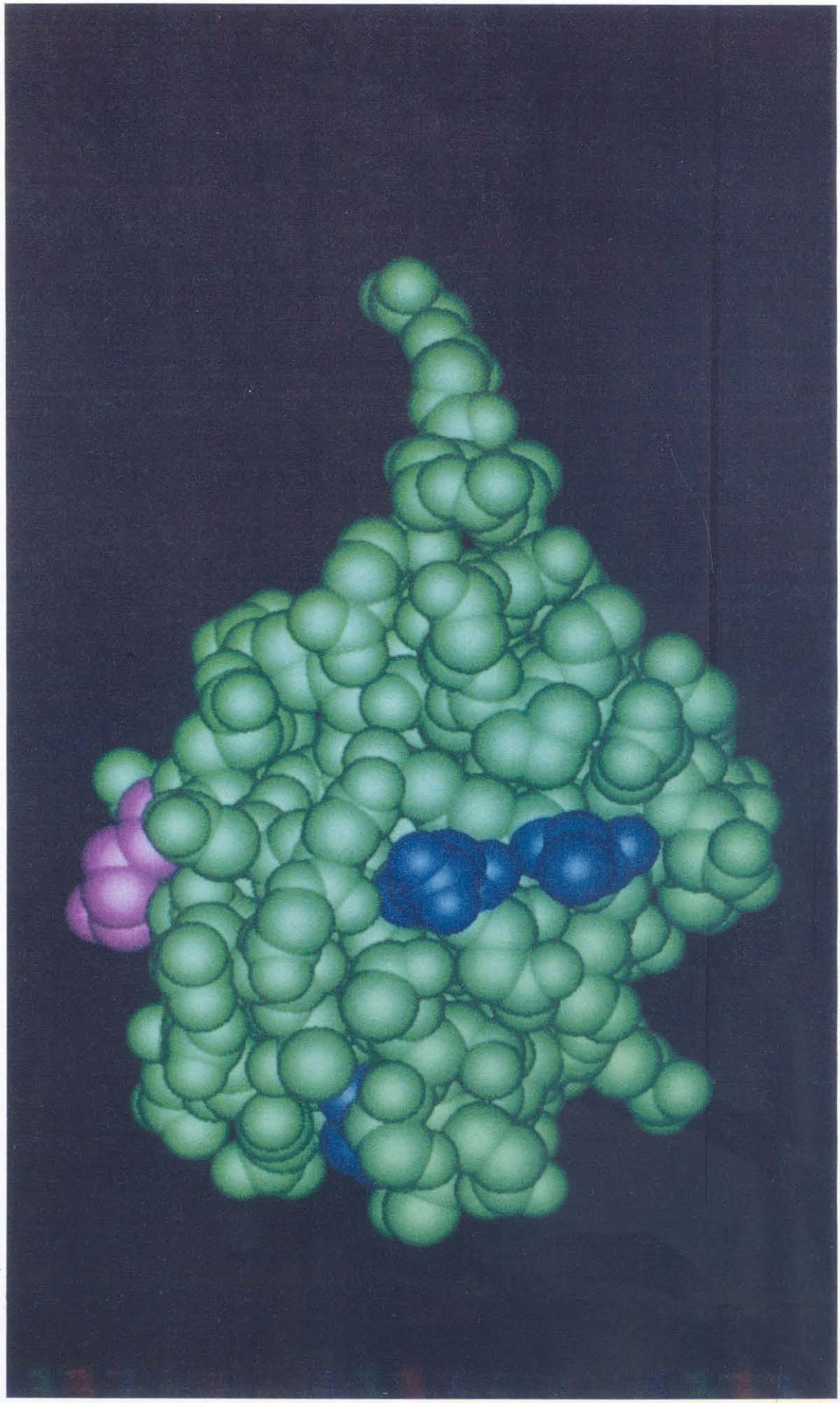


The World of Ubiquitin

The story of an old protein molecule is a tale of hazard and tear, of unceasing collisions with other molecules in the cell and assaults by a legion of highly reactive compounds that form in the process of metabolism.

by Alexander Varshavsky

The human ubiquitin molecule, shown here with its C-terminus at the top, differs from the yeast version by only the three amino acids rendered in blue. (The spheres represent individual atoms.) These three residues lie at positions 19, 24, and 28, as counted from the N-terminus. The pink atoms depict a lysine residue at position 48, through which another ubiquitin can attach itself to form a link in a multiubiquitin chain. (Ubiquitin's three-dimensional structure was determined by Senadhi Vijay-Kumar, Charles Bugg, and William Cook at the University of Alabama in Birmingham. Image courtesy of Michael Carson, Leigh Walter, and Cook.)

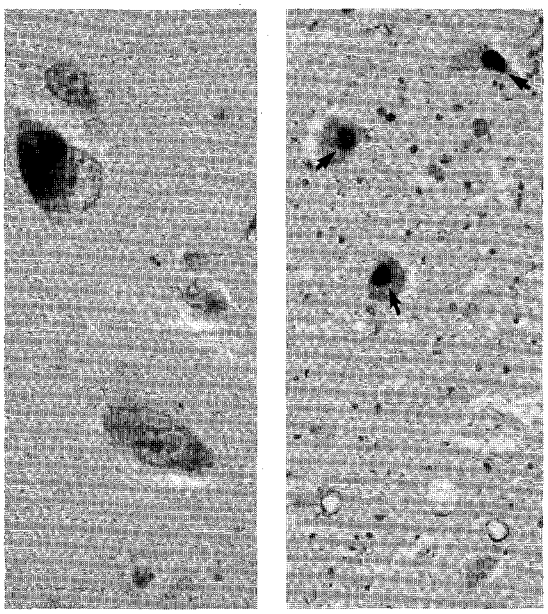
The pessimists have known it all along. Things of value in our eyes—fresh fruit, good weather, ourselves—tend to decay and fall apart. Proteins—the major constituents of living organisms—are no exception to this dreadful law. They are being destroyed inside and outside of cells, often in complicated ways, for a variety of reasons. The tale of protein degradation is a braid of interacting plots; in this article we focus on those that star a remarkable protein called ubiquitin.

But first, let's recall some basic molecular biology. Proteins are polymers, built from 20 different amino acids, which are assembled into linear chains according to instructions by segments of DNA called genes. The DNA's instructions are conveyed through messenger RNA to protein-making intracellular machines called ribosomes, which themselves are built from proteins and RNA. The protein's chain of amino acid residues (or simply residues) is called a polypeptide chain, and the residues are linked by chemical bonds called peptide bonds. The two distinct ends of a polypeptide chain are called the N-terminus and the C-terminus. The N-terminus bears a nitrogen-containing chemical group called the amino group, while the C-terminus bears the carbon-containing carboxyl group.

A newly formed protein, which emerges from the ribosome with its N-terminus first, faces a staggering variety of potential fates, one of which is degradation. Proteins are destroyed in a process called proteolysis, which may involve just a few cuts in a polypeptide chain, but can also result in the degradation of a protein all the way back to its constituent amino acids. Making proteins is an incredibly complex undertaking—why should they be destroyed at all? One reason for the existence of proteolysis is also kind of sad: proteins of a cell can be food for other cells, which often reside in a different organism. A lion dining on antelope looks utterly unlike a vegetarian munching a cucumber, but the strategy of both eaters is the same—to keep alive by subsisting on components of other living beings.

The enzymes (biological catalysts) that carry out proteolysis are a special class of proteins called proteases. Their size and complexity vary enormously—from relatively small proteases like trypsin and pepsin, which function outside of cells and digest proteins in food, to much larger ones called proteasomes, which consist of many protein subunits (polypeptide chains) and reside inside the cells.

Another function of proteolysis is the destruction of damaged or otherwise abnormal proteins. The story of an old protein molecule is a tale of hazard and tear, of unceasing collisions with other molecules in the cell and assaults by a legion of highly reactive compounds that form in the process of metabolism. Sometimes a protein molecule is abnormal from its very beginning, either because it is the product of a defective gene or because it failed to fold properly (folding properly is a complicated affair, assisted by special proteins). Yet another source of protein damage is environmental stress. Consider, for example, a yeast cell feeding on a grape at high noon. This cell has to cope, among other things,



Ubiquitin is a component of the intracellular protein aggregates characteristic of many human diseases, particularly those of the central nervous system. (The plaques that form in Alzheimer's disease are perhaps the best-known example.) An accumulation of these aggregates eventually kills the afflicted cell. Ubiquitin's role in aggregate formation (the aggregates also contain a variety of other proteins) remains unknown.

At left is a sample of brain cortex from a patient with Lewy Body dementia that has been stained with a ubiquitin-recognizing antibody (dark). A large ubiquitin-containing aggregate (called an inclusion body) is visible in one cell. The circular structures are cell nuclei.

At right, several ubiquitin-containing inclusion bodies (arrows) can be seen in the spinal cord neurons of a patient with amyotrophic lateral sclerosis, better known as Lou Gehrig's disease. Micrographs courtesy of John Mayer and colleagues at the University of Nottingham, England.

with the sun's heat—possibly a problem because the cell's temperature may become high enough to unfold and render inactive some of the yeast proteins.

Damaged proteins have to be repaired or eliminated. Protein repair systems (they do exist) are beyond the scope of our discussion. If repair fails or isn't attempted, a damaged protein has to be distinguished from its normal counterparts in the cell, singled out amidst the stir and bustle of other protein molecules, and then destroyed without perturbing nearby structures. We can now glimpse some of the reasons behind the complexity of the intracellular proteolytic machines—their task is vastly more subtle than the task of pepsin in the stomach, where every protein is fair game. The recognition and elimination of damaged proteins keeps a cell nearly, but not quite, free of them, because the surveillance mechanisms are blind to certain types of protein damage. In other cases, these mechanisms appear to recognize a damaged structure as such, but can't destroy it because it's protease-resistant or physically inaccessible—for example, by being a part of a huge protein aggregate, as happens in several chronic diseases. A damaged protein may also be difficult to reach in an otherwise normal structure. For example, the lenses of our eyes become more opaque with age, and often (if we live long enough) develop cataracts, in part because of a relatively inefficient protein turnover deep in the lenses, where the tightly packed lens proteins leave little room for anything else.

There exists yet another reason for a protein to be destroyed—if it *evolved* to be degraded quickly. Proteins like these often function as regulators—devices that control the activities of specific biological processes such as the transcription and replication of DNA, the life cycle of a virus inside its host, or the fluxes of specific compounds through the metabolic pathways of a cell. To understand the reason for making a regulator short-lived, imagine that a specific biochemical pathway, controlled by an activator protein, is required before but not after cell differentiation—a process in which a cell converts itself into a cell of another kind. Stopping the synthesis of the activator may not be a fast enough way to get rid of it: the activator would linger indefinitely in a nondividing cell (many differentiated cells no longer divide), and even a dividing cell would dilute the activator only twofold upon each division—too slowly for a good off-switch. But make the activator short-lived, and stopping its synthesis would result in a rapid decline in the activator's concentration, and therefore in a rapid shutoff of the no-longer-appropriate pathway.

Enter ubiquitin. Its saga began in 1975, when a group of scientists in New York reported the isolation of a 76-residue protein that was present in all tested organisms. The name proposed for the new molecule—"ubiquitin"—turned out to be remarkably apt, because later studies indicated that ubiquitin is one of the most highly conserved proteins among eukaryotes. (The eukaryotes include you, me, all other animals, plants, fungi, and everything else alive except bacteria. One characteristic feature of a eukaryotic cell is its nucleus—a membrane-enclosed compartment where the cell keeps most of its DNA in long, tightly coiled fibers called chromosomes.) "Highly conserved" means that the amino acid sequence (and hence the structure) of ubiquitin is nearly the same among different organisms. Since the sequences and, to a lesser extent, the structures of most proteins tend to change in the course of evolution, the sequence of a protein that performs a given function in one organism may be quite unlike the sequence of its functional "twin" in another organism. By contrast, the sequence of ubiquitin remained essentially unchanged in the course of roughly two billion years—the span of time since the nearest common ancestor of this writer and baker's yeast. This extraordinary evolutionary stability implies that almost the entire structure of the ubiquitin molecule participates in some extremely important cellular functions. But what those functions were was anybody's guess.

Two years later, scientists at the Baylor

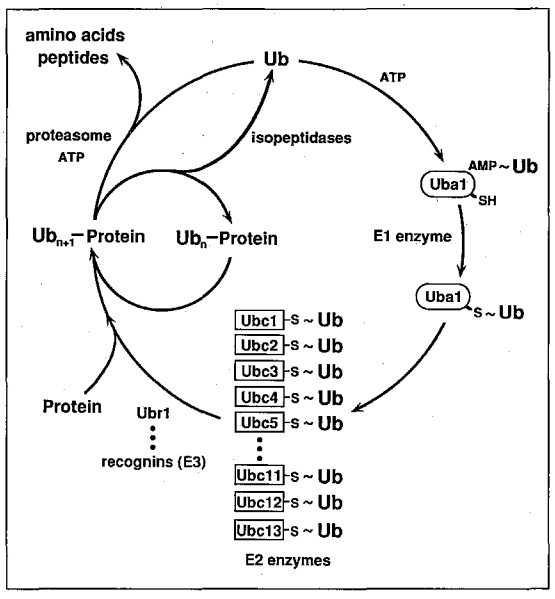
College of Medicine in Houston identified a mammalian protein of unusual structure, in which a chromosomal (DNA-bound) protein called H2A was linked to another protein—ubiquitin. In this "branched" protein, which they named ubiquitin-H2A or uH2A, ubiquitin was linked ("conjugated," as chemists say) to a lysine (an amino acid) within H2A, resulting in a protein with one C-terminus but two N-termini. The function of uH2A in chromosomes remains obscure to this day, but the branched structure of uH2A provided the first glimpse of a fundamental property of ubiquitin, soon to be encountered by scientists analyzing protein degradation.

The sequence of ubiquitin remained essentially unchanged in the course of roughly two billion years—the span of time since the nearest common ancestor of this writer and baker's yeast.

Many proteins that are slated for destruction meet their fate in specialized intracellular structures called lysosomes, but protein degradation also occurs elsewhere in a cell, including the cytosol and the nucleus. (Cytosol is the intracellular milieu outside of the many compartments that reside inside a cell. The nucleus is but one such compartment.) This extralysosomal protein degradation was found to require adenosine triphosphate (ATP), a universal source of chemical energy in living organisms. The ATP requirement for proteolysis was puzzling, because cleavage of the peptide bond between two residues in a polypeptide chain normally happens rapidly (and without a net input of energy) in the presence of a "simple" protease such as trypsin. (Try calling trypsin simple after you see its three-dimensional structure!) In 1978, Avram Hershko and his coworkers in Israel used an extract from reticulocytes (cells on their way to becoming red blood cells) in an attempt at understanding the mechanism of ATP-dependent protein degradation. They separated reticulocyte extract into two fractions that were inactive by themselves but became active when mixed together. The first fraction contained mostly hemoglobin and another, smaller protein, which was purified and shown to be the only factor that the second fraction required for ATP-dependent proteolysis; this protein was named APF ("ATP-dependent proteolysis factor").

At that time, it was unclear why some of the test proteins were degraded and some left intact in reticulocyte extract. So the strategy was simple—useful protein substrates were those that were short-lived in the extract, were degraded in an ATP-dependent manner, and were easy to obtain. Something unusual was happening to the short-lived proteins in these experiments: before disappearing, they temporarily became *larger.* A single species of the substrate—the protein about to be degraded—was observed in the extract samples that lacked ATP, whereas a set of larger substrate-containing molecules was formed in the presence of ATP. The researchers determined that these larger molecules were almost certainly those of the substrate conjugated to one or more APF molecules. The exploration of APF continued in Israel and the United States, and in 1980 APF was found to be—what else?—ubiquitin. This result brought together the study of ATP-dependent proteolysis and the earlier analysis of uH2A in chromosomes.

Meanwhile, my colleagues and I at MIT were studying chromosome replication and often discussed ubiquitin: what exactly is that branched protein, uH2A, doing in chromosomes? On a fateful day in 1981, I came across a paper from Tokyo University that described a mutant mouse cell line called ts85. The researchers showed that a specific nuclear protein disappeared at elevated temperatures from ts85 cells. They suggested that this protein might be uH2A. When I saw their data, I had to calm down to continue reading, because I *knew* that this protein *was* uH2A! If so, the ts85 mutant was a godsend to anyone who wanted to apply the power of genetics to the puzzle of ubiquitin. Like flipping a wall switch to see what lamp it controls, one could use ts85 cells to turn the conjugation of ubiquitin to other proteins on and off at will, and then observe what the cell did or didn't do. Daniel Finley (then a graduate student in my laboratory) and Aaron Ciechanover (then a postdoc at another MIT lab) started the analysis of ts85 and found that an extract from these mutant cells, in contrast to an extract from normal cells, produced ubiquitin-protein conjugates



The ubiquitin cycle. From the top, clockwise: In the presence of adenosine triphosphate (ATP), the last residue of a ubiquitin molecule (Ub) becomes joined through a high-energy bond (denoted by a ~) to a cysteine (an amino acid) of a ubiquitin-activating, or E1, enzyme (Uba1). This enzymatic reaction proceeds through an intermediate in which ubiquitin is joined to adenosine monophosphate (AMP). The activated ubiquitin is then transferred to another cysteine in one of several ubiquitin-conjugating, or E2, enzymes (Ubc1, etc.). An E2 enzyme, guided by an accessory protein called recognin, or E3, links the activated ubiquitin to its ultimate acceptor protein, whatever that may be. Many ubiquitin molecules can be linked, sequentially, to one molecule of the protein substrate, as shown by the subscript indicating the number of ubiquitins in a multiubiquitin chain. The substrate is then degraded, in yet another ATP-requiring step, by a protease called the proteasome. Ubiquitin molecules linked to the substrate are not degraded and reenter the free ubiquitin pool, after their liberation from the multiubiquitin chain by enzymes called isopeptidases.

The seemingly paradoxical idea—that ubiquitin may function as a protein stabilizer as well as a signal for protein degradation—was supported by other findings.

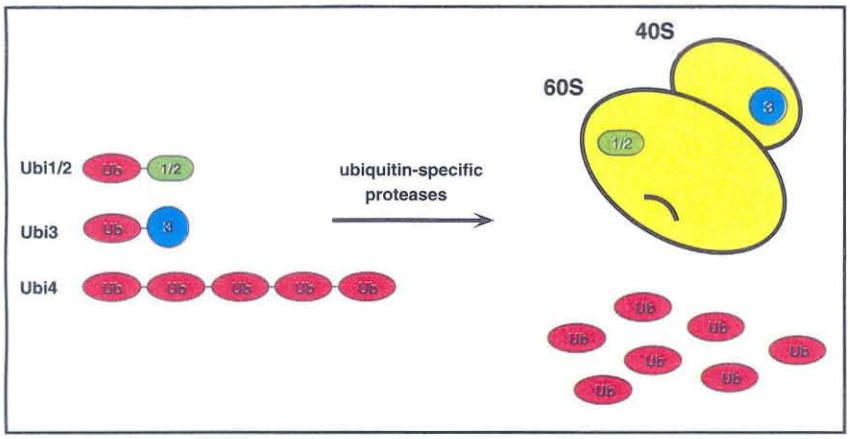
only at a relatively low temperature.

By then, the mammalian ubiquitin system had been resolved by other researchers into three components. The first of these was the ubiquitin-activating enzyme, or E1. This protein catalyzes an ATP-dependent reaction in which the C-terminal glycine residue of ubiquitin is joined to a specific cysteine residue in the E1 enzyme itself. The E1-ubiquitin complex then transfers this "activated" ubiquitin to a specific cysteine in another protein, called the ubiquitin-conjugating enzyme, or E2. The E2 enzyme, either by itself or in a complex with an accessory protein called recognin, or E3, forms ubiquitin-protein ligase—an enzyme that links ubiquitin to its ultimate acceptor proteins.

With this knowledge in mind, let us return to the adventure with ts85. We traced the heat sensitivity of ubiquitin conjugation in ts85 cells to the heat sensitivity of their mutant ubiquitin-activating (E1) enzyme. Since E1 is the first in the cascade of enzymes that prepare ubiquitin for its conjugation to other proteins, we could ask whether the ATP-dependent proteolysis I mentioned earlier also required E1. The results were striking: the degradation of short-lived proteins in ts85 cells was indistinguishable from that in normal cells at 30°C but nearly ceased at 39°C, whereas no inhibition of proteolysis was observed in normal cells at 39°C. These findings provided the first direct evidence that ubiquitin was required for protein degradation in living cells.

The study of ts85 cells was my first encounter with the power of approaches that bring together biochemical and genetic methods. But in the early eighties a sortie into mammalian genetics was still hampered by the impossibility of altering genes at will. (Things have improved greatly since then.) Therefore we embarked on a study of ubiquitin pathways in the species of yeast called *Saccharomyces cerevisiae.* This fungus was "domesticated" by humans eons ago for making bread and those mind-altering beverages called beer and wine. By 1983, when we started working with *S. cerevisiae,* it had already become a fair-haired eukaryote for genetic analysis, not only because of its rapid growth and simplicity (in comparison to plant and animal cells) but also because earlier work by geneticists had resulted in powerful techniques for manipulating yeast genes.

Our first target was the family of ubiquitin genes. Surprisingly, all of these genes were found to encode not the "mature" ubiquitin but precursor molecules that were enzymatically cleaved shortly after their synthesis, to yield ubiquitin and other proteins. One gene encoded a polyubiquitin, while the others encoded ubiquitin linked to unrelated ("tail") proteins. The mystery of the tails was solved in 1989, when Finley (by now a postdoc) and graduate student Bonnie Bartel in my lab, and Martin Rechsteiner's laboratory at the University of Utah, discovered that the free tails were components of the ribosome. We also showed that if the tail proteins were manufactured without ubiquitin, the assembly of ribosomes became inefficient, resulting in slowly growing cells. The likely explanation of this result stems from the fact that ubiquitin is an uncommonly stable and fast-folding protein.



Ubi1/2, Ubi3, and Ubi4 are ubiquitin precursor molecules in yeast. They are nearly identical to the precursors of human ubiquitin. These precursors are ubiquitin fusions—either to itself, as in the polyubiquitin protein Ubi4, or to other, "tail" proteins, as in Ubi1/2 and Ubi3. The precursors are cleaved by ubiquitin-specific processing proteases immediately after the ubiquitin's C-terminal residue, yielding free ubiquitin and (in the case of Ubi1/2 and Ubi3) the free tail proteins. These proteins were found to be components of the ribosome (shown in yellow). Tail 1/2 is a part of the small (40S) ribosomal subunit, whereas tail 3 resides in the large (60S) subunit. The tails' functions and exact locations within the ribosomal subunits are unknown.

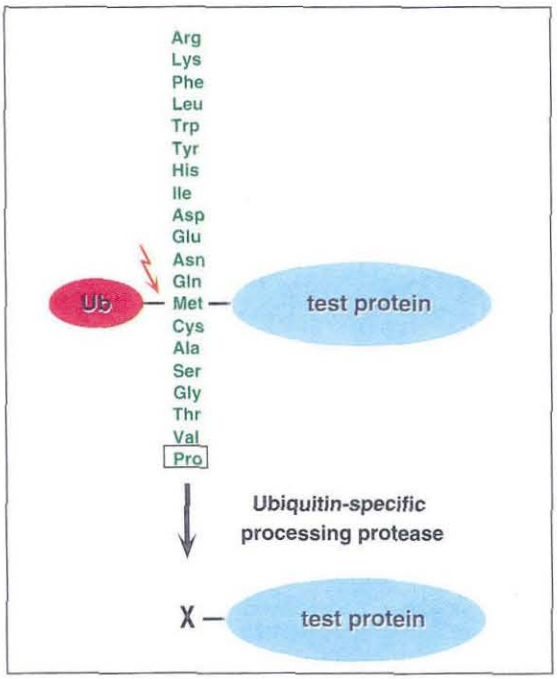
It may therefore protect the rest of a precursor protein from attack by the cell's ever-vigilant proteolytic systems. This protection is transient, because a newly formed ubiquitin precursor is cleaved at the junction of the ubiquitin and the tail. Since this cleavage is fast but not instantaneous, we suggested that ubiquitin's presence provides a partial protection to the ubiquitin-linked tail for the few fleeting seconds when the nascent tail is in gravest danger of being destroyed. As a result, a vulnerable tail-protein molecule may have a better chance of making it in one piece from the ribosome that produced it in the cytosol to an assembly site for ribosomes in the nucleus, where the tail is incorporated into a new ribosome.

Many if not all of the ribosomal proteins are short-lived in vivo unless they associate with each other and the ribosomal RNA to form the ribosome. This way of running the assembly of a multiprotein structure assures that any of its components produced in excess won't end up lingering in the cell. But why were only two of the many ribosomal proteins "chosen" to be produced as ubiquitin fusions during evolution, and why has this arrangement persisted in the course of the two billion years that separate fungi and humans from their nearest common ancestor? Here is a partial answer: the presence of ubiquitin and a ribosomal protein within a single precursor seems to be, among other things, the way to establish a coupling between the numbers of newly made ubiquitin molecules and the numbers of newly assembled ribosomes. An interdependence of this sort may be a useful homeostatic (order-maintaining) arrangement, because ribosomes are in the business of making proteins, whereas the ubiquitin system is about protein destruction—it would be helpful to the cell if these systems were sensitive to each other's abundance and activity.

The seemingly paradoxical idea—that ubiquitin may function as a protein stabilizer as well as a signal for protein degradation—was supported by other findings, which showed that if the gene for a protein that had been difficult to produce because of its rapid intracellular destruction was extended by adding a region that encoded ubiquitin, the yield of the resulting fusion of ubiquitin and the protein was often much higher than the yield of the initial protein.

What about the gene encoding polyubiquitin? Finley and I found that this gene was activated by just about every stressful treatment we could think of. For example, heating cells beyond their normal temperature range, starving them of nutrients, or exposing them to toxic compounds like hydrogen peroxide all resulted in the overproduction of ubiquitin by the polyubiquitin gene. Furthermore, a yeast mutant lacking the polyubiquitin gene was hypersensitive to the stresses that activated this gene in wild-type (normal) yeast. The mutant grew well in the absence of hardships, and seemed normal in other respects as well—until the going got tough. We concluded that ubiquitin, in addition to whatever else it does in a cell, functions as a stress protein—a member of the large class of proteins that all organisms produce, sometimes in copious amounts, in response to adversity. Many of these proteins are also present, at lower concentrations, in cells that are doing just fine, suggesting that stress-specific roles of these proteins are but enhanced versions of their functions in the absence of stress.

Why should a cell under stress overproduce ubiquitin? An oxidative or heat injury increases the amount of damaged proteins in the cell and therefore increases the demand for ubiquitin, whose conjugation to damaged proteins is required for their degradation. Interestingly, an overproduction of ubiquitin in stressed cells doesn't increase their level of *free* ubiquitin, suggesting that the essential function of the polyubiquitin gene is to maintain the cell's free-ubiquitin level in the face of the increased rate at which free ubiquitin is depleted through the formation of ubiquitin-protein conjugates. This property of being distributed between free and tightly protein-bound states is also characteristic of many stress proteins other than ubiquitin. Finley and I proposed that a stress-induced



Right: The ubiquitin fusion technique. Linear fusions of ubiquitin to other proteins are rapidly cleaved in vivo at the ubiquitin-protein junction, irrespective of the identity of the residue at the N-terminal side of the cleavage site. This feature of ubiquitin-specific proteases makes possible the generation of otherwise identical proteins bearing any of the twenty amino acids in the N-terminal position. Proline (Pro) is the only residue that partially inhibits cleavage at the ubiquitin-protein junction. Below: The standard one- and three-letter abbreviations for the amino acids.

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

increase in the *total* level of a stress protein is mediated by a regulatory mechanism that acts to maintain the required level of a *free* stress protein. Examples of such "feedback" circuits have recently been described for several stress proteins.

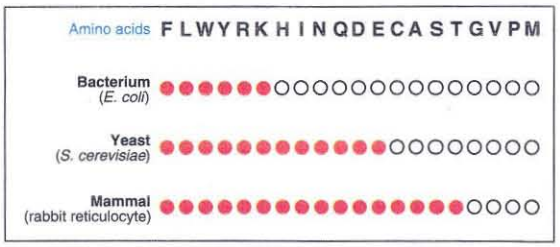
In 1987, Stefan Jentsch (then a postdoc in my lab) found that one of the ubiquitin-conjugating (E2) enzymes was encoded by a gene called *RAD6*. This gene has been known for many years, because mutations in *RAD6* perturb a number of processes, from sporulation to DNA repair. (Sporulation is one of the stress responses in yeast: when out of food, yeast cells form spores—small, tough, dormant cells ready to outlast the bad times until a wind or whatever transfers them onto anything edible.) Subsequent work greatly expanded the list of known E2 functions; it now includes the ability of cells to resist poisoning by toxic metals, the regulation of the cell cycle, and the control of protein transport across membranes. These remarkably diverse functions are probably underlain by a common mechanism—the degradation of specific proteins tagged by E2 enzymes.

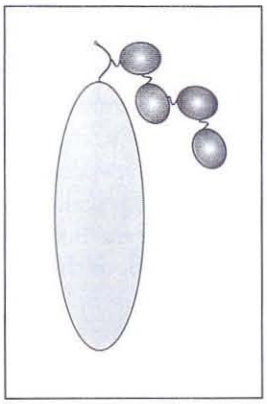
We are halfway through the story but quite a few things are still unexplained. For instance: why attach ubiquitin to a short-lived protein at all—why is this bulky and metabolically costly modification so necessary for the in vivo degradation of many proteins? And furthermore: what features of a protein make it a target of the ubiquitin system? Let us begin with the last problem.

There is no such thing as a totally nonspecific protease—a protease that can cleave any peptide bond with equal dexterity. Even "simple" extracellular proteases like trypsin or pepsin have their preferences, specific for each protease. Features of proteins that make them susceptible to proteolysis are called degradation signals, or degrons. In 1986, Andreas Bachmair and Finley (then postdocs in my lab) discovered the first intracellular degradation signal, and showed it to be recognized by a pathway that involves ubiquitin.

As often happens, the experiments that led to this insight were initially aimed at something else: we wanted to design a fusion protein whose ubiquitin component could not be removed by the ubiquitin-specific proteases that normally cleave a precursor protein at the junction between ubiquitin and a "downstream" polypeptide. To this end, a gene was constructed that encoded ubiquitin fused to an enzyme called β-galactosidase (βgal). (This enzyme was chosen because its fate in the cell could be followed in several convenient ways.) The gene was mutated to convert the methionine (Met) residue at the ubiquitin-βgal junction into a variety of other amino acids. Alas, the ubiquitin-specific proteases couldn't care less about these alterations of their substrate—they continued to cut ubiquitin off the ubiquitin-X-βgal fusion (X being the varied residue) as if nothing had happened.

This result proved to be good luck in disguise—we were thwarted, for a time, in making a fusion protein whose ubiquitin portion stays put, but the near indifference of the proteases to the identity of residue X yielded a method for producing, in a living cell, any residue at the N-terminus of any protein—until then an impossible feat. Why impossible? Because of the way the genetic code works: every mRNA molecule—the messenger that carries the protein's assembly instructions from the genes to the ribosomes, where the proteins are manufactured—*is "read"* starting from the codon (a unit of RNA encoding one amino acid) that specifies methionine. The ribosome needs some way of knowing where to begin, but why a methionine codon instead of a codon for another amino acid was chosen for this purpose at the dawn of earthly life is unclear. However, once this fundamental early choice had been made, it became "fixed" in the design of living cells. Thus all proteins produced in vivo start off with an N-terminal methionine. Lots of things can happen to this methionine later on—it's retained in many proteins, and it's chemically modified in others; it may even be removed by specific proteases, but the current understanding of these reactions is insufficient for their assured manipulation. Linking ubiquitin to the





Below: The N-end rule for yeast.
Top right: A comparison of the N-end rules in three organisms of increasing complexity. Open circles stand for stabilizing N-terminal residues; red circles are destabilizing ones. The N-end rule is actually more elaborate than is shown here, in that some destabilizing residues are recognized directly, whereas others undergo specific in vivo modifications before recognition.
Bottom right: This isn't a balloon animal, but a multiubiquitin chain containing four ubiquitins, drawn roughly to the scale of the X-βgal subunit to which the chain is attached. Multiubiquitin chains in vivo can contain more than 50 ubiquitins!

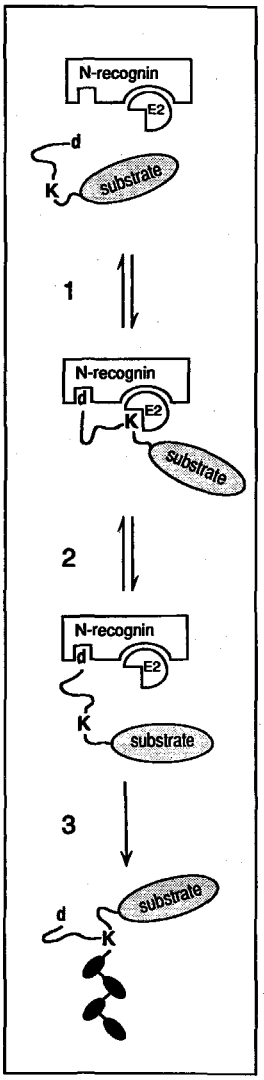
Residue X in X-βgal	In vivo half-life of X-βgal
Arg	2 min
Lys	3 min
Phe	3 min
Leu	3 min
Trp	3 min
Tyr	10 min
His	3 min
Ile	30 min
Asp	3 min
Glu	30 min
Asn	3 min
Gln	10 min
Cys	> 20 h
Ala	> 20 h
Ser	> 20 h
Thr	> 20 h
Gly	> 20 h
Val	> 20 h
Pro	> 20 h
Met	> 20 h

N-terminus of a protein bypasses this problem. The desired N-terminal residue can now be produced by the ubiquitin-specific proteases that cut the fusion protein after the last residue of ubiquitin—away from the initial N-terminus of a ubiquitin-protein fusion.

The new method in hand, we discovered something remarkable almost immediately: X-βgal proteins bearing different N-terminal residues had different in vivo half-lives. (The half-life of a protein is the time it takes for 50 percent of the protein molecules initially present to disappear.) For example, Met-βgal, (which bore N-terminal methionine) had a half-life of at least 30 hours—an eternity by the standards of short-lived proteins. In striking contrast, Arg-βgal (which bore N-terminal arginine) had a half-life of two minutes. One way to appreciate the fleetingness of this half-life is to consider that it takes the ribosome about two minutes to synthesize the approximately 1,100-residue Arg-βgal. In other words, a newly formed molecule of Arg-βgal is destroyed in about the time it took to make it in the first place!

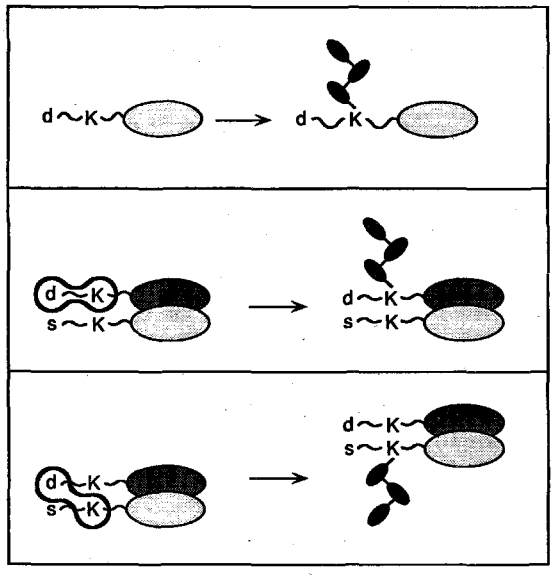
Measurements of degradation rates of X-βgal proteins in yeast yielded a relationship between the in vivo half-life of a protein and the identity of its N-terminal residue—a new, startlingly simple code. We named it the N-end rule and proceeded to explore the vistas opened up by this discovery. It was soon found that distinct versions of the N-end rule operated in all organisms examined, from bacteria to mammals. The three N-end rules in the illustration above are different but also hauntingly similar: the set of destabilizing residues in bacteria is a subset of the analogous set in yeast, and that, in turn, is a subset of the analogous set in mammalian reticulocytes—cells on their way to becoming red blood cells. We don't know the functional meaning of these differences, but it appears that the N-end rule book depends on the cell's physiological state. In other words, the N-end rule is a "soft-wired" code, in contrast, for example, to the genetic code, which is "hard-wired" in the sense that it is the same for all genes in all organisms. (There are, in fact, a few exceptions to the latter statement, as is the case with most statements in biology. Nearly every rule that can be broken in principle is actually violated somewhere in the world of living things, for evolution respects few constraints other than those imposed by physics.) The N-end rule is just beginning to yield its secrets—another story, to be described someday in an article of its own.

Central to understanding the N-end rule is the underlying degradation signal, which we named the N-degron. Is it actually as simple as a single residue at the N-terminus of a protein? What is the role of ubiquitin in the function of the N-degron? My colleagues and I addressed these questions by mutating N-end rule substrates (proteins that are degraded in accordance with the N-end rule) and determining their in vivo half-lives. By 1989, genetic analysis had shown that the N-degron consists of two components: a destabilizing N-terminal residue, and an amino acid residue called lysine at a specific position in the substrate. A parallel biochemical study indicated that multiple ubiquitin molecules



Left: The mechanism by which the N-end rule recognizes a substrate and prepares it for degradation. 1.) N-recognin binds to the substrate's destabilizing N-terminal residue (d). 2.) The relevant lysine (K) is captured by the ubiquitin-conjugating enzyme (E2) associated with the N-recognin. 3.) The lysine capture results in the synthesis of a lysine-linked multiubiquitin chain (black ovals) by the E2 enzyme.

Below: *Cis-trans* recognition and degradation of N-end rule substrates. The upper panel shows the single-subunit case, with d, K and the multiubiquitin chain as above. The middle panel illustrates *cis* recognition of a two-subunit protein, one subunit of which bears a stabilizing N-terminal residue (s). The bottom panel shows how the same protein can be recognized in *trans*. Note that the multiubiquitin chain is now linked to the other (lower) subunit.



become linked to an N-end rule substrate shortly before its degradation. Strikingly, all of these ubiquitin molecules were found to dangle from one lysine—the same one that had been pinpointed by genetic analysis. Thus, instead of being attached to several different lysines of a substrate such as Arg-βgal, the ubiquitin molecules formed a multiubiquitin chain.

What makes a lysine in an N-end rule substrate the site of ubiquitin conjugation? The relevant lysine must be located *spatially* close to the N-terminus—this requirement includes a proximity to the N-terminus *along* the substrate's polypeptide chain. The recognition of a short-lived protein by the proteolytic machinery starts with the binding of a protein, called N-recognin, to the substrate's N-terminal residue. This binding is reversible, and unless the E2 enzyme (which exists in a complex with N-recognin) binds rapidly to the proper internal lysine of the same substrate, the N-recognin-E2 complex "falls off" the substrate and has to start again. A critical lysine should be easy to find if it's positioned for a nearly simultaneous capture of both it and the substrate's N-terminal residue by the complex's two binding sites. Alternatively, the relevant lysine could be a part of a mobile region of the substrate that doesn't fold up into one preferred structure (or conformation, as we say in the trade). While flopping around, the substrate's lysine may approach the bound N-recognin-E2 complex often enough for the E2 enzyme to catch it before the entire complex dissociates from the substrate.

Now that we have gotten sophisticated about the recognition system, let's push it a little further. Thus far, the N-degron's two components have been assumed to reside in the same polypeptide; they are said, in this case, to be recognized in *cis*. But there's also an arrangement called *trans,* in which a destabilizing N-terminal residue and the relevant lysine are in two different subunits (polypeptide chains) of a multisubunit protein. Would such a "split" N-degron work? In 1990, Erica Johnson (then a graduate student in my lab) showed that it would. This discovery revealed a previously unsuspected ability of the N-end rule pathway: of the two subunits bearing the split N-degron, only one subunit—the one containing the relevant lysine—was degraded, whereas the other subunit was left unharmed. In other words, the destruction of a multisubunit N-end rule substrate is confined to those subunits that can be linked to a multiubiquitin chain.

How many distinct degrons (recognized by different recognins) are there in a cell? We don't

know, but "at least three" is a safe answer. One class contains the N-degrons I've already discussed. Another distinct class of degradation signals is present in proteins called cyclins, which function as subunits of cyclin-dependent kinases—enzymes that control cell growth and division. Several studies have shown that cyclin degradation is ubiquitin-dependent; moreover, a stretch of nine residues is conserved among many cyclins and is required for their destruction. Yet another class of degradation signals has been described by Martin Rechsteiner and coworkers at the University of Utah, who noticed that many short-lived proteins (including certain cyclins) contain sequences that are unusually rich in the amino acids proline, glutamate, serine, and threonine. Rechsteiner has suggested that some of these sequences may act as degrons. Indeed, deleting such a region from a short-lived protein often stabilizes the protein. And the end of the list of degrons is not yet in sight: for example, Mark Hochstrasser (then a postdoc in my lab) and I have described two distinct degradation signals in a single protein called Matα2—a repressor of RNA synthesis that regulates sexual differentiation in *S. cerevisiae* (yes, fungi have sex, but this story is about ubiquitin).

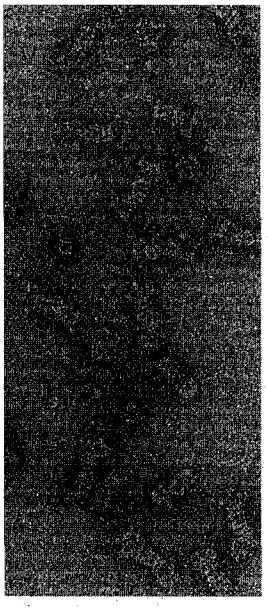
The two-component design established for the N-degron appears to be characteristic of other degradation signals as well. The first component of these signals is an internal region of a protein (instead of its N-terminus) that is specific for each degron, while the second component is likely to be a conformationally mobile lysine (or lysines). If so, these other, still dimly understood degradation signals may also exist in versions analogous to the "split" N-degron.

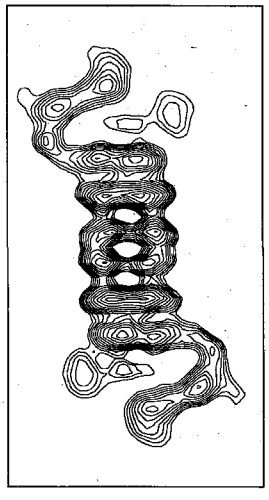
Indeed, Peter Howley and his colleagues at NIH and Harvard Medical School have shown that a protein called p53 can be marked for destruction as a result of its binding by a protein known as E6—a product of an oncogenic (cancer-causing) human papilloma virus. (Names such as "p53," "E6," and many others are often little more than labels used to distinguish one protein among the multitude of its fellows, which are often discovered before their functions are known. For instance, p53 means "a protein with a molecular mass of about 53,000 atomic mass units.") Oncogenic papilloma viruses, whose sexual transmission among humans increases the risk of certain cancers, are able to induce the proliferation of infected cells. The viruses achieve this in part by decreasing, through ubiquitin-dependent degradation, the level of the cellular regulatory protein p53, whose binding by the viral protein E6 destabilizes p53 *without* destabilizing E6 itself. There is a striking analogy between this effect (mediated by an unknown degradation signal in p53) and the mechanics of a split N-degron.

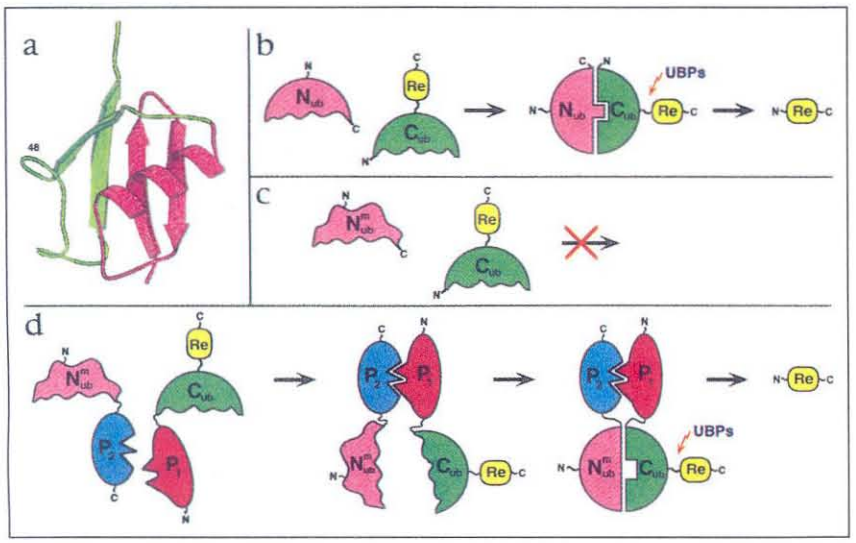
The protease that degrades ubiquitin-linked proteins is called the 26S proteasome; "26S" (26 Svedberg units) is shorthand for how rapidly this large particle sediments in a centrifuge. The 26S proteasome attacks a protein that bears a multiubiquitin chain in a reaction that requires ATP and the multiubiquitin itself. Thus ATP is used at least twice in ubiquitin-dependent proteolysis: first at the step of ubiquitin attachment (or rather ubiquitin activation), and then at a poorly understood step during the actual degradation of a substrate. The 26S proteasome contains some 40 distinct subunits and is unstable in the absence of ATP, dissociating into several components. One of them is called the 20S proteasome, a particle that can cleave a variety of peptide bonds in a reaction that doesn't require ATP. Biochemical studies of the 20S proteasome, and electron-microscopic observations of an analogous protease from bacteria, suggest that the proteasome destroys a protein substrate in a process that involves "threading" the substrate's polypeptide chain through a channel inside the cylinder-shaped proteasome.

Now that we have a nodding acquaintance with the gadgetry of the ubiquitin system, let us attempt an answer to the central question: what exactly is ubiquitin's function? One possibility is that the formation of a multiubiquitin chain linked to a substrate produces additional binding sites for the proteasome's components. As a result, the probability of the proteasome "falling off" the substrate would decrease, and *that* could facilitate the substrate's destruction. Here's why: the proteasome must at least partially unfold the protein in order to thread it through the channel where the proteolysis actually occurs. A folded protein molecule is not a static structure: its polypeptide chain moves about a bit, and sometimes quite a bit, as it gets kicked by packets of water molecules. If the proteasome can "catch" a mobile, relatively unstructured region that becomes exposed during these occasional partial unfoldings (called fluctuations), the substrate's conformation might be destabilized strongly enough for the proteasome to start its work. This implies that the proteasome is "waiting" for a fluctuation; the longer the wait, the greater the probability of a suitable unfolding event. And if the formation of a multiubiquitin chain retards the dissociation of the substrate from the proteasome, the allowed waiting time becomes longer, increasing in turn the probability of

Top: An electron-microscopic image of a crowd of 26S proteasome particles, magnified 300,000 times.
Bottom: A computer-enhanced image of a single 26S proteasome, magnified 1,800,000 times.
Electron micrographs courtesy of Wolfgang Baumeister and colleagues at the Max Planck Institute in Martinsried, Germany.







How to detect a protein interaction in vivo as it occurs. (a) This diagram illustrates the folding pattern of ubiquitin's polypeptide chain, without detailing the amino acids. The N- and C-terminal halves are shown in pink and green, respectively. (The 48 marks the lysine where other ubiquitins can attach.) (b) If a "reporter" protein (Re) is fused to a free C-terminal half (C_{ub}), ubiquitin-specific proteases (UBPs) won't cleave the fusion until the C-half associates with an N-half (N_{ub}) to form a nearly normal ubiquitin molecule. Once liberated, the reporter protein can be detected in several ways. (c) If the N-half is mutated (N^m_{ub}) in a way that weakens its interaction with the C-half, the reporter is not cleaved off. (d) But if the C-half and the mutant N-half are linked to proteins that interact in vivo (P_1 and P_2), the interaction will bring the two halves so close together that their residual affinity will be sufficient to form a functional ubiquitin anyhow, causing the reporter protein to be cut free.

catching a partially unfolded substrate.

Two results indicate that the unfolding of a protein substrate is indeed a prerequisite for its destruction by the proteasome, and that a multiubiquitin chain plays a role in the process. Jennifer Johnston (a postdoc in my lab) has found that the ubiquitin-dependent degradation of a protein—dihydrofolate reductase, or DHFR—by the N-end rule pathway can be inhibited by methotrexate, a small molecule that specifically binds to DHFR. This finding—that a modest increase in the conformational stability of DHFR as a result of its binding to methotrexate is sufficient to stop the proteasome juggernaut in its tracks—is consistent with the idea that a substrate's conformation is one major barrier faced by the proteasome. In addition, Tillmann Rümenapf (then a postdoc in my lab), James Strauss (PhD '67, Caltech's Bowles Professor of Biology), and I have found that the formation of a substrate-linked multiubiquitin chain is actually unnecessary for the substrate's degradation by the N-end rule pathway, provided that the substrate is conformationally unstable to start with. These findings are consistent with the model discussed above, but they are also consistent with another idea—that the substrate-linked multiubiquitin chain, by virtue of being in close proximity to the substrate, may interact with it and thereby play a direct role in destabilizing its conformation.

The above models are specific enough to make testable predictions, but barely begin to address the true range and subtlety of reactions at the proteasome. For example, we've discussed multiubiquitin chains as if they simply hang there—conjugated to a substrate, bound to the proteasome. In fact, a multiubiquitin chain has a life of its own: it folds in certain preferred ways; it also grows through the activity of E2 enzymes and shrinks through cuts made by ubiquitin-specific proteases, at least one of which is a component of the proteasome. These and other complexities are trying to speak to us and will be understood someday, when even a popular yarn about ubiquitin shall require a book to be told.

In the meantime, I shall mention just one instance of research on ubiquitin bearing fruit in other fields. Nils Johnsson, a postdoc in my laboratory, has found that the compact organization of ubiquitin belies a subtlety: the ubiquitin's N-terminal "half" retains elements of its folded structure even in the absence of the rest of the molecule. Moreover, the N-terminal half can bind in vivo to a *separately produced* C-terminal half, forming a nearly normal ubiquitin. This discovery has led to a new method for detecting protein interactions in living cells.

The growing understanding of intracellular proteolysis is providing us with powerful tools for manipulating the in vivo half-lives of intracellular proteins, including those whose malfunction or overproduction leads to cancer and other illnesses. Most drugs of today are incapable of altering the in vivo stability of a protein target. But this is likely to change, and when it does, an entirely new class of therapeutic agents will emerge, with exciting implications for the cure of currently intractable diseases. □

Alexander Varshavsky is the Smits Professor of Cell Biology at Caltech. He is also a member of the National Academy of Sciences and the American Academy of Arts and Sciences. Varshavsky was born and educated in Moscow, Russia. In 1977, he joined the faculty at the Massachusetts Institute of Technology in Cambridge and stayed there until 1992. Varshavsky and coworkers discovered the exposed regions in chromosomes (which form at the beginnings of active genes, at the origins of DNA replication, and at other sites of localized activity in the chromosomes), deciphered the mechanism of separation of intertwined sister DNA molecules during chromosome replication, and described the phenomenon of induced gene amplification that contributes to rapid evolution of cancer cells within a tumor. Varshavsky's initial interest in ubiquitin stemmed from its presence in chromosomes. His laboratory produced the first direct evidence that ubiquitin is required for protein degradation in living cells, and in 1986 discovered the first intracellular degradation signal.