# The Construction of a Database for Ubiquitin Signaling Cascade

**Tadahiko Sakiyama** sakiyama@kuicr.kyoto-u.ac.jp
**Shuichi Kawashima** shuichi@kuicr.kyoto-u.ac.jp
**Akiyasu C. Yoshizawa** acyshzw@kuicr.kyoto-u.ac.jp
**Minoru Kanehisa** kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

**Keywords:** ubiquitin, protein degradation, database

## 1 Introduction

Protein ubiquitination plays an important role in eukaryotic cellular processes. It mainly functions as a signal for 26S proteasome dependent protein degradation. Ubiquitin is a small protein, which consists of 76 residues, and is well conserved throughout all eukaryotes. The addition of ubiquitin to proteins being degraded is performed by a reaction cascade consisting of three enzymes, named E1 (ubiquitin activating enzyme), E2 (ubiquitin conjugating enzyme) and E3 (ubiquitin ligase). It is especially interesting that each E3 has specificity to its substrate (Fig. 1a) [3]. Corresponding to many proteins to be targeted by ubiquitination, many E3s are also discovered in eukaryotes. In addition, it is estimated that more E3s remained uncovered [1]. It is thought that the specificity and control of timing for the protein degradation are essential as much as the control of the gene transcription and translation to keep cellular functions normal. In order to investigate the protein degradation system regulated by ubiquitination, we have constructed the database of proteins that are involve in the ubiquitin signaling cascade, and have analyzed sequence similarities, protein domains and distribution across species.

(a) (b)

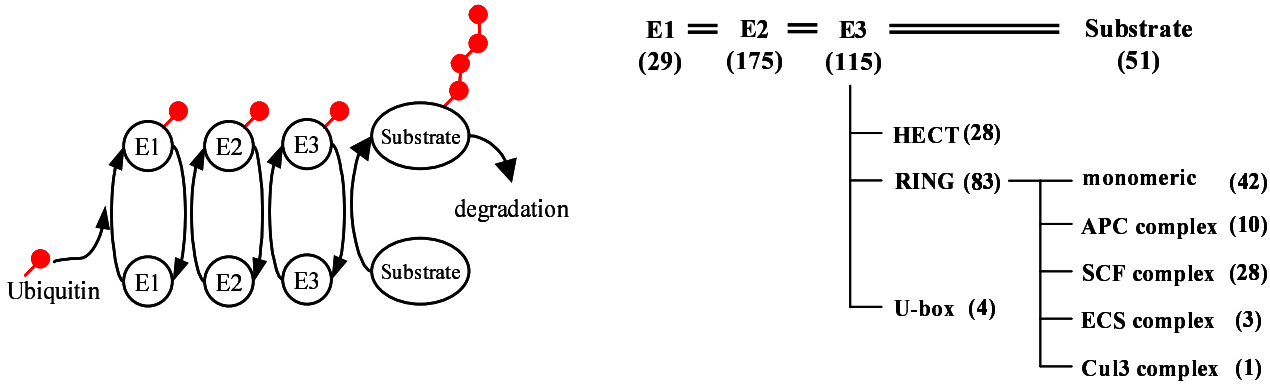


Figure 1: (a) Schematic diagram of ubiquitination cascade. (b) Number of collected sequences for each type of enzyme.

## 2 Methods

1. We have collected the information about enzymes which are confirmed to participate in ubiquitination from published papers and retrieved their sequences from the KEGG/GENES database

[2]. We identified 39 E1, 175 E2 and 115 E3 sequences. E3 enzymes are classified in several categories by their characteristic domains and structural forms (Fig 1b).

2. We have attempted to search similar sequences from nr-aa (non-redundant amino acid sequence) database in GenomeNet with BLAST2. Each known sequence was used as a query against nr-aa, and similar sequences were collected with the threshold of E-value $< e^{-5}$, removing redundant hits. We obtained similar sequence sets for E1, E2, and E3. The E3 dataset was split into each type of E3s.

3. The sequences in the dataset were checked with domain information in Pfam and PROSITE databases.

## 3 Result and Discussion

Both of E1 and E2 enzymes are well characterized experimentally, particularly in *S. cerevisiae.* In contrast to high sequence conservations in E1 and E2, the sequences of E3 enzymes are very diverged. The number of E3 similar sequence is 5409 entries in the BLAST2 search. It is much larger than 1592 entries for E1 and 1301 entries for E2 homolog. It indicates that many sequences have potentials to work as E3 ubiquitin ligase. Recently, a novel E3 complex of Cul3 was discovered, and it contains BTB domain protein as the substrate recognition subunit [4]. So, we surveyed KEGG/GENES to find the molecules that contain the domain "F-box" or "SOCS-box" or "BTB domain". These domains are part of the substrate recognition subunit of SCF, ECF and Cul3 complex, respectively, which are variations of the E3 complex of Cullin protein. The number of these molecules that we found is shown below (Table 1). Larger numbers of F-box and BTB proteins are observed in *A.thaliana* and *C.elegans*, and SOCS-box proteins are almost non-existent in these two species. The number of SOCS-box proteins tends to increase in higher eukaryotes. These proteins may be able to form a complex. To narrow the candidates for E3s, we used the information on substrate recognition domain. For instance, the number of BTB proteins reduced to 38 by considering "MATH domain" which are estimated to be substrate recognition domain. We think that the accumulation of such information is helpful to reveal proteins involved in ubiquitin signaling cascade.

Table 1: Number of the sequences with important domains for Cullin containing E3s.

| | *H.sapiens** | *M.muscles** | *D.melanogaster* | *A.thaliana* | *C.elegans* | *S.cerevisiae* | *S.pombe* |
|---|---|---|---|---|---|---|---|
| GENES entry | 16468 | 15919 | 15925 | 27762 | 21515 | 6332 | 5053 |
| F-box | 19 | 0 | 0 | 363 | 181 | 11 | 5 |
| SOCS-box | 30 | 32 | 17 | 0 | 3 | 0 | 0 |
| BTB | 51 | 0 | 0 | 63 | 102 | 1 | 1 |

(*Two species on the left side of the table is not associated with complete genome information yet.)

## Acknowledgments

This work was supported by grants from Ministry of Education, Science, Sports and Culture, the Science and Technology Agency, and the Japan Society for the Promotion of Science. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

## References

[1] Hershko, A. and Ciechanover, A., The ubiquitin system, *Annu. Rev. Biochem.*, 67:425–479, 1998.

[2] Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A., The KEGG database at GenomeNet, *Nucleic Acids Res.*, 30(1):42–46, 2002.

[3] Pickert, C.M., Mechanisms underlying ubiquitination, *Annu. Rev. Biochem.*, 70:503–533, 2001.

[4] Pintard, L., Willis, J.H., Willems, A., Johnson, J.L., Srayko, M., Kurz, T., Glaser, S., Mains, P. E., Tyers, M., Bowerman B., and Peter, M., The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase, *Nature*, 425:311–316, 2003.