemented with 10% (v/v) cosmic calf serum (Hyclone Laboratories) and 1% (v/v) penicillin-streptomycin (Gibco).

Transfections and luciferase reporter gene assays. HA-tagged Ci and its variants were expressed in Kc cells or S2 cells. Cells were plated in 12-well dishes and transfected with 5.5 µg expression plasmid (1.5 µg pPac5C-HA-Ci or its variants, 4 µg pPac5C control vector) together with 0.5 µg reporter plasmid using Cellfectin (Invitrogen) and allowed to express for 48 h at 24 °C.

Flag-tagged p105 and its variants were expressed in HEK 293T cells. Cells were seeded in 60-mm dishes and transfected the next day with 3 µg reporter plasmid together with 10 µg pCDNA3 control vector or different p105 constructs using the calcium phosphate method and allowed to express for 24 h at 37 °C.

Luciferase activity was measured using the Dual Luciferase Reporter Assay System (Promega). Luciferase activity measurements were normalized to cotransfected *Renilla* luciferase activity and expressed as a percentage of stimulated wild-type luciferase activity (set to 100%).

Western blot analysis. For western blot analysis, cells were harvested and lysed in whole-cell extract buffer (50 mM Tris-HCl (pH 8.0), 280 mM NaCl, 0.5% (v/v) NP-40, 0.2 mM EDTA, 10% (v/v) glycerol, 1 mM DTT and 1% protease inhibitor cocktail (Calbiochem)). After separation by SDS-PAGE, 20 µg protein from the whole-cell extract was transferred to a nitrocellulose membrane (Pall Life Science). HA-tagged Ci proteins and Flag-tagged p105 proteins were detected by rat monoclonal antibody to HA or mouse monoclonal antibody to Flag (Sigma), respectively. Proteins were visualized by chemiluminescence (Supersignal, Pierce).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

This work is dedicated to Irving M. Klotz. We thank R.W. Carthew and M. Feller for help and advice with the Ci cell culture experiments and C. Horvath and J.-P. Parisien for help with the NF-κB cell culture experiments. We also thank J. Widom, R.W. Carthew, C. Horvath, S. Prakash and S. Fishbain for advice. We are grateful to N. Shah and C. Malalis for constructing several plasmids. The work was supported by grant GM063004 from the US National Institutes of Health.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

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

1. Glickman, M.H. & Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol. Rev.* **82**, 373–428 (2002).
2. Pickart, C.M. & Cohen, R.E. Proteasomes and their kin: proteases in the machine age. *Nat. Rev. Mol. Cell Biol.* **5**, 177–187 (2004).
3. Larsen, C.N. & Finley, D. Protein translocation channels in the proteasome and other proteases. *Cell* **91**, 431–434 (1997).
4. Lee, C., Schwartz, M.P., Prakash, S., Iwakura, M. & Matouschek, A. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol. Cell* **7**, 627–637 (2001).
5. Prakash, S., Tian, L., Ratliff, K.S., Lehotzky, R.E. & Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat. Struct. Mol. Biol.* **11**, 830–837 (2004).
6. Nussbaum, A.K. *et al.* Cleavage motifs of the yeast 20S proteasome β subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci. USA* **95**, 12504–12509 (1998).
7. Bachmair, A., Finley, D. & Varshavsky, A. *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186 (1986).
8. Hoppe, T. *et al.* Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing. *Cell* **102**, 577–586 (2000).
9. Chen, C.H. *et al.* Nuclear trafficking of Cubitus interruptus in the transcriptional regulation of Hedgehog target gene expression. *Cell* **98**, 305–316 (1999).
10. Jia, J. *et al.* Shaggy/GSK3 antagonizes Hedgehog signaling by regulating Cubitus interruptus. *Nature* **416**, 548–552 (2002).
11. Jiang, J. & Struhl, G. Regulation of the Hedgehog and Wingless signalling pathways by the F-box/WD40-repeat protein Slimb. *Nature* **391**, 493–496 (1998).
12. Palombella, V.J., Rando, O.J., Goldberg, A.L. & Maniatis, T. The ubiquitin-proteasome pathway is required for processing the NF-κB1 precursor protein and the activation of NF-κB. *Cell* **78**, 773–785 (1994).
13. Lum, L. & Beachy, P.A. The Hedgehog response network: sensors, switches, and routers. *Science* **304**, 1755–1759 (2004).
14. Aza-Blanc, P., Ramirez-Weber, F.A., Laget, M.P., Schwartz, C. & Kornberg, T.B. Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* **89**, 1043–1053 (1997).
15. Chen, Y., Gallaher, N., Goodman, R.H. & Smolik, S.M. Protein kinase A directly regulates the activity and proteolysis of Cubitus interruptus. *Proc. Natl. Acad. Sci. USA* **95**, 2349–2354 (1998).
16. Price, M.A. & Kalderon, D. Proteolysis of the Hedgehog signaling effector Cubitus interruptus requires phosphorylation by glycogen synthase kinase 3 and casein kinase 1. *Cell* **108**, 823–835 (2002).
17. Fan, C.M. & Maniatis, T. Generation of p50 subunit of NF-κB by processing of p105 through an ATP-dependent pathway. *Nature* **354**, 395–398 (1991).
18. Henkel, T. *et al.* Intramolecular masking of the nuclear location signal and dimerization domain in the precursor for the p50 NF-κB subunit. *Cell* **68**, 1121–1133 (1992).
19. Rice, N.R., MacKichan, M.L. & Israel, A. The precursor of NF-κB p50 has IκB-like functions. *Cell* **71**, 243–253 (1992).

20. Orian, A. *et al.* SCF(β-TrCP) ubiquitin ligase-mediated processing of NF-κB p105 requires phosphorylation of its C-terminus by IκB kinase. *EMBO J.* **19**, 2580–2591 (2000).
21. Karin, M. & Ben-Neriah, Y. Phosphorylation meets ubiquitination: the control of NF-κB activity. *Annu. Rev. Immunol.* **18**, 621–663 (2000).
22. Rape, M. & Jentsch, S. Taking a bite: proteasomal protein processing. *Nat. Cell Biol.* **4**, E113–E116 (2002).
23. Lin, L. & Ghosh, S. A glycine-rich region in NF-κB p105 functions as a processing signal for the generation of the p50 subunit. *Mol. Cell. Biol.* **16**, 2248–2254 (1996).
24. Orian, A. *et al.* Structural motifs involved in ubiquitin-mediated processing of the NF-κB precursor p105: roles of the glycine-rich region and a downstream ubiquitination domain. *Mol. Cell. Biol.* **19**, 3664–3673 (1999).
25. Lin, L., DeMartino, G.N. & Greene, W.C. Cotranslational biogenesis of NF-κB p50 by the 26S proteasome. *Cell* **92**, 819–828 (1998).
26. Sears, C., Olesen, J., Rubin, D., Finley, D. & Maniatis, T. NF-κB p105 processing via the ubiquitin-proteasome pathway. *J. Biol. Chem.* **273**, 1409–1419 (1998).
27. Methot, N. & Basler, K. Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. *Cell* **96**, 819–831 (1999).
28. Johnston, J.A., Johnson, E.S., Waller, P.R.H. & Varshavsky, A. Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. *J. Biol. Chem.* **270**, 8172–8178 (1995).
29. Lin, L., DeMartino, G.N. & Greene, W.C. Cotranslational dimerization of the Rel homology domain of NF-κB1 generates p50-p105 heterodimers and is required for effective p50 production. *EMBO J.* **19**, 4712–4722 (2000).
30. Endo, T. & Schatz, G. Latent membrane perturbation activity of a mitochondrial precursor protein is exposed by unfolding. *EMBO J.* **7**, 1153–1158 (1988).
31. Iwakura, M., Nakamura, T., Yamane, C. & Maki, K. Systematic circular permutation of an entire protein reveals essential folding elements. *Nat. Struct. Biol.* **7**, 580–585 (2000).
32. Wootton, J.C. & Federhen, S. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **266**, 554–571 (1996).
33. Gonda, D.K. *et al.* Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700–16712 (1989).
34. Bachmair, A. & Varshavsky, A. The degradation signal in a short-lived protein. *Cell* **56**, 1019–1032 (1989).
35. Fukumoto, T., Watanabe-Fukunaga, R., Fujisawa, K., Nagata, S. & Fukunaga, R. The fused protein kinase regulates Hedgehog-stimulated transcriptional activation in *Drosophila* Schneider 2 cells. *J. Biol. Chem.* **276**, 38441–38448 (2001).
36. Sharipo, A., Imreh, M., Leonchiks, A., Imreh, S. & Masucci, M.G. A minimal glycine-alanine repeat prevents the interaction of ubiquitinated IκBα with the proteasome: a new mechanism for selective inhibition of proteolysis. *Nat. Med.* **4**, 939–944 (1998).
37. Zhang, M. & Coffino, P. Repeat sequence of Epstein-Barr virus–encoded nuclear antigen 1 protein interrupts proteasome substrate processing. *J. Biol. Chem.* **279**, 8635–8641 (2004).
38. Ou, C.Y., Lin, Y.F., Chen, Y.J. & Chien, C.T. Distinct protein degradation mechanisms mediated by Cul1 and Cul3 controlling Ci stability in *Drosophila* eye development. *Genes Dev.* **16**, 2403–2414 (2002).
39. Heusch, M., Lin, L., Geleziunas, R. & Greene, W.C. The generation of NFκB2 p52: mechanism and efficiency. *Oncogene* **18**, 6201–6208 (1999).
40. Sasaki, H., Nishizaki, Y., Hui, C., Nakafuku, M. & Kondoh, H. Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling. *Development* **126**, 3915–3924 (1999).
41. Rape, M. *et al.* Mobilization of processed, membrane-tethered SPT23 transcription factor by CDC48(UFD1/NPL4), a ubiquitin-selective chaperone. *Cell* **107**, 667–677 (2001).
42. Kenniston, J.A., Baker, T.A. & Sauer, R.T. Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. *Proc. Natl. Acad. Sci. USA* **102**, 1390–1395 (2005).
43. Kyte, J. & Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
44. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. in *The Proteomics Protocols Handbook* (ed. Walker, J.M.) 571–607 (Humana Press, Totowa, New Jersey, USA, 2005).
45. Fan, C.M. & Maniatis, T. Two different virus-inducible elements are required for human β-interferon gene regulation. *EMBO J.* **8**, 101–110 (1989).