A Data Set of Human Endogenous Protein Ubiquitination Sites*§

Yi Shi‡, Doug W. Chan§, Sung Yun Jung‡§, Anna Malovannaya§, Yi Wang‡§, and Jun Qin‡¶

Lysine ubiquitination is an important and versatile protein post-translational modification. Numerous cellular functions are regulated by ubiquitination, suggesting that extensive numbers of proteins, if not all, are modified with ubiquitin at certain times. However, proteome-wide profiling of ubiquitination sites in the mammalian system is technically challenging. We report the design and characterization of an engineered protein affinity reagent for the isolation of ubiquitinated proteins and the identification of ubiquitination sites with mass spectrometry. This recombinant protein consists of four tandem repeats of ubiquitin-associated domain from UBQLN1 fused to a GST tag. We used this GST-qUBA reagent to isolate polyubiquitinated proteins and identified 294 endogenous ubiquitination sites on 223 proteins from human 293T cells without proteasome inhibitors or overexpression of ubiquitin. Mitochondrial proteins constitute 14.7% of this data set, implicating ubiquitination in a wide range of mitochondrial functions. Molecular & Cellular Proteomics 10: 10.1074/mcp.M110.002089, 1–9, 2011.

Post-translational modification of proteins by ubiquitination at lysine residues plays regulatory roles in a broad spectrum of cellular processes including cell cycle progression, DNA damage, and immune response (1). Deregulation of ubiquitination is linked to many human diseases including cancer and neuronal disorders (2, 3). In the canonical ubiquitination reaction, a substrate is covalently conjugated with ubiquitin by an enzymatic cascade involving ubiquitin-activating enzyme (E1), ubiquitin conjugation enzyme (E2), and a ubiquitin ligase (E3). Conversely, the ubiquitin moiety can be cleaved off from the substrates by deubiquitinases (DUBs). The human genome encodes at least two E1, 53 E2, and ~500 E3 enzymes and more than 100 DUBs (4, 5); such diversity of the ubiquitination system rivals that of the kinome and offers multiple gateways for regulation of ubiquitination dynamics in a cell.

The diversity of the cellular ubiquitination system or ubiquitome is increased by formation of polyubiquitin (poly-Ub) chains with variable lengths and linkages. It is now known that all seven internal lysine residues and the N-terminus of ubiquitin can form poly-Ub chains that may impart different functions. For example, proteasome-mediated protein degradation is a well known consequence of Lys-48-linked poly-Ub modification (6). Other linkages may also lead to protein degradation (7) or may have non-proteolytic consequences. The Lys-63 linkage is best characterized in the inflammatory response where it aids in activation of kinase cascades (8, 9), the N-terminal linear poly-Ub has been shown to activate the IκB kinase (10), and Lys-33 linkage has recently been shown to regulate specific signal transduction in a proteolysis-independent manner (11).

Thorough understanding of this complex regulatory system requires identification of ubiquitinated substrates and, preferably, mapping of their ubiquitination sites, which is technically challenging. The most common approach used by laboratories to date is site-directed mutagenesis of putative lysine targets to infer the ubiquitination site(s). Although this is necessary for determining whether ubiquitination has a significant effect on a particular biology of the protein in question, it does not provide direct chemical evidence of ubiquitination. Errorneous conclusions can be made if mutation of lysine residues affects protein structure, folding, or docking site for E2 and E3 enzymes. Mass spectrometry (MS) has emerged as an indispensable tool for direct measurement of the lysine ubiquitination site(s) that complements mutagenesis studies. The ubiquitin conjugation generates an isopeptide bond between the ε-amine of the modified lysine on the substrate and the C-terminus of ubiquitin. After trypsin cleavage, ubiquitination can be detected by mass spectrometry as a 114.043-Da mass shift (from the Gly-Gly remnant of the ubiquitin C-terminus) on the modified peptides. The primary limitation in proteome-wide identification of ubiquitination sites is the lack of high affinity reagents for isolation of ubiquitinated peptides. Additionally, there are three confounding factors that limit our

From the 1Center for Molecular Discovery, Verna and Mars McLean Department of Biochemistry and Molecular Biology and 5Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030
Received, June 15, 2010, and in revised form, October 22, 2010
Published, MCP Papers in Press, October 24, 2010, DOI 10.1074/mcp.M110.002089
1 The abbreviations used are: DUB, deubiquitinase; Ub, ubiquitin; UBD, ubiquitin binding domain; UBA, ubiquitin-associated domain; qUBA, four tandem repeats of ubiquitin-associated domain; UBQLN1, ubiquitin-1; FA, formic acid; FDR, false discovery rate; SILAC, stable isotope labeling by amino acids in cell culture; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; MCODE, Molecular Complex Detection.

Molecular & Cellular Proteomics 10.5 10.1074/mcp.M110.002089–1

Want to cite this article? Please look on the last page for the proper citation format.
Eukaryotic cells have evolved protein domain structures, ubiquitin binding domains (UBDs), that can recognize and bind to ubiquitin modifications. More than 20 different families have been identified to date, and most of them bind poly-Ub relatively weakly (17, 18). The ubiquitin-associated domain (UBA) is the first identified UBD and is one of the best poly-Ub binders (19, 20). Although a single UBA domain has been successfully used as an affinity reagent to quantify the poly-Ub chains in a mouse model of Huntington disease (21), it bears moderate ubiquitin binding affinity that may not be sufficient for proteome-wide isolation of ubiquitinated proteins. Recently, several studies showed that tandem UBDs display avidity in poly-Ub binding (22–24). We have utilized this concept to develop and test the suitability of a tandem UBA protein for large scale characterization of ubiquitination substrates.

Here, we report using recombinant GST fusion of four tandem ubiquilin-1 (UBQLN1) UBA domains (GST-qUBA) as a bait to isolate endogenous ubiquitinated proteins from human 293T cells without ubiquitin overexpression and proteasome inhibition. We confirmed that tandem GST-qUBA is more efficient at binding poly-Ub chains than the single GST-UBA. Using GST-qUBA as an affinity purification reagent, we were able to detect ~300 lysine ubiquitination sites that are supported by high quality mass spectra. Both abundant and low abundance cellular proteins including tumor suppressors and regulators of apoptosis and NF-κB pathways were identified. We demonstrate that GST-qUBA is a powerful tool for the affinity purification and identification of endogenous protein ubiquitination substrates at the proteome scale.

**EXPERIMENTAL PROCEDURES**

**DNA Subcloning, Protein Purification, and GST-Protein Immobilization**—GST-UBA was subcloned by PCR amplification of the UBA domain (amino acids 540–589) of UBQLN1 and inserted into the pGEX-4T-1 vector using BamHl/EcoRI sites. To generate GST-qUBA, four repeats of the DNA sequence encoding the UBA domain were synthesized (Genscript) and subcloned into pGEX-4T-1 vector using BamHl/EcoRI sites. To generate GST-qUBA, a similar amount of recombinant proteins (5–10 mg) was used to isolate polyubiquitinated proteins from whole cell protein extracts (~0.4 mg/ml; 1 ml of protein extracts) of either “heavy” or “light” labeled cells. After binding, the beads were washed with NETN buffer and subjected to LC-MS/MS for quantification with the Orbitrap mass spectrometer.

**Sample Preparation for Mass Spectrometric Analysis**—One-half of the eluted proteins were resolved by 4–20% SDS-PAGE, and the gel was sliced into 22–24 bands based on molecular weight and subjected to in-gel trypsin digestion as described previously (25). The other half of the sample was dried in a Savant SpeedVac, reconstituted with 100 mM NH$_4$HCO$_3$, and digested by trypsin (1:50, w/w) at 37 °C overnight. The digested peptides were then centrifuged at 13,000 × g for 5 min, and the supernatant was subjected to isoelectric focusing (IEF) in an Agilent 3100 OFFGEL fractionator (Agilent, G3100AA) according to the manufacturer’s instructions. Peptide mixtures collected from the 12 IEF fractions were desalted and concentrated with C$_{18}$ stop and go extraction tips. Mass spectrometry was performed with an LTQ-Velos-Orbitrap mass spectrometer (Thermo Fisher) equipped with a nanospray ion source and HPLC system. Purified peptides were injected in an in-house-built C$_{18}$ column (75-μm inner diameter) and eluted with a 120-min linear gradient of 95% solvent A (0.1% FA in water) to 35% solvent B (0.1% FA in acetonitrile) with a flow rate of 400 nl/min. Raw data were acquired in a data-dependent mode in which full MS spectra (m/z = 350–1,300) were collected in the Orbitrap with targeted resolution of 100,000, and the top 20 most intense ions were fragmented by collision-induced dissociation (CID) and analyzed in the LTQ-Velos. The maximum ion injection time was set at 30 ms. A total of ~2 × 10$^6$ MS/MS spectra were recorded in which ~70,000 redundant spectra were identified at a 1% false discovery rate (FDR) by Mascot.

**Quantitative MS for Poly-Ub Quantification**—293T cells were cultured in stable isotope-labeled DMEM (Invitrogen; [²H$_6$]lysine and [³H$_6$]arginine) for six generations before harvesting (26). To compare relative ubiquitin binding efficiency between GST-qUBA and GST-UBA, a similar amount of recombinant proteins (5–10 mg) was used to isolate polyubiquitinated proteins from whole cell protein extracts (~0.4 mg/ml; 1 ml of protein extracts) of either “heavy” or “light” labeled cells. After binding, the beads were mixed and washed with NETN buffer. To account for potential difference in SILAC, the pulldown was repeated, but the order of the isotope label was switched. Proteins that bound the beads were resolved by SDS-PAGE, and the gel region above 50 kDa was digested with trypsin and subjected to LC-MS/MS for quantification with the Orbitrap mass spectrometer.
The experimental m/z value and retention time for the SILAC pairs of poly-Ub linkages including Lys-6 (m/z light/H11005 460.594; m/z heavy/H11005 464.607; z = 3; 29–31 min), Lys-11 (m/z light/H11005 801.425; m/z heavy/H11005 805.438; z = 3; 55–58 min), Lys-48 (m/z light/H11005 487.598; m/z heavy/H11005 491.611; z = 3; 38–41 min), and Lys-63 (m/z light/H11005 748.735; m/z heavy/H11005 752.748; z = 3; 57–61 min) were recorded for peak detection. The corresponding MS chromatographic peaks were automatically plotted, and peak intensities were calculated using the built-in area under the curve algorithm in the Xcalibur Qual Browser 2.1 with manual adjustment when the signal/noise ratio was too low.

Database Search of CID Spectra for Ubiquitinated Peptides and Protein Identification—Raw CID spectra were searched against a target-decoy (27) human RefSeq database (June 27, 2009; 39,165 entries) with the Mascot 2.2 (Matrix Science) algorithm embedded in the Proteome Discoverer 1.2 interface (Thermo Fisher). Key parameters used for the search were as follows: precursor mass tolerance was confined within 10 ppm with a fragment mass tolerance of 0.5 dalton and a maximum of two missed cleavages allowed. Dynamic modifications were allowed for lysine ubiquitination (+114.0429 Da) and methionine oxidation; fixed modification was set to carbamidomethylation of cysteine. Identified peptides were filtered with 5% FDR and subjected to rigorous manual verifications as described in the text. Identical parameters were used by Sequest (Thermo Fisher) for ubiquitination identification and compared with Mascot.

Bioinformatics Analysis—A list of 223 substrates was submitted to the Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.7) Bioinformatics Resources (28) for enrichment analysis. For ubiquitination interaction network analysis, the substrate proteins were first searched against the STRING database (29) (version 8.2) to identify protein-protein interaction network. The ubiquitination interaction networks extracted by STRING were then analyzed for densely connected regions with the MCODE algorithm (30). MCODE is a part of the toolkit for network analysis and visualization in the Cytoscape software (31) (version 2.63). Only the top four interaction modules were selected. The relative density map for ubiquitinated lysines was constructed in an in-house-built FileMaker-based application. Briefly, the identified lysine ubiquitination sites and the 6 flanking amino acid residues on either side of the modified lysine were used to generate a position-specific matrix. The ubiquitinated lysine was fixed at position 0. The frequency of amino acid occurrence at each position from −6 to +6 was calculated; enrichment against the natural abundance of amino acids in the human proteins was shown as an intensity map.

RESULTS AND DISCUSSION

Comparison of Poly-Ub Binding Efficiencies of qUBA and UBA with Quantitative Mass Spectrometry—We took advantage of the natural ability of UBA to bind poly-Ub with moderate affinity (19, 20) and made a recombinant affinity reagent for ubiquitin binding with tandem repeats of the UBA domain.
(qUBA), reasoning that it should have higher affinity to poly-Ub. We first compared relative binding efficiency of qUBA and of the single UBA domain to polyubiquitinated proteins from 293T cell lysate. To distinguish poly-Ub isolated with each UBA reagent by MS, we used similar amounts of heavy and light isotope-labeled cell lysates for binding with qUBA and UBA. We then mixed the samples isolated by the different UBAs, resolved them by SDS-PAGE, and in-gel-digested the region that is above 50 kDa with trypsin. The resulting peptides were measured by an LTQ-Velos-Orbitrap mass spectrometer for relative quantification of poly-Ubs. To check for a potential bias of heavy isotope labeling, we repeated the experiment with reversed heavy and light lysates for the UBAs (Fig. 1, A and B).

When a limited amount of cell lysate was used, Lys-6-, Lys-11-, Lys-48-, and Lys-63-linked poly-Ubs were readily detected from the sample pulled down by the GST-qUBA, whereas only the Lys-48 linkage was detectable for GST-UBA, and other Ub linkages were below the detection limit of the mass spectrometer. The enrichment for Lys-48 linkage by qUBA over UBA was ~500-fold as judged by the peak areas, whereas the enrichment for Lys-6, Lys-11, and Lys-63 linkages was estimated to be greater than a 1,000-fold (Fig. 1D). Other poly-Ub linkages (Lys-27, Lys-29, and Lys-33) were not detected in this experiment; however, they were detected when more cell lysate was used. We conclude that qUBA is more efficient in poly-Ub binding than a single UBA, consistent with the conclusion from the previous report using synthetic poly-Ub chains (24).

**Large Scale Identification of High Confidence Endogenous Ubiquitination Sites in 293T Cells Using Automated Search with FDR Threshold Followed by Manual Verification**—Next we sought to utilize qUBA for identification of ubiquitination sites in 293T cells that were not preincubated with proteasome inhibitors or overexpressing exogenous ubiquitin. To minimize the loss of ubiquitination by the DUB activity, both cysteine and serine protease inhibitors were included during cell lysis (32). Ubiquitinated proteins were isolated with the

---

**Fig. 2. Schematic representation of experimental procedure to identify endogenous ubiquitination sites.** Recombinant GST-qUBA is immobilized on beads and mixed with 293T cell lysate for pulldown of endogenous ubiquitinated substrates. Purified proteins are either in-gel-digested with trypsin or in-solution-digested followed by separation by IEF. Purified peptides are subjected to LC-MS/MS. Raw data are searched by Mascot for ubiquitination, and all ubiquitination spectra are manually verified.
GST–qUBA immobilized on GSH-agarose beads, digested in-gel after separation by SDS-PAGE, or digested in solution and further separated by IEF. Extracted tryptic peptides were subjected to LC-MS/MS analysis (Fig. 2).

Two LC-MS/MS runs were searched with Mascot and Sequest to compare various search parameters that may increase the reliability of ubiquitination identification. Using a target-decoy database search with 1 and 5% FDRs, Mascot identified 20 and 50 unique sites, respectively. Among the 20 sites identified at 1% FDR, 19 sites were manually verified, and one site was rejected. In contrast, among the 50 sites identified at 5% FDR, only 27 sites were manually verified, and 23 sites were rejected (Fig. 3). Thus, a 1% FDR search with Mascot may produce some false negative identification, whereas a 5% FDR search led to high levels of false positive identifications. Under the same FDR conditions, Sequest was able to identify more ubiquitination sites than Mascot; however, many of them could not be manually verified (Fig. 3). Potential differences exist between these two popular search engines as only 18 spectra were identified by both algorithms at 1% FDR (Fig. 3).

Theoretically, one can combine the results obtained by the two different algorithms: in our case, only 25 of the 46 combined spectra identified by Sequest and Mascot searched with a 1% FDR passed our manual verifications (Fig. 3D). Thus, it is unsafe to gauge the validity of ubiquitination site identification using a 1% FDR as the sole determinant. We conclude that manual verification is the most reliable means for the identification of ubiquitination sites, and the FDR alone might not be accurate enough to estimate the overall reliability.

Identification of 294 Unique Lysine Ubiquitination Sites in 293T Cells without Proteasome Inhibition or Ub Overexpression—A total of ∼70,000 redundant spectra were identified at a 1% FDR by Mascot in three replicate experiments in which 7.5% of the spectra were from ubiquitin itself. To identify “provisional” ubiquitinated peptides, we used the Mascot search engine with a 5% FDR and then subjected spectral matches to vigorous manual verification. Several criteria for manual verification were used to eliminate false positive assignments. First, in the correct assignment, major MS2 fragment peaks must be unequivocally assigned and follow a pattern of a continual stretch of y and b ions and the appearance of dominant fragment ions N-terminal to proline and C-terminal to aspartic acid for arginine-containing peptides (33–35). Second, only when both the ubiquitinated and unmodified peptides of the same protein are recovered is the ubiquitinated peptide assignment accepted. Lastly, we found that mass accuracy of the full mass spectra provides a good constraint on the fidelity of the ubiquitination assignment where good assignments that pass manual verification are within a ±2-ppm window from the median value of the mass accuracy distribution in the LC-MS run (supplemental Fig. 1). Thus, we rejected assignments that fall outside of this window.

A total of 294 endogenous ubiquitination sites on 223 substrates were identified and manually verified (supplemental Table 1) of which >95% of the ubiquitination sites...
were not previously reported in the Swiss-Prot database. Greater than 96% of the ubiquitinated peptides are either doubly or triply charged, and less than 4% of peptides with 4+ and 5+ charges were assigned with sufficient confidence (supplemental Fig. 2A). It is noteworthy that ubiquitination on internal, but not C-terminal, lysine was detected in our study, indicating that trypsin is indeed very specific and only cleaves the unmodified lysine residue. The 292 unique ubiquitination spectra (containing 294 lysine ubiquitination sites) were annotated and combined into a spectral library file provided as supplemental Fig. 5. Two substrates were randomly selected for biochemical validation (supplemental Fig. 2B). Ubiquitinated substrates were isolated by GST-qUBA from human 293T cells and immunoblotted with antibodies against Diablo or α-catenin. In agreement with MS identification, the detection of smears of these proteins on the SDS-PAGE gel is consistent with ubiquitination and further supports the conclusion that GST-qUBA specifically enriches the polyubiquitinated forms of these proteins.

Ubiquitination on seven internal lysines as well as Lys-6/ Lys-11 double linkage (or “forked structure”) was identified (supplemental Table 1), highlighting the complexity of poly-Ub topology in vivo. The N-terminal linear Ub linkage was not detected. Poly-Ub linkages on all internal lysine residues were abundantly detected on most regions of the SDS-PAGE gel, suggesting that modification of proteins with Ub linkages other than Lys-48 is widespread in the cells.

Motif-X analysis (36) identified a putative ubiquitination motif of KXXLXD (where X denotes any amino acid) (supplemental Fig. 3A). This sequence appears over 4,000 times in human proteome. Strikingly, ubiquitin itself contains this sequence at Lys-48. Although the appearance of this sequence motif is statistically significant, further tests will show whether it is indeed a functional motif used by cells to select the ubiquitination site to build a Lys-48 Ub linkage for protein degradation. We also constructed a density map to visualize distribution and overall enrichment of amino acid residues surrounding the ubiquitinated lysine (supplemental Fig. 3C).

Because both ubiquitination and acetylation are the major forms of lysine modification, we compared our data set with the recently published lysine acetylation sites (37) and found

---

Fig. 4. Diagrams that illustrate enrichment of ubiquitination sites in apoptotic and NF-κB pathways. A, ubiquitination sites on several key components in apoptosis pathways. Ubiquitinated lysine(s) identified in this work is marked on the substrate protein labeled by the gene symbol. B, enrichment of ubiquitination sites on the NF-κB pathway. XIAP, X-linked inhibitor of apoptosis protein; CYCS, cytochrome c, somatic; CASP, caspase; IKK, IκB kinase; FasL, Fas ligand.
### Table I

List of ubiquitination sites on mitochondrion-localized proteins

| No. | Protein name                                      | Protein gi no. | Lysine Ub site(s) | Localization       | Function                     |
|-----|---------------------------------------------------|----------------|-------------------|--------------------|------------------------------|
| 1   | Apoptosis-inducing factor, mitochondrion-associated, 1 | gi22202629     | Lys-109           | Envelope           | Apoptosis                    |
| 2   | Lipoyltransferase 1                               | gi21729878     | Lys-150           | Envelope/outer     | Lipid metabolism             |
| 3   | Caspase 8, apoptosis-related cysteine peptidase   | gi15718712     | Lys-148           | Envelope/outer     | Apoptosis                    |
| 4   | ATP-binding cassette, subfamily B (MDR/TAP), mem 6| gi9955963      | Lys-482           | Envelope/outer     | Mitochondrial respiratory chain |
| 5   | Hexokinase 2                                      | gi15553127     | Lys-62, Lys-323   | Envelope/outer     | Permeability transition pore |
| 6   | ATP synthase, H⁺-transporting, mitochondrial F1 complex, O subunit | gi4502303 | Lys-60            | Inner/envelope     | Mitochondrial respiratory chain |
| 7   | Cytochrome c oxidase subunit IV isoform 1         | gi4502981      | Lys-149           | Inner/envelope     | Mitochondrial respiratory chain |
| 8   | NADH dehydrogenase (ubiquinone) 1α subcomplex 4, 9 kDa | gi4505357     | Lys-73            | Inner/envelope     | Mitochondrial respiratory chain |
| 9   | NADH dehydrogenase (ubiquinone) Fe-S protein 5    | gi4758790      | Lys-28            | Inner/envelope     | Mitochondrial respiratory chain |
| 10  | Mitochondrial ribosomal protein L50               | gi21265096     | Lys-34            | Inner/envelope     | Mitochondrial ribosomes       |
| 11  | Mitochondrial ribosomal protein L45               | gi34304322     | Lys-71, Lys-102   | Inner/envelope     | Mitochondrial ribosomes       |
| 12  | Mitochondrial ribosomal protein L15               | gi7661806      | Lys-28            | Inner/envelope     | Mitochondrial ribosomes       |
| 13  | Translocase of inner mitochondrial membrane 8 homolog A (yeast) | gi4758152 | Lys-86, Lys-88    | Inner/envelope     | Permeability transition pore |
| 14  | Translocase of inner mitochondrial membrane 8 homolog B (yeast) | gi256773260 | Lys-80            | Inner/envelope     | Permeability transition pore |
| 15  | Cold shock domain-containing E1, RNA-binding      | gi56117850     | Lys-246           | Inner/matrix       | Virus infection               |
| 16  | Single-stranded DNA binding protein 1             | gi4507231      | Lys-51            | Inner/matrix       | Mitochondrial DNA replication |
| 17  | Ferrochelatase (protoporphyrila)                  | gi60499021     | Lys-57            | Inner/matrix/envelope | Heme synthesis               |
| 18  | Chaperonin                                        | gi31542947     | Lys-58, Lys-72, Lys-75 | Inner/matrix/envelope | Protein folding              |
| 19  | Diablo homolog (Drosophila)                       | gi21070976     | Lys-79, Lys-147   | Matrix/envelope     | Antia apoptosis               |
| 20  | Transcription factor A, mitochondrial precursor (Homo sapiens) | gi4507401 | Lys-76            | matrix             | Mitochondrial gene transcriptions |
| 21  | Cytochrome c, somatic                             | gi11128019     | Lys-40            | Matrix/envelope     | Apoptosis                    |
| 22  | Solute carrier family 25, member 5               | gi156071459    | Lys-33, Lys-63, Lys-147, Lys-272, Lys-295 | Matrix/envelope | ATP transport               |
| 23  | Citrate synthase                                  | gi38327625     | Lys-382           | Matrix/lumen       | Citric acid cycle             |
| 24  | Pyruvate dehydrogenase phosphatase catalytic subunit 2 | gi41349497 | Lys-459           | Matrix/lumen       | Mitochondrial respiratory chain |
| 25  | YME1-like 1 (Saccharomyces cerevisiae)            | gi7657689      | Lys-180           | N/A                | Mitochondrial protein metabolism |
| 26  | Ceroid-lipofuscinosis, neuronal 3                 | gi109698601    | Lys-262           | N/A                | Mitochondrial respiratory chain |
| 27  | Iron-sulfur cluster scaffold homolog (Escherichia coli) | gi24307953 | Lys-82            | N/A                | Mitochondrial respiratory chain |
| 28  | LYR motif-containing 4                            | gi38570054     | Lys-58            | N/A                | Mitochondrial respiratory chain |
| 29  | Heat shock 70-kDa protein 1A; heat shock 70-kDa protein 1B | gi167466173 | Lys-88, Lys-348   | N/A                | Protein folding              |
| 30  | Wolf-Hirschhorn syndrome candidate 1-like 1       | gi13699111     | Lys-1065          | N/A                | N/A                          |
| 31  | Tumor suppressor candidate 3                      | gi30410790     | Lys-180           | N/A                | N/A                          |
| 32  | Alkylglycerone-phosphate synthase                 | gi4501993      | Lys-108           | N/A                | N/A                          |
| 33  | KIAA0101                                         | gi71773819     | Lys-24            | N/A                | N/A                          |
that ∼15% (43 sites) of the ubiquitinated lysines are also acetylated. Such coincidence of ubiquitination and acetylation on the same lysine suggests a competitive relationship between the two modification pathways.

Abundant proteins that belong to protein translational factors, the Ub-proteasome system, cytoskeleton, heat shock proteins, RNA splicing, and histones are well represented in our dataset (supplemental Fig. 4 and Table 1). Importantly, low abundance proteins including a cell cycle regulator (cyclin B1), transcription factor (Id2), tumor suppressor (β-catenin), and oncprotein (c-Myc) were also detected (supplemental Table 1). The identification of low abundance proteins that are involved in tumorigenesis is particularly noteworthy. It is possible that these regulatory proteins are tightly regulated by the ubiquitination system for the maintenance of cellular homeostasis.

Approximately 20% of the substrates in our data set have been linked to various human diseases (supplemental Table 2). For example, ubiquitination sites of CASP8, a proapoptosis protein involved in multiple cancer types (38, 39), and UBE3A, whose mutations have been linked to Angelman syndrome (40), are among the substrates we identified. These observations further support that ubiquitination pathways play important roles in regulating various cell functions and that deregulation of these pathways may lead to human disease.

Ubiquitination Is Enriched in Apoptosis, NF-κB Pathway, and Mitochondria—As shown in Fig. 4A, ubiquitination sites for several critical components of both intrinsic and extrinsic apoptosis pathways were identified. There are three critical decision points in the intrinsic apoptosis pathway: 1) the release of cytochrome c, somatic (CYCS) to active the apoptosome (APAF1/CASP9), 2) the relief of inhibition by SMAC/DIABLO of X-linked inhibitor of apoptosis protein on caspase 3, and 3) the translocation of AIFM1 to the nucleus for DNA fragmentation (41). Surprisingly, ubiquitination sites on proteins in all three pathway components were mapped. Similarly, ubiquitination sites on the upstream regulators caspase 8 and protein kinase A that mediate apoptosis in response to survival and growth cues in the extrinsic pathway were also identified.

Proteins in the NF-κB pathway are well represented in the data set. As shown in Fig. 4B, multiple components proximal to the TNFα receptor are ubiquitinated at multiple lysine residues including several well known ubiquitin E3 ligases of both positive and negative regulators in the pathway; an E2 enzyme (UBE2N), which specifically processes a Lys-63 poly-Ub chain; and a DUB (OTUD5). Together, these observations highlight the importance of ubiquitination in the regulation of the NF-κB pathway and activation of the immune response (8). Intriguingly, most of the substrates identified in the NF-κB pathway are themselves enzymes that play regulatory roles in ubiquitination reaction, thereby demonstrating a delicate self-controlled system in this pathway.

Because 14.7% of the identified substrates (33 proteins) are mitochondrial, our data also reveal possible roles of ubiquitination in a wide range of mitochondrial functions. These proteins localize to inner and outer membranes as well as matrix of the mitochondria (Table I). Because the mitochondrion is well known as the hub of apoptosis (41) in which ubiquitination plays an essential role, we speculate that some substrates identified here regulate apoptosis. Our data raise several interesting questions regarding the mitochondrial ubiquitination reaction. For example, do the ubiquitination events happen inside or outside of mitochondria? Will there be any functional difference or poly-Ub linkage preferences of mitochondrial ubiquitination versus ubiquitination elsewhere in the cell? Answering these questions with further biological analysis will expand our understanding of mitochondrial functions.

In summary, we have provided evidence that the GST-qUBA binds poly-Ub chains more efficiently than GST-UBA and can be used as an affinity reagent to isolate ubiquitinated substrates. Using a combination of FDR and manual verification, we were able to map 294 endogenous lysine ubiquitination sites that are supported by high quality MS data. Ubiquitination sites on both abundant and scarce regulatory proteins were identified. Our ubiquitination site data set contains a significant number of substrates that are localized to the mitochondria, implicating a role of ubiquitination in a wide range of mitochondrial functions. Overall, our data set thus provides direct and accurate chemical evidence for ubiquitination on the substrates, and the information builds a solid foundation for future functional analysis of these substrates. Because this method does not rely on Ub overexpression or proteasome inhibition, it will find applications for ubiquitination profiling in physiological or pathological samples. Although this method is very efficient for isolation of ubiquitinated substrates, it does not directly enrich ubiquitinated peptides and is less than optimal for mapping ubiquitination sites. Affinity reagents that specifically recognize and enrich tryptic peptides containing a Gly-Gly isopeptide bond could be more useful for this purpose. We envision that the combination of qUBA pulldown followed by enrichment with the recently described ubiquitin antibody (42) will allow for an even deeper exploration of the ubiquitome.

Acknowledgments—We thank Drs. Zhou Songyang and Ron Kopito for providing plasmids and Drs. Jianhua Yang and Yaroslava A. Bulynko for critical reading of the manuscript and helpful suggestions.

* This work was supported, in whole or in part, by National Institutes of Health Grants CA84199 (to J. Q.) and GM080703 (to Y. W.). This work was also supported by the Welch Foundation (to J. Q.).

† This article contains supplemental Figs. 1–5 and Tables 1 and 2.
REFERENCES

1. Schwartz, A. L., and Ciechanover, A. (1999) The ubiquitin-proteasome pathway and pathogenesis of human diseases. Annu. Rev. Med. 50, 57–74
2. Dikic, I., Crosetto, N., Calatroni, S., and Bernasconi, P. (2006) Targeting ubiquitin in cancer. Eur. J. Cancer 42, 3095–3102
3. Shimura, H., Hattori, N., Kubo, S., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K., and Suzuki, T. (2000) Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nat. Genet. 25, 302–305
4. Jin, J., Li, X., Gygi, S. P., and Harper, J. W. (2007) Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. Nature 447, 1135–1138
5. Semple, C. A. (2003) The comparative proteomics of ubiquitination in mouse. Genome Res. 13, 1389–1394
6. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K., and Varshavsky, A. (1989) A ubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science 243, 1576–1583
7. Xu, P., Duong, D. M., Seyfried, N. T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M., Finley, D., and Peng, J. (2009) Quantitative proteomics reveals the function of unconventional ubiquitin chains in posttranslational degradation. Cell 137, 133–145
8. Skaug, B., Jiang, X., and Chen, Z. J. (2009) The role of ubiquitin in NF-kappaB regulatory pathways. Annu. Rev. Biochem. 78, 789–796
9. Xia, Z. P., Sun, L., Chen, X., Pineda, G., Jiang, X., Adhikari, A., Zeng, W., Cao, Y., Takao, T., Tanaka, K., and Iwai, K. (2009) Involvement of linear polyubiquitination of NEMO in NF-kappaB activation. Nat. Cell Biol. 11, 123–132
10. Huang, H., Jeon, M. S., Liao, X., Yang, C., Elly, C., Yates, J. R., 3rd, and Liu, Y. C. (2010) K33-linked polyubiquitination of T cell receptor-zeta regulates proteolysis-independent T cell signaling. Immunity 33, 60–70
11. Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Yue, B., Zhang, J., Crosetto, N., Calatroni, S., Bernasconi, P., and Cohen, R. E. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57
12. Cooper, E. M., Cutcliffe, G., Kristiansen, T. Z., Pabst, A., Pickart, C. M., and Cohen, R. E. (2009) K63-specific deubiquitination by two JAMM/MPN+ complexes: BRISC-associated Brcc36 and proteasomal Poh1. EMBO J. 28, 621–631
13. Tabb, D. L., Smith, L. L., Breci, L. A., Wysocki, V. H., Lin, D., and Yates, J. R., 3rd (2003) Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. Anal. Chem. 75, 1155–1163
14. Medzhiradszky, K. F. (2005) Peptide sequence analysis. Methods Enzymol. 402, 209–244
15. Crosetto, N., Calatroni, S., and Bernasconi, P. (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivo cross-linking. Mol. Cell. Proteomics 5, 737–748
16. Tagwerker, C., Flick, K., Oul, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huppa, L., and Kaiser, P. (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivo cross-linking. Mol. Cell. Proteomics 5, 737–748
17. Meierhofer, D., Wang, X., Huang, L., and Kaiser, P. (2008) Quantitative approach for understanding protein ubiquitination. Anal. Chem. 80, 4161–4169
18. Chipic, L., Schubert, H. L., and Hill, C. P. (2005) Ubiquitin-binding domains. Nat. Rev. Mol. Cell Biol. 6, 610–621
19. Dikic, I., Wakatsuki, S., and Walters, K. J. (2009) Ubiquitin-binding domains—from structures to functions. Nat. Rev. Mol. Cell Biol. 10, 659–671
20. Raasi, S., Yaradan, R., Fushman, D., and Pickart, C. M. (2005) Diverse polyubiquitin interaction properties of ubiquitin-associated domains. Nat. Struct. Mol. Biol. 12, 708–714
21. Raasi, S., Orlov, I., Fleming, K. G., and Pickart, C. M. (2004) Binding of polyubiquitin chains to ubiquitin-associated (UBA) domains of HHR23A. J. Mol. Biol. 341, 1367–1379
22. Bennett, E. J., Shaler, T. A., Woodman, B., Ryu, K. Y., Zaitseva, T. S., Becker, C. H., Bates, G. P., Schulman, H., and Kopito, R. R. (2007) Global changes to the ubiquitin system in Huntington’s disease. Nature 448, 704–708
23. Xu, G., Paige, J. S., and Jaffrey, S. R. (2010) Global analysis of lysine ubiquitination by ubiquitin remnant immunoadfinity profiling. Nat. Biotechnol. 28, 868–873
In order to cite this article properly, please include all of the following information: Shi, Y., Chan, D. W., Jung, S. Y., Malovannaya, A., Wang, Y., and Qin, J. (2011) A Data Set of Human Endogenous Protein Ubiquitination Sites. *Mol. Cell. Proteomics* 10(5):M110.002089. DOI: 10.1074/mcp.M110.002089.