Chapter 9
Ubiquitin and Ubiquitin-Like Conjugations in Complex Diseases: A Computational Perspective

Tianshun Gao, Zexian Liu, Yongbo Wang and Yu Xue

Abstract As one class of most essential and common post-translational modifications (PTMs), ubiquitin and ubiquitin-like (Ub/UBL) conjugations play an important role in almost all aspects of biological processes, and aberrances in the conjugation systems are highly involved in numerous complex diseases. Identification of the Ub/UBL-associated enzymes, substrates and sites is fundamental for understanding the molecular mechanisms of Ub/UBL conjugations, and provides a potential reservoir for discovering disease biomarkers and drug targets. Besides experimental identifications, computational analysis of Ub/UBL conjugations has also emerged as an attractive field. In this chapter, we first summarized the cutting-edge experimental techniques in the large-scale identification of Ub/UBL conjugation substrates, and further emphasized the importance of computational efforts by introducing online databases and predictors for Ub/UBL conjugations. Although computational analysis of Ub/UBL conjugations is still immature, we believe more and more efforts will be paid in the near future.

Keywords Ubiquitin and ubiquitin-like conjugation · Ubiquitination · Sumoylation · Proteomics · Small cell lung cancer

T. Gao · Z. Liu · Y. Wang · Y. Xue (✉)
Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, Luoyu road, 1037 Wuhan 430074, Hubei, China
e-mail: xueyu@hust.edu.cn

T. Gao
e-mail: gts.hust@gmail.com

Z. Liu
e-mail: lzx@mail.ustc.edu.cn

Y. Wang
e-mail: hust.wangyb@gmail.com

B. Shen (ed.), *Bioinformatics for Diagnosis, Prognosis and Treatment of Complex Diseases*, Translational Bioinformatics 4, DOI: 10.1007/978-94-007-7975-4_9,

9.1 Introduction

During the past three decades, the ubiquitin-proteasome system (UPS) has been demonstrated to be critical for protein degradation in most cellular processes (Ciechanover 1994; Bedford et al. 2011; Geng et al. 2012). Ubiquitin (Ub) is a small 76aa protein that binds to target proteins and takes them for destruction through Ubiquitination (Ciechanover 1994), which labels mono- or poly-ubiquitin proteins to substrates via an E1 (Ub-activating enzyme)-E2 (Ub-conjugating enzyme)-E3 (Ub-protein ligase) cascade mechanism (Fig. 9.1a). Recently, more than ten Ub-like modifiers (UBLs) have also been identified, such as SUMO, NEDD8, ISG15, Apg8/12, FAT10, Urm1, UFM1 and Hub1 in eukaryotes, prokaryotic Ub-like protein (Pup) and archaeal SAMPs (Hochstrasser 2009; van der Veen and Ploegh 2012). The prokaryotic homologs of Ub, ThiS and MoaD, are potential antecedents of all Ub/UBL modifiers in eukaryotes (Iyer et al. 2006; van der Veen and Ploegh 2012). Analogous to Ub, most UBLs share a β-grasp fold and a C-terminal diglycine motif, and their conjugation processes, such as sumoylation (Fig. 9.1b) and pupylation (Fig. 9.1c), have a conserved enzyme cascade

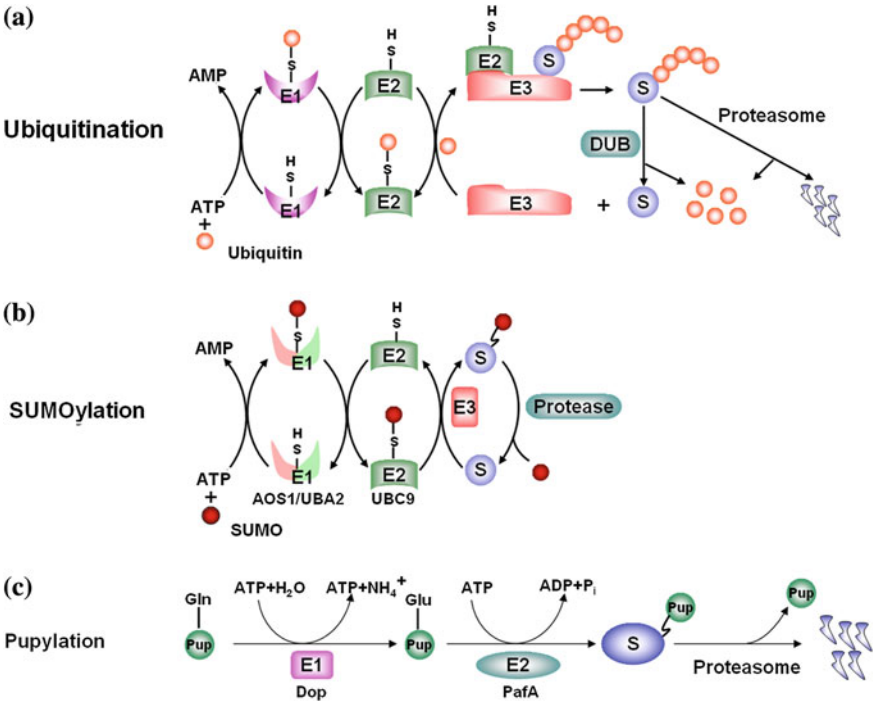

Fig. 9.1 The conjugation processes for **a** ubiquitination, **b** sumoylation, and **c** Pupylation. For ubiquitination, an E1-E2-E3 enzyme cascade mechanism was characterized, and Ub E3 ligases provide the major specificity for substrate recognition. However, SUMO E3 ligases are only cofactors that facilitate the sumoylation, while pupylation doesn't have E3s

mechanism (van der Veen and Ploegh 2012). Ub E3 ligases confer the major specificity of ubiquitination for recognizing substrates (Deshaies and Joazeiro 2009). However, SUMO E3 ligases are only cofactors that facilitate the conjugation of SUMO (Yunus and Lima 2006), and pupylation has only an E1–E2 cascade without any E3 ligases (Striebel et al. 2009). Substrates in the UPS pathway are ubiquitinated through three forms, mono-, multi- and poly-ubiquitination (Sadowski and Sarcevic 2010), while several UBL conjugations, such as SUMO, NEDD8 and SAMP, can also adopt analogical forms for targeting proteins (Ulrich 2008; Ohki et al. 2009; Humbard et al. 2010). Different forms can lead to different fates on substrates (Sadowski and Sarcevic 2010). Mono-ubiquitination affects the activity and location of substrates to be involved in histone regulation, endocytosis and membrane transport (Hicke 2001), while multi- and poly-ubiquitination mainly induce protein degradation as well as non-proteolytic functions (Ciechanover 1994; Rape et al. 2006; Chen and Sun 2009).

Ub and UBL conjugation pathways are implicated in diverse but essential biological functions. Cells usually use these pathways to select specific proteins for destruction, activation or other functions and ensure the fidelity of cellular processes (Ciechanover 1994; Chen and Sun 2009). Thus, aberrances in Ub/UBL conjugation pathways have been identified to be involved in numerous complex diseases (Dahlmann 2007; Bedford et al. 2011), including inflammation (Hochrainer and Lipp 2007; Coornaert et al. 2009), viral infection (Bogunovic et al. 2013), neurodegenerative disease (Hegde and Upadhya 2007; Lehman 2009; Mandel et al. 2009; Deng et al. 2013), cardiac disease (Sohns et al. 2010; Wang 2011), von Hippel-Lindau disease (Kaelin 2007) and several types of cancers (Bonacci et al. 2010; Irminger-Finger 2010; Linehan et al. 2010; Conrad et al. 2011; Escobar et al. 2011; Duncan et al. 2012). However, compared to phosphorylation, in which protein kinases occupied $\sim$30 % of the drug discovery programs in pharmaceutical research and development, ubiquitination owned less than 1 % of drug design (Cohen and Tcherpakov 2010), and only one proteasome Inhibitor Bortezomib was approved currently (Chen et al. 2011). To target complex diseases, theoretically, any components of the UPS and UBL conjugation pathways, including E1s, E2s, E3s, DUBs and proteasomes, can be selected for targeting by small-molecule inhibitors. For example, RING E3s including BARD1 and SIAH (Chasapis and Spyroulias 2009; Irminger-Finger 2010; Wong and Moller 2013), HECT E3 s such as ITCH and SMURF1 (Scheffner and Staub 2007; Melino et al. 2008; Lin et al. 2013), DUBs such as A20 and UCHL1 (Singhal et al. 2008; Coornaert et al. 2009; Day and Thompson 2010), and proteasome subunits such as PSMA7 (Du et al. 2009), had been identified as potential biomarkers of complex diseases. More, inhibitors of several SCF E3 complexes, such as SCF^{skp2}, $SCF^{\beta-TrCP1}$, SCF^{CDC4}, SCF^{Met30}, have also been identified (Chen et al. 2008; Nakajima et al. 2008; Aghajan et al. 2010; Orlicky et al. 2010). The rapid progresses suggested that Ub/UBL conjugation pathways can be a great reservoir for discovering potential biomarkers and drug targets (Cohen and Tcherpakov 2010).

9.2 Advances in High-Throughput Proteomic Analysis of Ub/UBL Conjugations

Because Ub E3 ligases bind substrates at distinct regions and modify specific lysine residues (Bustos et al. 2012), the Ub-mediated proteasomal substrates can be detected by mutating lysines for poly-ubiquitin chain (Chau et al. 1989), substituting E3-substrate binding site (House et al. 2006) or eliminating all lysines of substrate can disrupt the ubiquitination (Bourgeois-Daigneault and Thibodeau 2012). Since high-affinity Ub antibody, linkage specific antibodies and Ub epitope-tags were developed, further studies were focused on the detection of Ub-conjugated substrates (Muller et al. 1988; Newton et al. 2008). For a substrate containing only one ubiquitinated lysine, a single K to R mutation is enough for identifying the ubiquitination site (Flick et al. 2004). However, for multi-ubiquitinated substrates, accurate identification of all ubiquitination sites needs both individual and combinatorial mutations (Zhong et al. 2005). Reintroducing lysine residues one by one into the lysineless mutant (K0) is also an alternative method for identifying multiple ubiquitination sites (Rufini et al. 2011). However, any attempts based on the mutagenesis can only identify one substrate and several ubiquitination sites at most in a single study (Flick et al. 2004; Zhong et al. 2005; Rufini et al. 2011).

In contrast with conventional studies, high-throughput characterization of ubiquitinated substrates provides a more comprehensive understanding of the ubiquitination dynamics and potential relationships between ubiquitinaton and other important cellular processes. Recently, the technologies of mass spectrometry-based proteomics have a significant improvement for the identification of ubiquitination sites (Jeram et al. 2009; Bustos et al. 2012). In the presence of trypsin, Ub-conjugated substrates can be cleaved into K-GG modified peptides (Fig. 9.2a), which can be regarded as ubiquitination signatures (Denis et al. 2007). Thus, the liquid chromatography-mass spectrometry (LC/MS) analysis can detect a mass shift of 114.043 Da, which represents the diglycine (GG) remnant of Ub (Shi et al. 2011) (Fig. 9.2a).

Analogous to Ub, NEDD8, ISG15 and Pup can also produce K-GG remnants with their C-terminal (K/R) GG sequences by the trypsin cleavage, whereas SUMO can't because of the absence of a basic residue adjacent to the C-terminal GG motif (Kang and Yi 2011; Osula et al. 2012). Since the LC/MS identification can't distinguish among K-GGs of Ub, NEDD8 and ISG15, adding MLN4924 but not interferon can effectively block NEDD8ylation and ISG15ylation for exclusively identifying ubiquitinated substrates (Kim et al. 2011; Zhao et al. 2013). However, if Ub was not tagged, only one or several ubiquitination sites of one purified substrate can be identified in vitro (Wang et al. 2005). Thus, with the improvement of Ub epitope-tagging strategies, large-scale analysis of K-GG peptides can be available by the trypsin digestion of hundreds of epitope-tagging Ub-conjugated substrates after in vivo enrichment and purification of Ub-conjugated substrates (Peng et al. 2003; Maor et al. 2007; Danielsen et al. 2011; Kim et al. 2011; Lee et al. 2011; Shi et al. 2011; Oshikawa et al. 2012; Osula et al. 2012; Starita et al. 2012).

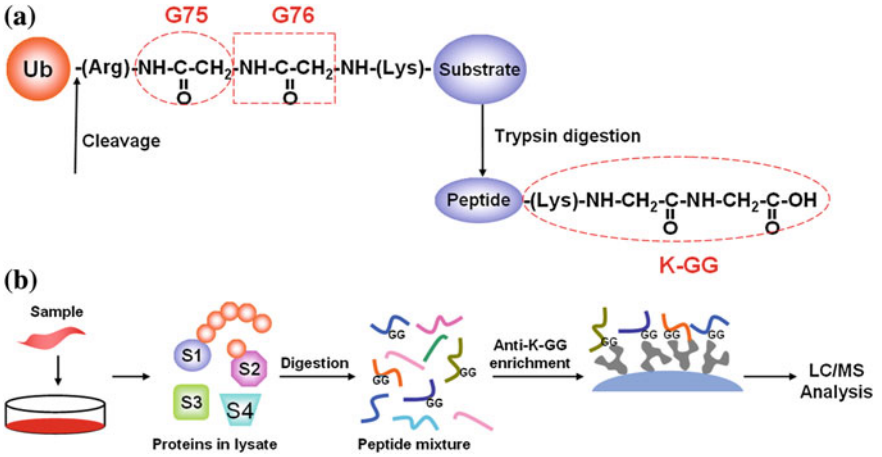

Fig. 9.2 Proteomic analysis of Ub/UBL conjugation substrates. **a** Ub-conjugated substrates can be cleaved into K-GG modified peptides by trypsin. **b** The direct enrichment of in vivo K-GG Peptides from samples has been an efficient approach for the large-scale identification of Ub/UBL conjugation sites

For example, Peng et al. (2003) identified 110 ubiquitination sites and 1,075 ubiquitinated substrates from yeast cells by using His_6-tagged Ub. Also, Maor et al. (2007) detected 85 ubiquitination sites and 294 Ub substrates from Arabidopsis cells with GST-tagged Ub. Furthermore, Meierhofer et al. (2008) characterized 44 ubiquitin acceptor sites and 669 ubiquitinated proteins in HeLa cells, using hexahistidine-biotin (HB)-fused Ub. In particular, Oshikawa et al. (2012) identified 1392 ubiquitination sites of 794 proteins in HEK293T cells, with His_6-tagged K0-Ub. Additionally, this strategy was also adopted for analyzing other UBL conjugations, such as pupylation, which can also generate -GG remnants for the high-throughput identification (Kang and Yi 2011). In fact, Festa et al. (2010) identified 55 pupylation sites from a single sample in *Mycobacterium tuberculosis* (Mtb). As the further improvement of MS techniques, the higher-throughput identification of K-GG peptides was achieved by the direct enrichment of K-GG Peptides in vivo from cells or tissues (Wagner et al. 2011, 2012; Udeshi et al. 2012, 2013) (Fig. 9.2b). For example, Wagner et al. characterized >20,000 ubiquitination sites of >5,200 proteins in murine tissues. In this regard, direct enrichment of K-GG peptides has attracted more attention for further large-scale assays.

9.3 Data Resources for Ub/UBL Conjugations

Currently, there are 13 databases available for Ub/UBL conjugations (Table 9.1). To circumvent competitions, most databases were focused on certain aspects. For example, Lee et al. (2008) developed a budding yeast-specific database SCUD, including 1 E1, 11 E2s, 42 E3s, 20 DUBs and 940 ubiquitinated substrates.

Table 9.1 A summary of Ub/UBL-related databases

Databases	Main propose	Species	Method[a]	Reference[b]	Number		
					Enzyme[c]	Substrate[d]	Sites
SCUD	Ubiquitin-associated enzymes and ubiquitinated substrates	*S. cerevisiae*	TO	Y	74	940	–
PlantsUPS	Ubiquitin-associated enzymes	Plants	TO	N	8,165	–	–
PlantsUBQ	Ubiquitin-associated enzymes	*A. thaliana*	SL, FA	Y	1,416	–	–
hUbiquitome	Ubiquitin-associated enzymes and ubiquitinated substrates	*H. sapiens*	SL	Y	168	279	36
E3Net	Ubiquitin E3 ligases and substrates	427 species	SL, TO	Y	2,201	4,896	–
UUCD	Ubiquitin-associated enzymes	70 eukaryotes	SL, FA	Y	56,949	–	–
DUDE-db	Ubiquitin-associated enzymes	50 eukaryotes	TO, FA	N	35,228	–	–
UbiProt	Ubiquitylated proteins	9 species	SL, TO	Y	–	1,104	222
UniProt	Ubiquitinated and sumoylated substrates	General	SL, OS	Y	–	–	2,502
SysPTM1.1	~50 PTMs including ubiquitinated substrates	General	MS, SL, TO	Y	–	699	1,164
dbPTM3.0	18 PTMs including ubiquitinated and sumoylated substrates	General	MS, SL, OS, FA	Y	–	–	48,781
mUbiSiDa	Mammalian ubiquitination sites	Mammalians	MS, SL	Y	–	27,272	79,425
PupDB	Pupylated proteins	Prokaryotes	SL, FA	Y	–	1305	215

[a] Method, methods used in collecting the data. *TO* taken from other databases or websites; *SL* manually curated from scientific literature; *PS* further computational analysis; *OS* orthologous sites of experimentally verified Ub/UBL conjugation sites; *MS* mass spectrometry-derived data

[b] Reference, whether the information provided in the databases is traceable to original publications

[c] Enzyme, ubiquitin and ubiquitin-like conjugation enzymes, including E1s, E2s, E3s and DUBs

[d] Substrate, substrates of ubiquitin and ubiquitin-like conjugations

Later, Du et al. (2009) constructed a ubiquitination-associated enzyme database plantsUPS, which contains 24 E1, 417 E2s and 7624 E3s from plants. Also, a similar database of PlantsUBQ was developed for plant Ub enzymes, with 2 E1s, 37 E2s, 1,326 E3s and 51 DUBs (http://plantsubq.genomics.purdue.edu/). Furthermore, the hUbiquitome was released for human ubiquitination, with 1 E1, 12 E2s, 138 E3s, 17 DUBs, 279 substrates and 36 ubiquitination sites (Du et al. 2011). In addition, by constructing the E3-mediated regulatory networks, Han et al. (2012) collected 2,201 E3s and 4,896 substrates. The above databases only contains enzyme information for Ub, while UBL conjugations were not included. Recently, we developed a comprehensive database Ubiquitin and Ubiquitin-like Conjugation Database (UUCD) that contains 738 E1s, 2,937 E2s and 46,631 E3s and 6,647 DUBs in 70 eukaryotic species (Gao et al. 2013). Later, Hutchins et al. (2013) also released a similar database DUDE-db for Ub/UBL conjugations, but only with 267 E1s, 2,095 E2s, 28,985 E3s and 3881 substrates in 50 eukaryotic species.

Additionally, several databases have developed exclusively for Ub/UBL conjugation substrates and sites (Table 9.1). The fist database only containing ubiquitinated substrates and sites was UbiProt, which collected 1,104 substrates and 222 ubiquitination sites (Chernorudskiy et al. 2007). The UniProt also contained substrates and sites for multiple post-translational modifications (PTMs), such as ubiquitination and sumoylation (Magrane and Consortium 2011). Since rapid progresses in MS-based proteomics have generated a large number of Ub/UBL conjugation substrates and sites, collection and integration these data sets will provide useful resources for further analysis. For example, Li et al. (2009) created SysPTM that contained modification information for nearly 50 types of PTMs, including 1,164 ubiquitination sites in 699 substrates. DbPTM 3.0, another PTM resource, contains 48,781 ubiquitination and sumoylation sites (Lu et al. 2013). Recently, Hui et al. provided a comprehensive database, including 79,425 mammalian ubiquitination sites of 27,272 proteins (http://222.193.31.35:8000/mUbiSiDa.php). In particular, a UBL conjugation database of PupDB was developed with 1,305 substrates and 215 pupylation sites (Tung 2012).

9.4 Prediction of Ub/UBL Conjugation Sites

Although more and more Ub/UBL conjugation substrates have been identified, accurate prediction of conjugation sites is still a great challenge. To date, although over 20 approaches have been developed for predicting Ub/UBL conjugation sites, only 13 applicable tools can be accessed (Table 9.2). In Tung and Ho (2008) used 531 physicochemical features and the support vector machines (SVMs) algorithm to develop the first predictor of UbiPred, with a training data set of 157 known ubiquitination sites. Using 442 positive sites, Lee et al. (2011) developed UbSite, which adopted a number of sequence features and the radial basis function networks (RBFNs) algorithm for training. Since different organisms may have different features in proteins selected for ubiquitination, the prediction accuracy might

Table 9.2 Predictors for non- or organism-specific Ub/UBL conjugation substrates and sites

Predictors	Training data set[a]	Specificity[b]	Method[c]
Ubiquitination			
UbiPred	157 ubiquitination sites	General	SVMs
UbSite	442 ubiquitination sites	General	RBFNs
UbPred	265 ubiquitination sites in S. cerevisiae	*S. cerevisiae*	RF
CKSAAP_UbSite	263 ubiquitination sites in S. cerevisiae	*S. cerevisiae*	SVMs
hCKSAAP_UbSite	6118 K sites in human	*H. sapiens*	SVMs
UbiProber	25,194 ubiquitination sites in *H. sapiens*, 5348 in *M. musculus* and 175 in *S. cerevisiae*	General and organism-specific	SVMs
GPS-ARM	74 D-box and 42 KEN-box motifs	General	GPS
Sumoylation			
SUMOplot	N/A	General	HS
SUMOsp1.0	239 sumoylation sites	General	GPS
SUMOpre	268 sumoylation sites	General	SM
SUMOsp2.0	279 sumoylation sites	General	GPS
seeSUMO	425 sumoylation sites	General	RF, SVMs
Pupylation			
GPS-PUP	127 pupylation sites	Prokaryotes	GPS

SVMs support vector machines, *RBFNs* radial basis function networks, *RF* random forest, *GPS* group-based prediction system, *HS* hydrophobic similarity; *SM* statistical method

[a] Training Data Set, the experimentally verified Ub/UBL sites were taken as the positive training data set

[b] Specificity, for general propose or organism-specific prediction

[c] Method, the computational methods used for training

be improved in organism-specific manner. For example, Radivojac et al. (2010) collected 265 yeast ubiquitination sites and developed the first organism-specific predictor of UbPred, with the random forest (RF) algorithm. Also, Chen et al. (2011) adopted the composition of *k*-spaced amino acid pairs (CKSAAPs) of lysine-centered peptides and SVMs algorithm to designed a yeast-specific predictor of CKSAAP_UbSite, with a training data set of 263 known ubiquitination sites. Later, they further constructed a human-specific predictor of hCKSAAP_UbSite with the same approaches (Chen et al. 2013). Recently, Chen et al. (2013) adopted a number of sequence features and used the SVMs algorithm to develop UbiProber, which can predict general or organism-specific ubiquitination sites. With the group-based prediction system (GPS) algorithm, we also developed GPS-ARM for the prediction of anaphase-promoting complex/cyclosome (APC/C) recognition motifs including D-box and KEN-box, which can be recognized by Cdh1 or Cdc20 for the protein degradation (Liu et al. 2012). Thus, the GPS-ARM predicts ubiquitinated substrates but not exact sites (Liu et al. 2012).

Beyond ubiquitination, there have been a considerable number of efforts taken for other UBL conjugations, such as sumoylation and pupylation. Because ~77 % of total sumoylation sites follow a canonical motif of Ψ-K-X-D/E (Ψ is a

hydrophobic residue, X is any amino acid) (Xue et al. 2006), the first predictor SUMOplot was developed by evaluating the hydrophobic similarity between given proteins and known sumoylation sites (http://www.abgent.com/sumoplot). Later, using 239 known sumoylation sites as positive samples, we developed SUMOsp1.0 with the GPS algorithm (Xue et al. 2006). With a statistical method, Xu et al. (2008) developed the SUMOpre, which was trained with 268 known sumoylation sites. In 2009, we greatly improved the GPS algorithm and released the SUMOsp 2.0 software package, with a superior performance than other existing tools (Ren et al. 2009). Recently, Teng et al. (2012) used RF and SVMs algorithms to develop the seeSUMO for predicting sumoylation sites. In addition, we also developed an accurate tool of GPS-PUP for the prediction of pupylation sites in prokaryotes (Liu et al. 2011). Due to the page limitation, the computational predictions of Ub/UBL conjugation sites without available programs were not summarized.

9.5 Computational Analysis of Disease-Associated Ub/UBL Conjugations Provides Potential Biomarkers and Drug Targets

To evaluate the importance of Ub/UBL conjugations in diseases and drug targets, we mapped Ub/UBL conjugation enzymes to other databases. First, we obtained 874 human Ub/UBL conjugation enzymes from the UUCD database (Gao et al. 2013), 474 known cancer genes from Cancer Gene Census (Forbes et al. 2011) and 4,096 well-characterized drug targets from Drugbank database (Knox et al. 2011). We mapped cancer genes and drug targets to the human proteomes and got 464 and 2,071 unique sequences, respectively. Also, we mapped all human Ub/UBL conjugation enzymes to the two data sets, and only identified 27 cancer genes and 16 drug targets. The statistical analyses with a hypergeometric distribution demonstrated that both known cancer genes and drug targets were not significantly enriched in Ub/UBL conjugation enzymes (p-value > 0.05). However, we further mapped all enzymes to the KEGG pathways (Kanehisa et al. 2012), and observed that Ub/UBL conjugations are significantly involved in a number of essential pathways (p-value $< 10^{-4}$), such as ubiquitin mediated proteolysis (hsa04120), protein processing in endoplasmic reticulum (hsa04141) and cell cycle (hsa04110) (Table 9.3). In particular, we revealed that Ub/UBL conjugation enzymes are over-represented in the pathway of small cell lung cancer (SCLC, hsa05222) (Table 9.3). Based on the results and KEGG annotations, we illustrated the pathway, and totally detected 12 E3s, 2 E3 complexes and 4 ubiquitinated substrates (Fig. 9.3). The results also demonstrated that ubiquitination plays an important role in SCLC-related PI3 K-Akt signaling, cell cycle, apoptosis and p53 signaling pathways (Fig. 9.3). In this regard, Ub/UBL conjugation enzymes and substrates can be a useful reservoir for further identifying potential biomarkers and drug targets.

Table 9.3 The enrichment analysis of KEGG pathways for 874 human Ub/UBL conjugation enzymes from the UUCD database (Gao et al. 2013) (the hypergeometric distribution, p-value $<10^{-4}$)

KEGG ID	Description	UUCD[a]		Proteome		E-ratio[d]	p-value
		Number[b]	Percentage[c]	Number	Percentage		
The most over-represent KEGG Pathway							
hsa04120	Ubiquitin mediated proteolysis	135	55.10	137	2.21	24.94	8.63E−206
hsa04141	Protein processing in endoplasmic reticulum	32	13.06	165	2.66	4.91	2.32E−14
hsa04110	Cell cycle	22	8.98	125	2.02	4.46	2.35E−09
hsa04114	Oocyte meiosis	20	8.16	110	1.77	4.60	7.24E−09
hsa05222	Small cell lung cancer	13	5.31	84	1.35	3.92	2.12E−05
hsa04914	Progesterone-mediated oocyte maturation	13	5.31	86	1.39	3.83	2.75E−05
hsa04330	Notch signaling pathway	9	3.67	47	0.76	4.85	7.39E−05
The most under-represent KEGG Pathway							
hsa01100	Metabolic pathways	2	0.82	1156	18.64	0.04	6.47E−20

[a] UUCD, proteins in the UUCD database
[b] Number, the number of proteins annotated with the KEGG ID
[c] Percentage the proportion of proteins annotated with the KEGG ID
[d] E-ratio, the enrichment ratio as the proportion of enzymes in UUCD divided by that in the proteome

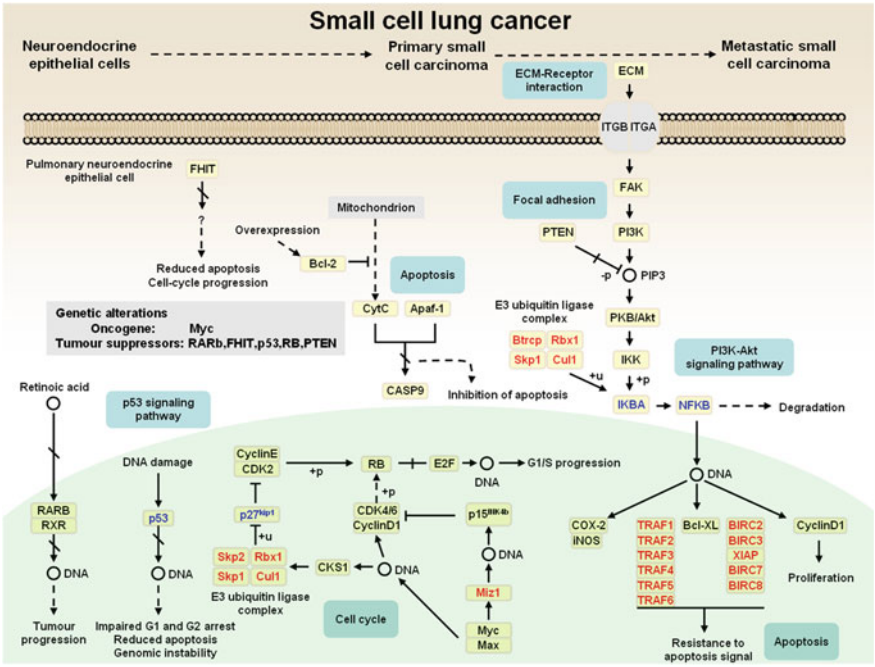

Fig. 9.3 The small cell lung cancer pathway (SCLC, hsa05222) adapted from the KEGG database. The known E3s were shown in *red*, whereas experimentally identified ubiquitinated substrates were shown in *blue*

9.6 Personal Perspectives on Further Computational Analysis of Ub/UBL Conjugations

In this chapter, we presented a brief summarization of current progresses especially computational efforts in Ub/UBL conjugations. Totally, there have been 13 online databases and 13 applicable predictors released for Ub/UBL conjugations. As more and more conjugation substrates and sites have been identified, we believed that more and more databases and tools will be developed in the near future. For further computational studies, we provided several personal perspectives as below:

1. Prediction of conjugation sites for more UBLs. Currently, most computational predictions were focused on ubiquitination and sumoylation, or in a less extent, pupylation. However, over ten UBLs have been characterized, while a number of proteomic analyses of substrates for these UBLs, such as Nedd8-mediated neddylation (Jones et al. 2008) and ISG15-mediated ISGylation (Giannakopoulos et al. 2005). The development of efficient algorithms and predictors can generate useful information for further experimental considerations.
2. Prediction of ubiquitinated substrates and sites in an E3-specific mode. For ubiquitination, the E3 ligases determined the specificity for substrate

recognition. Analogous to phosphorylation which can be catalyzed by ~520 kinases, there were 874 human Ub/UBL conjugation enzymes collected in the UUCD database (Gao et al. 2013). Because different kinases recognize different motifs for modification, we developed a kinase-specific predictor of GPS for the phosphorylation (Xue et al. 2005, 2008). Again, because different E3 ligases exhibited dramatically different sequence or structure profiles, it can be expected that different E3s can recognize distinct motifs for conjugations. In this regard, prediction of E3-specific ubiquitinated substrates and sites will achieve much better performance.

3. Re-construction of Ub/UBL-associated networks. Protein substrates can be modified by E1s, E2s, and E3s and de-modified by DUBs. Thus, the complex relations among Ub/UBL conjugation enzymes and substrates constitute the Ub/UBL-associated networks, which are fundamental for systematically understanding the molecular mechanisms and regulatory roles of Ub/UBL conjugations. Also, how to retrieve useful information from the networks will be a great challenge.

9.7 Conclusion

As a class of important and ubiquitous PTMs, Ub/UBL conjugations has attracted more and more attention to be potential biomarkers or drug targets. Besides both small- or large-scale experimental identifications, computational analysis of Ub/UBL conjugations has also emerged to a promising topic. However, the number of either databases or predictors for Ub/UBL conjugations is still limited, and more efforts should be paid in this field. We believed a better study will generate a deeper understanding on Ub/UBL conjugations and provide useful information for biomedical design.

Acknowledgments This work was supported by grants from the National Basic Research Program (973 project) (2012CB910101, and 2013CB933903), Natural Science Foundation of China (31171263 and 81272578), and International Science and Technology Cooperation Program of China (0S2013ZR0003).

References

Aghajan M, Jonai N, Flick K, Fu F, Luo M, Cai X, Ouni I, Pierce N, Tang X, Lomenick B, et al. Chemical genetics screen for enhancers of rapamycin identifies a specific inhibitor of an SCF family E3 ubiquitin ligase. Nat Biotechnol. 2010;28:738–42.

Bedford L, Lowe J, Dick LR, Mayer RJ, Brownell JE. Ubiquitin-like protein conjugation and the ubiquitin-proteasome system as drug targets. Nat Rev Drug Discov. 2011;10:29–46.

Bogunovic D, Boisson-Dupuis S, Casanova JL. ISG15: leading a double life as a secreted molecule. Exp Mol Med. 2013;45:e18.

Bonacci T, Roignot J, Soubeyran P. Protein ubiquitylation in pancreatic cancer. ScientificWorldJournal. 2010;10:1462–72.
Bourgeois-Daigneault MC, Thibodeau J. Autoregulation of MARCH1 expression by dimerization and autoubiquitination. J Immunol. 2012;188:4959–70.
Bustos D, Bakalarski CE, Yang Y, Peng J, Kirkpatrick DS. Characterizing ubiquitination sites by peptide based immunoaffinity enrichment. Mol Cell Proteomics. 2012.
Chasapis CT, Spyroulias GA. RING finger E(3) ubiquitin ligases: structure and drug discovery. Curr Pharm Des. 2009;15:3716–31.
Chau V, Tobias JW, Bachmair A, Marriott D, Ecker DJ, Gonda DK, Varshavsky A. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science. 1989;243:1576–83.
Chen D, Frezza M, Schmitt S, Kanwar J, Dou QP. Bortezomib as the first proteasome inhibitor anticancer drug: current status and future perspectives. Curr Cancer Drug Targets. 2011a;11:239–53.
Chen Q, Xie W, Kuhn DJ, Voorhees PM, Lopez-Girona A, Mendy D, Corral LG, Krenitsky VP, Xu W. Moutouh-de Parseval L et al. Targeting the p27 E3 ligase SCF(Skp2) results in p27- and Skp2-mediated cell-cycle arrest and activation of autophagy. Blood. 2008;111:4690–9.
Chen X, Qiu JD, Shi SP, Suo SB, Huang SY, Liang RP. Incorporating key position and amino acid residue features to identify general and species-specific ubiquitin conjugation sites. Bioinformatics. 2013.
Chen Z, Chen YZ, Wang XF, Wang C, Yan RX, Zhang Z. Prediction of ubiquitination sites by using the composition of k-spaced amino acid pairs. PLoS ONE. 2011b;6:e22930.
Chen Z, Zhou Y, Song J, Zhang Z. hCKSAAP_UbSite: Improved prediction of human ubiquitination sites by exploiting amino acid pattern and properties. Biochim Biophys Acta. 2013.
Chen ZJ, Sun LJ. Nonproteolytic functions of ubiquitin in cell signaling. Mol Cell. 2009;33:275–86.
Chernorudskiy AL, Garcia A, Eremin EV, Shorina AS, Kondratieva EV, Gainullin MR. UbiProt: a database of ubiquitylated proteins. BMC Bioinformatics. 2007;8:126.
Ciechanover A. The ubiquitin-proteasome proteolytic pathway. Cell. 1994;79:13–21.
Cohen P, Tcherpakov M. Will the ubiquitin system furnish as many drug targets as protein kinases? Cell. 2010;143:686–93.
Conrad C, Podolsky MJ, Cusack JC. Antiproteasomal agents in rectal cancer. Anticancer Drugs. 2011;22:341–50.
Coornaert B, Carpentier I, Beyaert R. A20: central gatekeeper in inflammation and immunity. J Biol Chem. 2009;284:8217–21.
Dahlmann B. Role of proteasomes in disease. BMC Biochem. 2007;8 Suppl 1:S3.
Danielsen JM, Sylvestersen KB, Bekker-Jensen S, Szklarczyk D, Poulsen JW, Horn H, Jensen LJ, Mailand N, Nielsen ML. Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. Mol Cell Proteomics. 2011;10:M110 003590.
Day IN, Thompson RJ. UCHL1 (PGP 9.5): neuronal biomarker and ubiquitin system protein. Prog Neurobiol. 2010;90:327–62.
Deng H, Liang H, Jankovic J. F-box only protein 7 gene in parkinsonian-pyramidal disease. JAMA Neurol. 2013;70:20–4.
Denis NJ, Vasilescu J, Lambert JP, Smith JC, Figeys D. Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. Proteomics. 2007;7:868–74.
Deshaies RJ, Joazeiro CA. RING domain E3 ubiquitin ligases. Annu Rev Biochem. 2009;78:399–434.
Du H, Huang X, Wang S, Wu Y, Xu W, Li M. PSMA7, a potential biomarker of diseases. Protein Pept Lett. 2009a;16:486–9.
Du Y, Xu N, Lu M, Li T. hUbiquitome: a database of experimentally verified ubiquitination cascades in humans. Database (Oxford). 2011;2011:bar055.

Du Z, Zhou X, Li L, Su Z. plantsUPS: a database of plants' ubiquitin proteasome system. BMC Genomics. 2009b;10:227.
Duncan K, Schafer G, Vava A, Parker MI, Zerbini LF. Targeting neddylation in cancer therapy. Future Oncol. 2012;8:1461–70.
Escobar M, Velez M, Belalcazar A, Santos ES, Raez LE. The role of proteasome inhibition in nonsmall cell lung cancer. J Biomed Biotechnol. 2011;2011:806506.
Festa RA, McAllister F, Pearce MJ, Mintseris J, Burns KE, Gygi SP, Darwin KH. Prokaryotic ubiquitin-like protein (Pup) proteome of Mycobacterium tuberculosis [corrected]. PLoS ONE. 2010;5:e8589.
Flick K, Ouni I, Wohlschlegel JA, Capati C, McDonald WH, Yates JR, Kaiser P. Proteolysis-independent regulation of the transcription factor Met4 by a single Lys 48-linked ubiquitin chain. Nat Cell Biol. 2004;6:634–41.
Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, et al. COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. Nucleic Acids Res. 2011;39:D945–50.
Gao T, Liu Z, Wang Y, Cheng H, Yang Q, Guo A, Ren J, Xue Y. UUCD: a family-based database of ubiquitin and ubiquitin-like conjugation. Nucleic Acids Res. 2013;41:D445–51.
Geng F, Wenzel S, Tansey WP. Ubiquitin and proteasomes in transcription. Annu Rev Biochem. 2012;81:177–201.
Giannakopoulos NV, Luo JK, Papov V, Zou W, Lenschow DJ, Jacobs BS, Borden EC, Li J, Virgin HW, Zhang DE. Proteomic identification of proteins conjugated to ISG15 in mouse and human cells. Biochem Biophys Res Commun. 2005;336:496–506.
Han Y, Lee H, Park JC, Yi GS. E3Net: a system for exploring E3-mediated regulatory networks of cellular functions. Mol Cell Proteomics. 2012;11:O111 014076.
Hegde AN, Upadhya SC. The ubiquitin-proteasome pathway in health and disease of the nervous system. Trends Neurosci. 2007;30:587–95.
Hicke L. Protein regulation by monoubiquitin. Nat Rev Mol Cell Biol. 2001;2:195–201.
Hochrainer K, Lipp J. Ubiquitylation within signaling pathways in- and outside of inflammation. Thromb Haemost. 2007;97:370–7.
Hochstrasser M. Origin and function of ubiquitin-like proteins. Nature. 2009;458:422–9.
House CM, Hancock NC, Moller A, Cromer BA, Fedorov V, Bowtell DD, Parker MW, Polekhina G. Elucidation of the substrate binding site of Siah ubiquitin ligase. Structure. 2006;14:695–701.
Humbard MA, Miranda HV, Lim JM, Krause DJ, Pritz JR, Zhou G, Chen S, Wells L, Maupin-Furlow JA. Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii. Nature. 2010;463:54–60.
Hutchins AP, Liu S, Diez D, Miranda-Saavedra D. The repertoires of ubiquitinating and deubiquitinating enzymes in eukaryotic genomes. Mol Biol Evol. 2013;30:1172–87.
Irminger-Finger I. BARD1, a possible biomarker for breast and ovarian cancer. Gynecol Oncol. 2010;117:211–5.
Iyer LM, Burroughs AM, Aravind L. The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. Genome Biol. 2006;7:R60.
Jeram SM, Srikumar T, Pedrioli PG, Raught B. Using mass spectrometry to identify ubiquitin and ubiquitin-like protein conjugation sites. Proteomics. 2009;9:922–34.
Jones J, Wu K, Yang Y, Guerrero C, Nillegoda N, Pan ZQ, Huang L. A targeted proteomic analysis of the ubiquitin-like modifier nedd8 and associated proteins. J Proteome Res. 2008;7:1274–87.
Kaelin WG. Von Hippel-Lindau disease. Annu Rev Pathol. 2007;2:145–73.
Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40:D109–14.

Kang C, Yi GS. Identification of ubiquitin/ubiquitin-like protein modification from tandem mass spectra with various PTMs. BMC Bioinform. 2011;12 Suppl 14:S8.

Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ, et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell. 2011;44:325–40.

Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res. 2011;39:D1035–41.

Lee KA, Hammerle LP, Andrews PS, Stokes MP, Mustelin T, Silva JC, Black RA, Doedens JR. Ubiquitin ligase substrate identification through quantitative proteomics at both the protein and peptide levels. J Biol Chem. 2011a;286:41530–8.

Lee TY, Chen SA, Hung HY, Ou YY. Incorporating distant sequence features and radial basis function networks to identify ubiquitin conjugation sites. PLoS ONE. 2011b;6:e17331.

Lee WC, Lee M, Jung JW, Kim KP, Kim D. SCUD: Saccharomyces cerevisiae ubiquitination database. BMC Genomics. 2008;9:440.

Lehman NL. The ubiquitin proteasome system in neuropathology. Acta Neuropathol. 2009;118:329–47.

Li H, Xing X, Ding G, Li Q, Wang C, Xie L, Zeng R, Li Y. SysPTM: a systematic resource for proteomic research on post-translational modifications. Mol Cell Proteomics. 2009;8:1839–49.

Lin H, Lin Q, Liu M, Lin Y, Wang X, Chen H, Xia Z, Lu B, Ding F, Wu Q et al. PKA/Smurf1 signaling-mediated stabilization of Nur77 is required for anticancer drug cisplatin-induced apoptosis. Oncogene. 2013.

Linehan WM, Bratslavsky G, Pinto PA, Schmidt LS, Neckers L, Bottaro DP, Srinivasan R. Molecular diagnosis and therapy of kidney cancer. Annu Rev Med. 2010;61:329–43.

Liu Z, Ma Q, Cao J, Gao X, Ren J, Xue Y. GPS-PUP: computational prediction of pupylation sites in prokaryotic proteins. Mol BioSyst. 2011;7:2737–40.

Liu Z, Yuan F, Ren J, Cao J, Zhou Y, Yang Q, Xue Y. GPS-ARM: computational analysis of the APC/C recognition motif by predicting D-boxes and KEN-boxes. PLoS ONE. 2012;7:e34370.

Lu CT, Huang KY, Su MG, Lee TY, Bretana NA, Chang WC, Chen YJ, Huang HD. DbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. Nucleic Acids Res. 2013;41:D295–305.

Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011;2011:bar009.

Mandel SA, Fishman-Jacob T, Youdim MB. Modeling sporadic Parkinson's disease by silencing the ubiquitin E3 ligase component, SKP1A. Parkinsonism Relat Disord. 2009;15(Suppl 3):S148–51.

Maor R, Jones A, Nuhse TS, Studholme DJ, Peck SC, Shirasu K. Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol Cell Proteomics. 2007;6:601–10.

Meierhofer D, Wang X, Huang L, Kaiser P. Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. J Proteome Res. 2008;7:4566–76.

Melino G, Gallagher E, Aqeilan RI, Knight R, Peschiaroli A, Rossi M, Scialpi F, Malatesta M, Zocchi L, Browne G, et al. Itch: a HECT-type E3 ligase regulating immunity, skin and cancer. Cell Death Differ. 2008;15:1103–12.

Muller S, Briand JP, Van Regenmortel MH. Presence of antibodies to ubiquitin during the autoimmune response associated with systemic lupus erythematosus. Proc Natl Acad Sci U S A. 1988;85:8176–80.

Nakajima H, Fujiwara H, Furuichi Y, Tanaka K, Shimbara N. A novel small-molecule inhibitor of NF-kappaB signaling. Biochem Biophys Res Commun. 2008;368:1007–13.

Newton K, Matsumoto ML, Wertz IE, Kirkpatrick DS, Lill JR, Tan J, Dugger D, Gordon N, Sidhu SS, Fellouse FA, et al. Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. Cell. 2008;134:668–78.

Ohki Y, Funatsu N, Konishi N, Chiba T. The mechanism of poly-NEDD8 chain formation in vitro. Biochem Biophys Res Commun. 2009;381:443–7.
Orlicky S, Tang X, Neduva V, Elowe N, Brown ED, Sicheri F, Tyers M. An allosteric inhibitor of substrate recognition by the SCF(Cdc4) ubiquitin ligase. Nat Biotechnol. 2010;28:733–7.
Oshikawa K, Matsumoto M, Oyamada K, Nakayama KI. Proteome-wide identification of ubiquitylation sites by conjugation of engineered lysine-less ubiquitin. J Proteome Res. 2012;11:796–807.
Osula O, Swatkoski S, Cotter RJ. Identification of protein SUMOylation sites by mass spectrometry using combined microwave-assisted aspartic acid cleavage and tryptic digestion. J Mass Spectrom. 2012;47:644–54.
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;21:921–6.
Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebl MG, Iakoucheva LM. Identification, analysis, and prediction of protein ubiquitination sites. Proteins. 2010;78:365–80.
Rape M, Reddy SK, Kirschner MW. The processivity of multiubiquitination by the APC determines the order of substrate degradation. Cell. 2006;124:89–103.
Ren J, Gao X, Jin C, Zhu M, Wang X, Shaw A, Wen L, Yao X, Xue Y. Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. Proteomics. 2009;9:3409–12.
Rufini A, Fortuni S, Arcuri G, Condo I, Serio D, Incani O, Malisan F, Ventura N, Testi R. Preventing the ubiquitin-proteasome-dependent degradation of frataxin, the protein defective in Friedreich's ataxia. Hum Mol Genet. 2011;20:1253–61.
Sadowski M, Sarcevic B. Mechanisms of mono- and poly-ubiquitination: Ubiquitination specificity depends on compatibility between the E2 catalytic core and amino acid residues proximal to the lysine. Cell Div. 2010;5:19.
Scheffner M, Staub O. HECT E3 s and human disease. BMC Biochem. 2007;8(Suppl 1):S6.
Shi Y, Chan DW, Jung SY, Malovannaya A, Wang Y, Qin J. A data set of human endogenous protein ubiquitination sites. Mol Cell Proteomics. 2011;10:M110 002089.
Shi Y, Xu P, Qin J. Ubiquitinated proteome: ready for global? Mol Cell Proteomics. 2011;10:R110 006882.
Singhal S, Taylor MC, Baker RT. Deubiquitylating enzymes and disease. BMC Biochem. 2008;9(Suppl 1):S3.
Sohns W, van Veen TA, van der Heyden MA. Regulatory roles of the ubiquitin-proteasome system in cardiomyocyte apoptosis. Curr Mol Med. 2010;10:1–13.
Starita LM, Lo RS, Eng JK, von Haller PD, Fields S. Sites of ubiquitin attachment in Saccharomyces cerevisiae. Proteomics. 2012;12:236–40.
Striebel F, Imkamp F, Sutter M, Steiner M, Mamedov A, Weber-Ban E. Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. Nat Struct Mol Biol. 2009;16:647–51.
Teng S, Luo H, Wang L. Predicting protein sumoylation sites from sequence features. Amino Acids. 2012;43:447–55.
Tung CW. PupDB: a database of pupylated proteins. BMC Bioinformatics. 2012;13:40.
Tung CW, Ho SY. Computational identification of ubiquitylation sites from protein sequences. BMC Bioinform. 2008;9:310.
Udeshi ND, Mani DR, Eisenhaure T, Mertins P, Jaffe JD, Clauser KR, Hacohen N, Carr SA. Methods for quantification of in vivo changes in protein ubiquitination following proteasome and deubiquitinase inhibition. Mol Cell Proteomics. 2012;11:148–59.
Udeshi ND, Svinkina T, Mertins P, Kuhn E, Mani DR, Qiao JW, Carr SA. Refined preparation and use of anti-diglycine remnant (K-epsilon-GG) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. Mol Cell Proteomics. 2013;12:825–31.

Ulrich HD. The fast-growing business of SUMO chains. Mol Cell. 2008;32:301–5.
van der Veen AG, Ploegh HL. Ubiquitin-like proteins. Annu Rev Biochem. 2012;81:323–57.
Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, Choudhary C. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. Mol Cell Proteomics. 2011;10:M111 013284.
Wagner SA, Beli P, Weinert BT, Scholz C, Kelstrup CD, Young C, Nielsen ML, Olsen JV, Brakebusch C, Choudhary C. Proteomic analyses reveal divergent ubiquitylation site patterns in murine tissues. Mol Cell Proteomics. 2012;11:1578–85.
Wang D, Xu W, McGrath SC, Patterson C, Neckers L, Cotter RJ. Direct identification of ubiquitination sites on ubiquitin-conjugated CHIP using MALDI mass spectrometry. J Proteome Res. 2005;4:1554–60.
Wang J. Cardiac function and disease: emerging role of small ubiquitin-related modifier. Wiley Interdiscip Rev Syst Biol Med. 2011;3:446–57.
Wong CS, Moller A. Siah: a promising anticancer target. Cancer Res. 2013;73:2400–6.
Xu J, He Y, Qiang B, Yuan J, Peng X, Pan XM. A novel method for high accuracy sumoylation site prediction from protein sequences. BMC Bioinform. 2008;9:8.
Xue Y, Ren J, Gao X, Jin C, Wen L, Yao X. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. Mol Cell Proteomics. 2008;7:1598–608.
Xue Y, Zhou F, Fu C, Xu Y, Yao X. SUMOsp: a web server for sumoylation site prediction. Nucleic Acids Res. 2006;34:W254–7.
Xue Y, Zhou F, Zhu M, Ahmed K, Chen G, Yao X. GPS: a comprehensive www server for phosphorylation sites prediction. Nucleic Acids Res. 2005;33:W184–7.
Yunus AA, Lima CD. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. Nat Struct Mol Biol. 2006;13:491–9.
Zhao C, Collins MN, Hsiang TY, Krug RM. Interferon-induced ISG15 pathway: an ongoing virus-host battle. Trends Microbiol. 2013;21:181–6.
Zhong Q, Gao W, Du F, Wang X. Mule/ARF-BP1, a BH3-only E3 ubiquitin ligase, catalyzes the polyubiquitination of Mcl-1 and regulates apoptosis. Cell. 2005;121:1085–95.