detach from the culture dish, washed with PBS twice and incubated with 1 μg/ml of anti-HE4 antibodies or isotype-matched control for 30 min at 4°C in PBS containing 0.5% BSA and 2 mM EDTA. Following washing with the above buffer twice, PE conjugate antimouse IgG (MBL, Nagoya, Japan) for monoclonal antibody and PE-conjugated anti-rabbit polyclonal antibody (MBL) were added and further incubated for 30 min at 4°C. All flow cytometry was performed on Cytomics FC500 (Beckman Coulter, Fullerton, CA).

Immunoprecipitation and Western blot

The reactivity of anti-HE4 antibodies to recombinant HE4 protein was confirmed by immunoprecipitation. Fifteen microliters of Protein G sepharose suspended in PBS containing 0.01% BSA (Sigma) was incubated with 5 μg of anti-HE4 antibodies for 2 h at 4°C with gentle rocking. During this step, 250 ng of the recombinant myc-His-tagged HE4 protein was incubated with Protein G beads for 30 min at 4°C with shaking to preclear the samples. The Protein G Sepharose incubated with the antibodies were centrifuged at 1,000*g* for 2 min and washed with PBS three times. Then, the precleared samples were added to the tube containing the washed Protein G sepharose and rotated overnight at 4°C. After the incubation, the beads were washed with PBS three times and boiled in 25 μl of 2× Laemmli's SDS sample buffer for 5 min. Proteins (20 μl of sample per lane) were separated by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on a 12.5% polyacrylamide gel and electrotransfered to a polyvinylidenefluoride (PVDF) membrane. The membrane blocked with 5% nonfat milk in PBS containing 0.05% Tween 20 (blocking buffer) was incubated with 1.0 μg/ml mouse monoclonal anti-Myc antibody (MBL) to react with the precipitated Myc-His-tagged HE4 for 1 h at room temperature. After three washes with PBS containing 0.05% Tween 20, the membrane was incubated with a horseradish peroxidase (HRP)-conjugated antimouse IgG (MBL) diluted 1:5,000 with the blocking buffer. Chemiluminescence was developed according to the manufacturer's procedure (ECL; GE Healthcare).

Sandwich ELISA

The concentration of HE4 in culture media of cancer cell lines and donor sera was measured by a HE4-specific sandwich ELISA constructed as follows: 96-well microtiter plates (Nalge Nunc International Corp., Rochester, NY) were coated with the capturing antibody clone 110-108 with carbonate buffer at 4°C overnight. The plates were blocked with 200 μl PBS containing 1.0% BSA for 2 h and then incubated for 1 h with culture media diluted to 1:2 with sample diluent which consists of PBS containing 1.0% BSA and 0.1% Tween 20 or serum samples diluted to 1:10 with the same diluent and HRP-conjugated antibody 128-119 diluted to 1:140,000 with PBS containing BSA. After washing the plates with PBS containing Tween20, 100 μl/well TMB (Moss Inc., Pasadena, MD) was added, and the plates were incubated for 30 min at room temperature. The color development was stopped by the addition of H_2SO_4. Color intensity was determined at a wavelength of 450 nm with a reference wavelength of 620 nm. Analyte concentrations were calculated by referring to the standard curve using serial diluted recombinant HE4 (Fig. 1d).

Immunohistochemistry

Paraffin-embedded cancer tissue slices or noncancer tissue slices derived from lung cancer patients (adenocarcinoma and squamous cell carcinoma) were purchased from Outdo, Shanghai, China. The tissue slices were deparaffinized by treatment with xylene for 5 min three times, 100% ethanol for 5 min twice, 90% ethanol for 5 min once, 80% ethanol for 5 min once, 70% ethanol for 5 min once and PBS for 5 min three times. Subsequently, for an antigen retrieval, the specimens were immersed in a citrate buffer and heated by microwave for 10 min twice. In order to inactive the endogenous peroxidase activity, the specimens were then treated with PBS containing a 3% hydrogen peroxide solution at room temperature for 10 min. After washing with PBS twice, the specimens were blocked with blocking buffer and incubated with a blocking buffer containing 1 μg/ml of polyclonal HE4 antibody for 1 h at room temperature. After washing with PBS twice, the specimens were incubated with HRP-conjugated second antibody (EnVision Dual Link, Dako, Denmark) for 1 h at room temperature. Subsequently, the specimens were washed with PBS twice and allowed to react with a DAB chromogen (Dako) for 10 min at room temperature. The reaction was stopped by washing with water. After counterstaining with hematoxylene, the tissue slices were dehydrated with ethanol and xylene and made into specimens using a mounting medium (Matsunami Glass, Osaka, Japan).

Statistical analysis

To test for statistically significant differences between two groups, an unpaired Student's *t*-test was used. For comparisons among three or more groups, the values were analyzed by one-way ANOVA followed by Scheffe's post hoc comparisons. Differences were considered significant at $P<0.05$. For drawing of receiver operating characteristic (ROC) curves and estimation of the area under the ROC curve (AUC) statistics software SPBS (Comworks, Saitama, Japan) was used to quantify the ability to differentiate between

healthy volunteers and patients with lung cancer. Analyses of the prognostic impact of serum HE4 levels on survival from response evaluation to death or last follow-up used the Kaplan–Meier method and logrank test.

Results

Generation of ELISA specific for HE4

To generate monoclonal antibodies that specifically react with HE4, recombinant HE4 protein corresponding to the amino acids 1–124 of the transcript for human HE4 was produced by 293 T transfectants (Fig. 1a) and immunized to BALB/c mice. Supernatants of obtained hybridomas were tested for binding activity to microplates coated with immunizing antigen and further examined by a competition assay for the immunogen (data not shown). The specific reactivities of selected clone 13F5G1, 99-44, 110-108, 128-119, 179-137 and 201-50 to HE4 were checked by immunoprecipitation using myc-His-tagged recombinant HE4 (Fig. 1b). The specificity of 110-108 and 128-119 to HE4 was tested using the 293 T transfectant-expressing HE4 by flowcytometory. Clones 110-108 and 128-119 detected the antigen on the cell surface of 293 T transfectant (Fig. 1c). A sandwich ELISA for HE4 was constructed using the anti-HE4 antibodies 110-108 and 128-119. The standard curve using purified recombinant HE4 is shown in Fig. 1d. The HE4 sandwich ELISA detected the antigen in the culture supernatant of the various types of cancer cell lines (Fig. 1e). We investigated immunohistochemical staining of lung cancer tissue using anti-HE4 antibody. HE4 expression was detected in lung cancer tissues but not in normal lung tissues. We compared HE4 staining between adenocarcinoma and squamous cell carcinoma lung tissue samples but could not detect significant differences in staining intensity. Furthermore, we could not find a correlation between differentiation and HE4 staining of lung cancer tissues. Strong HE4 staining was

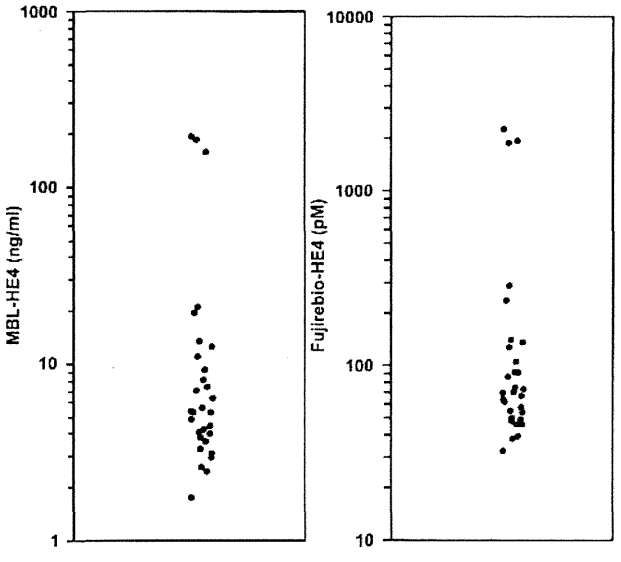


Fig. 2 HE4 levels in sera of ovarian cancer patients. HE4 levels determined by our ELISA system (*left*) and Fujirebio commercial ELISA kit (*right*). Each *dot* represents one patient

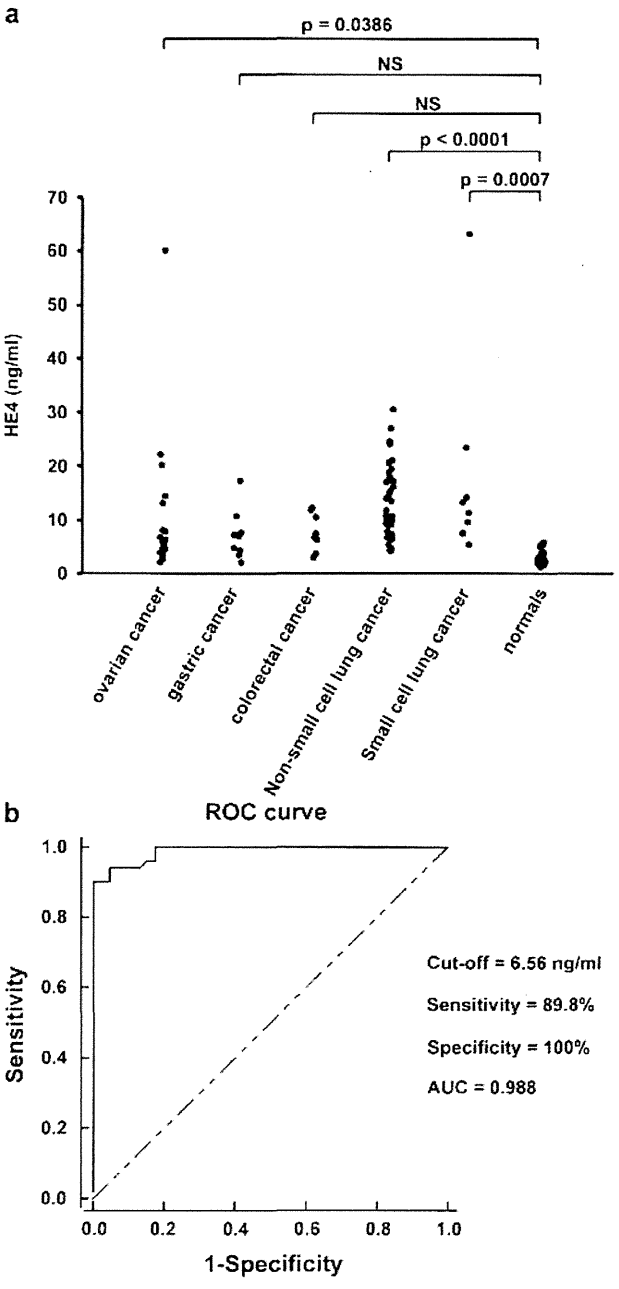


Fig. 3 a HE4 levels in the sera of cancer patients and normal controls. Each dot represents one patient. *NS* no significant. b Receiver operating characteristic (ROC) curves for HE4 for differentiation between lung cancer and healthy volunteers. The tables show the best statistical cutoff values for HE4 with pairs of sensitivity and specificity

Table 3 Pre- and post-treatment HE4 in lung cancer patients

Clinical response	Number of patients	Pre-treatment HE4 (ng/ml)			Post-treatment HE4 (ng/ml)		
		Mean (SD)	Median	Range	Mean (SD)	Median	Range
PR	8	23.0 (26.3)	14.2	8.3–86.5	12.6 (2.8)	12.9	8.2–16.1
SD	11	22.2 (32.3)	11.4	3.7–116.6	28.3 (44.4)	14.2	3.7–158.1
PD	5	21.3 (7.3)	22.8	9.3–28.1	23.8 (5.9)	25.9	14.2–29.1

detected in cytoplasmic and plasma membrane areas but not in nuclear area (Fig. 1f).

Evaluation of ELISA system

To assess the clinical potential of our ELISA system, we compared our ELISA system with an existing commercial ELISA kit (Fujirebio Diagnostics, Malvern, PA). We set the cutoff point of the existing ELISA kit as 150 pM based on the manufacturer's instructions (94.4 percentile of healthy individuals). In accordance with the existing ELISA kit, we set the cutoff point of our ELISA system as 5.5 ng/ml based on 94.4 percentile of 37 healthy individuals. We measured concentrations of HE4 of ovarian cancer patients by using two ELISA and found that our ELISA system shows better sensitivity for diagnosis of ovarian cancer than the existing ELISA kit by Fisher's exact probability test ($p<0.05$) (Fig. 2).

Diagnostic value of HE4

We measured serum HE4 levels in cancer patients and healthy controls by using our ELISA system. Mean serum HE4 levels in patients with non-small cell lung cancer, small cell lung cancer, ovarian cancer, gastric cancer, colorectal cancer and healthy adults were 13.3 ng/ml, 17.3 ng/ml, 10.9 ng/ml, 7.2 ng/ml, 7.7 ng/ml and 2.7 ng/ml (Table 1). Serum HE4 levels were significantly higher for non-small, small cell lung cancer and ovarian cancer patients than for healthy controls ($p<0.0001$, $p=0.0007$, and $p=0.0386$, respectively) (Fig. 3a).

To assess the clinical potential of HE4, we calculated sensitivities and specificities of HE4. The operating characteristics for HE4 with its cutoff points for achieving the best individual accuracy are shown in Fig. 3b. The AUC for serum HE4 was 0.988 for differentiating lung cancer patients from healthy adults, with a cutoff value of 6.56 ng/ml (sensitivity=89.8%, specificity=100%) (Fig. 3b). Serum HE4 levels were elevated in 36/40 (90.0%) non-small cell lung cancer patients, 8/9 (88.9%) small cell lung cancer patients, 8/18 (44.4%) ovarian cancer patients, 6/10 (60.0%) gastric cancer patients and 5/8 (62.5%) colorectal cancer patients (Table 1). These results suggest that the sensitivity of HE4 was high in lung cancer patients.

Prognostic value of HE4

Pre- and post-treatment mean HE4 were 23.0 ng/ml (range 8.3–86.5 ng/ml) and 12.6 ng/ml (range 8.2–16.1 ng/ml), respectively, for PR patients, whereas 21.3 ng/ml (range 9.3–28.1 ng/ml) and 23.8 ng/ml (range 14.2–29.1 ng/ml), respectively, for PD patients (Table 3). Post-treatment mean

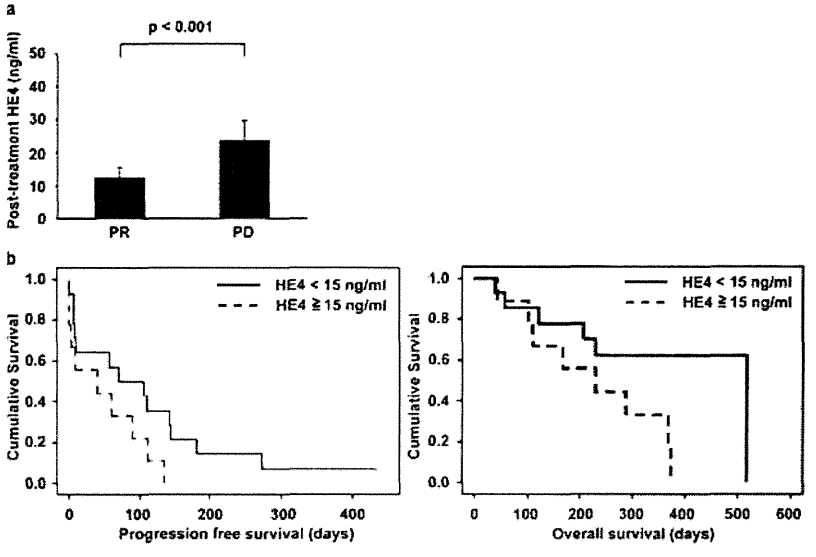


Fig. 4 **a** Post-treatment serum HE4 levels of lung cancer patients receiving chemotherapy. Figures show the average (*columns*) + SD (*bars*). **b** Kaplan–Meier plots of overall survival (*right*) and progression-free survival (*left*) after chemotherapy

HE4 in PR patients was significantly lower than that in PD patients ($p<0.001$) (Fig. 4a). Overall median survival after treatment was 9.6 months. Post-treatment HE4 levels were above the cutoff limit (15 ng/ml) in one of eight for PR and in four of five for PD patients. We divided patients into high (>15 ng/ml) and low (<15 ng/ml) HE4 groups post-treatment. We set the cutoff point at 15 ng/ml based on the median value of HE4 levels. Median overall survival of low HE4 group was significantly longer than that of high HE4 group (17.3 vs. 7.7 months; $p<0.05$) (Fig. 4b). Median progression-free survival after chemotherapy for low HE4 group was 2.4 months, compared with 1.4 months for high HE4 group ($p=0.083$) (Fig. 4b).

Discussion

In our study, we developed a novel ELISA system to detect serum HE4, and by using this system, we show that HE4 has potential as a diagnostic marker of lung cancer. Specifically, we found that serum HE4 levels in lung cancer patients after chemotherapy is strongly correlated with survival after the treatment.

Our HE4 ELISA shows better sensitivity for diagnosis of ovarian cancer than the existing HE4 ELISA kit. Furthermore, there are two advantages to our HE4 ELISA over the existing HE4 ELISA kit. First, while the existing ELISA kit is designed to use undiluted serum as the test sample, using our system, a small sample volume can be applied to our ELISA to detect HE4 in human serum. A diluted sample, maximally five times dilution, can be used in our assay system. This is of practical importance when simultaneous measures of other biomarkers are required from one serum sample. Second, a simpler and more rapid test is achieved using our HE4 ELISA. Test duration for our HE4 ELISA is approximately half of the time required by the existing HE4 ELISA kit. The existing HE4 ELISA kit requires a shaking procedure during reaction of antibody to the sample, while our new ELISA incubates the sample with antibody in a stationary condition. Thus, our HE4 ELISA does not require the purchase of specialized equipment for shaking incubation.

We evaluated the diagnostic efficacy of HE4 in lung, ovarian, gastric and colorectal cancer patients and found that HE4 indicated high sensitivity in lung cancer patients. Tissue expression of HE4 has been reported to be increased in pulmonary, ovarian and gastrointestinal carcinomas [15–18, 20, 21]. HE4 as a serum marker was mainly investigated in ovarian cancer patients and was shown to be a promising diagnostic marker [22]. We show that serum HE4 levels were significantly higher for not only ovarian cancer but also lung cancer patients than for healthy controls. Escudero et al. previously measured HE4 concentrations in patients with various types of malignant diseases and found that HE4 concentrations were abnormal primarily in gynecologic cancer and lung cancer [23]. Taken together, these data HE4 may be considered to be a potential diagnostic marker of lung cancer.

In this study, we found that post-treatment HE4 level is correlated with survival after chemotherapy. It was reported that high levels of serum HE4 is significantly correlated with worse prognosis in epithelial ovarian cancer patients [24]. Yamashita et al. reported that HE4 expression by immunohistochemistry staining is significantly correlated with prognosis in lung adenocarcinoma patients [25]. There is a growing need for diagnostic tools to estimate the prognosis of the patient, to monitor the treatment course and to early detect the response to therapy, which would help to optimize disease management on an individual basis. Our results suggest that serum HE4 is a promising prognostic marker. To our knowledge, this is the first time that the potential prognostic impact of serum HE4 in lung cancer patients has been investigated. We are aware of the small cohort of patients in this study, and thus, further study by larger scale prospective trials will be needed.

In conclusion, we used ELISA systems developed by us to detect significant differences in the levels of serum HE4 between lung cancer patients and normal controls. In addition, it is suggested that HE4 is correlated with prognosis after chemotherapy. We are planning a further study to evaluate serum HE4 for a diagnostic and prognostic marker by larger scale of patients.

Acknowledgments We appreciate Shintaro Nomura (Nagahama Institute of Bio-Science and Technology, Shiga, Japan) for providing helpful comments on immunohistochemical analysis, Barry Ripley for outstanding editing of the manuscript and Masako Ikeda for their technical assistance. We wish to thank Y. Ito, N. Kawakami and Y. Kanazawa for their secretarial assistance. This work was supported by a grant-in-aid from the Ministry of Health, Labour and Welfare, Japan.

Conflict of interest None.

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin. 2010;60:277–300.
2. Fruh M. The search for improved systemic therapy of non-small cell lung cancer—what are today's options? Lung Cancer. 2011;72:265–70.
3. Martini N, Kris MG, Gralla RJ, Bains MS, McCormack PM, Kaiser LR, Burt ME, Zaman MB. The effects of preoperative chemotherapy on the resectability of non-small cell lung carcinoma with mediastinal lymph node metastases (n2 m0). Ann Thorac Surg. 1988;45:370–9.
4. Shinkai T, Saijo N, Tominaga K, Eguchi K, Shimizu E, Sasaki Y, Fujita J, Futami H, Ohkura H, Suemasu K. Serial plasma carcinoembryonic antigen measurement for monitoring patients with advanced lung cancer during chemotherapy. Cancer. 1986;57:1318–23.

5. Pujol JL, Grenier J, Daures JP, Daver A, Pujol H, Michel FB. Serum fragment of cytokeratin subunit 19 measured by cyfra 21-1 immunoradiometric assay as a marker of lung cancer. Cancer Res. 1993;53:61–6.
6. Miyake Y, Kodama T, Yamaguchi K. Pro-gastrin-releasing peptide (31-98) is a specific tumor marker in patients with small cell lung carcinoma. Cancer Res. 1994;54:2136–40.
7. Kirchhoff C, Habben I, Ivell R, Krull N. A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors. Biol Reprod. 1991;45:350–7.
8. Kirchhoff C. Molecular characterization of epididymal proteins. Rev Reprod. 1998;3:86–95.
9. Wang K, Gan L, Jeffery E, Gayle M, Gown AM, Skelly M, Nelson PS, Ng WV, Schummer M, Hood L, Mulligan J. Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. Gene. 1999;229:101–8.
10. Schummer M, Ng WV, Bumgarner RE, Nelson PS, Schummer B, Bednarski DW, Hassell L, Baldwin RL, Karlan BY, Hood L. Comparative hybridization of an array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. Gene. 1999;238:375–85.
11. Hough CD, Sherman-Baust CA, Pizer ES, Montz FJ, Im DD, Rosenshein NB, Cho KR, Riggins GJ, Morin PJ. Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. Cancer Res. 2000;60:6281–7.
12. Ono K, Tanaka T, Tsunoda T, Kitahara O, Kihara C, Okamoto A, Ochiai K, Takagi T, Nakamura Y. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. Cancer Res. 2000;60:5007–11.
13. Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockhart DJ, Burger RA, Hampton GM. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci U S A. 2001;98:1176–81.
14. Shridhar V, Lee J, Pandita A, Iturria S, Avula R, Staub J, Morrissey M, Calhoun E, Sen A, Kalli K, Keeney G, Roche P, Cliby W, Lu K, Schmandt R, Mills GB, Bast Jr RC, James CD, Couch FJ, Hartmann LC, Lillie J, Smith DI. Genetic analysis of early- versus late-stage ovarian tumors. Cancer Res. 2001;61:5895–904.
15. Schaner ME, Ross DT, Ciaravino G, Sorlie T, Troyanskaya O, Diehn M, Wang YC, Duran GE, Sikic TL, Caldeira S, Skomedal H, Tu IP, Hernandez-Boussard T, Johnson SW, O'Dwyer PJ, Fero MJ, Kristensen GB, Borresen-Dale AL, Hastie T, Tibshirani R, van de Rijn M, Teng NN, Longacre TA, Botstein D, Brown PO, Sikic BI. Gene expression patterns in ovarian carcinomas. Mol Biol Cell. 2003;14:4376–86.
16. Lu KH, Patterson AP, Wang L, Marquez RT, Atkinson EN, Baggerly KA, Ramoth LR, Rosen DG, Liu J, Hellstrom I, Smith D, Hartmann L, Fishman D, Berchuck A, Schmandt R, Whitaker R, Gershenson DM, Mills GB, Bast Jr RC. Selection of potential markers for epithelial ovarian cancer with gene expression arrays and recursive descent partition analysis. Clin Cancer Res. 2004;10:3291–300.
17. Drapkin R, von Horsten HH, Lin Y, Mok SC, Crum CP, Welch WR, Hecht JL. Human epididymis protein 4 (he4) is a secreted glycoprotein that is overexpressed by serous and endometrioid ovarian carcinomas. Cancer Res. 2005;65:2162–9.
18. Rosen DG, Wang L, Atkinson JN, Yu Y, Lu KH, Diamandis EP, Hellstrom I, Mok SC, Liu J, Bast Jr RC. Potential markers that complement expression of ca125 in epithelial ovarian cancer. Gynecol Oncol. 2005;99:267–77.
19. Hellstrom I, Raycraft J, Hayden-Ledbetter M, Ledbetter JA, Schummer M, McIntosh M, Drescher C, Urban N, Hellstrom KE. The he4 (wfdc2) protein is a biomarker for ovarian carcinoma. Cancer Res. 2003;63:3695–700.
20. Bingle L, Cross SS, High AS, Wallace WA, Rassl D, Yuan G, Hellstrom I, Campos MA, Bingle CD. Wfdc2 (he4): a potential role in the innate immunity of the oral cavity and respiratory tract and the development of adenocarcinomas of the lung. Respir Res. 2006;7:61.
21. Galgano MT, Hampton GM, Frierson Jr HF. Comprehensive analysis of he4 expression in normal and malignant human tissues. Mod Pathol. 2006;19:847–53.
22. Li J, Dowdy S, Tipton T, Podratz K, Lu WG, Xie X, Jiang SW. He4 as a biomarker for ovarian and endometrial cancer management. Expert Rev Mol Diagn. 2009;9:555–66.
23. Escudero JM, Auge JM, Filella X, Torne A, Pahisa J, Molina R. Comparison of serum human epididymis protein 4 with cancer antigen 125 as a tumor marker in patients with malignant and nonmalignant diseases. Clin Chem. 2011;57:1534–44.
24. Steffensen KD, Waldstrom M, Brandslund I, Jakobsen A. Prognostic impact of prechemotherapy serum levels of her2, ca125, and he4 in ovarian cancer patients. Int J Gynecol Canc. 2011;21:1040–7.
25. Yamashita S, Tokuishi K, Hashimoto T, Moroga T, Kamei M, Ono K, Miyawaki M, Takeno S, Chujo M, Yamamoto S, Kawahara K. Prognostic significance of he4 expression in pulmonary adenocarcinoma. Tumour Biol. 2011;32:265–71.

Journal of proteome research

Article

pubs.acs.org/jpr

Comprehensive Identification of Substrates for F-box Proteins by Differential Proteomics Analysis

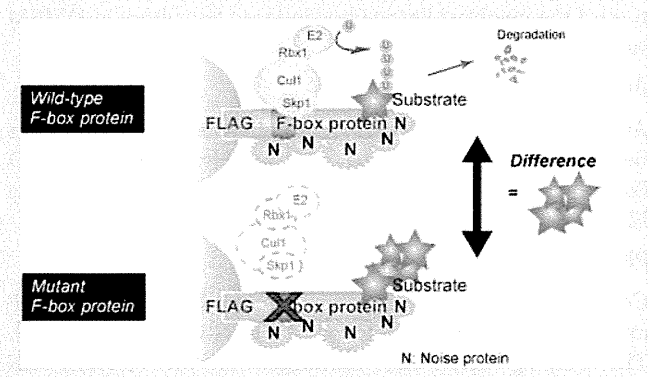
Kanae Yumimoto,[†,‡] Masaki Matsumoto,[†,‡] Koji Oyamada,[†,‡] Toshiro Moroishi,[†,‡] and Keiichi I. Nakayama*[,†,‡]

[†]Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan

[‡]CREST, Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan

Supporting Information

ABSTRACT: Although elucidation of enzyme–substrate relations is fundamental to the advancement of biology, universal approaches to the identification of substrates for a given enzyme have not been established. It is especially difficult to identify substrates for ubiquitin ligases, given that most such substrates are immediately ubiquitylated and degraded as a result of their association with the enzyme. We here describe the development of a new approach, DiPIUS (differential proteomics-based identification of ubiquitylation substrates), to the discovery of substrates for ubiquitin ligases. We applied DiPIUS to Fbxw7α, Skp2, and Fbxl5, three of the most well-characterized F-box proteins, and identified candidate substrates including previously known targets. DiPIUS is thus a powerful tool for unbiased and comprehensive screening for substrates of ubiquitin ligases.



KEYWORDS: *ubiquitin ligase, F-box protein, substrate identification, differential proteomics*

INTRODUCTION

Enzyme–substrate relations have been discovered by substrate-to-enzyme approaches in most instances.[1−5] Analysis of the genome sequences of various organisms has led to the identification of large numbers of enzymes such as protein kinases[6] and ubiquitin ligases[7] on the basis of their conserved catalytic domains. However, universal approaches to the identification of substrates for a given enzyme have not been established to date, and many enzymes remain "orphans". One of the challenges to substrate identification is detection of the interaction between enzymes and substrates, which is generally weak and transient and has therefore been referred to as "kiss and run". Although recent advances in mass spectrometry (MS) have allowed the detection of such weak interactions, elimination of the large number of proteins that bind nonspecifically to a bait protein remains problematic. It is also difficult to apply this approach to the identification of substrates for ubiquitin ligases, given that the ubiquitylation of most such substrates by the ubiquitin ligase promotes their degradation by the 26S proteasome.[8]

The Skp1–Cul1–F-box protein (SCF) complex is one of the most well-characterized types of mammalian ubiquitin ligase. Each SCF complex is composed of four subunits: Skp1, Cul1, and Rbx1 (also known as Roc1 or Hrt1) as invariable components and an F-box protein that primarily determines substrate specificity as a variable component.[9−11] Cul1 serves as a scaffold and interacts via its COOH terminus with the RING-finger protein Rbx1 to recruit a ubiquitin-conjugating enzyme, whereas its NH_2 terminus interacts with Skp1, an adaptor protein that binds to the F-box domain of an F-box protein, which in turn is responsible for substrate specification. The substrate specificity of each SCF complex is thus conferred by the incorporated F-box protein, with each such protein recognizing a different group of substrates. F-box proteins constitute a large family of eukaryotic proteins, with ~70 F-box proteins having been identified in humans.[12] Whereas some F-box proteins have been found to contribute to cellular activities such as cell cycle progression,[9] synapse formation,[13,14] plant hormone responses,[15] and the circadian clock,[16−18] the functions of most F-box proteins remain to be elucidated because their substrates are unknown.

We now describe the development of a general approach to the identification of substrates for SCF type ubiquitin ligases. The application of this approach, designated DiPIUS (differential proteomics-based identification of ubiquitylation substrates), to Fbxw7α, Skp2, and Fbxl5 resulted in the identification of many candidate substrates. These candidates include previously known targets, providing support for the fidelity of this method. We propose DiPIUS as a powerful tool

Received: December 10, 2011
Published: April 23, 2012

for the discovery of substrates for any ubiquitin ligase in cases in which the ubiquitylation of the substrate is followed by proteasomal degradation.

MATERIALS AND METHODS

Cell Culture

HEK293T cells, Neuro2A cells, mHepa cells, and mCAT-HeLa cells (HeLa cells stably expressing mCAT1, a transporter for basic amino acids, as well as a receptor for ecotropic retrovirus infection)[19] were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS; Invitrogen), 1 mM sodium pyruvate, penicillin (100 U/mL), streptomycin (100 mg/mL), 2 mM L-glutamine, and nonessential amino acids (10 mL/L; Gibco). The murine myoblast precursor cell line C2C12 was maintained in DMEM supplemented with 10% FBS. Differentiation of C2C12 cells toward the myoblast lineage was induced by culture in DMEM supplemented with 0.1% FBS.

Antibodies

Antibodies to c-Myc (N-262) were obtained from Santa Cruz Biotechnology; those to Skp1, p27, and Hsp90 were from BD Biosciences; those to Cul1 were from Zymed; and those to the FLAG epitope (M2) were from Sigma.

Plasmids

Complementary DNAs encoding mouse Fbxw7α or its ΔF mutant were subcloned into p3 × FLAG-CMV 7.1 (Sigma). The resulting vectors were introduced into mCAT-HeLa cells with the use of the FuGENE 6 or FuGENE HD transfection reagents (Roche). For retroviral expression, cDNAs for FLAG-tagged mouse Fbxw7α, Skp2, or corresponding mutants were subcloned into pMX-puro (kindly provided by T. Kitamura). The resulting vectors were introduced into Plat E cells with the use of FuGENE HD. The recombinant retroviruses thereby generated were used to infect mCAT-HeLa, Neuro2A, C2C12, or mHepa cells, which were then subjected to selection in medium containing puromycin (10 μg/mL). Complementary DNAs for FLAG-tagged mouse Fbxl5 or its PE/AA mutant were subcloned into pMX-puro-CMV. The resulting vectors together with pCL-Ampho (IMGENEX) were introduced into HEK293T cells with the use of FuGENE HD, and the recombinant retroviruses thereby generated were used to infect HEK293T cells, which were then subjected to selection in medium containing puromycin (5 μg/mL). These latter cells were treated with ferric ammonium citrate (100 μg/mL) for 24 h before sample preparation.

SILAC

SILAC (stable isotopic labeling using amino acids in cell culture) labeling medium (KOHJIN-bio) lacking lysine and arginine was reconstituted and supplemented with 10% dialyzed FBS and amino acid stocks prepared in phosphate-buffered saline (PBS). We added $^{13}C_6$-lysine and $^{13}C_6$-arginine (Cambridge Isotope Laboratories) to heavy medium and normal lysine and arginine (Sigma-Aldrich) to light medium. mCAT-HeLa or HEK293T cells were cultured in heavy medium for at least six cell doublings to allow adaptation and full incorporation of the stable isotope-containing amino acids. The wild-type F-box protein was expressed in light-labeled cells and the mutant protein in heavy-labeled cells.

Immunoaffinity Purification

Wild-type Fbxw7α, Skp2, or Fbxl5 tagged with the FLAG epitope at its NH_2 terminus was expressed in light-labeled mCAT-HeLa cells or HEK293T cells, and the corresponding mutant protein was expressed in heavy-labeled cells, for DiPIUS. Cells were incubated for 6 h in the presence of the proteasome inhibitor MG132 (10 μM; Peptide Institute) and were then lysed in 8 mL of a lysis buffer containing 20 mM HEPES-NaOH (pH 7.5), 150 mM NaCl, 1% digitonin, 10 mM NaF, 10 mM $Na_4P_2O_7$, 0.4 mM Na_3VO_4, 0.4 mM EDTA, leupeptin (20 μg/mL), aprotinin (10 μg/mL), and 1 mM phenylmethylsulfonyl fluoride. The lysates were centrifuged at 2200g for 20 min at 4 °C to remove debris, and the resulting supernatants were adjusted with lysis buffer to achieve a protein concentration of 2.5 mg/mL. The supernatants (20 mg of protein in 8 mL of solution) from light-labeled cells and heavy-labeled cells were combined for DiPIUS and incubated for 1 h at 4 °C with 120 μL of beads conjugated with M2 antibodies to FLAG (Sigma). The beads were washed three times with 4 mL of a solution containing 10 mM HEPES-NaOH (pH 7.5), 150 mM NaCl, and 0.1% Triton X-100, and bead-bound proteins were then eluted with the FLAG peptide (500 μg/mL; Sigma), precipitated with ice-cold 20% trichloroacetic acid, and washed with acetone. The concentrated proteins were then digested with Lys-C and trypsin, and the resulting peptides were analyzed with a QSTAR Elite Hybrid LC–MS/MS system (Applied Biosystems) for DiPIUS of Skp2 or with an LTQ Orbitrap Velos LC–MS/MS system (Thermo Finnigan) for DiPIUS of Fbxw7α or Fbxl5. For DiPIUS-NL, the concentrated proteins derived separately from lysates of nonlabeled cells expressing the wild-type or mutant F-box proteins were dissolved in SDS sample buffer, fractionated by SDS–polyacrylamide gel electrophoresis (PAGE), and stained with silver. Individual lanes of the stained gel were sliced into 8–16 pieces, and proteins within these pieces were subjected to in-gel digestion with trypsin as described previously.[20] The resulting peptides were dried and then dissolved in a solution containing 0.1% trifluoroacetic acid and 2% acetonitrile before analysis with an ion-trap mass spectrometer (LTQ-XL, Thermo Finnigan). The high molecular weight fractions were excluded from the analysis.

Spectral Counting

Peak lists were generated by lcq_dta.exe (Thermo Finnigan) and were compared with the use of the MASCOT algorithm (ver. 2.2.1) either with the "Target-decoy" Human IPI version 3.1.6 database [date of release, November 2006; with 62322 target sequences, searched against a total of 124644 sequences (target and reverse/decoy)] or with the "Target-decoy" Mouse IPI version 3.4.4 database [date of release, June 2008; with 55078 target sequences, searched against a total of 110156 sequences (target and reverse/decoy)], both maintained by the European Bioinformatics Institute. Trypsin was selected as the protease, the allowed number of missed cleavages was set to one, and carbamidomethylation of cysteine was selected as the fixed modification. Oxidized methionine and pyroglutamine were searched as variable modifications. The precursor mass tolerance was 1.5 Da, and the tolerance of MS/MS ions was 0.8 Da. Assigned high-scoring peptide sequences (MASCOT score >35) were processed by in-house software. If the MASCOT score was <45 (peptides for which the MS2 score was above the 95th percentile of significance), assigned sequences were manually confirmed by comparison with the corresponding

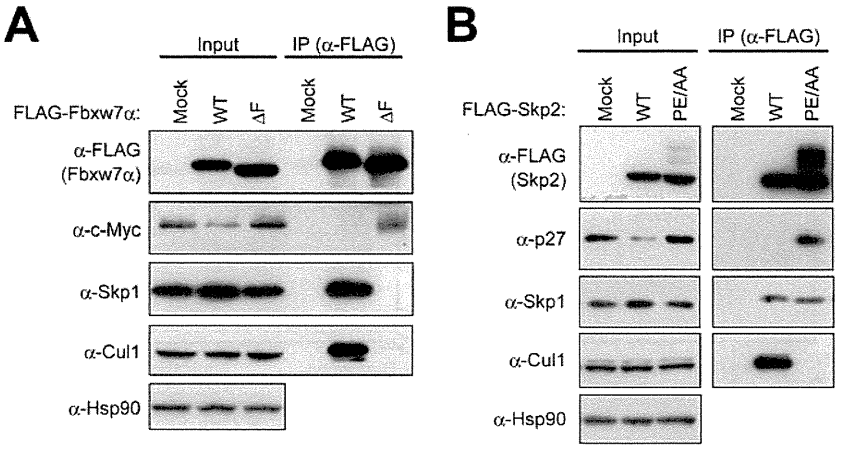


Figure 1. Stable interaction between mutant F-box proteins and their substrates. mCAT-HeLa cells transfected with expression vectors for FLAG-tagged wild-type (WT) or mutant (ΔF) Fbxw7α (A) or FLAG-tagged WT or mutant (PE/AA) Skp2 (B), or with the corresponding empty vector (Mock), were lysed and subjected to immunoprecipitation (IP) with antibodies to (α-) FLAG. The resulting precipitates as well as the original cell lysates (Input) were subjected to immunoblot analysis with the indicated antibodies.

collision-induced dissociation spectra on the basis of the following criteria: (i) a delta score of >15 or (ii) at least six successive matches for y- or b-ions or at least three blocks of three successive matches for y- or b-ions. Identified peptides from independent experiments were integrated and regrouped by IPI accession number. When multiple accession IDs were obtained with the same set of peptides, a representative IPI accession was chosen according to the following order of priorities: (i) IPI accession containing NCBI GeneID, (ii) IPI accession with the use of Swiss-Prot as the master sequence, and (iii) IPI accession with the use of TrEMBL as the master sequence. Proteins identified in only one experiment or with a single-peptide assignment were removed from spectral counting data. Estimated false discovery rates (FDRs) were zero at the protein level in all experiments with the exception of that involving the expression of Skp2 in mHepa cells (FDR = 0.26%) and that regarding the growth state of C2C12 cells (FDR = 0.51%). For derivation of the consensus Cdc4 phosphodegron (CPD) sequence, we aligned the CPDs of 16 known Fbxw7 (Cdc4) substrates in mammals or yeast[10] and found that positions −1 or +5 (or both) showed a preference for leucine or proline.

Quantification for SILAC

The peak lists were generated by Mascot Distiller (version 2.3.0) and compared with the "Target-decoy" Human IPI version 3.1.6 database with the use of the MASCOT algorithm (ver. 2.2.1). Trypsin was selected as the protease, the allowed number of missed cleavages was set to one, and carbamidomethylation of cysteine was selected as the fixed modification. Oxidized methionine, pyroglutamine, $^{13}C_6$-lysine, and $^{13}C_6$-arginine were searched as variable modifications. Precursor mass tolerance was 200 ppm for QSTAR Elite Hybrid and 20 ppm for LTQ Orbitrap Velos, and tolerance of MS/MS ions was 0.3 Da for QSTAR Elite Hybrid and 0.8 Da for LTQ Orbitrap Velos. Assigned high-scoring peptide sequences (MASCOT score >20) were processed by in-house software; the MASCOT score cutoff value for SILAC analysis was lower than that for spectral counting (MASCOT score >35) because of the use of high-resolution LC-MS/MS systems for SILAC. If the MASCOT score was <45, assigned sequences were manually confirmed as described above for spectral counting, and the FDR at the peptide level was estimated. SILAC quantitation was performed with the use of the Mascot Distiller Quantitation Tool (version 2.3.0). Extracted ion chromatogram (XIC) peak areas for the heavy and light peptides were measured, and the results were verified by manual inspection of MS spectra. SILAC ratios were calculated by comparison of the XIC peak areas of light peptides with those of the heavy peptides. The method "Auto" was selected for the removal of outlier peptide pairs (peptides with a number between 4 and 25 by Dixon's method or of >25 by Rosner's method). SILAC ratios for different peptides of a protein were averaged to give protein abundance ratios, which were then normalized with the heavy/light ratio for each bait protein (Fbxw7α, Skp2, or Fbxl5).

Functional Enrichment Analysis

We used a Web-based implementation of Database for Annotation, Visualization, and Integrated Discovery (DAVID)[21] for functional annotation of enriched proteins among specific binding proteins for each F-box protein and nonspecific binding proteins that interact with at least two of Fbxw7α, Skp2, and Fbxl5. We used the database of Gene Ontology (GO) level 2 for this analysis. Overrepresented functional categories among binding proteins were identified relative to the entire human genome with the use of Fisher's exact test. We processed the GO term lists with P values of <0.05 and contributing rate of proteins of >10%.

Immunoprecipitation and Immunoblot Analysis

Cells were incubated for 6 h in the presence of the proteasome inhibitor MG132 (10 μM) prior to harvest. Cell lysis and immunoprecipitation were performed as described.[22] Immunoprecipitates and cell lysates were subjected to immunoblot analysis as described,[23] with Hsp90 used as a loading control.

RESULTS

Development of the DiPIUS System

A substrate would be expected to be ubiquitylated and degraded on its recognition by a ubiquitin ligase, resulting in a decrease in its cellular concentration. In contrast, an F-box protein with a point mutation or deletion in the F-box domain might be expected to retain the ability to associate with a substrate but to have lost the ability to mediate ubiquitin conjugation, resulting in accumulation of the substrate in the

dx.doi.org/10.1021/pr201216u | J. Proteome Res. 2012, 11, 3175–3185

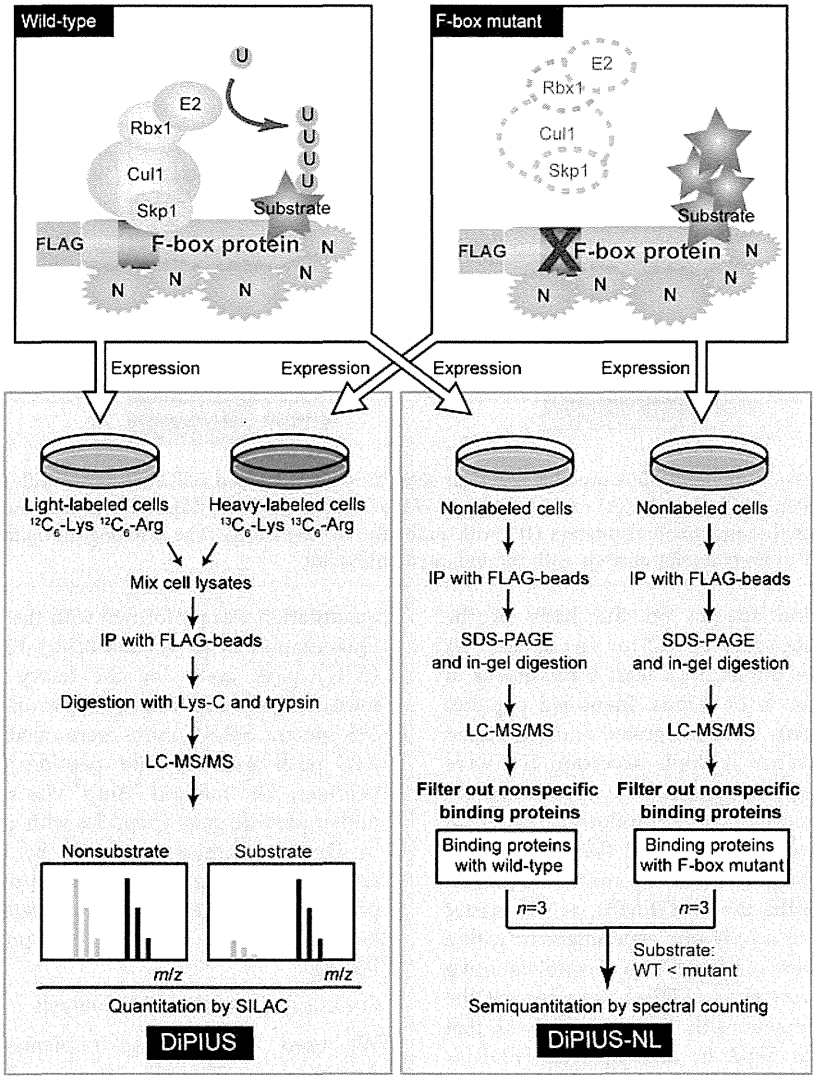


Figure 2. Strategy for comprehensive identification of substrates for an F-box protein. Schematic representations of DiPIUS and DiPIUS-NL are shown. Proteins that bind to WT or mutant F-box proteins are analyzed by quantitative SILAC (DiPIUS) or semiquantitative spectral counting (DiPIUS-NL). U, ubiquitin; N, nonspecific binding protein.

cell. We thus expected that whereas wild-type F-box proteins interact transiently with their substrates, mutant F-box proteins might stably associate with such substrates as the concentration of the latter would be increased. On the other hand, proteins that bind nonspecifically to other proteins would be expected to associate with wild-type and mutant F-box proteins to similar extents. To validate this hypothesis, we examined the interactions of Fbxw7α and Skp2 (also known as Fbxl1), two of the most well-characterized F-box proteins, with their representative substrates c-Myc and p27, respectively, as positive controls. We designed mutant Fbxw7α and Skp2 proteins that are unable to associate with the Cul1-Rbx1 module. In the case of Fbxw7α, the entire F-box domain was deleted (ΔF mutant), given that such deletion does not compromise binding to substrates. The COOH-terminal tail of Skp2 loops back to the F-box domain;[24] therefore, a large deletion within the F-box domain might be expected to affect the structure of the substrate-binding region, which is composed of many leucine-rich repeats. We thus replaced only two amino acids (P113 and E115) in the F-box domain that are essential for the association of Skp2 with Cul1 with alanine (PE/AA mutant).[14,25] As expected, coimmunoprecipitation analysis revealed that FLAG epitope-tagged wild-type Fbxw7α bound weakly to endogenous c-Myc, whereas more c-Myc was associated with the FLAG-tagged Fbxw7$\alpha(\Delta F)$ mutant (Figure 1A). Likewise, FLAG-tagged wild-type Skp2 associated weakly with endogenous p27, whereas the FLAG-tagged Skp2(PE/AA) mutant interacted with p27 to a greater extent (Figure 1B).

To isolate comprehensively the proteins that associate with mutant F-box proteins more efficiently than with wild-type F-box proteins, we developed the DiPIUS system (Figure 2). Wild-type or mutant forms of F-box proteins tagged at their NH_2 termini with the FLAG epitope were expressed in mCAT-HeLa or HEK293T cells and subjected to SILAC analysis.[26] Cells expressing the wild-type protein were cultured in regular (light) medium, whereas those expressing the mutant protein were cultured in medium containing stable isotope-labeled amino acids (heavy medium). Cell lysates were mixed and subjected to immunoprecipitation with antibodies to FLAG, and tryptic digests of the immunoprecipitates were analyzed by liquid chromatography and tandem MS (LC-MS/MS) (Figure 3 and Tables S1 and S2 in the Supporting Information). Alternatively, tryptic digests of immunoprecipitates derived

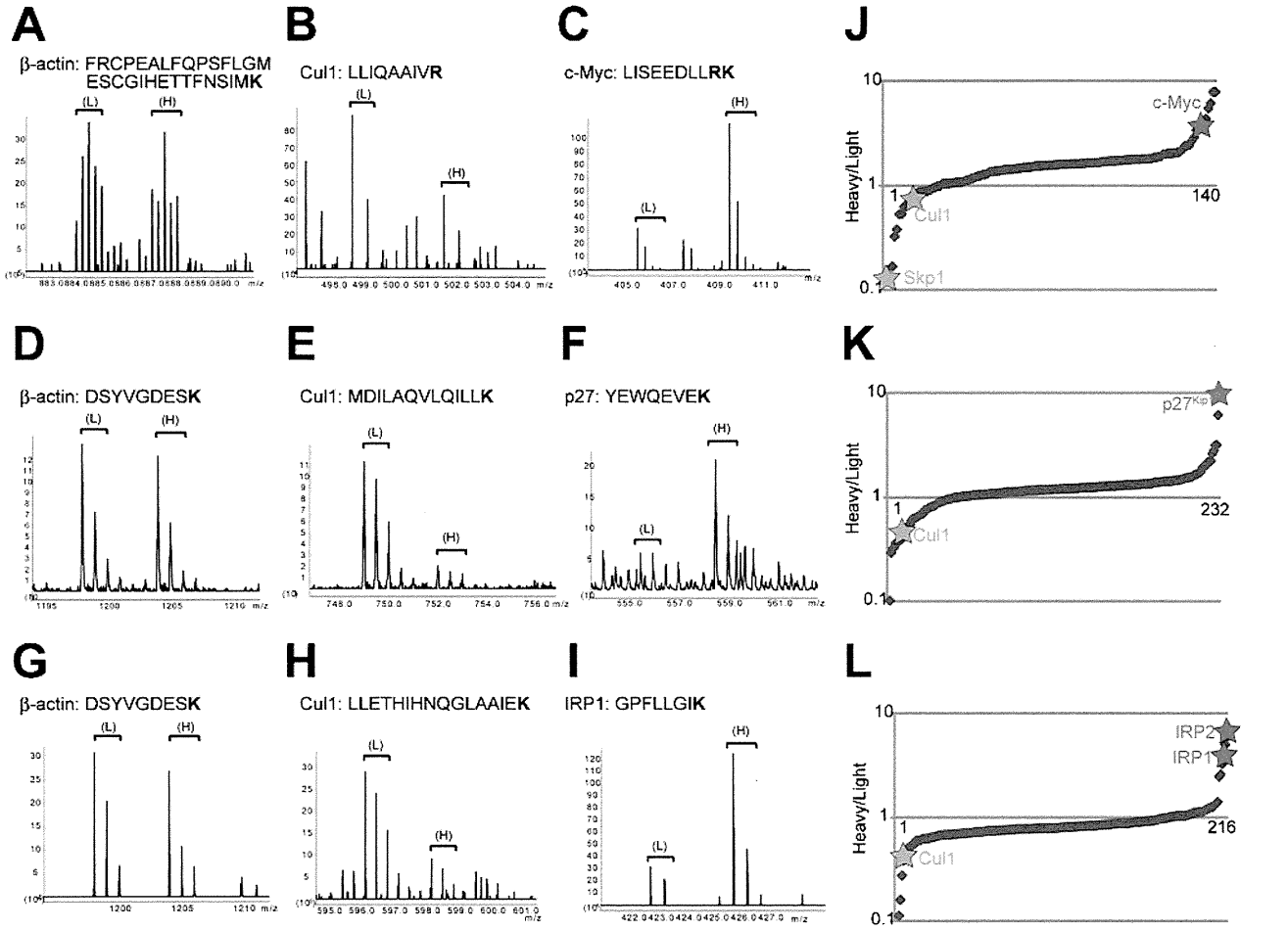


Figure 3. Candidate substrates for Fbxw7α, Skp2, and Fbxl5 analyzed by quantitative SILAC. (A–I) Representative mass spectra of SILAC analyses. Peak areas of light (L) and heavy (H) tryptic peptides derived from β-actin (A), Cul1 (B), or c-Myc (C) represent the abundance of proteins associated with WT and mutant Fbxw7α, respectively; those from β-actin (D), Cul1 (E), or p27 (F) represent the abundance of proteins associated with WT and mutant Skp2, respectively; and those from β-actin (G), Cul1 (H), or IRP1 (I) represent the abundance of proteins associated with WT and mutant Fbxl5, respectively. (J–L) Plot diagrams of heavy/light ratios for SILAC analysis of Fbxw7α (J), Skp2 (K), and Fbxl5 (L).

separately from lysates of nonlabeled cells expressing the wild-type or mutant F-box proteins were subjected to semi-quantitative spectral counting, a nonlabeling method hereafter referred to as DiPIUS-NL (Figure 4).

Identification of Candidate Substrates for F-box Proteins by DiPIUS

We applied DiPIUS to Fbxw7α, Skp2, and Fbxl5, three of the most well-characterized F-box proteins. SILAC ratios of proteins were normalized by the heavy/light ratio for each bait protein. We set the threshold for candidate substrates as a heavy/light ratio of $\geq$3 on the basis of the measurement of experimental error when equal amounts (1:1) of heavy- and light-labeled samples prepared from the same mCAT-HeLa cells expressing FLAG-tagged wild-type Skp2 were subjected to immunoprecipitation and analyzed; all of the detected proteins were included within the range of the ratio from 0.5 to 2.0 (Figure S1 in the Supporting Information). DiPIUS is based on the assumption that the expression levels of wild-type and mutant F-box proteins are identical. These levels are not perfectly identical in most cases, however, with small differences resulting in a slight shift in the baseline of the heavy/light ratio.

We identified many candidate proteins including previously known substrates such as c-Myc for Fbxw7α[9,10] (Figure 3C,J), p27 for Skp2[9,11] (Figure 3F,K), and IRP1 and IRP2 for Fbxl5[27,28] (Figure 3I,L). These substrates yielded almost the highest scores of SILAC analysis (heavy/light ratio) among the binding proteins. In contrast, β-actin, a representative non-specific protein, was found in almost equal amounts in the immunoprecipitates containing the wild-type or mutant F-box proteins (Figure 3A,D,G). Cul1, a component of SCF ubiquitin ligases, showed low scores on SILAC analysis in all experiments (Figure 3B,E,H), indicating that the mutant form of each F-box protein was unable to form a complete SCF complex. These results thus validated DiPIUS as a means to identify substrates for F-box proteins.

Given that only a few substrates for Fbxw7α, Skp2, and Fbxl5 were identified by the original DiPIUS system with SILAC, these F-box proteins were also subjected to DiPIUS-NL with spectral counting. With this approach, we identified from three independent experiments a total of 426, 379, and 515 proteins that associated with Fbxw7α, Skp2, and Fbxl5, respectively (Figure 4A and Tables S3 and S4 in the Supporting Information). Among these binding molecules, 405, 362, and 418 proteins, respectively, were eliminated because they were identified as proteins that interact with at least two of Fbxw7α, Skp2, and Fbxl5. The remaining 21, 17, and 97 proteins, respectively, were subjected to a more detailed analysis of the

dx.doi.org/10.1021/pr201216u | J. Proteome Res. 2012, 11, 3175–3185

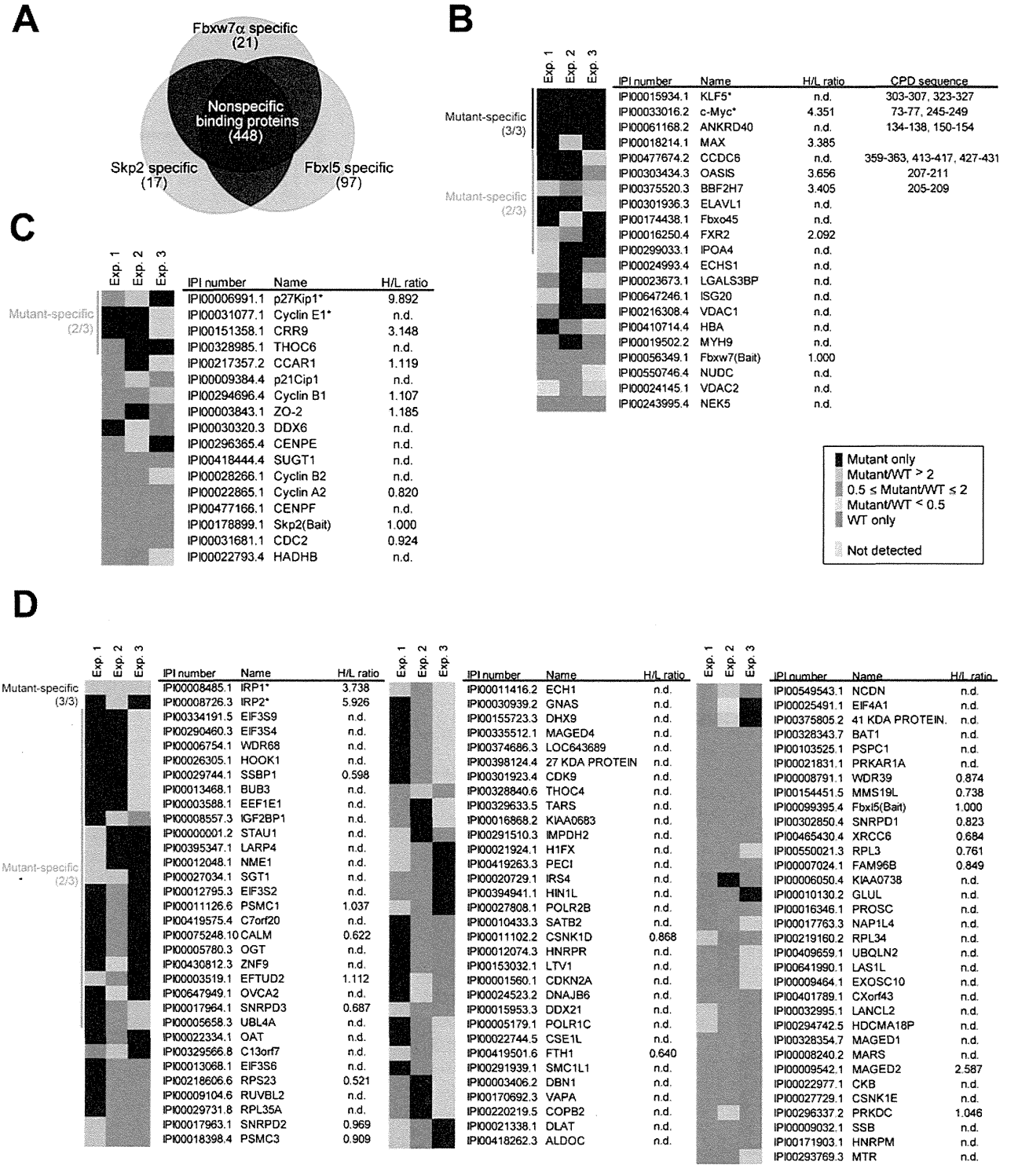


Figure 4. Candidate substrates for Fbxw7α, Skp2, and Fbxl5 analyzed by semiquantitative spectral counting. (A) Venn diagram for identified binding proteins for each F-box protein. (B–D) Specific binding proteins for Fbxw7α (B), Skp2 (C), and Fbxl5 (D) are shown. The ratios of the number of spectral counts for WT and mutant F-box proteins are indicated by the color scale. Proteins with a mutant/WT ratio of >2 in at least two of three independent experiments (Exp.) were considered binding proteins that associate to a greater extent with the mutant F-box protein than with the WT protein. The mutant (heavy)/WT (light) SILAC ratios (H/L ratios) obtained by original DiPIUS analysis are also shown. Proteins that contain a CPD sequence are indicated in (B). n.d., not detected. Asterisks indicate known substrates.

ratio of the spectral counts for the wild-type and mutant F-box proteins (spectral count score) in the individual experiments. We set the threshold as a score of >2 in at least two of three independent experiments for candidate substrates that preferentially bind to the mutant F-box protein. About 4.2% of proteins were classified as false positives at this threshold on the basis of the measurement of experimental error with mCAT-HeLa cells expressing wild-type Skp2 (Figure S2 in the Supporting Information). c-Myc was identified as a substrate for Fbxw7α (Figure 4B), consistent with our previous

observation that ablation of c-Myc reversed the overproliferation phenotype of Fbxw7-deficient thymocytes.[29] In addition, the transcription factor Krüppel-like factor 5 (KLF5), another known substrate for Fbxw7α,[30,31] yielded the highest score. For Skp2, both p27 and cyclin E1 were identified as candidate substrates (Figure 4C), consistent with previous observations.[32,33] IRP1 and IRP2 were identified as substrates for Fbxl5 with the highest scores (Figure 4D), consistent with the results of the original DiPIUS method (Figure 3L) as well as with our previous genetic data showing that IRP2 is the major substrate of Fbxl5 and that ablation of IRP2 prevents the embryonic death of Fbxl5-deficient mice.[34] SSBP1, PSMC1, CALM, EFTUD2, and SNRPD3 were also detected as candidate substrates of Fbxl5 by semiquantitative DiPIUS-NL, even though these proteins had a low heavy/light ratio in SILAC analysis (Figure 4D). DiPIUS-NL is therefore able to detect a wider spectrum of candidate substrates for F-box proteins but also yields more pseudopositive results compared with the original DiPIUS system with SILAC.

Among the 21 proteins remaining after elimination of 405 multibinders from the 426 total proteins that associated with Fbxw7α, 10 proteins were categorized as candidate substrates (score of >2) by DiPIUS-NL analysis. Six of these 10 proteins (60%) contain a consensus sequence for Fbxw7 binding, known as the Cdc4 phosphodegron (CPD) (Figure 5). In contrast,

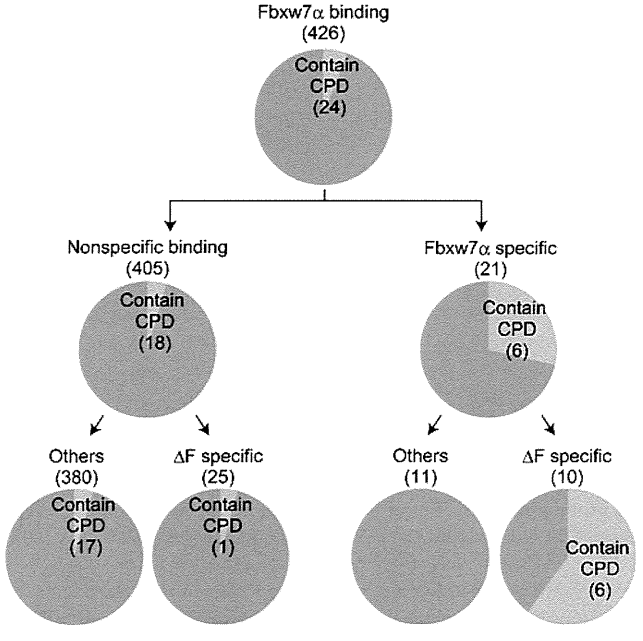


Figure 5. Number and distribution of Fbxw7α binding proteins in DiPIUS-NL analysis. Binding proteins for Fbxw7α were divided into two groups: "nonspecific binders", which also bound to Skp2 or Fbxl5 (or both), and "Fbxw7α-specific binders". The binding proteins were also classified as ΔF mutant-specific binders or not. Proteins that contain a CPD sequence are enriched among Fbxw7α(ΔF)-specific binders.

among the 405 nonspecific binders to Fbxw7α, 25 proteins were found to have a spectral count score of >2. However, only one (4%) of these 25 proteins contains the CPD sequence. These results suggest that elimination of the proteins that interact with multiple F-box proteins is effective for enrichment of real substrates. Proteins that do not contain the CPD sequence but which bound to the mutant form of Fbxw7α may thus represent false positives. For example, MAX, a binding partner of c-Myc, was detected by this approach as a false positive, probably because c-Myc preferentially associated with the Fbxw7α mutant. Together, these results suggested that DiPIUS is effective for the identification of the substrates of F-box proteins.

Enrichment Analysis of Specific Binding Proteins for F-box Proteins

We categorized the proteins identified by DiPIUS-NL as specific binding proteins for Fbxw7α, Skp2, or Fbxl5 with the use of the DAVID resource to highlight the most over-represented Gene Ontology (GO) annotation terms (>2-fold enrichment as compared with a *Homo sapiens* background; $P < 0.05$) for molecular functions (Figure 6A) or biological processes (Figure 6B). Such enrichment analysis revealed that the molecular function of "transcription factor activity" was significantly enriched (>4-fold) for the Fbxw7α-specific binding proteins (Figure 6A), consistent with the fact that known substrates for Fbxw7α including c-Myc, Notch, KLF5, c-Jun, TGIF, and SREBP are all classified in this category. These results suggested that Fbxw7α preferentially targets transcriptional factors for degradation. With regard to biological processes, the categories of "positive regulation of cellular process" and "catabolic process" were also enriched for Fbxw7α-specific binding proteins (Figure 6B). For the Skp2-specific binding proteins, the biological processes of "cell division", "cell cycle process", and "cell cycle" showed a >10-fold enrichment as compared with the general human background (Figure 6B), suggesting that Skp2 specifically associates with cell cycle regulators. This finding is also consistent with the fact that Skp2 targets the Cip/Kip family of cyclin-dependent kinase inhibitors (p21, p27, and p57) and G_1 cyclins (cyclins D1 and E1) for ubiquitylation.[9,11] For the Fbxl5-specific binding proteins, no annotation terms showed an enrichment of more than 3-fold (Figure 6). Together, these results suggested that some F-box proteins are specialized for specific cellular and molecular functions.

Identification of Different Sets of Substrates in Different Cell Types or under Different Conditions

Although many substrates have been previously identified for Fbxw7 and Skp2, only a subset of these substrates was identified by DiPIUS in mCAT-HeLa cells. We expected that if any of the substrates, corresponding kinases, or signaling molecules that activate the kinases are lacking in HeLa cells, then these substrates would not be detected by DiPIUS. To verify this assumption, we applied DiPIUS-NL to Fbxw7α in mouse cell lines including mHepa (hepatocyte cell line), Neuro2A (neuroblastoma cell line), and C2C12 (mesenchymal cell line). We identified other known substrates for Fbxw7α, including TGIF1, TGIF2, SREBP1, SREBP2, and c-Myb, in mHepa, Neuro2A, or C2C12 cells (Figure 7A and Table S5 in the Supporting Information), with none of these proteins having been identified in mCAT-HeLa cells. Whereas c-Myc was identified in all cell lines tested, KLF5 was detected by DiPIUS-NL only in mCAT-HeLa and C2C12 cells. c-Myb was identified only in Neuro2A cells.

We also applied DiPIUS-NL to Skp2 in mHepa and C2C12 cells, the latter of which were cultured under either growth- or differentiation-promoting conditions (Figure 7B and Tables S6 and S7 in the Supporting Information). Although p27 was detected in all cell lines tested, the spectrum of identified substrates differed substantially among mCAT-HeLa, mHepa,

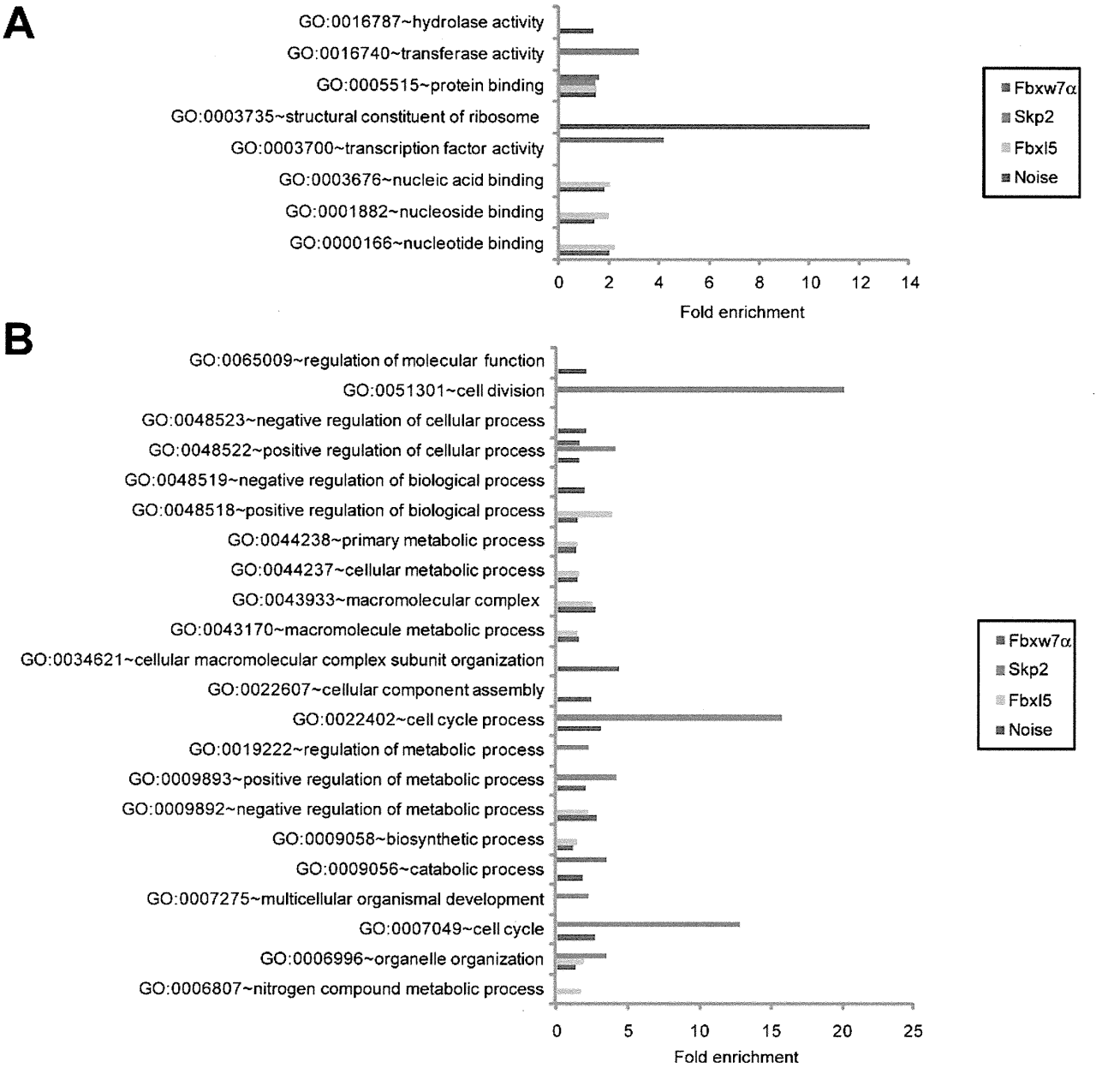


Figure 6. Functional analysis of differentially enriched proteins among binding proteins for each F-box protein. Fold enrichment among binding proteins for Fbxw7α, Skp2, or Fbxl5 of GO categories related to molecular functions (A) or biological processes (B) identified by DAVID ($P < 0.05$) are shown.

and C2C12 cells. Furthermore, cyclin E1, Cks1, Kif23, cyclin D3, Rnf12, and Mprip were identified as candidate substrates in C2C12 cells only under the growth condition, whereas Mypt1 and Gadd45α were identified as candidate substrates in these cells only under the differentiation (myoblast) condition. These results thus suggest that different sets of substrates are identified in different cell types or under different conditions by DiPIUS.

DISCUSSION

We have shown here that the DiPIUS approach identifies multiple substrates of a ubiquitin ligase in a comprehensive and unbiased manner. Several approaches to the identification of the substrates for mammalian ubiquitin ligases have been described.[35−37] However, there are several advantages of DiPIUS over such conventional methods that depend on simple purification of the substrates on the basis of their affinity for a given ubiquitin ligase. The principle of DiPIUS is based on a difference in the abundance of substrates between cells expressing wild-type or mutant F-box proteins. The many nonspecific proteins that are usually most problematic in conventional methods are efficiently eliminated by the DiPIUS system, given that such nonspecific proteins are expected to bind to wild-type and mutant F-box proteins to a similar extent. Our approach thus concentrates candidate substrates by eliminating many nonspecific proteins.

Most importantly, DiPIUS is applicable to any F-box protein. In addition to the SCF complex, DiPIUS has proved effective for the identification of substrates for the elongin B/C−Cul2−VHL-box protein (ECV) complex; we thus identified hypoxia-inducible factors (HIFs) as substrates of VHL[38,39] by DiPIUS (K. Y., M. M., and K. I. N., unpublished results). These

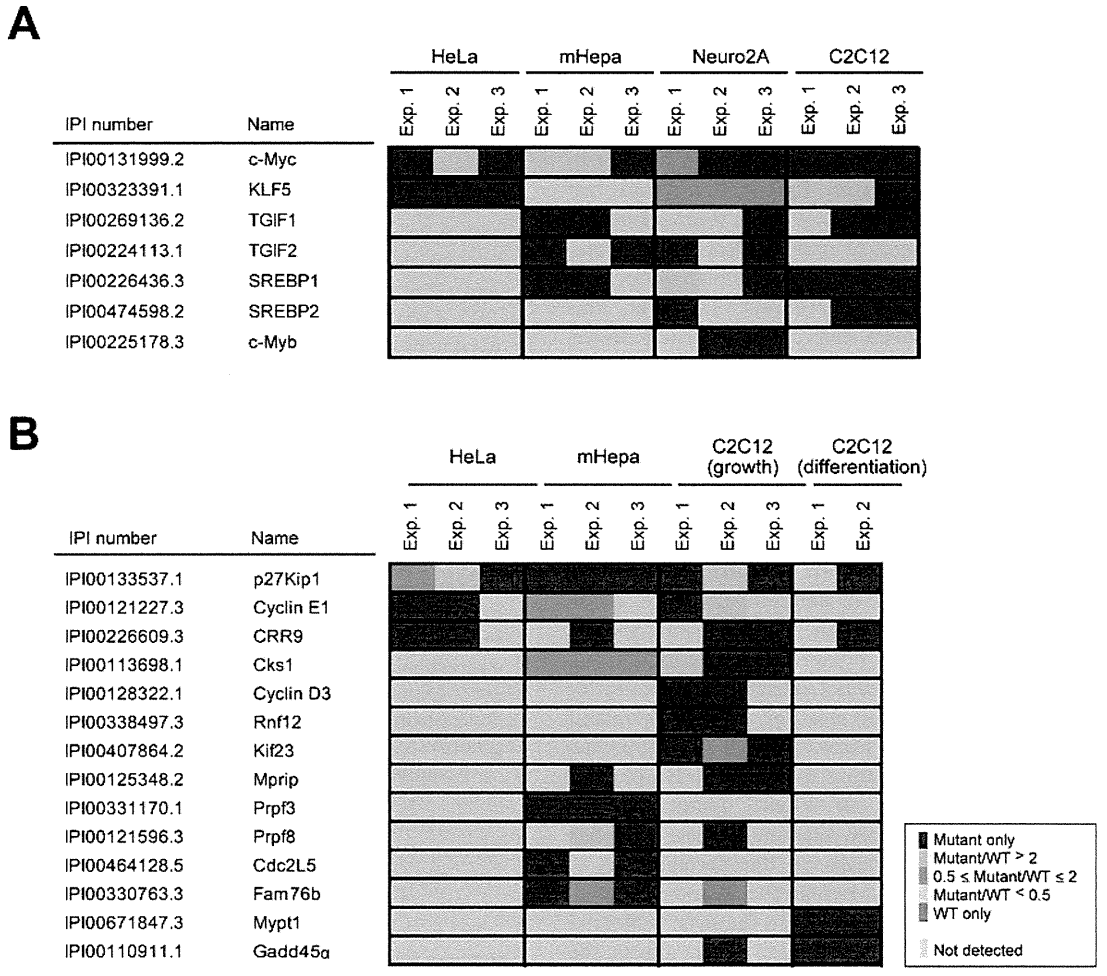


Figure 7. Known substrates of Fbxw7α and Skp2 analyzed by DiPIUS-NL in different cell types or under different conditions. (A) Known substrates for Fbxw7α detected by DiPIUS-NL in mCAT-HeLa (same as in Figure 4B), mHepa, Neuro2A, and C2C12 cells. The ratios of the number of spectral counts for WT and mutant F-box proteins are indicated by the color scale. Proteins with a mutant/WT ratio of >2 in at least two of three independent experiments (Exp.) were considered binding proteins that associate with the mutant F-box protein to a greater extent than with the WT protein. (B) Known and candidate substrates for Skp2 detected by DiPIUS-NL in mCAT-HeLa cells (same as in Figure 4C), mHepa cells, and C2C12 cells (cultured under growth- or myoblast differentiation-promoting conditions).

preliminary results suggest that DiPIUS may be applicable to all classes of cullin-based ubiquitin ligases. Moreover, in principle, DiPIUS could be applied to any ubiquitin ligase. A limitation of DiPIUS, which is based on protein degradation, is that it is unable to identify substrates for which ubiquitylation does not serve as a signal for degradation. DiPIUS appears to be highly sensitive, given that its application to F-box proteins has identified many low-abundance proteins including transcription factors. In addition, the number of cells required for DiPIUS is relatively small; we used 4×10^7 mCAT-HeLa cells per DiPIUS experiment. Furthermore, nonspecific binding proteins that usually render the discovery of bona fide substrates difficult in interaction-based technologies are efficiently removed by the comparison of protein profiles between wild-type and mutant ubiquitin ligases.

We have described two alternative forms of the DiPIUS system: original DiPIUS with SILAC and DiPIUS-NL with semiquantitative spectral counting. Original DiPIUS is effective for quantitation of the abundance of binding proteins for wild-type and mutant F-box proteins, although the number of such proteins identified is smaller than that detected by DiPIUS-NL. This latter problem is likely attributable to a greater complexity of the samples used for original DiPIUS as compared with those used for DiPIUS-NL; lysates of light isotope- and heavy isotope-labeled cells corresponding to proteins bound to wild-type and mutant F-box proteins, respectively, are thus mixed in original DiPIUS, and the immunoprecipitates derived therefrom are digested and analyzed directly without fractionation by SDS-PAGE and in-gel digestion. Although SDS-PAGE and in-gel digestion may reduce the complexity of samples, however, the increased processing may result in sample loss and a consequent reduction in both signal intensity and quantitative capacity. Furthermore, whereas the reduced sample complexity in DiPIUS-NL results in an increase in the number of candidate substrates identified, it also inevitably gives rise to a substantial number of false positives (~4.2%). We therefore propose that the two DiPIUS methods are complementary and should both be applied to obtain more comprehensive and precise results. Candidates identified by DiPIUS-NL analysis can be eliminated if their quantitation by original DiPIUS does not yield a high heavy/light score. Validation studies are necessary to verify that the candidates identified by DiPIUS-NL are indeed substrates of the ubiquitin ligase of interest.

Given that many substrates have been previously identified for Fbxw7 and Skp2, it was unexpected that only a subset of these substrates was identified by DiPIUS. There are several

dx.doi.org/10.1021/pr201216u | J. Proteome Res. 2012, 11, 3175–3185

possible explanations for this outcome. First, ubiquitylation requires that various conditions be satisfied, including that the substrate, ubiquitin ligase, corresponding kinase, and signals to activate the kinase be present in a cell (or a limited compartment of the cell) at the same time. The performance of DiPIUS in different cell types might thus be expected to identify different sets of substrates. Indeed, we found that different sets of substrates for Fbxw7α or Skp2 were discovered by DiPIUS in different cell types or under different conditions. Second, the sensitivity of DiPIUS might not be high enough for the detection of low-abundance proteins. However, this possibility seems unlikely given that, as mentioned above, DiPIUS identified many transcription factors that are expressed at a low level. Third, ubiquitylation of some of the proteins that were designated substrates on the basis of previous results might occur infrequently in the physiological setting. Consistent with this notion, some of these putative substrates were found not to accumulate in mice or cells that lack Skp2.[9,11]

As mentioned above, the application of DiPIUS to different cell types led to the identification of different substrates. These observations appear consistent with our previous genetic analyses of mice in which the Fbxw7 gene is conditionally ablated in a variety of tissues. Distinct subsets of substrates for Fbxw7 accumulate in such mutant mice in a tissue-dependent manner. For example, loss of Fbxw7 in the hematopoietic system results in c-Myc accumulation that leads to lymphomagenesis.[29,40] Hepatic ablation of *Fbxw7* results in the accumulation of sterol regulatory element-binding proteins (SREBPs) and Notch1, which leads to the development of steatohepatitis and hamartoma, whereas the abundance of c-Myc is not affected by the loss of Fbxw7 in the liver.[41] Notch1 accumulates in the Fbxw7-deficient brain, which affects the maintenance of neural stem cells,[42] whereas the abundance of c-Myc and SREBPs in the brain is not increased in this mutant during embryogenesis. Collectively, our biochemical and genetic studies suggest that a given ubiquitin ligase targets a different set of substrates in each cell or tissue type.

Comprehensive identification of substrates for a given ubiquitin ligase may provide valuable clues to the physiological functions of the enzyme. The DiPIUS system will thus likely provide important insight into the functions of ubiquitin ligases, the number of genes for which is estimated to be >500 in the human genome.[7]

■ ASSOCIATED CONTENT

Supporting Information

Supplementary Figures S1 and S2 and Tables S1–S7. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +81-92-642-6815. Fax: +81-92-642-6819. E-mail: nakayak1@bioreg.kyushu-u.ac.jp.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank T. Kitamura for pMX-puro; J. M. Cunningham and K. Hanada for the mCAT-1 plasmid; M. Oda, E. Koba, N. Nishimura, T. Takami, and other laboratory members for technical assistance; and M. Kimura and A. Ohta for help with preparation of the manuscript. This work was supported in part by a grant from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

■ REFERENCES

(1) Payen, A.; Persoz, J. F. Memoire sur la Diastase, les Principaux Produits de ses Reactions, et leurs Applications aux arts Industriels. *Ann. Chim. (Phys.)* **1833**, *53*, 73–92.

(2) Lohmann, K. Über die enzymatische Aufspaltung der Kreatinphosphorsäure; zugleich ein Beitrag zum Chemismus der Muskelkontraktion. *Biochem. Z* **1934**, *271*, 264–277.

(3) Irniger, S.; Piatti, S.; Michaelis, C.; Nasmyth, K. Genes involved in sister chromatid separation are needed for B-type cyclin proteolysis in budding yeast. *Cell* **1995**, *81* (2), 269–278.

(4) King, R. W.; Peters, J. M.; Tugendreich, S.; Rolfe, M.; Hieter, P.; Kirschner, M. W. A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. *Cell* **1995**, *81* (2), 279–288.

(5) Sudakin, V.; Ganoth, D.; Dahan, A.; Heller, H.; Hershko, J.; Luca, F. C.; Ruderman, J. V.; Hershko, A. The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol. Biol. Cell* **1995**, *6* (2), 185–197.

(6) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298* (5600), 1912–1934.

(7) Semple, C. A. The comparative proteomics of ubiquitination in mouse. *Genome Res.* **2003**, *13* (6B), 1389–1394.

(8) Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **1998**, *67*, 425–479.

(9) Nakayama, K. I.; Nakayama, K. Ubiquitin ligases: Cell-cycle control and cancer. *Nat. Rev. Cancer* **2006**, *6* (5), 369–381.

(10) Welcker, M.; Clurman, B. E. FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. *Nat. Rev. Cancer* **2008**, *8* (2), 83–93.

(11) Frescas, D.; Pagano, M. Deregulated proteolysis by the F-box proteins SKP2 and β-TrCP: Tipping the scales of cancer. *Nat. Rev. Cancer* **2008**, *8* (6), 438–449.

(12) Jin, J.; Cardozo, T.; Lovering, R. C.; Elledge, S. J.; Pagano, M.; Harper, J. W. Systematic analysis and nomenclature of mammalian F-box proteins. *Genes Dev.* **2004**, *18* (21), 2573–2580.

(13) Liao, E. H.; Hung, W.; Abrams, B.; Zhen, M. An SCF-like ubiquitin ligase complex that controls presynaptic differentiation. *Nature* **2004**, *430* (6997), 345–350.

(14) Saiga, T.; Fukuda, T.; Matsumoto, M.; Tada, H.; Okano, H. J.; Okano, H.; Nakayama, K. I. Fbxo45 forms a novel ubiquitin ligase complex and is required for neuronal development. *Mol. Cell. Biol.* **2009**, *29* (13), 3529–3543.

(15) Wolters, H.; Jurgens, G. Survival of the flexible: hormonal growth control and adaptation in plant development. *Nat. Rev. Genet.* **2009**, *10* (5), 305–317.

(16) Godinho, S. I.; Maywood, E. S.; Shaw, L.; Tucci, V.; Barnard, A. R.; Busino, L.; Pagano, M.; Kendall, R.; Quwailid, M. M.; Romero, M. R.; O'Neill, J.; Chesham, J. E.; Brooker, D.; Lalanne, Z.; Hastings, M. H.; Nolan, P. M. The after-hours mutant reveals a role for Fbxl3 in determining mammalian circadian period. *Science* **2007**, *316* (5826), 897–900.

(17) Siepka, S. M.; Yoo, S. H.; Park, J.; Song, W.; Kumar, V.; Hu, Y.; Lee, C.; Takahashi, J. S. Circadian mutant Overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression. *Cell* **2007**, *129* (5), 1011–1023.

(18) Busino, L.; Bassermann, F.; Maiolica, A.; Lee, C.; Nolan, P. M.; Godinho, S. I.; Draetta, G. F.; Pagano, M. SCF^{Fbxl3} controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins. *Science* **2007**, *316* (5826), 900–904.

(19) Albritton, L. M.; Tseng, L.; Scadden, D.; Cunningham, J. M. A putative murine ecotropic retrovirus receptor gene encodes a multiple

membrane-spanning protein and confers susceptibility to virus infection. *Cell* **1989**, *57* (4), 659–666.
(20) Matsumoto, M.; Hatakeyama, S.; Oyamada, K.; Oda, Y.; Nishimura, T.; Nakayama, K. I. Large-scale analysis of the human ubiquitin-related proteome. *Proteomics* **2005**, *5* (16), 4145–4151.
(21) Dennis, G., Jr.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **2003**, *4* (5), P3.
(22) Kitagawa, M.; Hatakeyama, S.; Shirane, M.; Matsumoto, M.; Ishida, N.; Hattori, K.; Nakamichi, I.; Kikuchi, A.; Nakayama, K.; Nakayama, K. An F-box protein, FWD1, mediates ubiquitin-dependent proteolysis of β-catenin. *EMBO J.* **1999**, *18* (9), 2401–2410.
(23) Kamura, T.; Hara, T.; Kotoshiba, S.; Yada, M.; Ishida, N.; Imaki, H.; Hatakeyama, S.; Nakayama, K.; Nakayama, K. I. Degradation of $p57^{Kip2}$ mediated by SCF^{Skp2}-dependent ubiquitylation. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (18), 10231–10236.
(24) Schulman, B. A.; Carrano, A. C.; Jeffrey, P. D.; Bowen, Z.; Kinnucan, E. R.; Finnin, M. S.; Elledge, S. J.; Harper, J. W.; Pagano, M.; Pavletich, N. P. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* **2000**, *408* (6810), 381–386.
(25) Kamura, T.; Maenaka, K.; Kotoshiba, S.; Matsumoto, M.; Kohda, D.; Conaway, R. C.; Conaway, J. W.; Nakayama, K. I. VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. *Genes Dev.* **2004**, *18* (24), 3055–3065.
(26) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376–386.
(27) Salahudeen, A. A.; Thompson, J. W.; Ruiz, J. C.; Ma, H. W.; Kinch, L. N.; Li, Q.; Grishin, N. V.; Bruick, R. K. An E3 ligase possessing an iron-responsive hemerythrin domain is a regulator of iron homeostasis. *Science* **2009**, *326* (5953), 722–726.
(28) Vashisht, A. A.; Zumbrennen, K. B.; Huang, X.; Powers, D. N.; Durazo, A.; Sun, D.; Bhaskaran, N.; Persson, A.; Uhlen, M.; Sangfelt, O.; Spruck, C.; Leibold, E. A.; Wohlschlegel, J. A. Control of iron homeostasis by an iron-regulated ubiquitin ligase. *Science* **2009**, *326* (5953), 718–721.
(29) Onoyama, I.; Tsunematsu, R.; Matsumoto, A.; Kimura, T.; de Alboran, I. M.; Nakayama, K.; Nakayama, K. I. Conditional inactivation of Fbxw7 impairs cell-cycle exit during T cell differentiation and results in lymphomatogenesis. *J. Exp. Med.* **2007**, *204* (12), 2875–2888.
(30) Liu, N.; Li, H.; Li, S.; Shen, M.; Xiao, N.; Chen, Y.; Wang, Y.; Wang, W.; Wang, R.; Wang, Q.; Sun, J.; Wang, P. The Fbw7/human CDC4 tumor suppressor targets proproliferative factor KLF5 for ubiquitination and degradation through multiple phosphodegron motifs. *J. Biol. Chem.* **2010**, *285* (24), 18858–18867.
(31) Zhao, D.; Zheng, H. Q.; Zhou, Z.; Chen, C. The Fbw7 tumor suppressor targets KLF5 for ubiquitin-mediated degradation and suppresses breast cell proliferation. *Cancer Res.* **2010**, *70* (11), 4728–4738.
(32) Nakayama, K.; Nagahama, H.; Minamishima, Y. A.; Matsumoto, M.; Nakamichi, I.; Kitagawa, K.; Shirane, M.; Tsunematsu, R.; Tsukiyama, T.; Ishida, N.; Kitagawa, M.; Nakayama, K. I.; Hatakeyama, S. Targeted disruption of *Skp2* results in accumulation of cyclin E and $p27^{Kip1}$, polyploidy and centrosome overduplication. *EMBO J.* **2000**, *19* (9), 2069–2081.
(33) Nakayama, K.; Nagahama, H.; Minamishima, Y. A.; Miyake, S.; Ishida, N.; Hatakeyama, S.; Kitagawa, M.; Iemura, S.; Natsume, T.; Nakayama, K. I. Skp2-mediated degradation of p27 regulates progression into mitosis. *Dev. Cell* **2004**, *6* (5), 661–672.
(34) Moroishi, T.; Nishiyama, M.; Takeda, Y.; Iwai, K.; Nakayama, K. I. The FBXL5-IRP2 axis is integral to control of iron metabolism in vivo. *Cell Metab.* **2011**, *14* (3), 339–351.
(35) Yen, H. C.; Elledge, S. J. Identification of SCF ubiquitin ligase substrates by global protein stability profiling. *Science* **2008**, *322* (5903), 923–929.
(36) Merbl, Y.; Kirschner, M. W. Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (8), 2543–2548.
(37) Burande, C. F.; Heuze, M. L.; Lamsoul, I.; Monsarrat, B.; Uttenweiler-Joseph, S.; Lutz, P. G. A label-free quantitative proteomics strategy to identify E3 ubiquitin ligase substrates targeted to proteasome degradation. *Mol. Cell Proteomics* **2009**, *8* (7), 1719–1727.
(38) Jaakkola, P.; Mole, D. R.; Tian, Y. M.; Wilson, M. I.; Gielbert, J.; Gaskell, S. J.; Kriegsheim, A.; Hebestreit, H. F.; Mukherji, M.; Schofield, C. J.; Maxwell, P. H.; Pugh, C. W.; Ratcliffe, P. J. Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O_2-regulated prolyl hydroxylation. *Science* **2001**, *292* (5516), 468–472.
(39) Ivan, M.; Kondo, K.; Yang, H.; Kim, W.; Valiando, J.; Ohh, M.; Salic, A.; Asara, J. M.; Lane, W. S.; Kaelin, W. G., Jr. HIFα targeted for VHL-mediated destruction by proline hydroxylation: Implications for O_2 sensing. *Science* **2001**, *292* (5516), 464–468.
(40) Matsuoka, S.; Oike, Y.; Onoyama, I.; Iwama, A.; Arai, F.; Takubo, K.; Mashimo, Y.; Oguro, H.; Nitta, E.; Ito, K.; Miyamoto, K.; Yoshiwara, H.; Hosokawa, K.; Nakamura, Y.; Gomei, Y.; Iwasaki, H.; Hayashi, Y.; Matsuzaki, Y.; Nakayama, K.; Ikeda, Y.; Hata, A.; Chiba, S.; Nakayama, K. I.; Suda, T. Fbxw7 acts as a critical fail-safe against premature loss of hematopoietic stem cells and development of T-ALL. *Genes Dev.* **2008**, *22* (8), 986–991.
(41) Onoyama, I.; Suzuki, A.; Matsumoto, A.; Tomita, K.; Katagiri, H.; Oike, Y.; Nakayama, K.; Nakayama, K. I. Fbxw7 regulates lipid metabolism and cell fate decisions in the mouse liver. *J. Clin. Invest.* **2011**, *121* (1), 342–354.
(42) Matsumoto, A.; Onoyama, I.; Sunabori, T.; Kageyama, R.; Okano, H.; Nakayama, K. I. Fbxw7-dependent degradation of Notch is required for control of "stemness" and neuronal-glial differentiation in neural stem cells. *J. Biol. Chem.* **2011**, *286* (15), 13754–13764.

Journal of proteome research

ARTICLE

pubs.acs.org/jpr

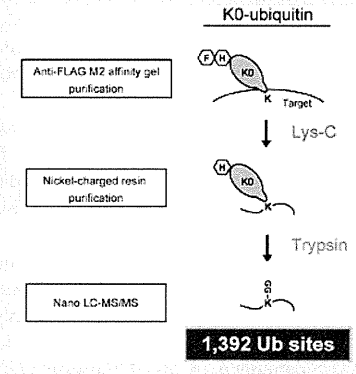
Proteome-wide Identification of Ubiquitylation Sites by Conjugation of Engineered Lysine-less Ubiquitin

Kiyotaka Oshikawa, Masaki Matsumoto, Koji Oyamada, and Keiichi I. Nakayama*

Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan, and CREST, Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan

Supporting Information

ABSTRACT: Ubiquitin conjugation (ubiquitylation) plays important roles not only in protein degradation but also in many other cellular functions. However, the sites of proteins that are targeted for such modification have remained poorly characterized at the proteomic level. We have now developed a method for the efficient identification of ubiquitylation sites in target proteins with the use of an engineered form of ubiquitin (K0-Ub), in which all seven lysine residues are replaced with arginine. K0-Ub is covalently attached to lysine residues of target proteins via an isopeptide bond, but further formation of a polyubiquitin chain does not occur on K0-Ub. We identified a total of 1392 ubiquitylation sites of 794 proteins from HEK293T cells. Profiling of ubiquitylation sites indicated that the sequences surrounding lysine residues targeted for ubiquitin conjugation do not share a common motif or structural feature. Furthermore, we identified a critical ubiquitylation site of the cyclin-dependent kinase inhibitor $p27^{Kip1}$. Mutation of this site thus inhibited ubiquitylation of and stabilized $p27^{Kip1}$, suggesting that this lysine residue is the target site of $p27^{Kip1}$ for ubiquitin conjugation in vivo. In conclusion, our method based on K0-Ub is a powerful tool for proteome-wide identification of ubiquitylation sites of target proteins.



KEYWORDS: ubiquitylation, posttranslational modification, ubiquitin profiling, lysine-less ubiquitin, mass spectrometry, $p27^{Kip1}$

INTRODUCTION

Conjugation of ubiquitin (ubiquitylation) to protein substrates plays an integral role in regulation of diverse physiological processes and in maintenance of cellular homeostasis in eukaryotes.[1−3] Ubiquitin is a highly conserved protein of 76 amino acids and is covalently conjugated to substrates by a cascade of reactions catalyzed by a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2), and a ubiquitin-protein isopeptide ligase (E3).[4] The COOH-terminal glycine residue of ubiquitin is covalently linked through an isopeptide bond to the ε-amino group of a lysine residue (or residues) in the substrate. It is likely that many thousands of human proteins are ubiquitylated in vivo in a highly regulated and spatiotemporal-dependent manner.

Recent technological advances in mass spectrometry (MS)-based protein identification have made it possible to detect various types of posttranslational modifications.[5−7] In the case of ubiquitylation, trypsin digestion of a ubiquitin-conjugated protein yields a "ubiquitin signature peptide," in which tandem glycine residues derived from the COOH-terminus of ubiquitin are conjugated at the lysine residue of the substrate.[8,9] Conjugation of ubiquitin to the lysine residue gives rise to a mass shift (+114.0429 Da) and miscleavage by trypsin, resulting in the generation of a T-shaped peptide.[10] On the basis of these characteristics, ubiquitin modification sites are identified as unique tandem MS (MS/MS) spectra in database searches. Approaches to the comprehensive identification of protein ubiquitylation sites have included those based on the use of (i) affinity-tagged ubiquitin,[9,11−14] (ii) antibodies to (anti-) ubiquitin,[15−18] (iii) ubiquitin-binding proteins,[11,19−21] and (iv) antibodies specific for the ubiquitin remnant.[22] However, large-scale identification of ubiquitylated sites has been unsuccessful, because ubiquitin forms polyubiquitin chains that are heterogeneous in length and in topology, making it difficult to identify ubiquitylated sites of target proteins. Moreover, many ubiquitylated proteins undergo degradation by the ubiquitin−proteasome system; the turnover of these unstable intermediates is rapid, with the result that their steady-state levels are low. Despite the biological and clinical importance of ubiquitylation, our knowledge of ubiquitylated proteins and their modification sites thus remains limited. An effective method to enrich ubiquitin-conjugated proteins and to identify the ubiquitylation sites therefore remains highly desirable.

Many groups including ours have attempted to develop MS-based approaches to the comprehensive identification of ubiquitylation sites. Peng et al. identified 110 ubiquitylation sites in yeast with the use of hexahistidine (His_6)-tagged ubiquitin.[9] Recent development of anti-diglycine markedly improved the efficiency of identification of ubiquitylation sites, but it has still been difficult to identify >1000 ubiquitylation sites at a time. Furthermore, the anti-diglycine method is unable to discriminate ubiquitylation from other modifications such as conjugation with the ubiquitin-like protein Nedd8 (neddylation).[22]

We have now developed an effective method for the identification of ubiquitylation sites based on an engineered lysine-less

Received: July 16, 2011
Published: November 05, 2011

form of ubiquitin (K0-Ub), and with this approach we have determined 1392 ubiquitylation sites of proteins from HEK293T cells. Wild-type ubiquitin (WT-Ub) contains seven lysine residues, all of which can be used for the formation of diverse ubiquitin chain linkages. Whereas K0-Ub is unable to form ubiquitin chains because of the lack of the ε-amino group of lysine, it can be linked to lysine residues of target proteins via an isopeptide bond. Two-step digestion of K0-Ub-modified proteins with lysyl endopeptidase (Lys-C) and trypsin coupled with affinity purification allowed us to efficiently enrich the ubiquitin signature peptides derived from target proteins. This method thus allows the large-scale and precise identification of ubiquitylation sites. Our results also provide insight into the characteristics of ubiquitylation sites, and we validated the approach by analysis of cyclin-dependent kinase (CDK) inhibitor $p27^{Kip1}$.

METHODS

Reagents

Ubiquitin aldehyde (Ub-CHO), leupeptin, and MG132 were obtained from Peptide Institute (Osaka, Japan). Iodoacetamide (IAA), aprotinin, and phenylmethylsulfonyl fluoride (PMSF) were from Wako (Osaka, Japan). Thymidine, cycloheximide (CHX), and other reagents, unless indicated otherwise, were obtained from Sigma (St. Louis, MO).

Antibodies

Antibodies to ubiquitin were obtained from Abcam (Cambridge, United Kingdom); those to the FLAG epitope (M2) were from Sigma; those to cyclin D1 and to glutathione *S*-transferase (GST) were from MBL (Nagoya, Japan); those to $p27^{Kip1}$ and to HSP90 were from BD Biosciences (San Jose, CA); and those to the hemagglutinin epitope (HA11) were from Covance (Princeton, NJ).

Cell Culture and Transfection

Human embryonic kidney 293T (HEK293T) and HeLa cells were cultured under an atmosphere of 5% CO_2 at 37 °C in Dulbecco's modified Eagle's medium (DMEM, Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (Invitrogen). Cell transfection was performed with the use of the FuGENE 6 reagent (Roche, Indianapolis, IN).

Immunoblot Analysis

Cell lysis and immunoblot analysis were performed as described.[23] Briefly, cells were lysed in buffer containing 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.5% Triton X-100, aprotinin (10 μg/mL), leupeptin (10 μg/mL), 1 mM PMSF, 0.4 mM Na_3VO_4, 0.4 mM EDTA, 10 mM NaF, and 10 mM sodium pyrophosphate. The lysate was centrifuged at 20000*g* for 15 min at 4 °C, and the resulting supernatant was separated by SDS–polyacrylamide gel electrophoresis (PAGE), transferred to a Immobilon-P membrane (Millipore, Billerica, MA), and subjected to immunoblot analysis. Immune complexes were detected with Supersignal West Pico chemiluminescence reagent (Pierce, Waltham, MA).

Preparation of Ubiquitin Signature Peptides

The cDNAs encoding FLAG- and His_6-tagged WT-Ub or K0-Ub were subcloned into the pcDNA3 mammalian expression plasmid (Invitrogen). Transfected HEK293T cells from 10 15-cm Petri dishes were harvested, washed twice with ice-cold phosphate-buffered saline, and lysed in 20 mL of buffer A [50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.5% Triton X-100, 25 mM IAA, aprotinin (10 μg/mL), leupeptin (20 μg/mL), 1 mM PMSF, 0.4 mM Na_3VO_4, 0.4 mM EDTA, 10 mM NaF, 10 mM sodium pyrophosphate]. The lysate was centrifuged at 20 000*g* for 15 min at 4 °C, and the resulting supernatant was mixed with anti-FLAG M2 affinity gel (500 μL, Sigma) with rotation for 30 min at 4 °C. The affinity gel was then washed with buffer B [25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.25% Triton X-100, 1 mM PMSF, 0.4 mM Na_3VO_4, 0.4 mM EDTA, 10 mM NaF, 10 mM sodium pyrophosphate] and subjected to elution with 2 mL of the FLAG peptide (0.5 mg/mL, Sigma) in buffer B. Proteins in the eluate were denatured with 8 M urea, after which the sample was diluted to 4 M urea and Lys-C (1:100 w/w, Wako) was added. The digested sample (4 mL) was mixed with 50 μL of nickel-charged resin (ProBond resin, Invitrogen) by rotation for 1 h at 4 °C, after which the resin was washed with buffer C [25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 100 mM imidazole, 50% acetonitrile] and then suspended directly in SDS sample buffer. The His_6-K0-Ub conjugates were separated by SDS–PAGE on a 15% gel and stained with Coomassie brilliant blue G-250. Gel pieces containing the stained proteins were treated first with 10 mM dithiothreitol at 56 °C for 45 min and then with 55 mM IAA at 25 °C for 30 min. In-gel digestion by trypsin (Promega, Madison, WI) and peptide extraction were performed as described.[24] The samples were then desalted with the use of StageTips with C18 Empore disk membranes (3M, St. Paul, MN).[25]

MS Analysis and Database Search

The obtained peptides were dried and then dissolved in a solution containing 0.1% trifluoroacetic acid and 2% acetonitrile before nanoscale liquid chromatography (nanoLC)–MS/MS analysis with a system consisting of an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Waltham, MA) coupled with a nanoLC instrument (Advance, Michrom BioResources, Auburn, CA) and HTC-PAL autosampler (CTC Analytics, Zwingen, Switzerland). Peptide separation was performed with an in-house pulled fused silica capillary (internal diameter, 0.1 mm; length, 15 cm; tip internal diameter, 0.05 mm) packed with 3-μm C_{18} L-column (Chemicals Evaluation and Research Institute, Japan). The mobile phases consisted of 0.1% formic acid (A) and 100% acetonitrile (B). Peptides were eluted with a gradient of 5–35% B for 180 min at a flow rate of 200 nL/min. Collision-induced dissociation (CID) spectra were acquired automatically in the data-dependent scan mode with the dynamic exclusion option. Full MS spectra were obtained with Orbitrap in the mass/charge (m/z) range of 300–2000 with a resolution of 30 000. The nine most intense precursor ions in the full MS spectra were selected for subsequent ion-trap MS/MS analysis with the automated gain control mode. Lock mass function was activated to minimize mass error during analysis. The peak lists were generated by MSn.exe (Thermo Fisher Scientific) with a minimum scan/group value of 1 and were compared with in-house-curated target/decoy Human International Protein Index version 3.16 database (IPI, 57 366 protein sequences; European Bioinformatics Institute) with the use of the MASCOT algorithm (ver. 2.1.4). Trypsin was selected as the enzyme used, the allowed number of missed cleavages was set at 2, and carbamidomethylation on cysteine was selected as the fixed modification. Oxidized methionine, NH_2-pyroglutamine, lysine modified with diglycine, and protein modified with diglycine at its NH_2-terminus were searched as variable modifications. Precursor mass tolerance was 10 ppm, and tolerance of MS/MS ions was 0.5 Da. The threshold used for peptide identification was a MASCOT

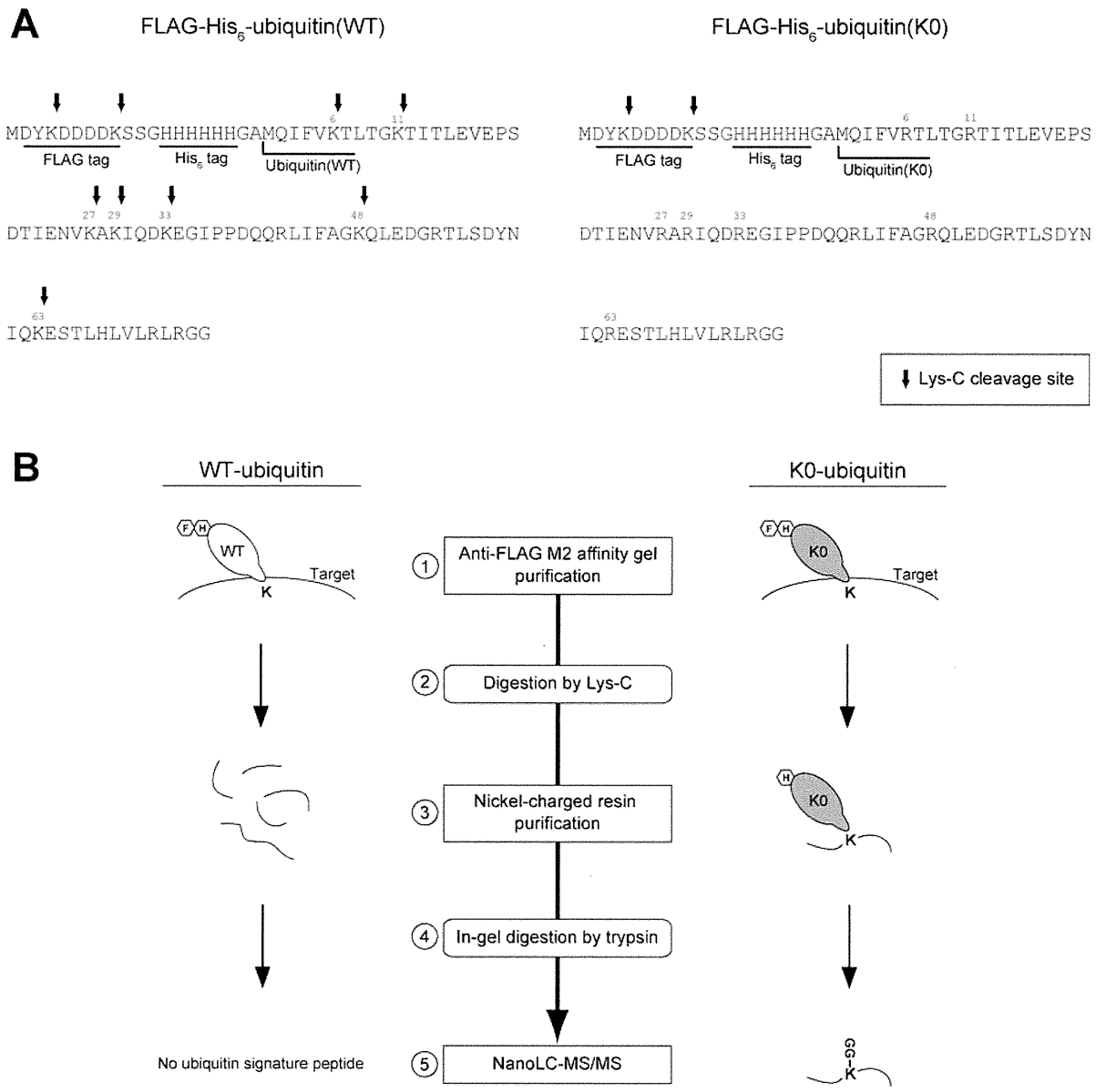


Figure 1. Strategy to identify ubiquitylation sites with K0-Ub. (A) Amino acid sequences of wild-type ubiquitin (WT-Ub) and lysine-less mutant ubiquitin (K0-Ub) tagged with tandem FLAG and His_6 epitopes. Arrows indicate the positions of potential Lys-C cleavage sites. (B) Cells are transfected with plasmids for expression of either WT-Ub (control) or K0-Ub tagged with the FLAG-His_6 tandem epitopes. The cells are then lysed, and ubiquitylated proteins are purified with an anti-FLAG column. Enriched ubiquitylated proteins are digested with Lys-C under denaturing conditions. WT-Ub is digested by Lys-C, and so no ubiquitin signature peptides should be recovered. On the other hand, Lys-C-resistant K0-Ub and its conjugated target protein fragments are further purified with nickel-charged resin. Enriched K0-Ub conjugates are separated by SDS-PAGE, subjected to in-gel digestion with trypsin, and analyzed by nanoLC-MS/MS.

score of ≥30 and a delta score (the difference between first-assigned and second-assigned peptides) of ≥18. With these criteria, the false positive rate was <1%.

Identification of Abundant Proteins with emPAI

To compare ubiquitylated proteins identified in the study with abundant proteins of HEK293T cells, we constructed an abundant protein list with the use of the exponentially modified protein abundance index (emPAI). HEK293T cells were lysed in a solution containing 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.5% Triton X-100, aprotinin (10 μg/mL), leupeptin (20 μg/mL), 1 mM PMSF, 0.4 mM Na_3VO_4, 0.4 mM EDTA, 10 mM NaF, and 10 mM sodium pyrophosphate. The lysate was centrifuged at 20000*g* for 15 min at 4 °C, and proteins in the resulting supernatant were fractionated by SDS-PAGE on a 12% gel and stained with Coomassie brilliant blue G-250. Each lane of the gel was cut into 16 slices, which were treated first with 10 mM DTT at 56 °C for 45 min and then with 55 mM IAA at 25 °C for 30 min. In-gel digestion with trypsin and peptide extraction were performed, and the resulting peptides were analyzed by MS. With the use of a MASCOT database search, we counted the number of observed unique parent ions per protein. The emPAI value was calculated from the number of observed unique parent ions per protein and the number of observable peptides per protein. Ishihama et al.

previously defined emPAI and showed that it is directly proportional to protein content.[26,27]

Bioinformatics Analysis

The sequences for 10 amino acids on each side of ubiquitylated lysines were extracted from the entire protein sequences. To compare the sequence logo around the identified ubiquitylation sites with the ~10 000 lysine residues randomly assigned from human IPI database version 3.16, we used WebLogo (http://weblogo.berkeley.edu).

To access the secondary structure types for ubiquitylated lysine residues, we searched for crystal structures of all the ubiquitylated proteins in the Protein Database Bank (PDB). A total of 151 PDB structures contained lysines that we found are susceptible to ubiquitylation (290 ubiquitylated lysines and 2463 total lysines). In cases where multiple PDB structures for a ubiquitylated protein were described, the structure of the best quality was used.

Subcellular localization of ubiquitylated proteins was analyzed and clustered with the use of the Gene Ontology (GO) term mapper after conversion of the Entrez Gene ID numbers of identified ubiquitylated proteins to UniProt Knowledgebase numbers by the UniProt ID mapping tool. A total of 1152 cellular components was found in the database, and the category was further grouped into five classes. Some proteins were predicted to have multiple subcellular localizations, whereas others had no subcellular information available in the database.

Statistical Analysis

To determine whether the frequencies of ubiquitylation sites and ubiquitylated proteins were significantly higher than the frequencies of total lysine sites and proteins, we compared the frequencies using the normal approximation to the binomial distribution. A p value of <0.05 was considered statistically significant.

In Vitro Ubiquitylation Assay with Biotinylated K0-Ub

To prepare proteins modified with biotinylated ubiquitin, we subcloned the cDNA for WT-Ub or K0-Ub with an added NH_2-terminal cysteine residue as a biotinylation site into the pGEX-6P vector (GE Healthcare, Little Chalfont, United Kingdom). The encoded GST-tagged WT-Ub or K0-Ub proteins were expressed in *Escherichia coli* strain BL21(DE3) pLys(S) (Novagen, Darmstadt, Germany) cultured in the presence of 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG). The bacterial cells were resuspended in 50 mM Tris-HCl (pH 7.5) containing 500 mM NaCl and then lysed by ultrasonic treatment (Branson Sonifier, Teltow, Germany). The lysate was centrifuged at 20 000g for 10 min at 4 °C to remove debris, and the resulting supernatant was mixed with glutathione-Sepharose 4B (GE Healthcare) with rotation at 4 °C for 30 min. The beads were then washed with 50 mM Tris-HCl (pH 7.5) containing 500 mM NaCl, and the GST-tagged proteins were eluted with 50 mM Tris-HCl (pH 7.5) containing 10 mM reduced glutathione. The GST tag was cleaved from the isolated proteins with PreScission Protease (GE Healthcare). The purity of the recombinant proteins was confirmed by SDS-PAGE and staining with Coomassie brilliant blue G-250 (data not shown). The purified samples were biotinylated with the use of EZ-Link iodoacetyl-PEO_2-biotin (Pierce). To confirm the biotinylation, we measured the molecular mass of biotinylated WT-Ub and K0-Ub by MALDI-TOF MS analysis (data not shown).

Biotinylated WT-Ub or K0-Ub (500 ng) was incubated for various times at 30 °C in a final volume of 20 μL with E1 (100 ng,

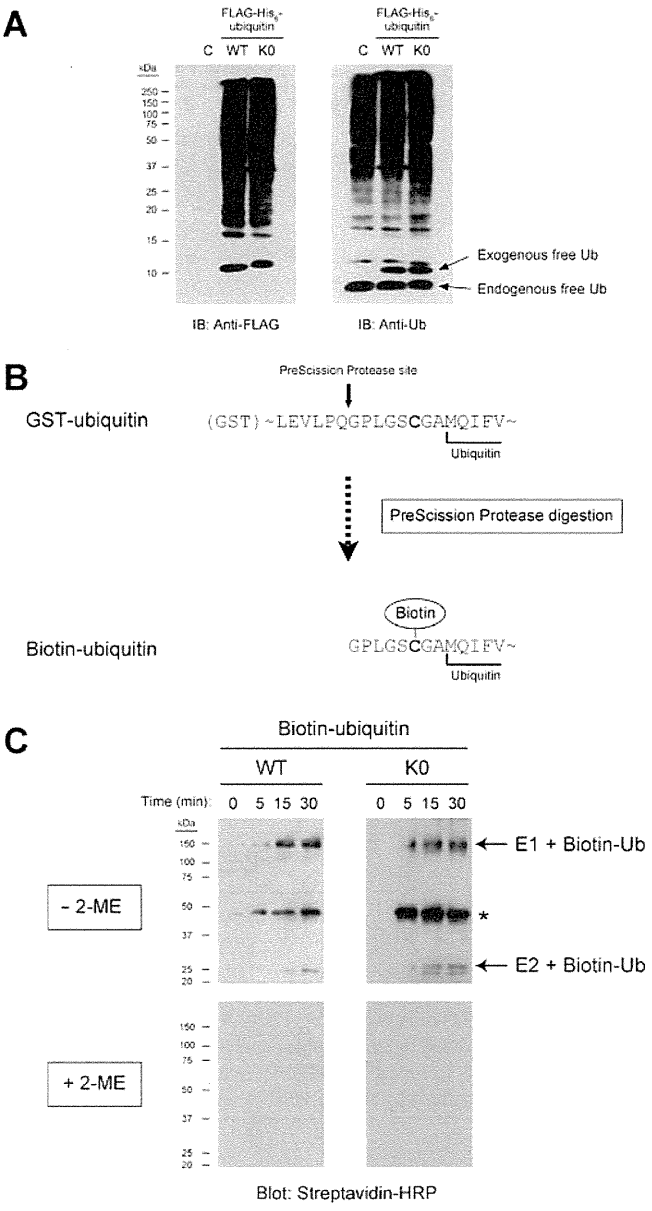


Figure 2. Effect of K0-Ub expression on cellular ubiquitylation. (A) HEK293T cells were transfected with plasmids encoding FLAG-His_6-tagged forms of WT-Ub or K0-Ub, or with the corresponding empty vector (C), after which cell lysates were subjected to immunoblot (IB) analysis with anti-FLAG and antiubiquitin. (B) Preparation of biotinylated ubiquitin. Ubiquitin fused with GST as well as an intervening PreScission Protease site and additional cysteine residue as a biotinylation target was expressed in *E. coli* and purified with glutathione–Sepharose 4B. The GST tag was cleaved with PreScission Protease, and the remaining protein was biotinylated with the use of EZ-Link iodoacetyl-PEO_2-biotin. (C) Biotinylated WT-Ub or K0-Ub was incubated with E1, E2, and an ATP-regenerating system for the indicated times, after which the reaction mixture was subjected to SDS-PAGE in the absence or presence of 2-ME followed by blot analysis with streptavidin-HRP. Asterisk indicates a nonspecific band.

Boston Biochem, Cambridge, MA) and E2 (200 ng of UbcH5c[28] in the presence of an ATP-regenerating system [25 mM Tris-HCl (pH 7.5), 120 mM NaCl, 1 mM $MgCl_2$, 2 mM ATP, 1 mM creatine phosphate, phosphocreatine kinase (0.5 U/mL)]. The reaction was terminated by the addition of SDS sample buffer,