BMC Bioinformatics

UbiProt: a database of ubiquitylated proteins

Alexander L Chernorudskiy[1], Alejandro Garcia[2], Eugene V Eremin[3], Anastasia S Shorina[2], Ekaterina V Kondratieva[2] and Murat R Gainullin*[2]

Address: [1]Nizhny Novgorod State University, 23 Gagarin Ave., Nizhny Novgorod, Russia, [2]Central Research Laboratory, Nizhny Novgorod State Medical Academy, 10/1 Minin Sq., Nizhny Novgorod, Russia and [3]Institute of Applied Physics RAS, 46 Ul'yanov St., Nizhny Novgorod, Russia

Email: Alexander L Chernorudskiy - chalbio@mail.ru; Alejandro Garcia - garcia@rbcmail.ru; Eugene V Eremin - eugene@awp.nnov.ru; Anastasia S Shorina - anastasia@gma.nnov.ru; Ekaterina V Kondratieva - kondratyeva_k@mail.ru; Murat R Gainullin* - biochem@gma.nnov.ru

* Corresponding author

Abstract

Background: Post-translational protein modification with ubiquitin, or ubiquitylation, is one of the hottest topics in a modern biology due to a dramatic impact on diverse metabolic pathways and involvement in pathogenesis of severe human diseases. A great number of eukaryotic proteins was found to be ubiquitylated. However, data about particular ubiquitylated proteins are rather disembodied.

Description: To fill a general need for collecting and systematizing experimental data concerning ubiquitylation we have developed a new resource, UbiProt Database, a knowledgebase of ubiquitylated proteins. The database contains retrievable information about overall characteristics of a particular protein, ubiquitylation features, related ubiquitylation and de-ubiquitylation machinery and literature references reflecting experimental evidence of ubiquitylation. UbiProt is available at http://ubiprot.org.ru for free.

Conclusion: UbiProt Database is a public resource offering comprehensive information on ubiquitylated proteins. The resource can serve as a general reference source both for researchers in ubiquitin field and those who deal with particular ubiquitylated proteins which are of their interest. Further development of the UbiProt Database is expected to be of common interest for research groups involved in studies of the ubiquitin system.

Background

Ubiquitin is a small (76 amino acids) protein that has an ability to be linked to other intracellular proteins via a covalent isopeptide bond [1]. Such a posttranslational modification of target proteins named "ubiquitylation" leads to a manifestation of various biological effects of ubiquitin. Ubiquitylation is implicated in protein degradation via an intracellular ATP-dependent proteolytic system [2]. Additionally, it participates in several non-proteolytic events. It is common knowledge that the ubiquitin system mediates throng of essential cellular outcomes, including a cell cycle control, an inflammatory response, carcinogenesis and many others, therefore ubiquitylation is a phenomenon of great importance for cell vital functions [3].

A diversity of known biological effects of ubiquitylation is realized through modification of a huge variety of target proteins followed by alteration of their functions. A rapidly growing number of experimental evidence on protein

ubiquitylation determines a general need for such data collecting and systematization. We try to solve this problem by elaborating a new resource, the UbiProt Database, a knowledgebase of ubiquitylated proteins. This project aims to summarize a significant volume of data concerning ubiquitylation and to provide essential information on target proteins.

Construction and content

UbiProt is developed and deployed with an open source software. A database management system is MySQL 4.0. The software is developed on the basis of PHP 4.0.3 including some additional modules like SMARTY and PEAR. Web interface software uses PHP+SMARTY template framework.

All data included were experimentally obtained by various research groups and can be verified using respective references. The main content was obtained from several large-scale proteomic studies [4-7]. The rest of target proteins were acquired from original research articles containing direct evidence for ubiquitylation of particular proteins. A manual annotation was performed for all database entries.

Each entry possesses a block structure summarized in a single page (Figure 1) to provide highly structured and easy-retrievable information on each ubiquitylated protein. The first information block contains overall characteristics of a particular protein, i.e. a name and synonyms, genes, a source organism and an expression system in case of recombinant proteins, as well as some basic properties such as molecular weight, full length and mature form length. The next block describes protein ubiquitylation features and includes a description of modified lysine residue(s) and an ubiquitylation type: a number of attached ubiquitins (mono-ubiquitylation, multi-ubiquitylation, pluri-ubiquitylation) and a structure of the branched multi-ubiquitin chains. These details are especially important because different kinds of ubiquitylation have a different impact on a fate of the modified protein. For instance, a multi-ubiquitin chain linked by the Lys48-Gly76 isopeptide bond targets a protein for proteasome-mediated degradation, while the Lys63 assembled chains participate in a DNA repair and other non-catabolic processes [8]. On the other hand, mono-ubiquitylation often serves as an endocytic signal for a diverse set of membrane proteins [3]. A diversity of ubiquitin forms and ubiquitin-protein conjugates is illustrated in Figure 2. It also reflects specific terms used in UbiProt in accordance with the terminology offered by H.P. Jennissen [1]. We should also note that there is no universal terminology for ubiquitin system, and ubiquitylation-related terms vary from paper to paper. For example, some authors use the term "poly-

ubiquitylation" for a type of modification defined as "multi-ubiquitylation" in Jennissen's terminology.

The following information block provides information on the respective ubiquitylation and deubiquitylation machinery: cognate ubiquitin-conjugating components E2 and E3, deubiquitylating enzyme(s) (DUB), non-enzymatic adaptor proteins (E4/AP), if these components are known.

Each protein entry also contains literature references with the PubMed links [9] denoting primary research articles from which ubiquitylation details have been obtained. Cross-reference to the respective Swiss-Prot/TrEMBL entry is also available [10]. A link to the Protein Data Bank entry is added in case when 3D structure of the protein from the same source is available [11]. The dataset is supplied with a full-text search tool with a possibility of using Boolean operators. An example of search results is given in Figure 3. Additionally, the whole dataset can be viewed using a [View all proteins] option.

Utility and discussion

UbiProt Database is considered to be a useful tool for different purposes, but mainly for identification of proteins posttranslationally modified by ubiquitin in high-throughput studies. A computational biology (e.g. comparative and evolutionary analysis of protein ubiquitylation) also offers a promising field of a database application. The dataset collected will provide an insight into the ubiquitin-dependent mechanisms controlling essential cellular events and, in future, help to develop new compounds for treating ubiquitin-associated human diseases.

At the present time the information about 400 individual proteins from different organisms is collected and this work is still in progress. Our group permanently analyses a broad array of literature data (approximately 40 journal articles per week) in order to collect new information on the target proteins, identified ubiquitylation sites, structures of multi-ubiquitin chains and features of the ubiquitylation machinery. A continuous renewal of the corresponding fields as well as an insertion of new information blocks, especially concerning a biological function of ubiquitylation, are provided. Our database undergoes updating as soon as appropriate confident information becomes available. We also plan to extend our dataset with a detailed description of up-stream and down-stream components of ubiquitin system, together with comprehensive information on the domain structure of target proteins including ubiquitin-binding domains (reviewed in [12]).

UbiProt database V 1.0

[Main]   [Search]   [Logout]   [Contacts]

Protein information

Protein name	NF-kappa-B essential modulator
Synonyms	NEMO NF-kappa-B essential modifier Inhibitor of nuclear factor kappa-B kinase gamma subunit IkB kinase gamma subunit I-kappa-B kinase gamma IKK-gamma IKKG IkB kinase-associated protein 1 IKKAP1 FIP-3
Gene name	IKBKG , FIP3 , NEMO .
Source	Homo sapiens (Human)
Expression system	Human Jurkat T cells
Full length	419
Precursor	no
Mature form length	419
Molecular weight (Da)	48198
Features	
Comments	Activation of the "classical" NF-k B pathway converges on the Ik B kinase (IKK) signalosome, a protein complex composed of two kinase subunits (IKK-a and IKK-b) and a non-catalytic subunit NF-k B essential modulator (NEMO)

Cross-references

Swiss-Prot/TrEMBL	NEMO_HUMAN
PDB	

Ubiquitylation features

Ubiquitylation site	K399
Ubiquitylation type	Lys63-linked multi-ubiquitin chain
Features	The adaptor protein Bcl10 promotes activation of NF-k B transcription factors by T- and B-cell receptors through paracaspase-and UBC13-dependent ubiquitination of NEMO.

Ubiquitylation/deubiquitylation machinery

E2	E2 complex UBC13-MMS2
E3	Paracaspase
E4/AP	Bcl10
DUB	ND
Comments	Paracaspase lacks a RING or HECT domain normally found in ubiquitin ligases, but instead uses its C-terminal caspase-like domain to catalyse ubiquitination

References

[1]	*Zhou H, Wertz I, O'Rourke K, Ultsch M, Seshagiri S, Eby M, Xiao W, Dixit VM* Bcl10 activates the NF-kappaB pathway through ubiquitination of NEMO. Nature 2004, 427:167-171 [14695475]

[Back]

Figure 1
Structure of a typical UbiProt entry. This entry contains information about an ubiquitylated NEMO protein.
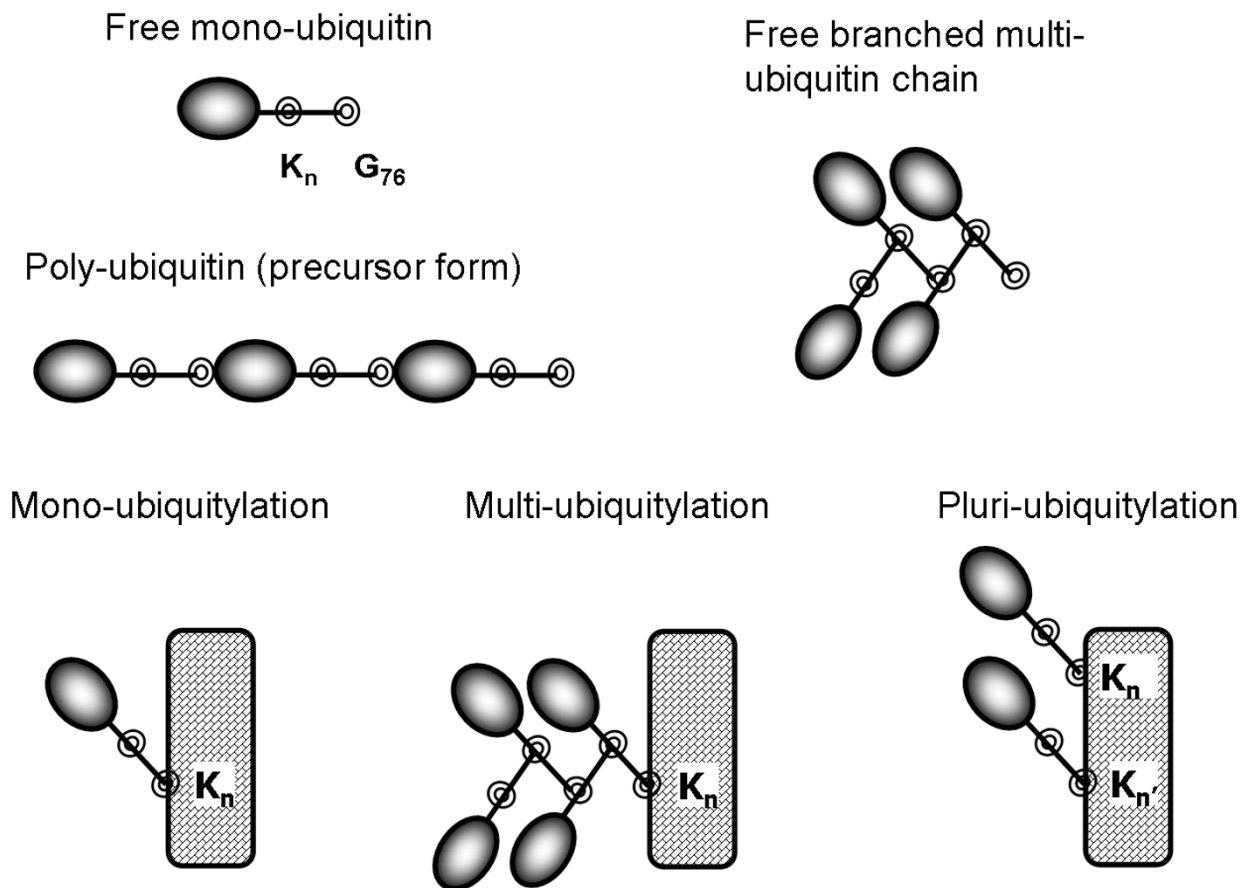
Figure 2
Different forms of ubiquitin and ubiquitin-modified proteins. Ubiquitylation may result in addition of a single ubiquitin moiety or a branched multi-ubiquitin chain to the target protein lysine(s). Note that a ubiquitin molecule possesses 7 inner lysine residues that can serve as attachment sites of the next ubiquitin moiety, resulting in the formation of the chains that have a different structure and topology. Functionally significant amino acids are marked as follows: K_n and $K_{n'}$ – lysine residue(s) that can serve as attachment sites of the ubiquitin moiety; G_{76} – ubiquitin C-terminal glycine residue participating in the isopeptide bond formation. Poly-ubiquitin refers to a ubiquitin fusion protein, a precursor form polymerized "head-to-tail".

UbiProt is more convenient in terms of search of information about ubiquitylated proteins than more general databases. Although databases like Swiss-Prot and Human Protein Reference Database [13] also contain data about ubiquitylation, there are several problems with a retrieval of the information of interest. First of all, one trying to obtain information about ubiquitylation from Swiss-Prot using the SRS system [14] will face some difficulties with a query formulation, as far as the SRS poorly works with complex queries [15]. A query should be formulated very precisely, so it is necessary to find at least one entry prior to the main search. In addition, even precise queries work well only for proteins with an identified ubiquitylation site(s). An amount of proteins with unknown ubiquityla-

tion sites is much more bigger, and in this case query formulation rules may differ. Another problem is search redundancy. A search for ubiquitylated proteins in Swiss-Prot most likely will return not only pure ubiquitylated proteins, but also numerous enzymes of the ubiquitylation cascade.

Besides several search problems mentioned above, there is lack of data in the established databases. Many known substrates of ubiquitylation do not appear as ubiquitylated proteins in Swiss-Prot (e.g. p53, BRCA1, MDM2, Ymer and others), for many other proteins precise ubiquitylation sites are not designated. Human Protein Reference Database also cannot serve as a complex reference
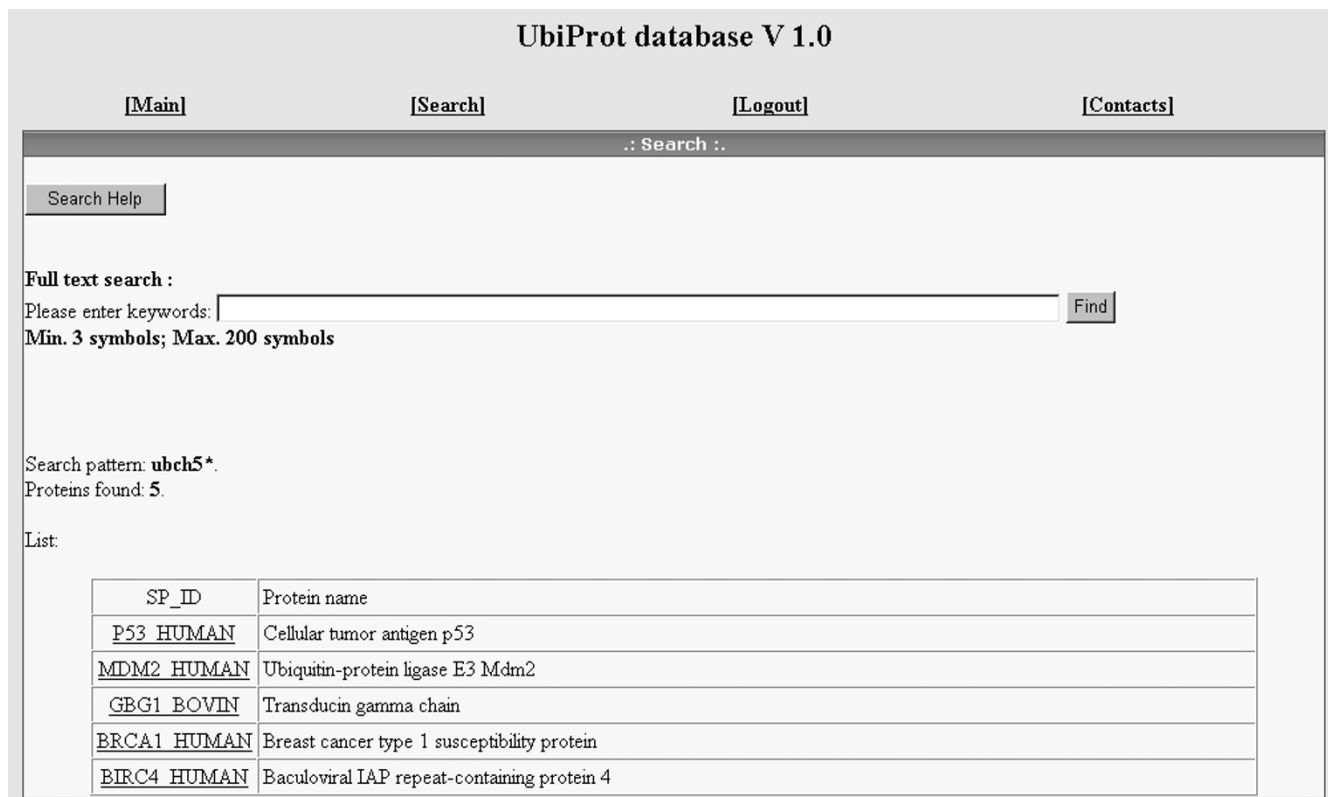
Figure 3
Example of full-text search results. The search pattern is "ubch5*" that corresponds to the particular E2 ubiquitin-conjugation enzyme (UbcH5). The result shows proteins to be modified by the specified enzyme.

source for ubiquitylation, because it contains data only about 18 ubiquitylated proteins, without details about an ubiquitylation type and respective enzymes. This can be due to the insufficient use of data from the proteomic studies dedicated to ubiquitylation. Only 2 recent papers presenting results of high-throughput analysis [4,5] are reviewed in Swiss-Prot at the moment. Our dataset is based also on another proteomic works published so far [6,7,16].

UbiProt aims to collect data from a number of resources including cited databases to make the information easily accessible after validation and annotation. It is more specific and comprehensive comparing to general sequence databases such as Swiss-Prot.

All scientists working on protein ubiquitylation are encouraged to join collaboration in keeping the database up-to-date by submitting additional information and comments. A downloadable Excel form can be used for submitting new data.

Conclusion

UbiProt Database is a comprehensive resource on ubiquitylated proteins, aimed to systematize information on protein ubiquitylation and to make it available for further analysis and use. The resource is considered to be useful as a general reference source both for researchers in the ubiquitin field and those who deal with particular ubiquitylated proteins of interest. Its biological utility and application also includes identification of proteins post-translationally modified by the ubiquitin in high-throughput studies and bio-computational analysis of the ubiquitin-protein conjugates. It may help to understand the mechanisms of ubiquitylation and to develop new compounds for manipulating the ubiquitin-dependent pathways.

Further development of the UbiProt Database is expected to be of common interest for research groups involved in studies of the ubiquitin system.

Availability and requirements

UbiProt can be accessed on a public Apache powered website at http://ubiprot.org.ru. The website was tested with the most of commonly used Web browsers. It is freely

available for non-commercial use. For any questions regarding commercial use please contact the corresponding author.

Abbreviations

ATP – adenosine triphosphate

Lys – lysine

Gly – glycine

DUB – deubiquitylating enzyme

SRS – Sequence Retrieval System

Authors' contributions

AC has carried out data analysis, annotation process and linked the dataset with the related protein databases, and drafted the manuscript.

AG has participated in data analysis and coordination of the study and helped to draft the manuscript.

EE has designed the database structure and built the website.

Both AS and EK have participated in the development and maintenance of the database, and carried out data annotation.

MG has conceived of the study, participated in its design and coordination and helped to draft the manuscript.

All the authors have read and approved the final manuscript.

References

1. Jennissen HP: **Ubiquitin and the enigma of intracellular protein degradation.** *Eur J Biochem* 1995, **231**:1-30.
2. Glickman MH, Ciechanover A: **The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction.** *Physiol Rev* 2002, **82**:373-428.
3. Haglund K, Dikic I: **Ubiquitylation and cell signaling.** *EMBO J* 2005, **24**:3353-3359.
4. Hitchcock AL, Auld K, Gygi SP, Silver PA: **A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery.** *Proc Natl Acad Sci USA* 2003, **100**:12735-12740.
5. Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP: **A proteomics approach to understanding protein ubiquitination.** *Nat Biotechnol* 2003, **21**:921-926.
6. Kirkpatrick DS, Weldon SF, Tsaprailis G, Liebler DC, Gandolfi AJ: **Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin.** *Proteomics* 2005, **5**:2104-2111.
7. Mayor T, Lipford JR, Graumann J, Smith GT, Deshaies RJ: **Analysis of poly-ubiquitin conjugates reveals that the Rpn10 substrate receptor contributes to the turnover of multiple proteasome targets.** *Mol Cell Proteomics* 2005, **4**:741-751.
8. Pickart CM: **Ubiquitin in chains.** *Trends Biochem Sci* 2000, **25**:544-548.
9. **Entrez PubMed** [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed]
10. **UniProt Knowledgebase (Swiss-Prot and TrEMBL)** [http://www.expasy.org/sprot/]
11. **RCSB Protein Data Bank** [http://www.pdb.org/pdb/]
12. Hicke L, Schubert HL, Hill CP: **Ubiquitin-binding domains.** *Nat Rev Mol Cell Biol* 2005, **6**:610-621.
13. **Human Protein Reference Database** [http://www.hprd.org/]
14. **SRSWWW at ExPASy** [http://www.expasy.org/srs5/]
15. Croce O, Lamarre M, Christen R: **Querying the public databases for sequences using complex keywords contained in the feature lines.** *BMC Bioinformatics* 2006, **7**:45.
16. Vasilescu J, Smith JC, Ethier M, Figeys D: **Proteomic analysis of ubiquitinated proteins from human mcf-7 breast cancer cells by immunoaffinity purification and mass spectrometry.** *J Proteome Res* 2005, **4**:2192-2200.