

# Global Analysis of Gel Mobility of Proteins and Its Use in Target Identification<sup>\*S</sup>

Received for publication, November 9, 2007, and in revised form, February 20, 2008. Published, JBC Papers in Press, February 21, 2008, DOI 10.1074/jbc.M709211200

Atsuko Shirai<sup>‡§¶1</sup>, Akihisa Matsuyama<sup>‡¶1,2</sup>, Yoko Yashiroda<sup>‡¶</sup>, Atsushi Hashimoto<sup>‡¶</sup>, Yumi Kawamura<sup>‡¶</sup>, Ritsuko Arai<sup>‡¶</sup>, Yasuhiko Komatsu<sup>||3</sup>, Sueharu Horinouchi<sup>§</sup>, and Minoru Yoshida<sup>‡§¶</sup>

From the <sup>‡</sup>Chemical Genetics Laboratory, RIKEN, Wako, Saitama 351-0198, the <sup>§</sup>Department of Biotechnology, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, the <sup>¶</sup>CREST Research Project, Japan Science and Technology Agency, Saitama 332-0012, and the <sup>||</sup>Advanced Life Science Institute Inc., Wako, Saitama 351-0012, Japan

SDS-PAGE is a basic method that has long been used for separation of proteins according to their molecular sizes. Despite its simplicity, it provides information on characteristics of proteins beyond their molecular masses because gel mobility of proteins often reflects their physicochemical properties and post-translational modifications. Here we report on a global analysis of gel mobility of the proteome, which we term the “mobilitome,” covering 93.4% of the fission yeast proteome. To our surprise, more than 40% of proteins did not migrate to their calculated positions. Statistical analyses revealed that the discrepancy was largely dependent on the hydrophobicity of proteins. This experimental data set, with a high coverage rate of real mobility, made it feasible to identify proteins detected on the gel without using any specialized techniques. This approach enabled us to detect previously unknown post-translational modifications of a protein; for example, we revealed that eIF5A is novel substrate of a Sir2-related deacetylase Hst2. Furthermore, we concomitantly identified twelve acetylated and eight methylated proteins using specific anti-acetylated and anti-methylated lysine antibodies, most of which had not been known to be subject to the modifications. Thus, we propose the general usefulness of the mobilitome and electrophoresis-based methodology for the identification and characterization of proteins detected on the gel.

Gel mobility analysis of proteins often provides useful information on their physicochemical properties and post-translational modifications and can sometimes provide important insights into the function and regulation of a given protein. SDS-PAGE has been long and widely used for separation and size determination of proteins. In the early 1960s, polyacrylamide gels became appreciated as a convenient and versatile alter-

native to starch gels (1, 2). It was subsequently demonstrated that in the presence of SDS, proteins migrate according to the molecular weight of their polypeptide chains (3, 4). However, the band resolution of the technique was relatively low. Therefore, this system, which uses only one phosphate buffer for the tanks and the gel, has lost its popularity despite its simplicity. Meanwhile, Laemmli modified the Ornstein and Davis system by adding SDS to the Tris/glycine buffer, in which a nonrestrictive large pore gel called a stacking gel is put on the top of a separating gel, and showed that proteins could be reliably separated with high resolution (1, 2, 5). As is well known, the Laemmli modification of the Ornstein and Davis system has become a mainstream technique for protein separation.

Identification of a protein detected on an SDS-polyacrylamide gel is a critical step for biochemical studies. Recent advances in analytical techniques, such as mass spectrometry, as well as the accumulation of genome information, have enabled systematic identification of proteins on a gel. Despite the high sensitivity of mass spectrometry, however, it is still difficult to identify the proteins when they are present in low abundance, highly modified, or impure. Therefore, in addition to these forward proteomic techniques, a reverse proteomic approach based on the set of all identified genes would be useful.

The genome sequence of the fission yeast *Schizosaccharomyces pombe* was completed in 2002 (6), officially putting *S. pombe* into the post-genomic era. Although *S. pombe* is a well studied model eukaryote, experimental resources and information for functional genomics are so far limited. Recently, we established a proteome resource by expressing the ORFeome<sup>4</sup> of *S. pombe* at moderate to high levels under the control of the inducible promoter *nmt1* (7). In the present study, we report electrophoresis-based analyses of the fission yeast proteome, including the collection of information regarding gel mobility of the proteome. This unique data set enables one to identify an unknown target protein detected on a gel, even if it is hard to identify by classical purification-based methods. Furthermore, we identified several novel acetylated or methylated proteins, using a global screen based on induced expression of the tagged proteins in combination with Western blotting (WB).

<sup>4</sup> The abbreviations used are: ORF, open reading frame; WB, Western blotting; GRAVY, grand average hydrophobicity; ER, endoplasmic reticulum; HA, hemagglutinin; MM, minimal medium; EF, elongation factor; eIF, eukaryotic translation initiation factor.

<sup>\*</sup> This work was supported in part by the CREST Research Project, Japan Science and Technology Agency, the Bioarchitect Research, RIKEN, and a Grant-in-Aid for Scientific Research on Priority Area “Cancer” from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

<sup>S</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1–S7 and supplemental Tables S1–S3.

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Supported in part by Grant-in-Aid for Young Scientists (A) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. To whom correspondence should be addressed: Chemical Genetics Laboratory, RIKEN, Wako, Saitama 351-0198, Japan. Tel.: 81-48-462-1335; Fax: 81-48-462-1336; E-mail: akihisa@riken.jp.

<sup>3</sup> Present address: Signal Creation Inc., Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan.

## Global Analysis of Gel Mobility of Proteins

### EXPERIMENTAL PROCEDURES

***S. pombe* Strains and Media**—*S. pombe* strains used in this study are listed in supplemental Table S1. A series of *S. pombe* strains in which each ORF is fused with two FLAG epitopes and a hexahistidine tag was integrated at the *leu1* locus were derived from AM2 (7). The construction of these strains is described elsewhere (7). AM5 was used for multicopy overexpression of an ORF. Complete medium YE (0.5% yeast extract, 2% glucose, 5  $\mu$ g/ml adenine) was used for routine culture of *S. pombe* strains. Selective minimal medium SD (8) containing only 1% glucose was used for routine culture of *S. pombe* transformants. Minimal medium MM (9) was used for expression of genes driven by the *nmt1* promoter (10). Adenine, leucine, and uracil were added to the medium (final concentration, 50  $\mu$ g/ml) when needed.

**General Methods and Transformation of *S. pombe***—General methods to handle fission yeast cells were as described (11). A high efficiency protocol for transformation of *S. pombe* cells, as described previously (12), was modified to the 96-well plate format. The cells were freshly grown in complete medium YE, harvested, and washed with and then resuspended in 0.1 M lithium acetate solution (pH 5.0). One microgram of salmon sperm DNA was added as carrier DNA into each 30- $\mu$ l cell suspension. For chromosomal integration of pDUAL-FFH1 (13) and its derivatives, each DNA was digested with NotI, SacII, or ApaI in a volume of 12  $\mu$ l, and 4  $\mu$ l of the resultant solution was used directly. Sample DNA, together with 90  $\mu$ l of 40% (w/v) polyethylene glycol 4000, was added immediately to the cell suspension, and the mix was incubated at room temperature for at least 30 min. After the addition of 12  $\mu$ l of dimethyl sulfoxide, the cells were heated at 42  $^{\circ}$ C for 15 min, spun down, resuspended in water, and plated on the SD selective medium. For gene induction, the cells harboring a testing plasmid or integrants were pregrown on the SD plate, and parts of them were subsequently streaked on the MM plate and grown for 22 h at 30  $^{\circ}$ C. The cells were then harvested and used for the subsequent experiments.

**Measurement of Gel Mobility of Proteins in SDS-PAGE**—Total cell lysates were prepared by the post-alkaline method described by V. V. Kushnirov (14) with slight modification (15). Cultures of tagged strains grown on MM plates for 22 h at 30  $^{\circ}$ C were harvested in 25  $\mu$ l of MM medium and placed in 96-well plates. Forty microliters of 0.4 N NaOH were added into the suspension. The cells were incubated for 5 min at room temperature, pelleted, resuspended in 30  $\mu$ l of SDS gel loading buffer, and boiled for 5 min. Aliquots of 10  $\mu$ l of the SDS-lysed extracts were loaded on a 28-well, 9  $\times$  15 cm, 5–20% gradient polyacrylamide gel (Bio-Craft). The proteins were electrotransferred to polyvinylidene difluoride membranes and then blocked for 1 h with 2% bovine serum albumin in PBST (1 $\times$  phosphate-buffered saline with 0.1% Tween 20) before primary antibodies were applied. A polyclonal anti-His tag antibody (Medical and Biological Laboratories Co., Ltd.) or a monoclonal anti-FLAG antibody M2 (Sigma), diluted with PBST (1:6,000) was incubated as a primary antibody with the membrane overnight. The membrane was washed with PBST three times and incubated for 1 h with a secondary antibody, an

IRDye<sup>TM</sup>800CW-conjugated anti-rabbit (-mouse) IgG antibody (Rockland), diluted 1:10,000 with PBST. After washing with PBST, the His<sub>6</sub> (FLAG)-tagged proteins were detected using the Odyssey<sup>®</sup> imaging system (LI-COR).

As a control, a protein marker lane containing a mixture of Precision Plus Protein<sup>TM</sup> Standards, all blue (Bio-Rad), and BenchMark<sup>TM</sup> Protein Ladder (Invitrogen) was included in each gel. Before measurement of gel mobility of proteins in SDS-PAGE, we analyzed the correlation between gel mobility of each band of the BenchMark<sup>TM</sup> Protein Ladder visualized by an anti-His tag antibody and its molecular mass in each membrane (Pearson product moment correlation coefficient  $r_{xy} > 0.99$ ) and calculated the regression equation for each membrane. The apparent molecular masses of almost all proteins predicted in fission yeast were measured from gel mobility on SDS-PAGE according to these regression equations.

The gel mobility of each gene product was initially determined using one integrant. When no positive band was detected in WB, another integrant was analyzed. Cells transformed with a multicopy plasmid were used only when no signal was obtained in WB using integrants.

**Antibodies Used in This Study**—The rabbit polyclonal antibodies against the His<sub>6</sub> tag and the HA epitope were purchased from MBL. The mouse monoclonal anti-FLAG antibody M2 was purchased from Sigma. The mouse monoclonal antibodies against an acetylated lysine and a methylated lysine (clones AKL5C1 and MEK3D7, respectively) were described elsewhere (16–18).

**Data Source and Bioinformatics Analyses**—Protein sequence information was retrieved from the *S. pombe* GeneDB data base (The Sanger Centre). The presence of signal sequences was calculated by using PSORT II (19). Grouping by subcellular localization of proteins was performed according to the analysis of Matsuyama *et al.* (7). SYSTAT 10.2 (Systat Software Inc., Richmond, CA) was used for statistical analyses. We calculated *p* values using the Mann-Whitney test for analyzing the relationship between gel mobility and the GRAVY score or isoelectric point of proteins, as well as the Kolmogorov-Smirnov test for analyzing the relationship between gel mobility of proteins and their subcellular localization. The results were further corrected for multiple hypothesis testing by applying the Bonferroni correction. Meanwhile, the relationship between the calculated and apparent molecular masses was analyzed by the Bonferroni-adjusted Wilcoxon matched pairs signed rank test.

**Website**—Information about gel mobility and basic data of proteins is available on our website. Users are free to browse and search the data base.

### RESULTS

**The Mobilitome, an “-omic” Data Set Based on Gel Mobility of Proteins**—To set up a strategy for global analysis of gel mobility of proteins in SDS-PAGE, we constructed *S. pombe* integrants in which each of the ORFs fused with two copies of the FLAG epitope and six consecutive histidine residues tag (His<sub>6</sub>) at the 3' terminus is integrated at the chromosomal *leu1* locus (13) (supplemental Fig. S1A). This strain collection was established based on a Gateway-cloned ORFeome resource that covered more than 99% of genes predicted by the genome project (7). All

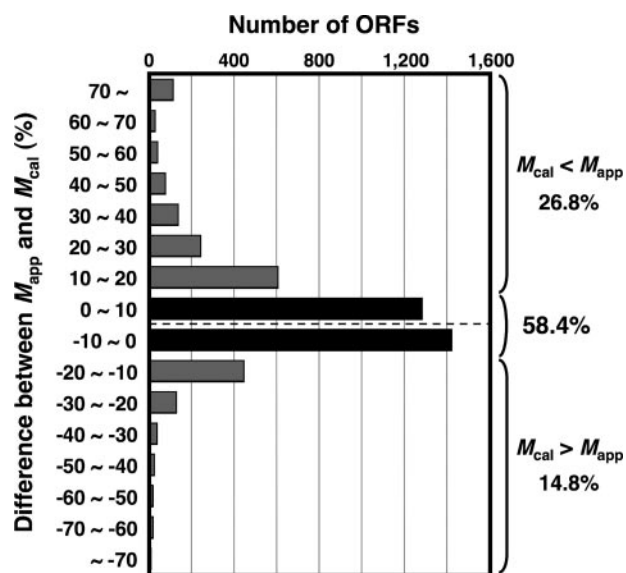


FIGURE 1. Overview of gel mobility of the proteome. Shown is the difference between  $M_{app}$  and  $M_{cal}$  of proteins. The errors of  $M_{app}$  from  $M_{cal}$  are calculated as  $y = (M_{app} - M_{cal})/M_{cal}$ .

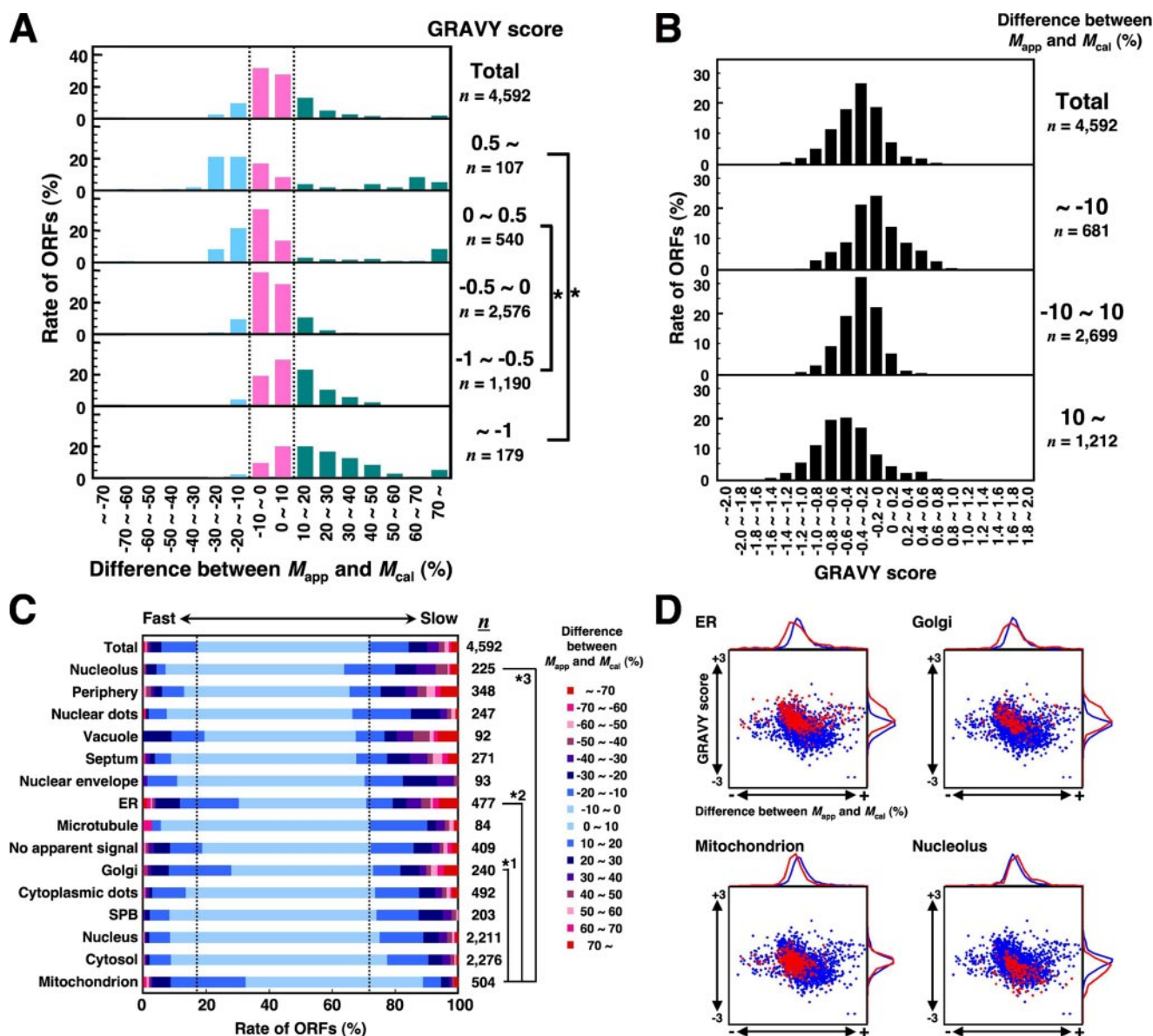
ORFs were expressed under the thiamine-regulatable *nmt1* promoter (20). Each protein, expressed in cells grown in MM, was separated by SDS-PAGE using a 5–20% gradient gel. Gel mobility was measured by WB, using either an anti-His tag antibody (supplemental Fig. S1B) or an anti-FLAG antibody. In those cases where the signal was below detectable levels, we overexpressed the ORFs from multicopy plasmids to improve detection (13). Thus far, we have determined the gel mobility of 4,622 proteins, corresponding to 93.4% of all gene products predicted in the genome (supplemental Table S2), thereby establishing the “mobilitome,” a novel “-omic” data set of gel mobility of the proteome. The mobilitome includes not only the gel mobilities of main populations of proteins but also those of additional minor populations when observed.

The overall size distribution of proteins measured by gel mobility was not significantly different from that of proteins calculated from their amino acid sequences ( $p = 0.060$ , as determined by the Mann-Whitney  $U$  test) (supplemental Fig. S2). When the sizes were compared one by one, however, more than 40% of the gene products did not migrate within 10% of the predicted position (calculated each ORF size plus 4.48 kDa for the tag) (Fig. 1), and there was a significant difference between the calculated molecular mass ( $M_{cal}$ ) and the apparent molecular mass ( $M_{app}$ ) ( $p < 1 \times 10^{-8}$ , as determined by the Wilcoxon signed rank test). This result indicates that it is difficult to predict gel mobility of proteins in SDS-PAGE solely from their amino acid sequences.

Gel mobility is influenced by various properties of proteins. Among them, the hydrophobicity denoted by the GRAVY score (21) had the greatest influence on gel mobility. Statistical analysis revealed that hydrophilic proteins tend to slow migration, whereas gel mobility of a protein became steadily greater as its GRAVY score increased (Fig. 2A). Reciprocally, proteins with greater mobility tend to have a high GRAVY score, whereas those with lower mobility have a low score ( $p < 1 \times 10^{-8}$ ) (Fig. 2B). This tendency is probably due to the manner in which SDS

binds to proteins. In general, SDS binds to a protein at a ratio of approximately one molecule/two amino acid residues (22, 23). However, because this binding is mediated via the hydrophobic interaction, SDS preferentially binds to the hydrophobic region of proteins (24). As a result, it is likely that hydrophobic proteins bind higher amounts of SDS than hydrophilic ones and therefore migrate faster in SDS-PAGE. Furthermore, the isoelectric points of proteins also influenced gel mobility to some extent. Acidic proteins (those for which isoelectric points are less than 6.0) exhibited mobility smaller than calculated ( $p < 1 \times 10^{-7}$ ) (supplemental Fig. S3); this may be due to negative charge repulsion with SDS. Although proteins showed these physicochemical properties on the whole, there were a significant number of proteins that did not match this pattern. For instance, there was a population of hydrophobic proteins with much lower mobility than predicted. It seems possible that these proteins are subject to post-translational modifications that affect their mobility. Indeed, of seven hydrophobic proteins (GRAVY score  $> 0.5$ ) that have so far been reported to be glycosylated in fission yeast, five proteins showed much lower mobility than predicted. Although the remaining two did not show such the slow migration, both proteins were accompanied by additional slowly migrating minor bands, which might be derived from glycosylation (data not shown). In addition to these characteristics, the molecular or functional characteristics of proteins such as subcellular localization seemed to have a significant influence on gel mobility. For instance, cytosolic proteins migrated in accordance with the prediction, whereas most nucleolar proteins were likely to migrate more slowly than expected. In contrast, proteins localized in the endoplasmic reticulum (ER) or the Golgi apparatus showed a tendency to migrate faster than calculated (Fig. 2C). Those proteins for which gel mobility was hard to predict were biased toward proteins that localize at the cell surface or membrane. It seems possible that this inconsistency could be mainly ascribed to the hydrophobicity of the proteins and the post-translational modifications as mentioned above. More remarkably, many proteins that are localized in the ER or the Golgi apparatus had a high GRAVY score, thereby migrating fast in SDS-PAGE, whereas nucleolar proteins are generally hydrophilic, which may lead to slower migration (Fig. 2D). It is known that many ER proteins have hydrophobic signal sequences that are cleaved off prior to translocation. Therefore, one would predict that the high mobility of ER proteins is likely to be caused by their signal sequences rather than their hydrophobicity. However, statistical analysis revealed that distribution of gel mobilities of ER proteins having predicted signal sequences was similar to that of ER proteins lacking them ( $p = 0.23$ , as determined by the Mann-Whitney  $U$  test) (supplemental Fig. S4A). Similarly, mitochondrial proteins migrated relatively quickly compared with their predicted mobilities, although these proteins are not as hydrophobic as ER and Golgi proteins. Furthermore, the distribution of gel mobilities of mitochondrial proteins having predicted signal sequences was similar to that of the mitochondrial proteins lacking them ( $p = 0.10$ , as determined by the Mann-Whitney  $U$  test) (supplemental Fig. S4B). The reason for this difference is still unclear, but it seems possible that

## Global Analysis of Gel Mobility of Proteins

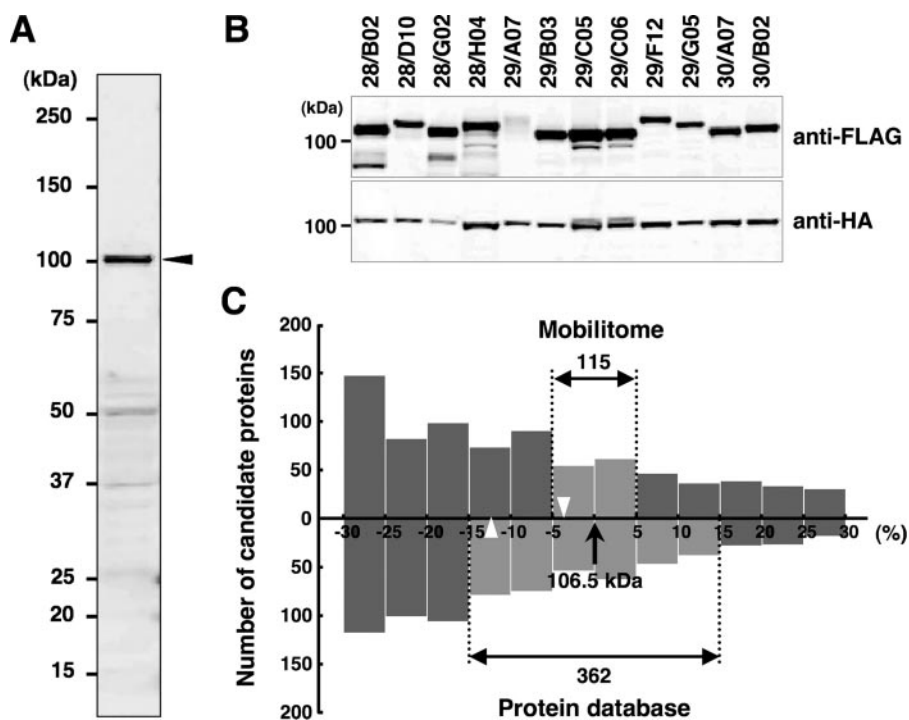


**FIGURE 2. Relationship between gel mobility and characteristics of proteins.** *A*, correlation between gel mobility and the hydrophobicity of proteins. The distribution of the errors between  $M_{app}$  and  $M_{cal}$  was examined according to the GRAVY score. Asterisks,  $p < 1 \times 10^{-8}$  by the Mann-Whitney  $U$  test. *B*, the distribution of proteins grouped by difference between  $M_{app}$  and  $M_{cal}$ , binned by the GRAVY score, is shown using intervals of 0.2. The y axis shows the ratio of proteins in each category of gel mobility. The difference between  $M_{app}$  and  $M_{cal}$  was calculated as described in the legend to Fig. 1.  $p < 1 \times 10^{-8}$  for all combinations of each group by the Mann-Whitney  $U$  test. *C*, relationship between  $M_{app}$  and localization of proteins. The percentage (x axis) is the ratio of proteins. The percentage of the error margin is indicated as colors. The difference between  $M_{app}$  and  $M_{cal}$  was calculated as described in the legend to Fig. 1. \*1,  $p = 1.6 \times 10^{-4}$ ; \*2,  $p = 2.5 \times 10^{-6}$ ; \*3,  $p = 1.2 \times 10^{-7}$ , by the Kolmogorov-Smirnov test. *D*, correlation between  $M_{app}$  and hydrophobicity of proteins localized in each compartment. The line graphs show the density of proteins. Proteins localized in each compartment are plotted in red, and the others are in blue.

unknown mitochondrion-specific protein processing or post-translational modification could be involved.

In addition to the variation of gel mobility of the main bands, we detected 1,304 proteins (28.2% of proteins with positive signals) that exhibited additional minor populations migrating slower than the main bands. Slowly migrating populations showed various patterns that contained either a discrete band, a smear, or a ladder, as well as combinations of these simpler patterns, suggestive of post-translational modifications such as ubiquitination and glycosylation (25–27) (data not shown). On the other hand, minor bands supposed to be degraded intermediates that migrated faster than the main populations were

observed in 2,339 proteins (48.1%) (data not shown). Of these quickly migrating populations, some may have a significant role in the biological process, although it is difficult to distinguish them from merely degraded intermediates, unlike the slowly migrating populations. The use of the strong *nmt1* promoter for gene expression could raise a risk of forced expression of proteins that are not abundant in normal physiological conditions, which could induce unexpected proteolytic cleavage of expressed proteins. Therefore, it seemed possible that the high mobility of some proteins is caused by their unexpected degradation. To assess this possibility, we examined the relationship between the presence of quickly migrating minor bands and gel



**FIGURE 3. Identification of proteins reactive with an anti-HA antibody using the mobilome data set.** *A*, total cell lysates blotted with an anti-HA antibody (MBL). An arrowhead indicates the target protein with a size of  $\sim 106.5$  kDa. *B*, identification of the target using the mobilome data base. FLAG<sub>2</sub>-His<sub>6</sub>-tagged candidate proteins expressed from the chromosomal *leu1* locus were detected by WB using an anti-FLAG antibody (*upper panel*) and were examined to see whether or not they cross-react with the anti-HA antibody using the same blot. Only 12 representative samples of 115 candidates are shown. ORFs are denoted as an individual ID managed in our laboratory (see our website for details). *C*, comparison between the mobilome and the protein sequence data base. Size distribution of proteins in the mobilome data base (*upper bars*) and the existing sequence-based data base (*lower bars*), binned by errors against  $M_{app}$  of the target proteins, is shown using 5% intervals.  $M_{app}$  of the target protein is set to the position of 0% (indicated as an arrow). The molecular masses of the identified target protein in the mobilome data base and the existing sequence-based data base are indicated by white arrowheads.

mobility of the main bands. Unexpectedly, however, statistical analysis revealed that proteins exhibiting additional quickly migrating minor populations migrated slightly slower than proteins with no such populations ( $p = 1.4 \times 10^{-7}$ , as determined by the Mann-Whitney *U* test, supplemental Fig. S5), although the difference was very small. These results suggest that it is unlikely that the unexpected degradation caused by forced expression is the major reason for the higher mobility of proteins than predicted.

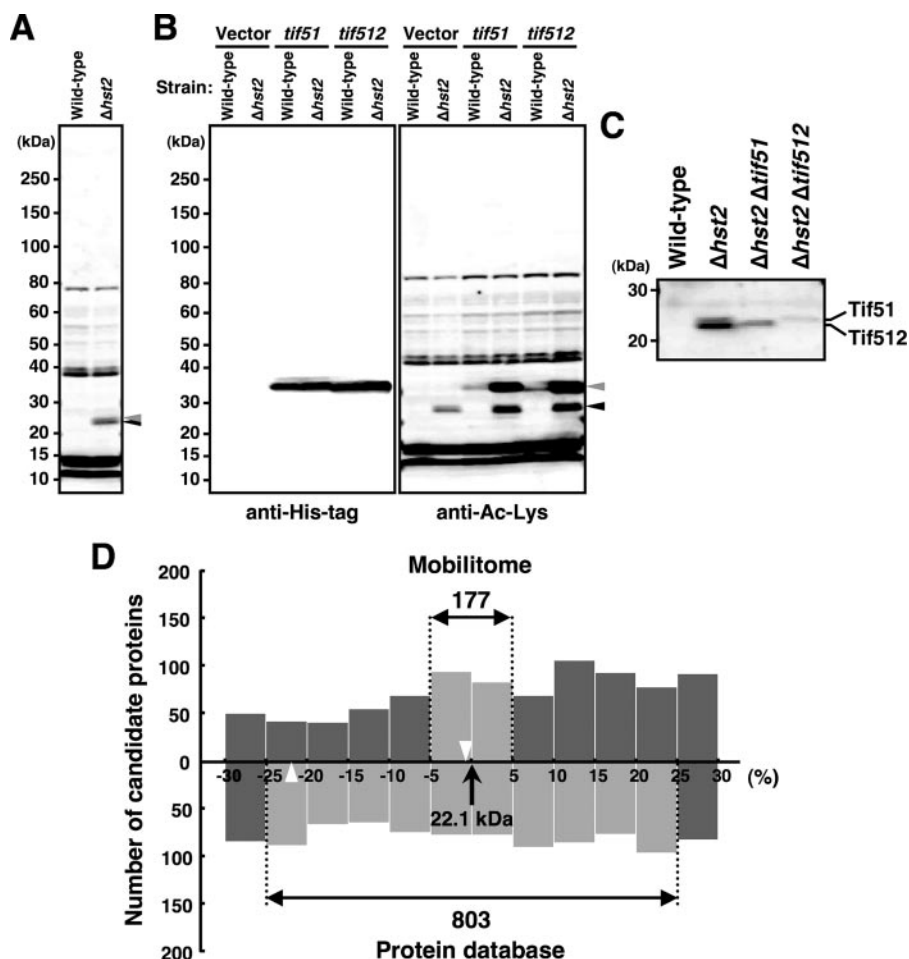
**Identification of Target Proteins Using the Mobilome Data Set**—The global mobility data base we developed enables one to rapidly identify a protein on the gel. Given an unknown protein of interest detected by a standard procedure, candidates for this protein can be chosen using the mobilome, based on the real value of gel mobility, and the transformants expressing tagged candidates can be then analyzed. Analyzing each candidate one by one in the same way as the first detection will easily lead to the identification of the target, because the sample containing the target protein is expected to show twin bands when detected, one of which is the endogenous protein, and the other of which is the transgene product, for which mobility is slightly shifted by the addition of the small C-terminal tags. For this methodology, we set up a website that allows searching for candidates by using gel mobility data of target proteins. As a proof of principle, we tried to identify an *S. pombe* protein detected by

an anti-HA tag antibody. This protein was observed as a background band with a size of  $\sim 106.5$  kDa when the total cell lysate was blotted with anti-HA (Fig. 3A). We searched for candidates for this 106.5 kDa ( $\pm 5\%$ ) in the mobilome data base and retrieved 115 proteins. We next analyzed the total cell lysate of each candidate strain by the same way as the mobilome but instead using the anti-HA antibody. As a result of the screen of 115 candidates, we rapidly identified two ORFs that gave twin bands (Fig. 3B). These two clones contained ORFs encoding the same protein, translation elongation factor 2 (EF2). We found a partially matched HA epitope sequence (VPDY) in EF2. We next tested whether the sequence is responsible for the antibody reactivity by introducing a mutation. The mutant EF2-GA having a sequence (VGAY) lost the reactivity to the antibody (supplemental Fig. S6), showing that the VPDY sequence acts as the HA epitope in EF2. Although it seems possible that a similar methodology could be applied to the existing sequence data base, it would not be practical, because of the low coincidence

between  $M_{cal}$  and  $M_{app}$ . Indeed, proteins that migrated with  $M_{app}$  within 5% of  $M_{cal}$  account for only 33.6% of the proteome, indicating that only one-third of targets could be identified. If a similar methodology were to be used for the identification of this protein based on existing sequence data bases and predicted ORF sizes alone, at least 306 clones would need to be tested, because of the low coincidence between  $M_{cal}$  and  $M_{app}$  (Fig. 3C).

To apply this methodology to the identification of novel acetylated proteins, we screened for proteins responsible for the two bands found only in the disruptant of the *hst2* gene, encoding a NAD-dependent histone deacetylase (28), using an anti-acetyllsine antibody (Fig. 4A).  $M_{app}$  of these acetylated proteins were estimated as 22.1 and 22.0 kDa, respectively, in WB using an anti-acetyllsine antibody. A screen of 177 candidates from the mobilome data base with mobilities within  $\pm 5\%$  of 22.0–22.1 kDa (Fig. 4D) resulted in the identification of two proteins, Tif51 and Tif512 (29), which are closely related paralogs of eukaryotic translation initiation factor 5A (eIF5A) (Fig. 4, B and C). In this case, to identify Tif51 and Tif512 using the sequence data base, it would have been necessary to search for proteins within  $\pm 25\%$  of  $M_{app}$ , which corresponds to nearly one-sixth of the proteome (Fig. 4D). Although acetylation of eIF5A has been previously reported in other organisms (30, 31), the precise role of acetylation remains unknown. However, the

## Global Analysis of Gel Mobility of Proteins



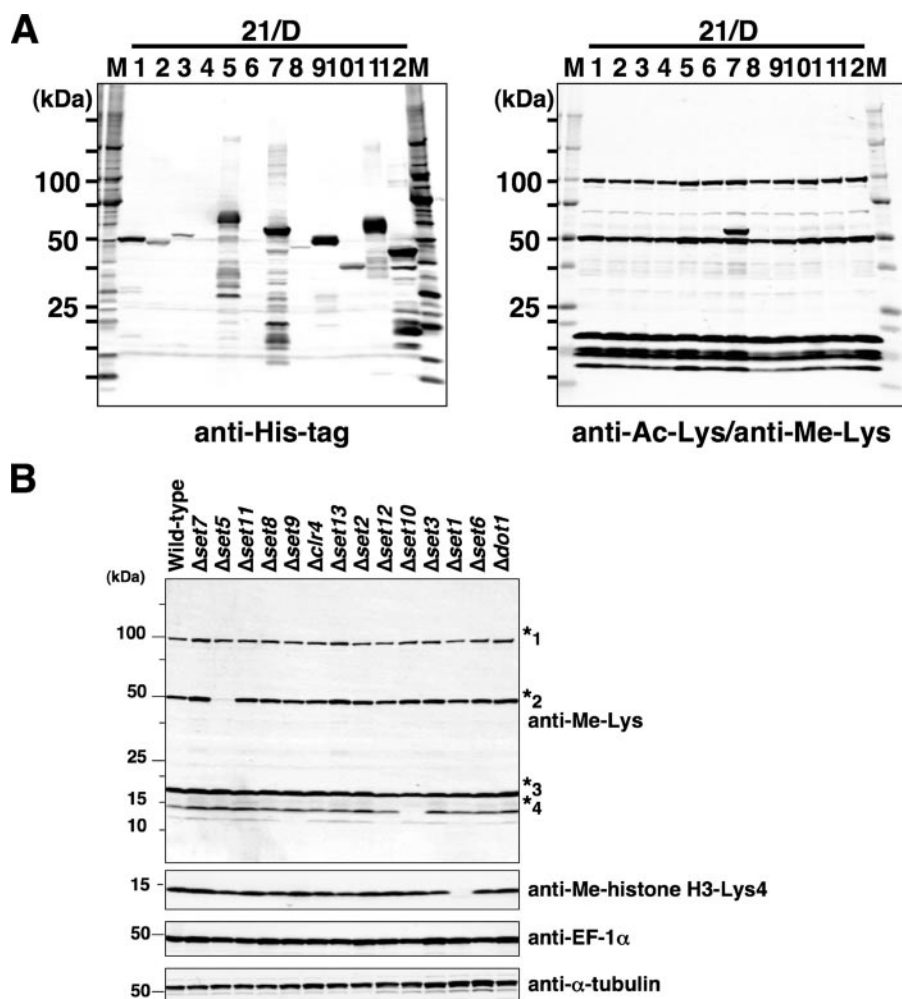
**FIGURE 4. Identification of eIF5A as a substrate of Hst2.** *A*, total cell lysates prepared from a wild-type strain and the *hst2* disruptant (*hst2*Δ) blotted with an anti-acetyllysine antibody. Two bands with  $M_{app}$  of ~22 kDa observed only in *hst2*Δ are denoted as *gray* and *black arrowheads*, respectively. *B*, acetylation of eIF5A in fission yeast. Total cell lysate of each pDUAL-FFH1-integrand was blotted with an anti-His tag antibody (*left*) and an anti-acetyllysine antibody (*right*), respectively. A *black arrowhead* indicates the endogenous acetylated protein identified as eIF5A, whereas a *gray arrowhead* indicates two paralogs of fission yeast eIF5A tagged with FLAG<sub>2</sub>-His<sub>6</sub>. *C*, acetylation of the two endogenous paralogs of eIF5A. Total cell lysates prepared from the indicated strains were blotted with an anti-acetyllysine antibody. *D*, comparison between the mobilome and protein sequence data base. Size distribution of proteins in the mobilome data base (*upper bars*) and the existing sequence-based data base (*lower bars*), binned by errors against  $M_{app}$  of one of the target proteins (indicated by a *gray arrowhead* in *A*), is shown using 5% intervals.  $M_{app}$  of the target protein Tif51 is set to the position of 0% (indicated as an *arrow*). The *white arrowheads* indicate the molecular masses of Tif51 in the mobilome data base and the existing sequence-based data base.

fact that eIF5A is deacetylated by the NAD-dependent histone deacetylase sheds light on the as-yet-unknown role of protein acetylation.

**Proteome Scale Screen for Acetylated or Methylated Proteins—**Identification of eIF5A in the screen for acetylated proteins prompted us to screen for a certain type of modified proteins. As indicated in Fig. 4*A*, neither endogenous Tif51 nor Tif512 was detected with an anti-acetyllysine antibody in whole cell lysates unless the *hst2* gene was deleted. However, when they were overexpressed, acetylation of these proteins became visible even in wild-type cells (Fig. 4*B*). These results suggest that simple overexpression leads to detection of modified proteins that are not easily detected in the normal conditions. Based on this idea, we screened all the overexpressed proteins for acetylated or methylated proteins by SDS-PAGE followed by WB using anti-acetyllysine and anti-methyllysine antibodies, which are context-independent antibodies having a wide reactivity

reported so far in any organisms. Assigned by gel mobility data, methylation of EF2, EF1α, and ribosomal proteins L12 and L23 are readily detected in normal conditions (Fig. 5*B*). To test whether these proteins are substrates for predicted methyltransferases, we deleted 13 genes that encode HMT domain-containing methyltransferases, as well as the *SPBC1539.02* gene, which encodes a homolog of the budding yeast Dot1 methyltransferase for lysine 79 of histone H3 (37) (Fig. 5*B*). Western analysis of total lysates showed that in the *set5* and the *set10* disruptants, methylation of EF1α and the ribosomal protein L23, respectively, was lost, indicating that the SET proteins are involved in methylation of non-histone proteins as well as histones (Fig. 5*B*). This finding was confirmed by expressing the tagged EF1α or the ribosomal protein L23 in the SET mutants (supplemental Fig. S7). Recently, Rkm1, a budding yeast homolog of Set10/Spbc1709.13c, was reported to methylate the ribosomal protein L23 (32). Thus, the regulation of this protein

with acetyllysine- or methyllysine-containing proteins (16, 18). Only a few bands were seen even when these antibodies were simultaneously used against the total cell lysate of wild-type cells (Fig. 5*A*). As expected, however, by testing 4,622 tagged proteins one by one, we identified twelve and eight proteins (ORFs having the same amino acid sequence were counted as one) that were supposed to be acetylated or methylated, respectively (Fig. 5*A* and supplemental Table S3). The number of these modified proteins exceeds the number of the bands seen in WB using total cell lysates. Histone H3, a protein known to be both acetylated and methylated in fission yeast, was also identified in this screen, serving as a proof of principle. In this screen, ribosomal proteins L12 and L23, as well as eukaryotic translation elongation factors EF1α and EF2, were identified as lysine-methylated proteins, implying that protein-lysine methylation is deeply involved in translation, whereas protein acetylation seemed to have relatively wider involvement in cellular metabolic activity. Ribosomal proteins are methylated in some organisms (32–35), and lysine methylation of EF1α has been widely observed among eukaryotes (36). Although methylation of the fission yeast EF1α had not been known, our results demonstrated that this is the case also in fission yeast. On the other hand, methylation of EF2 has not been



**FIGURE 5. Genome-wide screen of methylated and acetylated proteins.** *A*, screening of methylated and/or acetylated proteins using modification-specific antibodies. Total cell lysate of each pDUAL-integrand was blotted with an anti-His tag antibody (*left*) to confirm the position of each tagged protein. The same blot was simultaneously stained with anti-methyllysine (*Me-Lys*) and anti-acetyllysine (*Ac-Lys*) antibodies (*right*). *B*, the methylation levels of proteins in methyltransferase-encoding gene disruptants. Total cell lysate of each disruptant was blotted with anti-methyllysine, anti-methylated histone H3 Lys4, anti-EF1 $\alpha$ , and anti- $\alpha$ -tubulin antibodies, respectively. The numbered bands are identified as EF2 (\*1), EF1 $\alpha$  (\*2), and the ribosomal proteins L12 (\*3) and L23 (\*4).

by methylation may be conserved between the two yeasts. It has also been reported that the ribosomal protein L12 is methylated by Rkm2 (38) in *S. cerevisiae*, which is a homolog of Set11/Spcc1223.04c. However, methylation of EF2 and the ribosomal protein L12 was not impaired in these single SET mutants, including the *set11* disruptant. These results suggest that multiple methyltransferases are involved in methylation of these substrates.

## DISCUSSION

In this paper, we describe an electrophoresis-based analysis of the fission yeast proteome, including establishment of the mobilitome and a genome-wide screen for acetylated or methylated proteins. The mobilitome data set showed two unexpected results. First, more than 40% of the proteins in *S. pombe* did not migrate to the positions predicted from their amino acid sequences, although it was recently reported that most proteins migrated at the calculated positions in the budding

yeast (39). SDS-PAGE is the most widely used method for determination of the molecular size of a polypeptide. However, SDS might preferentially bind to the hydrophobic region of proteins and less to a protein with negative charges according to its chemical properties, which may cause inconsistencies with theoretical mobility. Indeed, the statistical analyses used in this study revealed that gel mobility of proteins was influenced both by the hydrophobicity and isoelectric points of proteins. Furthermore, ~30% of the proteome had additional slower migrating bands, showing the post-translational modifications that retard gel mobility. These modifications probably include phosphorylation, glycosylation, ubiquitination, and ubiquitin-like modifications.

The data set of gel mobility of the fission yeast proteome provides not only useful information on physicochemical properties but also tools for functional proteomics. One of such applications is the use of the mobilitome for the systematic identification of a target protein of interest detected on a gel. Despite its simplicity, this methodology has great advantages for the identification of target proteins. For example, because both pan-reactive anti-acetyllysine and anti-methyllysine antibodies used in this study show high detection sensitivity in WB but do not have enough affinity to pre-

cipitate lysine-acetylated or -methylated proteins (18), it was essentially impossible to perform one-step purification-based approaches. In addition, because most post-translational modifications, especially those involved in signal transduction events such as phosphorylation, take place in only a small population of molecules within the cells, the amount of the target suffering modifications is sometimes insufficient for identification by classical methods. However, the mobilitome, with its high coverage rate, can overcome these difficulties, because it only requires "detection" of the target and does not necessarily depend on its amount. In principle, any protein we were able to detect on a gel with a standard procedure can be identified using the mobilitome, unless the protein is not listed in the mobilitome data base. The high coverage rate, more than 93% of the proteome, ensures the effectiveness of this methodology. In addition, we have included information in the current data base on the positions of more slowly migrating minor bands in addition to the main band. Modifications of proteins, which

## Global Analysis of Gel Mobility of Proteins

sometimes show the greatly altered gel mobility, hamper the prediction of the target using existing protein data bases. Our collection of minor band information may help identify such modified proteins by the mobilitome-based approach.

To date, over 200 different post-translational modifications have been described (40), which implies the existence of an unexplored field of global research that might be termed the "modificome," the whole of the post-translational modifications within the cell. For instance, a proteomics survey using mass spectrometry identified a large number of lysine-acetylated proteins in mammalian cells (30). In this study, we identified methylated and acetylated proteins using a reverse proteomics approach, and most of the proteins were not known to be subject to such the modifications. Most importantly, however, almost all the acetylated proteins would never be identified unless they were overexpressed, because acetylation of proteins other than histones is not observed in WB of total cell lysates of wild-type cells. It is still possible that overexpression could cause unexpected modifications; thus, the results should be interpreted with caution. Nonetheless, an overexpression-based screen for modified proteins seems effective as the first step to analyze the physiological role of a modification of interest. Thus, in combination with the reverse proteomic approach, the mobilitome and the methodology described herein will also serve as a platform to reclaim the world of post-translational modifications.

*Acknowledgment*—We thank Dr. Masaaki Matsuura at the Cancer Institute of the Japanese Foundation for Cancer Research for useful comments in statistical analyses.

### REFERENCES

- Ornstein, L. (1964) *Ann. N. Y. Acad. Sci.* **121**, 321–349
- Davis, B. J. (1964) *Ann. N. Y. Acad. Sci.* **121**, 404–427
- Shapiro, A. L., Vinuela, E., and Maizel, J. V., Jr. (1967) *Biochem. Biophys. Res. Commun.* **28**, 815–820
- Weber, K., and Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406–4412
- Laemmli, U. K. (1970) *Nature* **227**, 680–685
- Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E. J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R. G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymoprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritz, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z., Hunt, C., Moore, K., Hurst, S. M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, L., Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerutti, L., Lowe, T., McCombie, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V., Ussery, D., Barrell, B. G., and Nurse, P. (2002) *Nature* **415**, 871–880
- Matsuyama, A., Arai, R., Yashiroda, Y., Shirai, A., Kamata, A., Sekido, S., Kobayashi, Y., Hashimoto, A., Hamamoto, M., Hiraoka, Y., Horinouchi, S., and Yoshida, M. (2006) *Nat. Biotechnol.* **24**, 841–847
- Sherman, F., Fink, G., and Hick, J. (1986) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Moreno, S., Klar, A., and Nurse, P. (1991) *Methods Enzymol.* **194**, 795–823
- Maundrell, K. (1990) *J. Biol. Chem.* **265**, 10857–10864
- Gutz, H., Heslot, H., Leupold, U., and Loprieno, N. (1974) in *Handbook of Genetics* (King, R., ed) pp. 395–446, Plenum Publishing Corp., New York
- Matsuyama, A., Yabana, N., Watanabe, Y., and Yamamoto, M. (2000) *Genetics* **155**, 539–549
- Matsuyama, A., Shirai, A., Yashiroda, Y., Kamata, A., Horinouchi, S., and Yoshida, M. (2004) *Yeast* **21**, 1289–1305
- Kushnirov, V. V. (2000) *Yeast* **16**, 857–860
- Shirai, A., Matsuyama, A., Yashiroda, Y., Arai, R., and Yoshida, M. (2006) *Nat. Protocols* **10.1038/nprot.2006.191**
- Komatsu, Y., Yukutake, Y., and Yoshida, M. (2003) *J. Immunol. Methods* **272**, 161–175
- Matsuyama, A., Shimazu, T., Sumida, Y., Saito, A., Yoshimatsu, Y., Seigneurin-Berny, D., Osada, H., Komatsu, Y., Nishino, N., Khochbin, S., Horinouchi, S., and Yoshida, M. (2002) *EMBO J.* **21**, 6820–6831
- Iwabata, H., Yoshida, M., and Komatsu, Y. (2005) *Proteomics* **5**, 4653–4664
- Teasdale, R. D., and Jackson, M. R. (1996) *Annu. Rev. Cell Dev. Biol.* **12**, 27–54
- Maundrell, K. (1993) *Gene (Amst.)* **123**, 127–130
- Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Pitt-Rivers, R., and Impiombato, F. S. (1968) *Biochem. J.* **109**, 825–830
- Reynolds, J. A., and Tanford, C. (1970) *J. Biol. Chem.* **245**, 5161–5165
- Robinson, N. C., and Tanford, C. (1975) *Biochemistry* **14**, 369–378
- Cook, L. B., Zhu, C. C., and Hinkle, P. M. (2003) *Mol. Endocrinol.* **17**, 1777–1791
- Poly, W. J. (1997) *Comp Biochem. Physiol. A Physiol.* **118**, 551–572
- Zhu, C. C., Cook, L. B., and Hinkle, P. M. (2002) *J. Biol. Chem.* **277**, 28228–28237
- North, B. J., and Verdin, E. (2004) *Genome Biol.* **5**, 224
- Linder, P., Vornlocher, H. P., Hershey, J. W., and McCarthy, J. E. (1999) *Yeast* **15**, 865–872
- Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., Grishin, N. V., White, M., Yang, X. J., and Zhao, Y. (2006) *Mol. Cell* **23**, 607–618
- Klier, H., Csonga, R., Joao, H. C., Eckerskorn, C., Auer, M., Lottspeich, F., and Eder, J. (1995) *Biochemistry* **34**, 14693–14702
- Porras-Yakushi, T. R., Whitelegge, J. P., Miranda, T. B., and Clarke, S. (2005) *J. Biol. Chem.* **280**, 34590–34598
- Lee, S. W., Berger, S. J., Martinovic, S., Pasa-Tolic, L., Anderson, G. A., Shen, Y., Zhao, R., and Smith, R. D. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5942–5947
- Williamson, N. A., Raligh, J., Morrice, N. A., and Wettenhall, R. E. (1997) *Eur. J. Biochem.* **246**, 786–793
- Kruiswijk, T., Kunst, A., Planta, R. J., and Mager, W. H. (1978) *Biochem. J.* **175**, 221–225
- Cavallius, J., Zoll, W., Chakraburty, K., and Merrick, W. C. (1993) *Biochim. Biophys. Acta* **1163**, 75–80
- Ng, H. H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y., and Struhl, K. (2002) *Genes Dev.* **16**, 1518–1527
- Porras-Yakushi, T. R., Whitelegge, J. P., and Clarke, S. (2006) *J. Biol. Chem.* **281**, 35835–35845
- Gelperin, D. M., White, M. A., Wilkinson, M. L., Kon, Y., Kung, L. A., Wise, K. J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H., Gerstein, M., Dumont, M. E., Phizicky, E. M., Snyder, M., and Grayhack, E. J. (2005) *Genes Dev.* **19**, 2816–2826
- Radha, K., and Finn, W. (1993) *Adv. Enzymol. Relat. Areas Mol. Biol.* **67**, 265–298