

COMMUNICATION

Comparative Genomics and Protein Domain Graph Analyses Link Ubiquitination and RNA Metabolism

J. Ignasi Lucas¹, Vicente Arnau² and Ignacio Marín^{1*}

¹Departamento de Genética
Universidad de Valencia
Burjassot, Spain

²Departamento de Informática
Universidad de Valencia
Burjassot, Spain

The human gene *parkin*, known to cause familial Parkinson disease, as well as several other genes, likely involved in other neurodegenerative diseases or in cancer, encode proteins of the RBR family of ubiquitin ligases. Here, we describe the structural diversity of the RBR family in order to infer their functional roles. Of particular interest is a relationship detected between RBR-mediated ubiquitination and RNA metabolism: a few RBR proteins contain RNA binding domains and DEAH-box RNA helicase domains. Global protein domain graph analyses demonstrate that this connection is not RBR-specific, but instead many other proteins contain both ubiquitination and RNA-related domains. These proteins are present in animals, plants and fungi, suggesting that the link between these two cellular processes is ancient. Our results show that global bioinformatic approaches, involving comparative genomics and domain network analyses, may unearth novel functional relationships involving well-known and thoroughly studied groups of proteins.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: comparative genomics; ubiquitin ligases; Parkinson disease; domain graph

*Corresponding author

Most proteins contain domains, characteristic units that provide important clues about their functions. Generally, adscription of functions to novel proteins is easy when known catalytic domains are present. In addition, genome projects have provided enough information as to predict the functions of many proteins by examining their non-catalytic domains. A particularly fruitful approach has been to analyze the relationships among domains found together in multidomain proteins.^{1–4} It has been shown that multidomain proteins tend to include domains that are functionally related.^{5,6} Thus, we can associate an unknown domain with a particular biochemical role if it appears together with other domains of known function. This rationale has led to a considerable interest for developing global analyses of protein domains. These global approaches start by building domain graphs. In them, domains are treated as nodes. Whenever two or more domains are found together in a protein, edges are traced to connect them in the graph. The study of those graphs has provided interesting insights on how multidomain

proteins evolve.^{7–10} They have also been shown to be useful to establish the existence of lineage-specific combinations of domains, and a direct relationship between protein complexity and organism complexity.^{5,8,11–13} However, domain graph analyses have just been started to be explored as a method of assigning functional roles to proteins encoded by complex gene families.⁶ In part, this is due to limitations in the way graphs are analyzed. For example, so far there are only a few strategies to extract clusters of highly connected nodes, that may be functionally related, from complex graphs. We tackled this problem in a recent work in which we described a new tool, UVCLUSTER, specifically devised to solve the problem of finding clusters in complex graphs.¹⁴ UVCLUSTER performs a large number (10^3 – 10^6) of hierarchical clustering analyses in order to establish the average strength of the links among all nodes in a graph. Average strength means in this context the probability of two nodes being part of the same cluster when many, equally valid, clustering solutions are obtained. Originally devised to explore protein–protein interaction graphs, UVCLUSTER can be easily adapted to other types of graphs, such as domain graphs. The search strategy implemented in our program has advantages over previous tools. Particularly,

E-mail address of the corresponding author:
ignacio.marin@uv.es

UVCLUSTER allows the selection of one or a few starting nodes and then detects additional nodes that are highly connected with the selected ones. In this way, regions of graphs that contain particularly significant nodes can be explored. This allows for rapid, highly focused analyses that were impossible to perform before.

In eukaryotes, the control of protein degradation by the ubiquitin-proteasome system is crucial for many cellular functions.^{15,16} Ubiquitination is also involved in other processes, such as ribosome function, regulation of transcription, membrane protein transport or protein translocation to and from the nucleus in ways that are often unrelated to protein degradation.^{17–22} The importance and diversity of its functions makes the study of the ubiquitination enzymatic machinery a very active field of research.^{16,23} Three kinds of enzymes are involved in the cascade of reactions leading to ubiquitination: ubiquitin-activating enzymes (also known as E1s), ubiquitin-conjugating enzymes (E2s) and ubiquitin-ligases (E3s).^{23,24} E3s are by far the most diverse group and the main determinants of the specificity and regulation of ubiquitination. They can be classified into three main types. The first are characterized by containing HECT domains.²⁵ The second type of E3s contains RING fingers, characteristic cysteine-rich regions that mediate protein–protein interactions.²⁶ Structural data suggest that the RING finger domain acts as a scaffold that positions the ubiquitin-conjugating enzyme and the substrate.^{27,28} Finally, a third type of E3s is characterized by having a U-box, a domain structurally related to the RING finger.²⁹

RING finger E3s may act alone or in multi-subunit E3 complexes.²⁴ They are the most abundant type of ubiquitin ligases in eukaryotic genomes. Out of 442 murine proteins predicted to be E3 enzymes, 305 contained RING fingers, while 38 were HECT E3s, the remaining being mostly members of multi-subunit E3 complexes also containing RING finger proteins.³⁰ Some time ago, we described the evolutionary history of a peculiar family of RING finger-containing E3s, that we named RBR.³¹ Proteins of the RBR family are characterized by a specific arrangement of domains, namely two RING fingers (called RING1 and RING2) separated by an IBR (“in between rings”) domain.³² Recently it has been shown that RING2, which we already showed does not fit the traditional RING finger sequence pattern,³¹ has a three-dimensional structure that is very different from that of a canonical RING finger.³³ In any case, the RING/IBR/RING or RBR signature is a characteristic supra-domain,³⁴ which defines this particular family of proteins and is not present in any other RING finger-containing E3s. There is a considerable interest in determining the cellular roles of RBR proteins. Particularly, several members of the RBR family are being extensively studied because of their involvement in human diseases. Mutations in the RBR gene *parkin* cause familiar Parkinson disease³⁵ and this gene has been shown

recently to be also involved in susceptibility to leprosy.³⁶ Another RBR gene, *dorfin*, may be involved in several neurodegenerative diseases, most especially amyotrophic lateral sclerosis.^{37,38} On the other hand, several RBRs are regulators of tumor suppressors or act as tumor suppressors themselves, and thus may be involved in cancer.³⁹

In order to characterize the structural diversity of the RBR proteins, and following the strategies described in Supplementary Data, we generated a database containing 347 RBR sequences. We performed a much more extensive analysis than the one presented before,³¹ in which only 74 sequences were considered. RBRs were found in many different eukaryotic organisms, including animals, plants, fungi and protozoa. No prokaryotic sequences were detected but a single viral sequence that most likely co-opted an RBR domain from an eukaryotic host was found (see details in Supplementary Data). Phylogenetic and structural analyses defined 14 subfamilies (Supplementary Data Figure 1; the corresponding alignment is publicly available at the EMBL-Align database, with accession number ALIGN_000850). We confirmed the existence of the seven subfamilies previously described: Ariadne, Plant I, Plant II, Dorfin, ARA54, XAP3 and Parkin.³¹ Seven new ones, that we have named Protozoan I, Fungal I, Fungal II, RNF144, Triad3, Paul, and IBRDC1, became evident in this new analysis. All the details of the evolutionary range and phylogeny of the subfamilies derived from these analyses can be found in Supplementary Data that accompanies this work (see Supplementary Data and Supplementary Figures 1–3).

Our extensive searches demonstrated that structural diversification has occurred in parallel with sequence diversification and the emergence of RBR subfamilies. So far, we have found 16 different structures in RBR proteins (Figure 1). Structural data provide a general view of the pattern of RBR genes diversification in different evolutionary lineages. Diversification of subfamilies and appearance of novel domain architectures have been especially prevalent in animals. We detected six animal-specific subfamilies (Parkin, Dorfin, Paul, Xap3, IBRDC1, RNF144) and the first four of them have novel domain architectures. Meanwhile, both in plants and fungi, only two lineage-specific subfamilies were found (Fungal I and Fungal II and Plant I and Plant II, respectively), and only one of them in each lineage (Plant I and Fungal I) are structurally novel. Both Plant I and Fungal I proteins have particularly interesting structures (Figure 1). They contain domains involved in RNA binding (RRM_1, KH_1) or typical of RNA helicases (DEAD, Helicase_C). These results suggest that these proteins, and perhaps other RBRs, are involved in some type of function that implies the participation of RNAs. This would be a novel role for RBR proteins.

We can establish which protein domains are closely related to the RBR supra-domain by

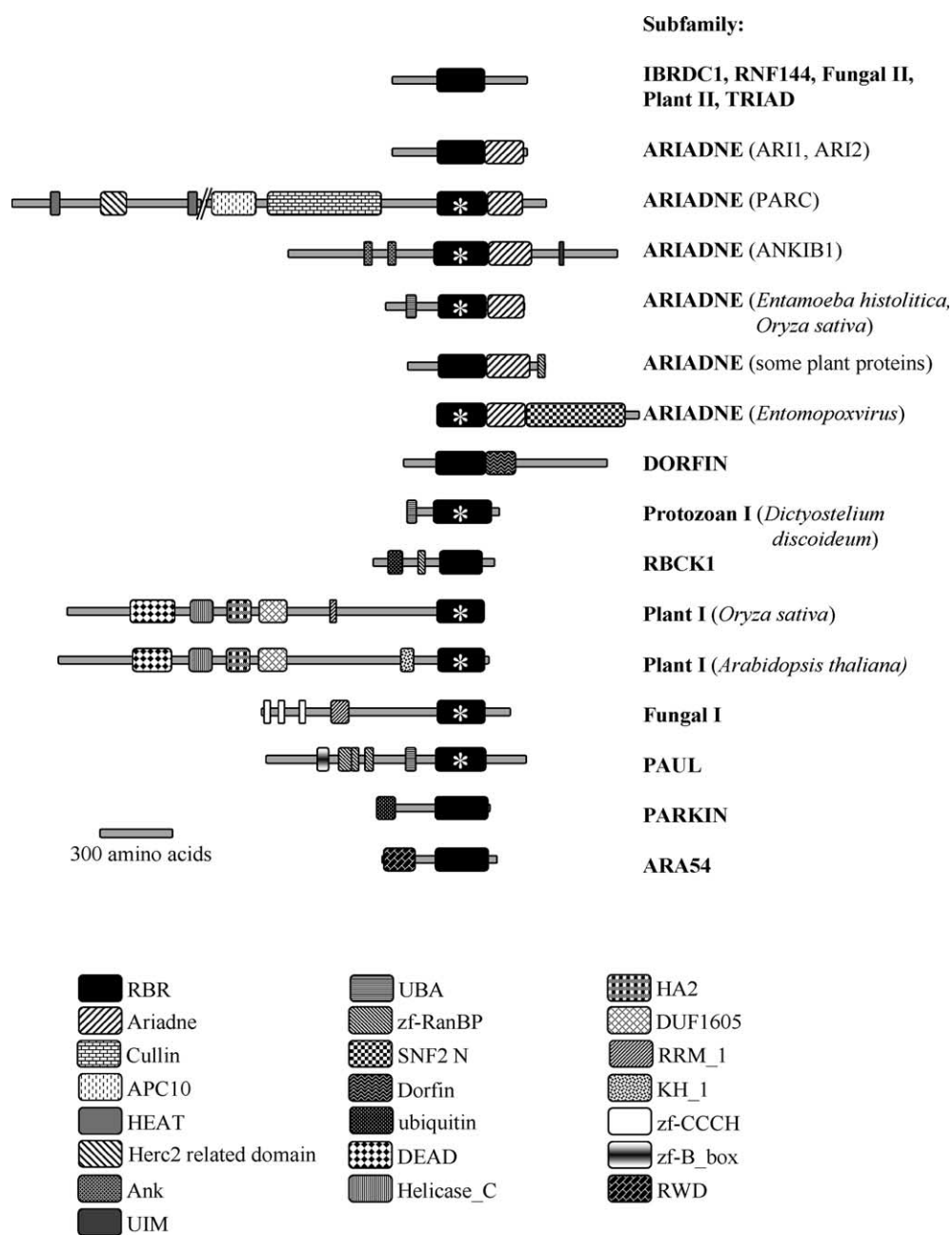


Figure 1. Domain architectures of the RBR family. Most RBR subfamilies are characterized by specific architectures. Asterisks show structures that are described here for the first time. To generate this Figure, we explored the Conserved Domain Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>), and the Pfam database (<http://www.sanger.ac.uk/Software/Pfam/>), as well as other InterPro member databases by means of the InterProScan tool (<http://www.ebi.ac.uk/InterProScan/>). In some particular cases, BLASTP or PSI-BLAST⁵³ searches were used to solve ambiguities or disagreements among databases.

analyzing a global domain-association graph, using the domains present in RBR proteins as a starting point. We therefore extracted from Pfam-A all domains with average distances $d \leq 2$ from the 22 domains present in RBR proteins and at the same time directly connected by an edge with at least two of those domains. We found a total of 77 domains with those features. This is only about 1% of the total number of domains present in Pfam-A (for details see Supplementary Data) showing that the criteria used for selection were very restrictive.

Iterative clustering using UVCLUSTER and the primary distance data for those 77 domains plus the other 22 present in RBR proteins established their relative proximity. Figure 2 shows an UPGMA-based tree established with the secondary distances generated with UVCLUSTER for those 99 domains. There are several main conclusions that can be established based on this tree. First, many domains present in ubiquitination-related proteins and especially the three ubiquitin ligase-specific domains (RING finger (or zf-C3HC4, according to

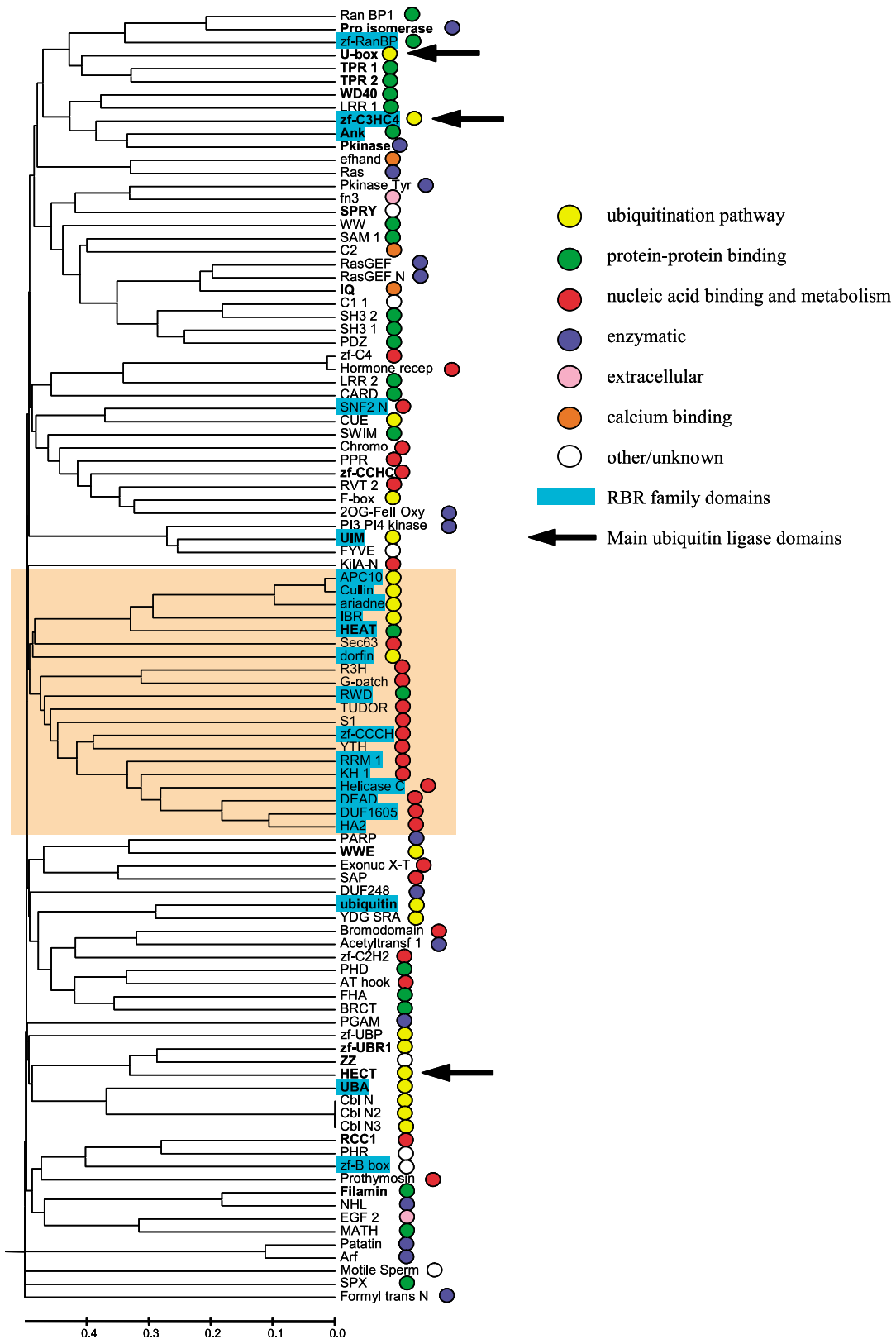


Figure 2. UVCLUSTER-based tree relating the protein domains that are present in the RBR proteins (marked in blue) with those that are, on average, the closest to them. Domains are color coded according to their most likely function, as described in Pfam.⁵⁴ The cluster in which the IBR domain is present, together with other domains involved in ubiquitination and RNA metabolism, is highlighted. Bold letters indicate those domains that are present in at least two families of ubiquitin ligases (see Figure 3 and the text for the details).

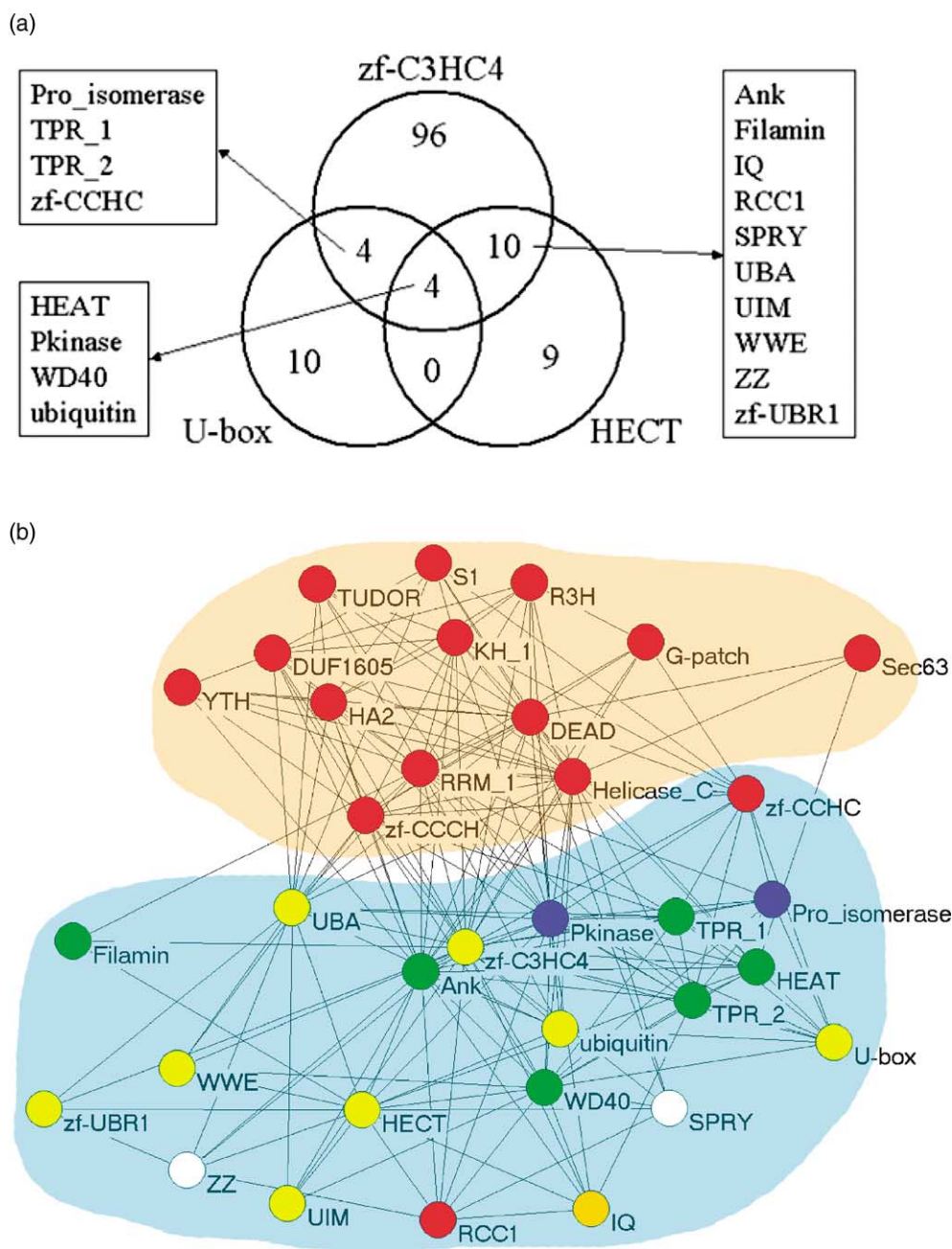


Figure 3. (a) The 18 domains present in at least two of the three main classes of ubiquitin ligases (characterized by having RING finger/zf-C3HC4, U-box or HECT domains). (b) Graph showing the connections among 21 ubiquitination domains (the 18 domains described in (a), above, plus the RING finger/zf-C3HC4, U-box and HECT domains) and the RNA binding and RNA metabolism domains detected by UVCLUSTER. Color codes for nodes are those detailed in Figure 2. Distances to build this Figure were extracted from the global graph data using Matlab 7.0, and the graph was drawn using PAJEK 1.09 (available at <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>).

Pfam nomenclature), U-box and HECT) appear in the tree. This result suggests that many proteins related to the ubiquitination process contain domains that are closely linked in the global domain graph and also that the three families of ubiquitin ligases must share a significant number of domains. This point is explored further below. Second, none of those ubiquitin-ligase specific domains are very close to the IBR domain, which, being specific of RBR proteins may be considered

for our purposes as the central node of the graph. This is easily understood considering that the RING finger, which should be the most closely linked to the IBR of the three domains, appears in many other types of proteins, some of them unrelated with ubiquitination. Third, the tree includes only a few domains of unknown function. Finally, and most interestingly, the IBR domain and other domains present in RBR proteins appear in a cluster that contains a high number of domains characteristic of

RNA binding or RNA metabolism. According to Pfam, all the domains marked with red dots in the cluster highlighted in Figure 2 are found in RNA-interacting proteins. Because only a few of them (R3H, Helicase_C, DEAD) are present also in DNA-interacting proteins, this result must be largely explained by the existence of a link between RBR proteins and RNA metabolism. We described above the particular RBR proteins that have RNA-related domains. However, such a strong connection was totally unexpected, because UVCLUSTER results imply that RNA-related domains must be highly connected with many of the domains present in RBR proteins. Otherwise, it would be impossible for the program to determine such a close relationship. An approximate estimation of the likelihood of this cluster arising by chance can be obtained using data from Pfam. There are 648 domains in Pfam that are included under the keyword “RNA” and ten of them are found in the cluster highlighted in Figure 2. Using the hypergeometric distribution, we can estimate the probability of this distribution of data occurring by chance to be 1.4×10^{-4} . The Pfam keyword search is not very precise (e.g. we have found a few domains described as RNA-interacting that not are included under the RNA keyword), but in any case this result confirms the significance of the observed cluster. In Supplementary Figure 4, we show the edges connecting

domains present in RBR proteins and the RNA-related domains detected in the UVCLUSTER analysis. In agreement with UVCLUSTER results, both types of domains are often connected, that is, there must be multiple proteins that contain at the same time the domains present in RBR proteins, in general typical of ubiquitin metabolism proteins, and the RNA metabolism or RNA binding-related domains.

We decided to explore whether this link between ubiquitination and RNA metabolism is RBR family-specific or, on the contrary, proteins involved in ubiquitination other than the RBRs have also RNA-binding domains. To do so, we decided to select a set of domains involved in the ubiquitination process. We chose those that are present in at least two of the three families of ubiquitin ligases (18 in total) plus the three domains that characterize each E3 family (RING finger, HECT, U-box) (for details see Figure 3(a)). We may expect to find this set enriched in domains typically involved in ubiquitin metabolism. When we checked whether these domains are connected to the RNA metabolism group found in the RBR cluster observed in Figure 2, we detected again numerous edges among domains in both groups (Figure 3(b)). Because only six of the 21 ubiquitination-related domains shown in Figure 3(a) and (b) are present in RBR proteins, it is obvious that the link between ubiquitination and

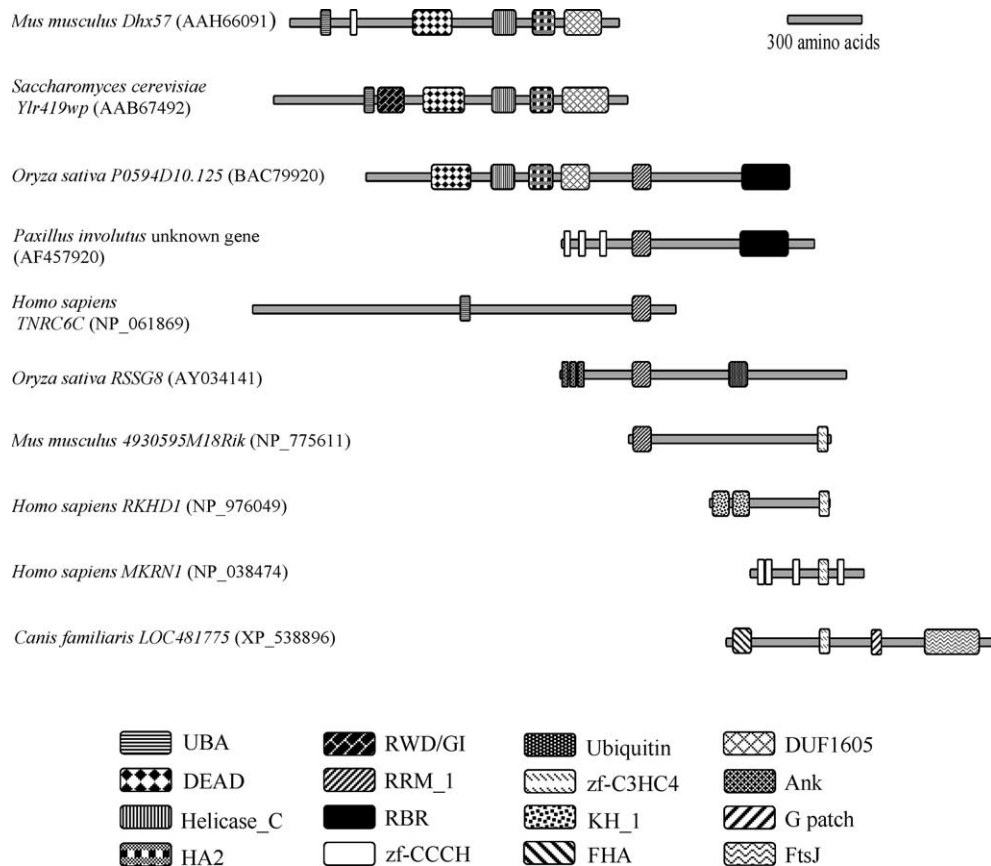


Figure 4. Representative members of the types of DEAH-box RNA helicases or proteins with RNA binding domains that also contain domains typically involved in ubiquitination. Data were obtained from Pfam.

RNA metabolism is not a peculiarity of this family, but it must be a general feature of ubiquitination-related proteins. In Figure 4, we show proteins detected in Pfam that contain one or several of the RNA binding or RNA metabolism domains detected in our global searches plus typical ubiquitination domains. For example, we found three distinct structural types of putative RNA helicases that contain ubiquitination-related domains, one of them being the RBR protein already characterized before as belonging to the Plant I subfamily. Sequence analyses showed that these proteins unambiguously belong to the DEAH-box class of RNA helicases (i.e. they are clearly not related to DNA helicases, again demonstrating that the results found are due to a specific RNA/ubiquitination relationship). Interestingly, the proteins that contain both ubiquitination and RNA-related domains are present in animals, plants or fungi, suggesting that connections between RNA metabolism and ubiquitination are ancient and must be widespread in eukaryotes. The available information for the proteins shown in Figure 4 is very scarce. Regarding the putative DEAH RNA helicases, our group described in a previous study *Ylr419w* as the most likely ortholog in *S. cerevisiae* of the *Drosophila* dosage compensation gene *maleless* and the mammalian gene *RNA helicase A*.⁴⁰ Human *RKHD1* was shown to encode an RNA binding protein that interacts with the mRNA of the anti-apoptotic gene *Bcl-2*.⁴¹ Finally, the protein encoded by *Homo sapiens MKRN1* (Makorin ring finger protein 1) has been shown very recently to act as a ubiquitin ligase, having as a substrate the catalytic subunit of the telomerase, hTERT.⁴² The Makorin gene family is widespread in animals.⁴³ This result is particularly significant, because it is the first time that a protein with demonstrated ubiquitin ligase activity was shown to possess RNA binding domains. The fact that telomerase includes both protein and RNA components may explain the features of the MRKN1 protein.

In summary, we have shown that RBR proteins often contain multiple closely related domains, typical of ubiquitin ligases. This is why we found the HECT and U-box domains, characteristic of the two families of ubiquitin ligases that lack RING fingers, as well as many other domains present in different families of ubiquitin ligases among those most closely connected to those present in RBR proteins (Figure 2). The second main conclusion is that a few RBRs have incorporated domains that apparently have no relationship with ubiquitination. Particularly, the structures of RBRs of the Plant I and Fungal I subfamilies, which contain DEAD, Helicase_C, HA2, RRM_1 or KH_1 domains, strongly suggest an involvement of those proteins in RNA metabolism. In the last part of this study, we have conclusively shown that this is not an exceptional feature of RBR proteins. On the contrary, a large number of edges among domains typical of ubiquitination and domains involved in RNA metabolism have been found in the domain

graph (Figure 3). Among the proteins that contain both types of domains and are included in Pfam, we have selected some examples (Figure 4). The detection of ubiquitin-related and RNA-related domains in many different proteins and in multiple distant lineages implies a general and ancient connection between ubiquitination and RNA metabolism. Although it is true that isolated domain associations found in multi-domain proteins may not imply a functional connection, the finding of multiple unrelated proteins in which the same type of association is present is strong evidence for a functional link to exist. Our finding of a few RNA helicases that contain ubiquitination-related domains is particularly significant. In spite of the diverse roles of DEAD/DEAH-box helicases^{44–46} and also the increasing diversity of roles, often not related to proteolysis, performed by the ubiquitin system,^{19,47} evidence linking ubiquitination and RNA metabolism is currently very weak. Spence *et al.*¹⁷ showed that in yeast ribosomal protein L28 is ubiquitinated when part of the ribosome and that ubiquitination may contribute to regulation of protein synthesis. Very recently, Brenner & Guthrie⁴⁸ detected genetic interactions between splicing and ubiquitination genes while Vinuesa *et al.*⁴⁹ found a ubiquitin ligase (Roquin) with an RNA-binding domain (zf-CCHC, precisely one of the domains connected to both RING fingers and U-box domains; see Figure 3(a)), speculating that this may mean that Roquin acts on RNA-associated substrates. In our opinion those results, the ones presented here and others such as the description of a DNA helicase that contains an F-box, another domain typical of proteins involved in ubiquitination⁵⁰ or the description of another likely DNA helicase with an UBA domain,⁵¹ suggest that it is likely that the relationships of diverse aspects of the metabolism of nucleic acids with the ubiquitination system remain to be characterized.

This was the first time that UVCLUSTER was used to extract significant biological information from graphs unrelated to protein–protein interaction data, although we already mentioned in our description of the program that analyses involving protein domains could be easily performed using it.¹⁴ The advantages of UVCLUSTER-based graph analyses are obvious when we compare our results with those obtained by other authors. Both Ye & Godzik⁵ and Liu *et al.*⁵² detected some edges that connected ubiquitin and nucleic acid metabolism domains in their analyses of eukaryotic domain graphs. However, because their approaches were able to determine the existence of connections but not the strength of the connections (i.e. the importance of the particular edges in the context of all the edges), it was difficult for them to draw any conclusions. Thus, Ye & Godzik⁵ suggested that the connection may be due to a putative role of the RING finger in DNA binding (so far this remains a speculation) while Liu *et al.*⁵² suggested that the connection may be related to some general role of

ubiquitination in gene regulation. The basic of the strategy implemented in UVCLUSTER is, however, to analyze multiple times the graph in order to gather an estimation of the relative strength of the connections among the nodes. In this way when we observe, as happened in Figure 2, that two apparently unrelated types of domains are closely linked, this necessarily means that there must be not just some but indeed abundant direct connections among the two classes of domains. It is thus strong proof of a functional relationship. This was clearly demonstrated in our subsequent analyses. In summary, we have shown how UVCLUSTER may be used to provide novel insights about the precise meaning of the connections found in particular regions of complex graphs.

Acknowledgements

Our group is supported by grants GEN2001-4851-C06-02 and SAF2003-09506 (Ministerio de Educación y Ciencia; Spain) and grant GV04B-141 (Generalitat Valenciana, Spain). J. I. L. is the recipient of a predoctoral fellowship (Generalitat Valenciana).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.12.068

References

- Koonin, E. V., Wolf, Y. I. & Karev, G. P. (2002). The structure of the protein universe and genome evolution. *Nature*, **420**, 218–223.
- Chothia, C., Gough, J., Vogel, C. & Teichmann, S. A. (2003). Evolution of the protein repertoire. *Science*, **300**, 1701–1703.
- Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C. & Teichmann, S. A. (2004). Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* **14**, 208–216.
- Bornberg-Bauer, E., Beaussart, F., Kummerfeld, S. K., Teichmann, S. A. & Weiner, J., III (2005). The evolution of domain arrangements in proteins and interaction networks. *Cell. Mol. Life Sci.* **62**, 435–445.
- Ye, Y. & Godzik, A. (2004). Comparative analysis of protein domain organization. *Genome Res.* **14**, 343–353.
- Björklund, A. K., Ekman, D., Light, S., Frey-Skött, J. & Elofsson, A. (2005). Domain rearrangements in protein evolution. *J. Mol. Biol.* **353**, 911–923.
- Wuchty, S. (2001). Scale-free behavior in protein domain networks. *Mol. Biol. Evol.* **18**, 1694–1702.
- Apic, G., Gough, J. & Teichmann, S. A. (2001). Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **310**, 311–325.
- Kummerfeld, S. K. & Teichmann, S. A. (2005). Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* **21**, 25–30.
- Vogel, C., Teichmann, S. A. & Pereira-Leal, J. (2005). The relationship between domain duplication and recombination. *J. Mol. Biol.* **346**, 355–365.
- Wolf, Y. I., Brenner, S. E., Bash, P. A. & Koonin, E. V. (1999). Distribution of protein folds in the three superkingdoms of life. *Genome Res.* **9**, 17–26.
- Wuchty, S. & Almaas, E. (2005). Evolutionary cores of domain co-occurrence networks. *BMC Evol. Biol.* **5**, 24.
- Tordai, H., Nagy, A., Farkas, K., Bányai, L. & Patthy, L. (2005). Modules, multidomain proteins and organismic complexity. *FEBS J.* **272**, 5064–5078.
- Arnau, V., Mars, S. & Marín, I. (2005). Iterative cluster analysis of protein interaction data. *Bioinformatics*, **21**, 364–378.
- Weissman, A. M. (2001). Themes and variations on ubiquitylation. *Nature Rev. Mol. Cell. Biol.* **2**, 169–178.
- Ciechanover, A. & Brundin, P. (2003). The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron*, **40**, 427–446.
- Spence, J., Gali, R. R., Dittmar, G., Sherman, F., Karin, M. & Finley, D. (2000). Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell*, **102**, 67–76.
- Conaway, R. C., Brower, C. S. & Conaway, J. W. (2002). Emerging roles of ubiquitin in transcription regulation. *Science*, **296**, 1254–1258.
- Muratani, M. & Tansey, W. P. (2003). How the ubiquitin-proteasome system controls transcription. *Nature Rev. Mol. Cell. Biol.* **4**, 192–201.
- Hicke, L. & Dunn, R. (2003). Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annu. Rev. Cell Dev. Biol.* **19**, 141–172.
- Li, M., Brooks, C. L., Wu-Baer, F., Chen, D., Baer, R. & Gu, W. (2003). Mono- versus polyubiquitination: differential control of p53 fate by Mdm2. *Science*, **302**, 1972–1975.
- Shcherbik, N. & Haines, D. S. (2004). Ub on the move. *J. Cell Biochem.* **93**, 11–19.
- Glickman, M. H. & Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol. Rev.* **82**, 373–428.
- Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533.
- Huibregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. M. (1995). A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc. Natl Acad. Sci. USA*, **92**, 2563–2567.
- Borden, K. L. B. & Freemont, P. S. (1996). The RING finger domain: a recent example of a sequence-structure family. *Curr. Opin. Struct. Biol.* **6**, 395–401.
- Cardozo, T. & Pagano, M. (2004). The SCF ubiquitin ligase: insights into a molecular machine. *Nature Rev. Mol. Cell. Biol.* **5**, 739–751.
- Passmore, L. A. & Barford, D. (2004). Getting into position: the catalytic mechanisms of protein ubiquitylation. *Biochem. J.* **379**, 513–525.
- Hatakeyama, S. & Nakayama, K. I. (2003). U-box proteins as a new family of ubiquitin ligases. *Biochem. Biophys. Res. Commun.* **302**, 635–645.
- Semple, C. A. M. & RIKEN GER Group, GSL Members (2003). The comparative proteomics of ubiquitination in mouse. *Genome Res.* **13**, 1389–1394.
- Marín, I. & Ferrús, A. (2002). Comparative genomics of the RBR family, including the Parkinson's disease-related gene *Parkin* and the genes of the *Ariadne* subfamily. *Mol. Biol. Evol.* **19**, 2039–2050.

32. Morett, E. & Bork, P. (1999). A novel transactivation domain in parkin. *Trends Biochem. Sci.* **24**, 229–231.
33. Capili, A. D., Edghill, E. L., Wu, K. & Borden, K. L. B. (2004). Structure of the C-terminal RING finger from a RING-IBR-RING/TRIAD motif reveals a novel zinc-binding domain distinct from a RING. *J. Mol. Biol.* **340**, 1117–1129.
34. Vogel, C., Berzuini, C., Bashton, M., Gough, J. & Teichmann, S. A. (2004). Supra-domains: evolutionary units larger than single protein domains. *J. Mol. Biol.* **336**, 809–823.
35. Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S. *et al.* (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*, **392**, 605–608.
36. Mira, M. T., Alcais, A., Nguyen, V. T., Moraes, M. O., Di Flumeri, C., Vu, H. T. *et al.* (2004). Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature*, **427**, 636–640.
37. Niwa, J., Ishigaki, S., Hishikawa, N., Yamamoto, M., Doyu, M., Murata, S. *et al.* (2002). Dornin ubiquitylates mutant SOD1 and prevents mutant SOD1-mediated neurotoxicity. *J. Biol. Chem.* **277**, 36793–36798.
38. Takeuchi, H., Niwa, J., Hishikawa, N., Ishigaki, S., Tanaka, F., Doyu, M. & Sobue, G. (2004). Dornin prevents cell death by reducing mitochondrial localizing mutant superoxide dismutase 1 in a neuronal cell model of familial amyotrophic lateral sclerosis. *J. Neurochem.* **89**, 64–72.
39. Marín, I., Lucas, J. I., Gradilla, A. & Ferrús, A. (2004). Parkin and relatives: the RBR family of ubiquitin ligases. *Physiol. Genomics*, **17**, 253–263.
40. Sanjuán, R. & Marín, I. (2001). Tracing the origin of the compensasome: evolutionary history of DEAH helicase and MYST acetyltransferase gene families. *Mol. Biol. Evol.* **18**, 330–343.
41. Donnini, M., Lapucci, A., Papucci, L., Witort, E., Jacquier, A., Brewer, G. *et al.* (2004). Identification of TINO. A new evolutionary conserved Bcl-2 AU-rich element RNA-binding protein. *J. Biol. Chem.* **279**, 20154–20166.
42. Kim, J. H., Park, S. M., Kang, M. R., Oh, S. Y., Lee, T. H., Muller, M. T. & Chung, I. K. (2005). Ubiquitin ligase MKRN1 modulates telomere length homeostasis through a proteolysis of hTERT. *Genes Dev.* **19**, 776–781.
43. Gray, T. A., Hernandez, L., Carey, A. H., Schaldach, M. A., Smithwick, M. J., Rus, K. *et al.* (2000). The ancient source of a distinct gene family encoding proteins featuring RING and C₂H zinc-finger motifs with abundant expression in developing brain and nervous system. *Genomics*, **66**, 76–86.
44. Tanner, N. K. & Linder, P. (2001). DexD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell.* **8**, 251–262.
45. Silverman, E., Edwalds-Gilbert, G. & Lin, R. J. (2003). DexD/H-box proteins and their partners: helping RNA helicases unwind. *Gene*, **312**, 1–16.
46. Rocak, S. & Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. *Nature Rev. Mol. Cell. Biol.* **5**, 232–241.
47. Sun, L. & Chen, Z. J. (2004). The novel functions of ubiquitination in signaling. *Curr. Opin. Cell Biol.* **16**, 119–126.
48. Brenner, T. J. & Guthrie, C. (2005). Genetic analysis reveals a role for the C terminus of the *Saccharomyces cerevisiae* GTPase Snu114 during spliceosome activation. *Genetics*, **170**, 1063–1080.
49. Vinuesa, C. G., Cook, M. C., Angelucci, C., Athanasopoulos, V., Rui, L., Hill, K. M. *et al.* (2005). A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity. *Nature*, **435**, 452–458.
50. Kim, J., Kim, J. H., Lee, S. H., Kim, D. H., Kang, H. Y., Bae, S. H. *et al.* (2002). The novel human DNA helicase hFBH1 is an F-box protein. *J. Biol. Chem.* **277**, 24530–24537.
51. Bagherieh-Najjar, M. B., de Vries, O. M. H., Kroon, J. T. M., Wright, E. L., Elborough, K. M., Hille, J. & Dijkwel, P. P. (2003). *Arabidopsis* RecQsim, a plant-specific member of the RecQ helicase family, can suppress the MMS hypersensitivity of the yeast *sgs1* mutant. *Plant Mol. Biol.* **52**, 273–284.
52. Liu, S., Zhang, C. & Zhou, Y. (2005). Domain graph of *Arabidopsis* proteome by comparative analysis. *J. Proteome Res.* **4**, 435–444.
53. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
54. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S. *et al.* (2004). The Pfam protein families database. *Nucl. Acids Res.* **32**, D138–D141.

Edited by D. E. Draper

(Received 9 November 2005; received in revised form 18 December 2005; accepted 20 December 2005)
Available online 9 January 2006