Proteases: Multifunctional Enzymes in Life and Disease

Carlos López-Otin ¹ and Judith S. Bond ²

From the ¹Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain; and ²Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA.

Running Head: Proteases: Multifunctional Enzymes

Address correspondence to: Judith S. Bond, Department of Biochemistry and Molecular Biology, Pennsylvania College of Medicine, Hershey, Pennsylvania 17033; Tel: 717-531-8586; Fax: 717-531-7072; email: jbond@psu.edu

Our view of proteases has come a long way since P.A. Levene reported his studies on “the cleavage products of proteases” in the first issue of The Journal of Biological Chemistry published on October 1, 1905 (1). Today, after more than 100 years and 350,000 articles on these enzymes in the scientific literature, proteases remain at the cutting edge of biological research.

Proteases likely arose at the earliest stages of protein evolution as simple destructive enzymes necessary for protein catabolism and the generation of amino acids in primitive organisms. For many years, studies on proteases focused on their original roles as blunt aggressors associated with protein demolition. However, the realization that, beyond these nonspecific degradative functions, proteases act as sharp scissors and catalyze highly specific reactions of proteolytic processing, producing new protein products, inaugurated a new era in protease research (2). The current success of research in this group of ancient enzymes mainly derives from the large collection of findings demonstrating their relevance in the control of multiple biological processes in all living organisms (3-11). Thus, proteases regulate the fate, localization and activity of many proteins, modulate protein-protein interactions, create new bioactive molecules, contribute to the processing of cellular information, and generate, transduce, and amplify molecular signals. As a direct result of these multiple actions, proteases influence DNA replication and transcription, cell proliferation and differentiation, tissue morphogenesis and remodeling, heat shock and unfolded protein responses, angiogenesis, neurogenesis, ovulation, fertilization, wound repair, stem cell mobilization, haemostasis, blood coagulation, inflammation, immunity, autophagy, senescence, necrosis and apoptosis. Consistent with these essential roles of proteases for cell behavior, survival and death of all organisms, alterations in proteolytic systems underlie multiple pathological conditions, such as cancer, neurodegenerative disorders, and inflammatory and cardiovascular diseases. Accordingly, many proteases are a major focus of attention for the pharmaceutical industry as potential drug targets or as diagnostic and prognostic biomarkers (12). Proteases also play key roles in plants and contribute to the processing, maturation or destruction of specific sets of proteins in response to developmental cues or to variations in environmental conditions (13). Likewise, many infectious microorganisms require proteases for replication or use proteases as virulence factors, which has facilitated the
development of protease-targeted therapies for diseases of great relevance to human life such as AIDS (12). Lastly, proteases are also important tools of the biotechnological industry because of their usefulness as biochemical reagents or in the manufacture of numerous products (e.g., 14).

This outstanding diversity in protease functions directly results from the evolutionary invention of a multiplicity of enzymes which exhibit a variety of sizes and shapes. Thus, the architectural design of proteases ranges from small enzymes made up of simple catalytic units (~20 kDa), to sophisticated protein processing and degradation machines, like the proteasome and meprin metalloproteinase isoforms (0.7 to 6 MDa) (15). In terms of specificity, diversity is also a common rule. Thus, some proteases exhibit an exquisite specificity towards a unique peptide bond of a single protein (e.g. angiotensin-converting enzyme); however, most proteases are relatively non-specific for substrates and some are overtly promiscuous and target multiple substrates in an indiscriminate manner (e.g., proteinase K). Proteases also follow different strategies to establish their appropriate location in the cellular geography, and in most cases operate in the context of complex networks comprising distinct proteases, substrates, cofactors, inhibitors, adaptors, receptors and binding proteins, which provide an additional level of interest but also complexity to the study of proteolytic enzymes.

This work aims at serving as a primer to a minireview series on proteases to be published in forthcoming issues of this journal. This introductory article will focus on the discussion of the large and growing complexity of proteolytic enzymes present in all organisms, from bacteria to man. We will first show the results of comparative genomic analysis which have shed light on the real dimensions of the proteolytic space. The levels of protease complexity and mechanisms of protease regulation will then be addressed. Finally, we will discuss current frontiers and future perspectives in protease research.

**The vast proteolytic landscape**

Proteases are the efficient executioners of a common chemical reaction: the hydrolysis of peptide bonds (16). Most proteolytic enzymes cleave α-peptide bonds between naturally occurring amino acids but there are some proteases that perform slightly different reactions. Thus, a large group of enzymes known as DUBs (deubiquitylating enzymes) can hydrolyze isopeptide bonds in ubiquitin and ubiquitin-like protein conjugates; γ-glutamyl hydrolase and glutamate carboxypeptidase target γ-glutamyl bonds; γ-glutamyltransferases both transfer and cleave peptide bonds; and intramolecular autoproteases – such as nucleoporin and polycystin-1 – only hydrolyze a single bond on their own polypeptide chain but then lose their proteolytic activity. Notably, and under some conditions, proteases can also synthesize peptide bonds.

Proteases were initially classified into endopeptidases, which target internal peptide bonds, and exopeptidases (aminopeptidases and carboxypeptidases), whose action is directed by the NH₂ and COOH-terminus of their corresponding substrates. However, the availability of structural and mechanistic information on these enzymes facilitated new classification schemes. Based on the mechanism of catalysis, proteases are classified into six distinct classes: aspartic-, glutamic-, metallo-, cysteine-, serine-, and threonine-proteases, although glutamic proteases have not been found in mammals so far. The first three classes utilize an activated water molecule as a nucleophile...
to attack the peptide bond of the substrate, whereas in the remaining enzymes the nucleophile is an amino acid residue (Cys, Ser or Thr, respectively) located in the active site from which the class names derive (Fig. 1). Proteases of the different classes can be further grouped into families on the basis of amino acid sequence comparison, and families assembled into clans based on similarities in their three-dimensional structures. Bioinformatic analysis of genome sequences has been decisive for establishing the dimensions of the complexity of proteolytic systems operating in different organisms (Fig. 2). The last release of MEROPS (http://merops.sanger.ac.uk) – a comprehensive database of proteases and inhibitors – annotates 1008 entries for human proteases and homologues, although it includes a large number of pseudogenes and protease-related sequences derived from endogenous retroviral elements embedded in our genome. A highly curated database (http://www.uniovi.es/degradome), which does not incorporate protease pseudogenes or these retrovirus-derived sequences, lists 569 human proteases and homologues classified in 68 families (17). Metalloproteases and serine proteases are the most densely populated classes, with 194 and 176 members, respectively, followed by 150 cysteine proteases, whereas threonine and aspartic proteases only contain 28 and 21 members, respectively.

The recent availability of the genome sequence of different mammals has allowed the identification of their entire protease complement (termed degradome), and their detailed comparison with that of humans (Fig. 2). The chimpanzee degradome is very similar to the human degradome although it exhibits some remarkable differences in immune defense proteases like caspase-12 (18). Interestingly, mouse and rat contain more protease genes (644 and 629, respectively) than humans despite the fact that their genomes are smaller (19,20). These differences mainly derive from the expansion in rodents or the inactivation in humans of members of protease families – such as kallikreins and placental cathepsins – involved in immunological and reproductive functions (21,22). The recent analysis of the degradome of other mammals such as the duck-billed platypus (Ornithorhynchus anatinus) has revealed some interesting findings on protease evolution. This fascinating monotreme also has more than 500 protease genes, but lacks all genes encoding gastric pepsins, which are the archetypal digestive proteases widely conserved in all mammals (unpublished data). Birds, amphibians, and fish also contain large numbers of protease genes (382 in Gallus gallus, 278 in Xenopus tropicalis, and 503 in Danio rerio), although the protease-annotation work in these species has not been as detailed as in mammals. Surprisingly, analysis of the protease content of invertebrates such as Drosophila melanogaster – a model organism with a gene content considerably lower than vertebrates – has shown the presence of more than 600 protease genes (23). Similar results have been reported for other insects such as the malaria mosquito Anopheles gambiae. This relatively large number of insect proteases mainly derives from the amplification of a family of serine proteases genes associated with embryonic development and immune defense. Cross-genome comparisons have also facilitated the analysis of the protease complement of other model organisms such as Caenorhabditis elegans and the characterization of proteolytic enzymes involved in processes of notable biological interest such as the nematode moultng cycle (24). Likewise, the completion of several genome-sequencing projects in plants has provided new opportunities to appreciate the complexity of protease systems (13) (Fig. 2). Thus, the model
plant *Arabidopsis thaliana* contains at least 723 protease-encoding genes, whereas a total of 955 protease genes have been annotated in the tree *Populus trichocarpa*. These marked differences are linked to the expansion of some protease families in *Populus*, especially the copia-transposon endopeptidase family of aspartic proteases which has 20 components in *Arabidopsis* and 123 in *Populus* (13). Genomic analyses have also shown that plants share with prokaryotes a set of serine proteases absent in other eukaryotes, which may be an indication of ancient endosymbiotic events leading to evolution of chloroplasts (25). Finally, there is a growing interest in analyzing the degradome of bacteria, viruses, fungi and parasites as part of strategies aimed to define novel targets for therapeutic intervention (26-28). In this regard, the MEROPS database annotates more than 100 protease genes in the genome of bacteria such as *Yersinia pestis* and *Legionella pneumophila* or in the malaria parasite *Plasmodium falciparum*, which cause devastating human diseases.

In summary, the emerging pattern derived from the global analysis of proteolytic systems is one of diversity and multiplicity. These comparative genomic studies have also provided valuable insights into the conservation, evolution, and functional relevance of this group of enzymes. Thus, it has become evident that, in addition to proteolytic routines conserved in all organisms, there are also specific roles played by unique proteases in different species. Nevertheless, further studies will be necessary to clarify the genetic and molecular basis underlying the evolutionary differences between the complex protease repertoire of all living forms.

**Levels of protease complexity**

Proteolytic enzymes are not mere catalytic devices working in isolation in their search for substrates to be hydrolyzed. Thus, many proteases link their catalytic domains to a variety of specialized functional modules or domains that provide substrate specificity, guide their cellular localization, modify their kinetic properties, and change their sensitivity to endogenous inhibitors. These non-catalytic domains include archetypal sorting signals that direct these enzymes to their proper location, auto-inhibitory prodomains that prevent premature activation, and ancillary domains that facilitate homotypic interactions or heterotypic contacts with other proteins, substrates, receptors or inhibitors. Some of these ancillary domains, like the EGF domains, have been very successful in their incorporation into proteases and are present in a variety of enzymes from different families, whereas other domains – such as the thrombospondin repeats of ADAMTSs – have expanded within the same enzyme forming long tandem repeats (29). Other proteases, including diverse members of the type II transmembrane serine protease (TTSP) family, exhibit a complex mosaic structure with up to 6 distinct ancillary domains located within a single polypeptide chain (30). This exuberant strategy of domain accretion and shuffling has also led to the creation of very peculiar structures including protease-inhibitor chimeras or proteases with different catalytic units embedded in the same polypeptide chain (31). It is very likely that the substantial combinatorial activity observed in protease genes has been a driving force in the protease transition from non-specific primitive enzymes to highly selective catalysts responsible for subtle proteolytic events which are at the heart of multiple biological processes.

The complexity of proteases is further increased through post-transcriptional events such as alternative splicing and differential polyadenylation of genes encoding proteases (32,33), by the
The occurrence of gene copy number variations or polymorphic variants that may contribute to the modification of protease functions or alter their regulatory mechanisms (34,35), or by post-translational modifications such as glycosylation or phosphorylation. Finally, we must emphasize that in many cases proteases act in the context of complex cascades, pathways, circuits and networks, comprising many protein partners that dynamically interact to form the so-called protease web (36). Accordingly, to understand the role of a certain protease in a given biological or pathological process we must identify the mechanisms that regulate the expression and activity of the different enzymes and try to place them in the context of the multiple components that can influence its activity.

**Mechanisms of protease regulation**

Proteolytic processing represents an excellent strategy for increasing the diversity of the limited protein repertoire encoded in the genome of any living system. However, in contrast to enzymes involved in other post-translational modifications, proteases catalyze essentially irreversible hydrolytic reactions and, consequently, they must be strictly regulated. The action of proteases can be controlled in vivo by several mechanisms: regulation of gene expression; activation of their inactive zymogens; blockade by endogenous inhibitors; targeting to specific compartments such as lysosomes, mitochondria and specific apical membranes; and post-translational modifications, such as glycosylation, metal binding, S-S bridging, proteolysis and degradation.

To date, transcriptional mechanisms regulating gene expression are largely unknown for most proteases, although in some specific protease families of great relevance for human disease, such as MMPs, detailed information is already available about the variety of hormones, growth factors, cytokines, and chemokines controlling their expression in both normal and pathological conditions (37). The promoter regions of some of these genes have also been characterized, which has facilitated the identification of transcription factors such as Fos, Jun, NFκB, and Cbfa1, and epigenetic mechanisms that mediate changes in the expression levels of these enzymes (38).

The activation of inactive protease precursors can be either autocatalytic or catalyzed by other proteases, although in some cases, protease activation requires additional factors or platforms, such as the apoptosome, which mediates the activation of proapoptotic caspases (39). Protease activation may also be modulated by protein cofactors, such as the tissue factor glycoprotein that binds to the serine protease factor VIIa and initiates the coagulation cascade (40). Substrate-driven allosteric mechanisms of protease activation without prodomain cleavage have also been proposed for some metalloproteases (41).

All known endogenous protease inhibitors are proteins, although some microorganisms produce small non-protein inhibitors which block the proteolytic activity of host proteases. To date, the number of identified endogenous inhibitors is considerably lower than that of proteases. As an illustrative example, a total of 183 genes encoding protease inhibitors have been annotated in the rat genome, which markedly contrasts with the more than 600 protease genes present in this species (20). This unbalanced situation derives in part from the relaxed specificity of several inhibitors towards their target proteases, although there are also many proteases which are not blocked by any endogenous inhibitor, as their proteolytic activities are regulated at other
Protease inhibitors have been classified in families of structurally related members or according to the catalytic class of proteases targeted by them. Nevertheless, this classification is hampered by the occurrence of both compound inhibitors that contain inhibitor units of different protease classes, and pan-inhibitors – such as α2-macroglobulin – that target enzymes of different classes through a trapping reaction induced after inhibitor cleavage by the targeted protease (42). Protease inhibitors can also be classified into four groups according to their mechanism of inhibition (12,43). The canonical inhibitors, including serpins (serine protease inhibitors) block the active site of their target proteases through binding in a virtually substrate-like manner. By contrast, exosite-binding inhibitors like cystatins and some thrombin inhibitors, bind a region adjacent to the active site, thereby preventing substrate access to this centre, but without directly blocking the catalytic residues. A third group of protease inhibitors including TIMPs (tissue inhibitors of metalloproteases), use an intermediate mechanism based on a combination of the canonical and exosite-binding mechanisms. Finally, allosteric inhibitors – such as X-linked inhibitor of apoptosis protein, a caspase inhibitor – bind a region which is distantly located from the active site but this binding prevents dimerization of the target protease and blocks its activity.

In addition to these main regulatory mechanisms, proteolysis may also be regulated or fine-tuned by epigenetic changes in the promoter regions of protease genes, control of mRNA stability, translation and degradation by trans-acting factors such as RNA-binding proteins and microRNAs, spatial and temporal protease compartmentalization, substrate interaction with inactive protease homologs which act as protease antagonists, shedding of substrate-binding domains, oligomerization, cellular internalization, and finally, autolysis reactions which lead to the termination of proteolytic activities. All these mechanisms must operate in a coordinate manner to assure that the correct substrates are processed at the right moment and in the appropriate environment, thereby preventing the potentially harmful actions of uncontrolled proteases on living systems. Over the past years, our understanding of regulatory mechanisms acting at the level of individual proteases has considerably improved, but limited information is available on the global regulation of proteolytic systems. The emergence of high-throughput methodologies for profiling proteases in different organisms will contribute to define the regulatory mechanisms operating in the precise and dynamic control of the protease web.

**Frontiers and Perspectives**

At the beginning of the post-genome era, a large body of information is available about the composition and organization of proteolytic systems in many living organisms. These global genomic views have revealed that the protease landscape is vast and quite unexplored. Therefore, it is very likely that the size of the different degradomes will grow in the near future, as new enzymes with unusual structural designs and catalytic mechanisms are identified and characterized. The recent finding of two novel and evolutionary conserved cysteine proteases called UfSP1 and UfSP2 represents an example of experimental work which has led to the unmasking of “hidden proteases” that had remained invisible to homology-based screening methods (44). Clearly, the rapid flow of genomic data has widely expanded the proteolytic universe but many newly identified proteases remain as *in silico*
predictions without experimental evidence for enzymatic activity. A major challenge for the future will be to demonstrate enzymatic properties for these predicted proteases, especially in the case of non-mammalian proteases whose functional characterization is proceeding at a relative low pace. These comparative genomic studies have also provided interesting information about conservation, neo-functionalization and sub-functionalization events in the protease field. Thus, the lineage-specific expansion of reproductive proteases in rodents may help to explain some of the pronounced reproductive differences between mammalian species, whereas changes in immune-related proteases may reflect evolutionary diversification of host defense mechanisms in response to new environmental conditions (45). The ongoing and future analysis of additional degradomes will shed new light on the molecular changes that have facilitated the development of species-specific evolutionary innovations in protease-mediated functions.

In relation to the relevance of proteases for human disease, genomic studies will contribute to the elucidation of genetic diseases caused by mutations in protease loci as well as to the identification of protease gene polymorphisms associated with an increased susceptibility to certain diseases. Likewise, the availability of a complete human protease catalog will be a building framework for the evaluation of new proteases as drug targets or prognostic markers (12). The future implementation of protease-chips for global analysis of expression and activity of human proteases will be very helpful for this purpose (3). Peptidome analysis in which protease-generated peptides are identified in blood or other complex biological samples, might also be an alternative approach to measure the activity of proteolytic systems dysregulated in human disease (46). Functional and clinical studies will likely extend the emerging idea that, beyond their classical detrimental roles in human disease, proteases may also have beneficial roles in different pathological conditions (17). The future availability of three-dimensional structures of proteases determined by high-throughput crystallographic methods, ab initio predictions or homology modelling approaches will provide excellent opportunities to design new generations of therapeutic inhibitors, such as those recently developed for targeting the proteasome in myeloma multiple (47), or dipeptidyl peptidase IV in type II diabetes (48).

Recent advances in different fields have also converged in the development of innovative strategies to profile the expression and activity levels of the multiple proteases present in complex cellular samples. Oligonucleotide microarrays, activity-based probes and different fluorescent-based assays including quantum dot-peptide conjugates have been recently used for profiling and monitoring protease levels and activity. Novel methods are also being introduced to identify the in vivo substrates targeted by individual enzymes, a crucial step towards the functional characterization of those orphan proteases whose biological role is yet unknown. The currently available strategies for de-orphaning proteases and identify their in vivo functions are as diverse as the protease themselves, although a first step towards this goal is frequently based on the determination of consensus cleavage sites for a protease by using phage-displayed peptide libraries, combinatorial fluorogenic substrate libraries, positional scanning synthetic libraries, or mRNA-displayed protein libraries (49). Nevertheless, these methods only provide information about peptide sequences that can be cleaved but do not demonstrate that they are actually cleaved in their natural context, thus making necessary the utilization of additional approaches for linking a
protease to its specific substrates. These approaches can be classified in two general categories: ex vivo proteomic-based methods and in vivo genetic-based methods (49). The first group of techniques includes exosite scanning, inactive catalytic domain capture (ICDC), or methods such as ICAT (isotope-coded affinity tag) and iTRAQ (isobaric tagging for relative and absolute quantification), which use affinity-tagged mass spectrometric labeling of newly cleaved substrates (49). The genetic studies to identify in vivo protease substrates are usually based on the detection of non-processed substrates accumulated in tissues of knockout mice deficient in specific proteases. Similar studies in other model organisms such as C. elegans, D. melanogaster, and A. thaliana have also allowed the identification of in vivo substrates of proteases (49), although these genetic strategies are hampered by the occurrence in most proteolytic systems of redundant and compensatory activities. A powerful alternative to loss-of-function animal models for substrate identification derives from the application of RNA interference techniques to the protease field (50). Nevertheless, it is unlikely that a single methodology will be sufficient to identify the substrates targeted by specific proteases under in vivo conditions. A system-wide approach termed degradomics (3) and involving the combination of biochemical studies, genetic tactics, cell-based assays and proteomic methods will be necessary in the quest for the natural substrates of the multiple orphan proteases still present in all organisms. Degradomic studies will also be essential to define the regulatory and functional connections between all different components of proteolytic systems that form the protease web.

Finally, the detailed analysis of complex protease-mediated processes such as proteolytic regulation of transcription factor activity, protein ectodomain shedding, or regulated intramembrane proteolysis, are among the challenges to be addressed in the near future. Hopefully, through a series of articles focused on structural and functional aspects of proteolytic systems as well as on the analysis of relevant biological processes regulated by them, this minireview series will provide a current view of this complex group of protein sculptors that decisively influence the rhythms of cell life and death in all living forms.
Acknowledgements. We thank all members of our labs for their helpful comments on the manuscript and apologize for omission of relevant works due to space constraints. C.L-O. is supported by grants from Ministerio de Educación y Ciencia, European Union, Fundación M. Botín, and Fundación Lilly. The Instituto Universitario de Oncología is supported by Obra Social Cajastur-Asturias, Spain. J.S.B. is supported by grants from the National Institutes of Health (DK-19691) and the Pennsylvania Department of Health Tobacco Settlement Funds.

REFERENCES


FIGURE LEGENDS

Fig. 1. Representative structures of the protease domains of different catalytic classes, and their mechanisms of proteolysis.

Panel A: Crystal structures of representative members of the five major mechanistic classes of proteases: Serine (S) protease, human matriptase (PDB entry 2gv6); Threonine (T) protease, yeast 20S proteasome containing 28 subunits (PDB entry 2fak); Cysteine (C) protease, human bleomycin hydrolase (PDB entry 1cb5); Aspartic (A) protease, HIV-1 protease (PDB entry 830c); Metalloprotease (M), human collagenase-1 (PDB entry 830c). The α-helices are displayed in blue and β-sheets in yellow. Side chains of the catalytic residues are shown in red. Atomic coordinates were downloaded from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb), and rendered with molmol (http://hugin.ethz.ch/wuthrich/software/molmol/index.html).

Panel B: Schematic representation of two mechanisms of proteolysis. One (upper scheme) for enzymes that form covalent complexes with substrates (Serine, Threonine, Cysteine proteases) through a nucleophilic attack by a catalytic residue; and the other (lower scheme) for those enzymes that do not form covalent complexes with substrates (Aspartic and Metalloproteases). Upper scheme, example of proteolysis mediated by a serine protease. Side chains of the catalytic triad (Ser-Asp-His residues) of the protease are shown. The activated catalytic serine side chain attacks the carbonyl group between the P1 and P1’ residues of the substrate (nomenclature of Schecter & Berger; see Appendix of ref. 16). This attack produces a covalent protease-substrate (acyl-enzyme) intermediate. The rearrangement of this intermediate and attack by water results in hydrolysis of the substrate. For threonyl-and cysteine-proteases, the catalytic nucleophile is either oxygen of threonine or sulfur of cysteine, but the basic mechanism is similar to that shown for serine proteases. Lower scheme, proteolysis mediated by metalloproteases. The catalytic glutamate side chain assisted by a metal ion (usually zinc) activates a water molecule which attacks the carbonyl group between the P1 and P1’ residues of the substrate. A reaction intermediate is formed, in which protease and substrate interact non-covalently. The rearrangement of this intermediate, catalyzed by the protease, results in hydrolysis of the substrate. In aspartyl-proteases, a similar mechanism is proposed, with the water molecule activated in concert by two catalytic aspartate side chains.

Fig. 2. A global view of the proteolytic landscape in representative eukaryotic genomes. Proteases from human, mouse, Drosophila and Arabidopsis are shown distributed in catalytic classes and families of related members. Catalytic classes are indicated as (A) aspartic, (C) cysteine, (M) metallo, (T) threonine and (S) serine proteases, whereas the associated numbers correspond to the different families of proteases belonging to each catalytic Class. For example, there are only five distinct evolutionary families within in the aspartic protease Class, whereas there are 27 families in the metalloprotease Class. The Y-axis shows the number of individual proteases identified thus far within a family for each specie. For example, for the human aspartic protease family A01 there are 5 proteases, whereas in Arabidopsis there are 55 members in this family. Data for human and mouse proteases are from ref. 17 and http://uniovi.es/ degradome; data for Arabidopsis are from ref. 13; and data for Drosophila from (http://merops.sanger.ac.uk).