

# Comparative proteomic and transcriptomic profiling of the fission yeast *Schizosaccharomyces pombe*

Michael W Schmidt<sup>1,2,3</sup>, Andres Houseman<sup>4,5</sup>, Alexander R Ivanov<sup>1,2,\*</sup> and Dieter A Wolf<sup>1,2,\*</sup>

<sup>1</sup> NIEHS Center for Environmental Health Proteomics Facility, Harvard School of Public Health, Boston, MA, USA, <sup>2</sup> Department of Genetics and Complex Diseases, Harvard School of Public Health, Boston, MA, USA, <sup>3</sup> Institute for Biochemistry, University of Stuttgart, Stuttgart, Germany and <sup>4</sup> Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

<sup>5</sup> Present Address: Department of Work Environment, University of Massachusetts Lowell, One University Avenue, Lowell, MA 01854, USA.

\* Corresponding authors: DA Wolf or AR Ivanov, Department of Genetics and Complex Diseases, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA. Tel.: +1 617 432 2093; Fax: +1 617 432 2059; E-mail: dwolf@hsph.harvard.edu or aivanov@hsph.harvard.edu

Received 15.8.06; accepted 13.12.06

The fission yeast *Schizosaccharomyces pombe* is a widely used model organism to study basic mechanisms of eukaryotic biology, but unlike other model organisms, its proteome remains largely uncharacterized. Using a shotgun proteomics approach based on multidimensional prefractionation and tandem mass spectrometry, we have detected ~30% of the theoretical fission yeast proteome. Applying statistical modelling to normalize spectral counts to the number of predicted tryptic peptides, we have performed label-free quantification of 1465 proteins. The fission yeast protein data showed considerable correlations with mRNA levels and with the abundance of orthologous proteins in budding yeast. Functional pathway analysis indicated that the mRNA–protein correlation is strong for proteins involved in signalling and metabolic processes, but increasingly discordant for components of protein complexes, which clustered in groups with similar mRNA–protein ratios. Self-organizing map clustering of large-scale protein and mRNA data from fission and budding yeast revealed coordinate but not always concordant expression of components of functional pathways and protein complexes. This finding reaffirms at the protein level the considerable divergence in gene expression patterns of the two model organisms that was noticed in previous transcriptomic studies.

*Molecular Systems Biology* 13 February 2007; doi:10.1038/msb4100117

*Subject Categories:* proteomics

*Keywords:* fission yeast; LC-MS/MS; mRNA–protein correlation; relative protein quantification; protein profiling

## Introduction

*Schizosaccharomyces pombe* is a unicellular archiascomycete fungus displaying many properties of more complex eukaryotes. It has been estimated that fission yeast diverged from budding yeast ~1100 million years ago (Heckman *et al*, 2001), thus accounting for their considerable divergence in genome organization (Wood *et al*, 2002). Despite differences in the number of genes, the number of introns, and centromere size, basic cellular processes are highly conserved between the two yeasts with ~3600 proteins being predicted or confirmed orthologs (Wood, 2006). However, it is still unclear to what extent mechanisms of gene expression in fission yeast overlap with those in budding yeast.

*S. pombe* is a well-established model organism for the study of cell-cycle regulation, cytokinesis, DNA repair and recombination, and checkpoint pathways, but only ~1500 of its predicted ~4900 genes and proteins have been experimentally characterized. Although mRNA profiling has begun to address functional aspects of the fission yeast genome (Mata *et al*, 2002; Chen *et al*, 2003; Rustici *et al*, 2004; Oliva *et al*, 2005;

Peng *et al*, 2005; Marguerat *et al*, 2006), the notion was expressed that mRNA levels are only a partial reflection of the functional state of an organism (Greenbaum *et al*, 2003). It is widely accepted that a comprehensive understanding of the genomic information will require, besides other strategies, means of analyzing quantitative differences in protein expression on a proteome-wide scale (Anderson *et al*, 2000; Bakhtiar and Tse, 2000; Yates, 2000).

Several quantitative methods, including ICAT (Gygi *et al*, 1999), iTRAQ (Ross *et al*, 2004), stable isotope labelling (Ong *et al*, 2002; Washburn *et al*, 2002), AQUA (Gerber *et al*, 2003), spectral sampling (Liu *et al*, 2004; Kislinger *et al*, 2006), protein abundance indexing (Ishihama *et al*, 2005), and whole-genome ORF epitope tagging (Ghaemmaghami *et al*, 2003; Matsuyama *et al*, 2006), have been employed for proteomic analyses of model organisms, in particular budding yeast. All of these techniques have their intrinsic strengths and limitations, including the bias of mass spectrometry-based methods toward proteins of medium to high abundance, and the potential for interference of epitope tags with endogenous protein function, expression, and localization. In addition,

epitope tagging can only interrogate putative known ORFs and is only applicable to organisms that are readily amenable to genetic manipulation in a high-throughput format. Mass spectrometry, on the contrary, can potentially identify new proteins and is broadly applicable to any proteome for which a corresponding genome sequence is available.

Weighing the advantages and disadvantages of currently available methods, we have embarked on a mass spectrometry-based approach for relative quantitation of unmodified fission yeast proteins. In addition, we have compared mRNA and protein expression profiles in fission yeast and budding yeast to assess the overall protein–mRNA correlation in these related organisms.

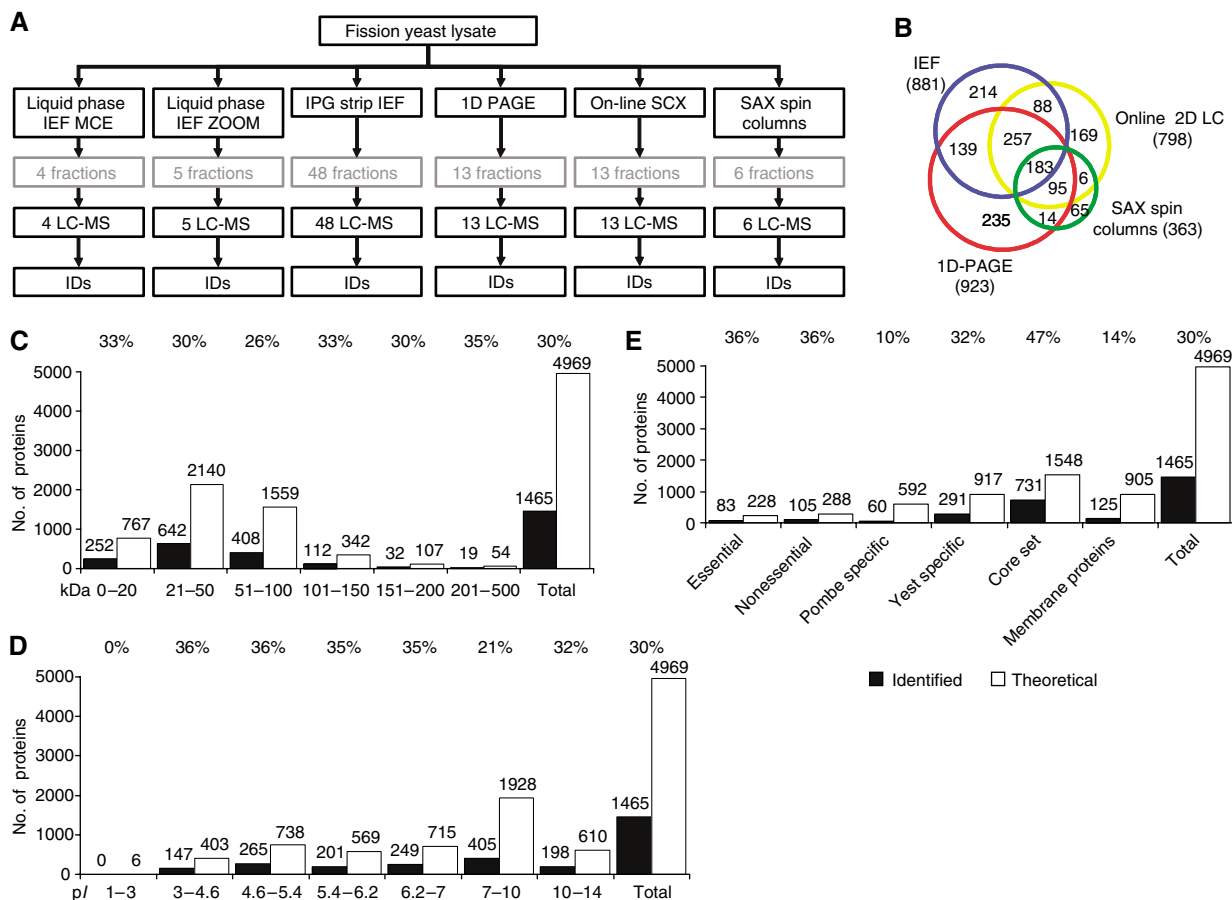
## Results and discussion

### Analysis of the *S. pombe* proteome by multidimensional prefractionation and LC ESI MS/MS

We devised the extensive multidimensional biochemical prefractionation scheme outlined in Figure 1A, starting with

total cell lysate from wild-type fission yeast cells growing vegetatively in mid-log phase in rich media. Aliquots of the lysate were fractionated by preparative isoelectric focusing (IEF) on immobilized pH gradients, or in two different liquid-phase formats, by one-dimensional (1D) gel electrophoresis, and by strong ion-exchange chromatography in a spin column format (Doud *et al*, 2004), followed by analysis of individual fractions by 1D liquid chromatography coupled with electrospray ionization tandem mass spectrometry (LC ESI MS/MS). In parallel, total fission yeast lysate was subjected to on-line 2D LC ESI MS/MS (=‘MudPIT’; Washburn *et al*, 2001), upon in-solution digestion into tryptic peptides.

Altogether, ~3 million mass spectra were collected and rigorously searched against the fission yeast protein database using the SEQUEST algorithm (Eng *et al*, 1994). Mass spectra were matched to 12 413 nonredundant peptides (Supplementary Data File 1), resulting in the identification of 1465 proteins (Supplementary Data File 2) with a predicted false-positive peptide identification rate of 1.05%, as determined by searching against a combined forward and reverse protein database (Peng *et al*, 2003a). The identified proteins cover ~29.5% of the predicted fission yeast proteome. To our



**Figure 1** Analysis of the *S. pombe* proteome by multidimensional prefractionation and LC ESI MS/MS. (A) Flow chart of sample prefractionation. IEF=isoelectric focusing; ZOOM, MCE (multicompartment electrolizer)=liquid-phase IEF devices, IPG=immobilized pH gradient strips, SAX=strong anion exchange, LC-MS=liquid chromatography and mass spectrometry. (B) Summary of the number of proteins identified with each prefractionation method. The overlap between fractions is indicated. (C, D) Molecular weight and pI distribution of the identified proteins compared to the theoretical proteome. (E) Fractions of proteins identified that belong to the indicated categories.

knowledge, this represents the highest percent coverage of native, unmodified proteins reported to date for any eukaryotic proteome. We also confirmed 40 predicted sequence orphans as well as five hypothetical proteins, and identified three new proteins, which were listed as dubious ORFs (SPAC13G6.13, SPBC354.04) or pseudogenes (SPBC16E9.16c) in the *S. pombe* genome database.

Although the individual pre-fractionation techniques contributed to the total protein count to different extents (Figure 1B), the extensive scale of the combined approaches identified a list of proteins that was representative of the whole proteome across the entire range of molecular weights and isoelectric points (Figure 1C and D). Most major Gene Ontology (GO) attributes for *S. pombe* were represented, indicating that our study broadly sampled across cell functions (Supplementary Data File 1). For example, we identified 132 of 141 ribosomal proteins and all subunits of the 26S proteasome and the CCT chaperonin complex. We also identified all enzymes of the cysteine, glutamate, glycine, isoleucine, leucine, proline, threonine, valine, aspartate, adenine and aromatic amino-acid biosynthesis pathways as well as 45 kinases (23% of all kinases predicted from the genome sequence), 20 predicted transcriptional regulators (14%), and 21 mitochondrial proteins (15%).

More detailed analysis revealed equal identification rates for essential and non-essential proteins (both 36%; Figure 1E) based on 187 proteins present in our data set for which information on essentiality is available in fission yeast (83 essential, 104 nonessential genes/proteins). Similarly, yeast-specific proteins were represented at the same rate as the entire proteome (30%; Figure 1E). Metazoan 'core' proteins (proteins common to *S. pombe*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster*; see Supplementary information), were overrepresented (47%; Figure 1E), a finding that is consistent with their higher mRNA levels (Mata and Bahler, 2003). In contrast, we undersampled proteins containing predicted transmembrane domains (14%) and *S. pombe*-specific proteins (10%; Figure 1E). Although not all membrane proteins may be equally amenable to extraction under our sample preparation conditions, the underrepresentation of *S. pombe*-specific proteins is mostly due to their specialized functions in the sexual differentiation pathway (data not shown), which cannot be effectively sampled in the vegetatively growing cells used here.

### Label-free relative quantitation of *S. pombe* proteins

To quantitatively rank the identified proteins relative to each other, we used spectral counts. Spectral counts represent the number of nonredundant mass spectra identifying the same protein. Whereas spectral counts are predicted to increase linearly with protein abundance (Liu *et al*, 2004), this relationship is amended by protein size, with larger proteins having a statistically higher probability of being detected. The relationship is further modified by the sequence-dependent number of peptides produced by the tryptic cleavage. Finally, an allowance for up to three enzymatic miscleavages is often granted during the SEQUEST database search, thus further

distorting the theoretical linear relationship between spectral count and protein abundance.

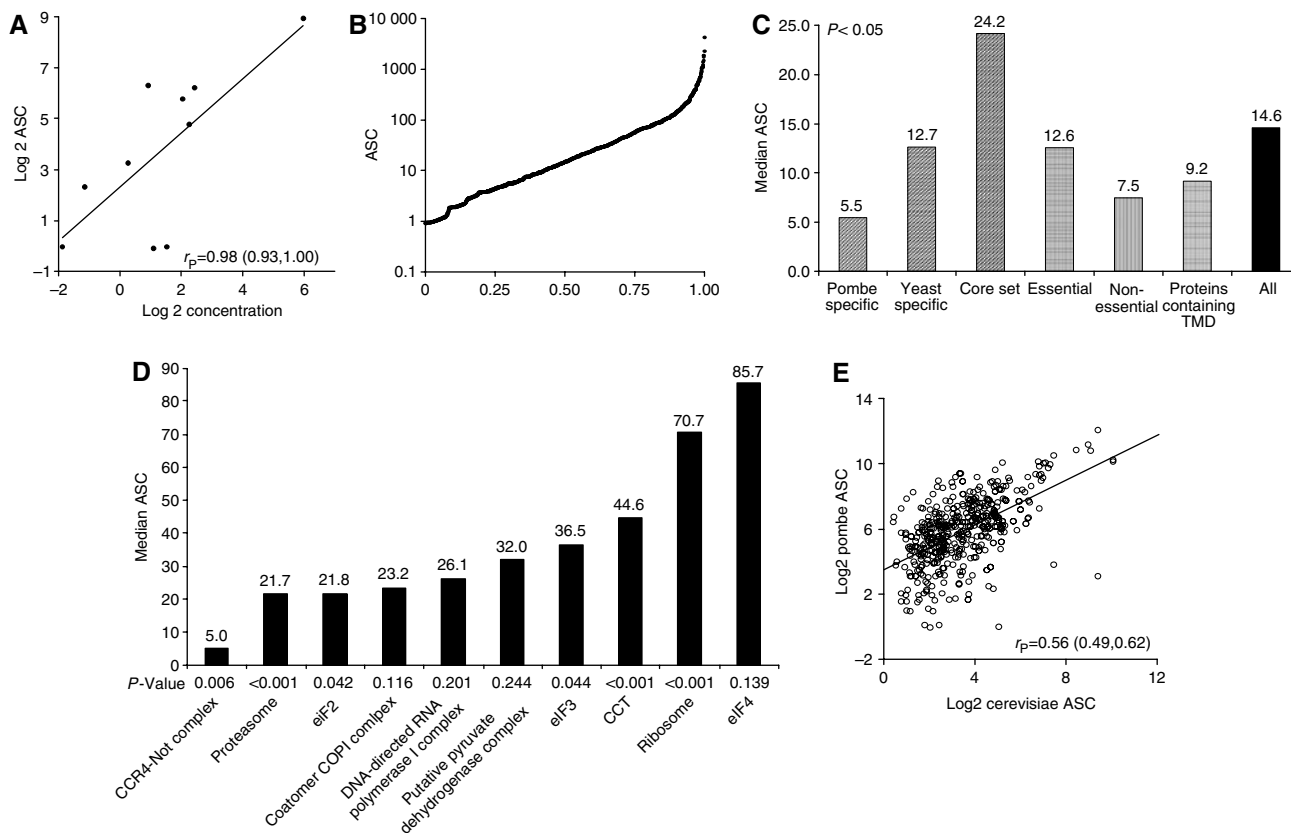
To apply an appropriate adjustment of spectral counts to a measure of protein size, we compared goodness-of-fit statistics applied to negative binomial regression models to determine which of the above parameters (number of amino acids, number of tryptic peptides, miscleavages) figured most prominently. The models revealed that adjustment to the number of tryptic peptides with one miscleavage resulted in the most optimal fit statistics for the experimental LC ESI MS/MS data (Supplementary Figure 4 and Supplementary Table 1).

Based on adjusted spectral counts (ASCs), we assembled a ranked list of all 1465 proteins identified (Supplementary Data File 2). This quantitative ranking reflects the abundance of each protein relative to all others and their quantitative distances. The ranked list was validated by comparing it to absolute quantitation data established for a series of 27 cytokinesis-related fission yeast proteins (Wu and Pollard, 2005). While these absolute measurements rely on epitope tagging, the tagged alleles were extensively validated for functionality under various conditions and in various genetic backgrounds, thus suggesting that tagging did not interfere with normal protein expression (Wu and Pollard, 2005). Of the 27 cytokinesis proteins, 10 were represented on our list. Plotting our ASC data versus the absolute quantitation data revealed a close correlation ( $r_p=0.98$ ; Figure 2A), suggesting that ASCs provide a good approximation of relative protein abundance.

The range of ASCs spanned more than three orders of magnitude (Figure 2B). The mean ASC was 68.0, whereas the median was 14.6, indicating that the vast majority of the 1465 proteins identified are of relatively low abundance compared to a small number of hyperabundant proteins (Figure 2B). The group of the 30 most abundant proteins (ASC between 584 and 4269) contained proteins of which all but three have orthologs in budding yeast that were also detected by whole-genome TAP tagging (Ghaemmaghami *et al*, 2003). This group includes eight glycolytic enzymes, six enzymes involved in biosynthetic pathways, seven translation factors, five heat-shock proteins, as well as two thioredoxin peroxidases (Supplementary Data File 2). The most abundant fission yeast protein is Eno101, a subunit of the phosphopyruvate hydratase complex (ASC=4269), followed by phosphoglycerate kinase (Pkg1, ASC=2301) as a distant second.

The group of the 30 least abundant proteins detected (ASC=0.93–0.95) contains a variety of enzymes involved in RNA metabolism (two helicases, Argonaute 1, two RNA-binding proteins) and ubiquitin-mediated proteolysis, two SH3 domain proteins, three kinases, as well as eight proteins of unknown function (Supplementary Data File 2). Notably, 10 out of these 30 proteins do not have orthologs in budding yeast. In addition, seven out of those 20 that do have orthologs did not give signals in the TAP-tagging approach (Ghaemmaghami *et al*, 2003).

Our quantitative data also indicated that the median abundance of metazoan core proteins (ASC=24.2) is significantly higher than that of all proteins detected (ASC=14.6,  $P<0.05$ ), whereas the abundance of *S. pombe*-specific proteins is considerably lower (ASC=5.5; Figure 2C). This finding is consistent with the higher representation of core proteins in



**Figure 2** Label-free relative quantification of *S. pombe* proteins. **(A)** Correlation of published absolute quantitation data for several cytokinesis proteins with their corresponding ASCs. **(B)** ASCs for each of the 1465 identified proteins plotted on a log scale. **(C)** Median ASCs for proteins belonging to the indicated categories. All numbers were statistically different at  $P < 0.05$  (TMD=transmembrane domain). **(D)** Median ASCs for subunits of the indicated protein complexes. For protein complexes with few subunits,  $P$ -values are not always  $< 0.05$  owing to some outliers (see Supplementary Data File 4). **(E)** Correlation between the Pombe-ASC and the *Cerevisiae*-ASC data sets.

our data set (Figure 1E) and with their higher mRNA levels as reported previously (Mata and Bahler, 2003). In addition, essential proteins are considerably more abundant (median ASC=12.6) than non-essential proteins (ASC=7.5). This finding can be rationalized by the enrichment of highly expressed core proteins in the set of essential proteins (Supplementary Data File 2).

Analysis of 10 protein complexes for which we identified greater than 80% of their known or predicted subunits, and which are involved in a large variety of cellular processes, indicated that the translation initiation factor eIF4 is the most abundant protein complex in *S. pombe* (median ASC=85.7; Figure 2D). eIF4 is similar in abundance to the ribosome (ASC=70.7), but three- to four-fold more abundant than eIF2 (ASC=21.7) and eIF3 (ASC=32.0), two other translation initiation factor complexes (Figure 2D). Although, during the process of translation initiation, all of these eIFs are known to join a stoichiometric 43S initiation complex, it is thought that eIF2 and eIF3, but not eIF4, dissociate from the mRNA upon successful scanning for the initiator AUG codon (Gebauer and Hentze, 2004). Our finding that eIF2 and eIF3 are considerably less abundant than eIF4 and the ribosome therefore, underpins the concept that the former eIFs are only transiently involved during the initiation reaction, whereas the cap-binding eIF4 complex and the ribosome stay on the mRNA during

translation. Our data also indicate that the protein synthesis machinery (ribosome, eIFs) and the protein folding and degradation machinery (CCT chaperonin, proteasome) are among the most abundant molecular modules in fission yeast and perhaps other eukaryotes (Figure 2D).

### Comparison of *S. pombe* proteome data with *S. cerevisiae*

We compared the abundance ranked list of *S. pombe* proteins with similar lists of *S. cerevisiae* proteins. This was carried out for the subset of proteins that have known or predicted orthologs in budding yeast (1285 of 1465 proteins based on ortholog mapping information in *S. pombe* GeneDB ([www.genedb.org/genedb/pombe/index.jsp](http://www.genedb.org/genedb/pombe/index.jsp))). Two data sets of *S. cerevisiae* proteins were used. The first was derived from published 2D LC ESI MS/MS data (Liu *et al*, 2004) that we subjected to our adjustment of spectral counts to the number of tryptic peptides (=Cerevisiae-ASC data set). This set contained 473 pairs of orthologous proteins that were detected in both studies. The second list was assembled from the absolute quantitation data derived from whole-genome ORF tagging with the TAP epitope (Cerevisiae-TAP data set; Ghaemmaghami *et al*, 2003). This data set contained 1033

orthologs, 252 fewer than the theoretically possible 1285, because ~20% of the native fission yeast proteins we detected by 2D LC ESI MS/MS could not be quantified when TAP tagged in budding yeast (Ghaemmaghami *et al*, 2003). For example, our data set contained all 32 subunits of the 26S proteasome (Finley *et al*, 1998; Supplementary Data File 4), whereas only 25 of these subunits were detected in the ORF tagging approach (Ghaemmaghami *et al*, 2003). Similarly, we identified 94% of all cytosolic ribosomal subunits, whereas only 76% were identified by TAP tagging in budding yeast (Supplementary Data File 4).

Both budding yeast data sets correlated with the fission yeast protein list as indicated by Pearson correlation coefficients of 0.56 and 0.45 and Spearman rank correlation coefficients  $r_s$  of 0.55 and 0.42, respectively (Figure 2E; Supplementary Figure 1). Notably, our data showed an overall stronger correlation with the budding yeast 2D LC ESI MS/MS data presented by Liu *et al* (2004) (=Cerevisiae-ASC). This finding reinforces previously expressed notions regarding the limitations of comparing mass spectrometry-based proteomics data to absolute quantitation based on ORF tagging (Liu *et al*, 2004). However, organism-specific differences in protein expression are also expected to distort the correlation (see Figure 5).

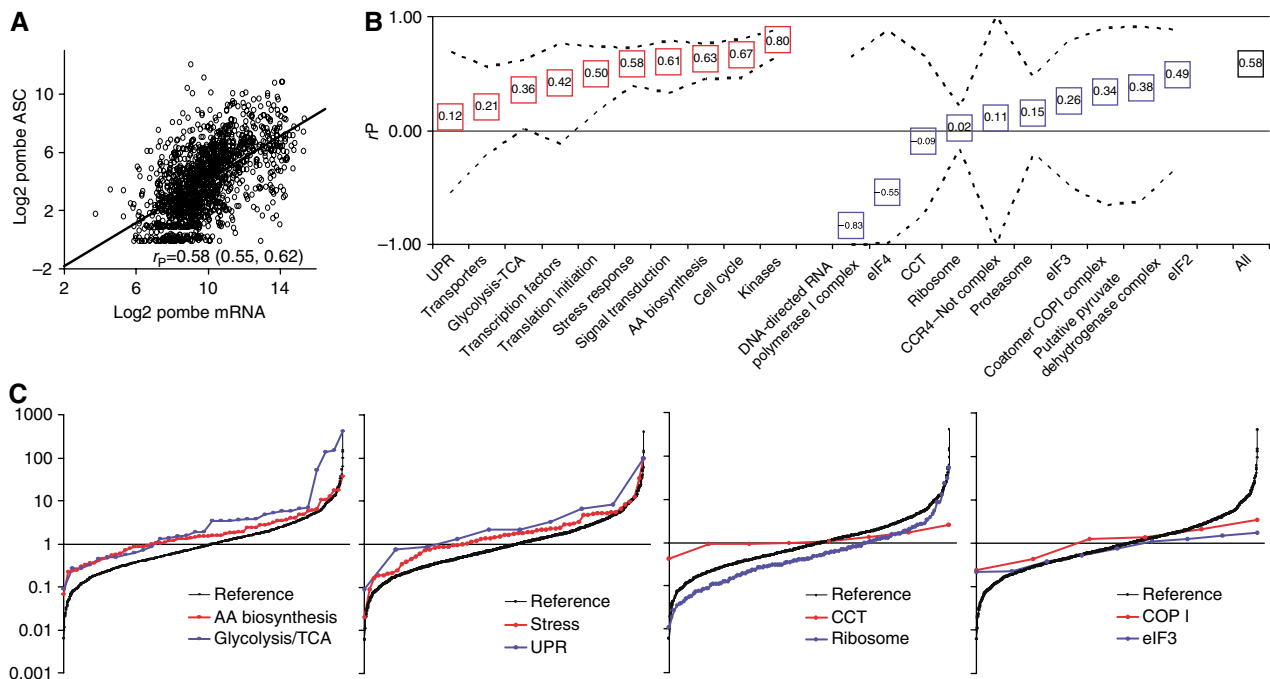
Nonetheless, our LC ESI MS/MS data showed a remarkable overlap with the Cerevisiae-TAP data set in the relative frequency distribution of the detected proteins across the entire dynamic range (Supplementary Figure 2). For example, 88% of the 1033 budding yeast proteins, for which we have identified the fission yeast orthologs, are present at under 50 000 molecules/cell, 62% are under 10 000 molecules/cell, and 11% are under 1000 mole-

cules/cell. This finding suggests that the dynamic range of multidimensional prefractionation and LC ESI MS/MS analysis is not necessarily inferior to that of the wholeORF tagging approach.

### Correlation of protein and mRNA levels in fission yeast

We next determined the overall correlation of our protein data set with mRNA abundance as estimated by cDNA microarray analysis. Total RNA was prepared from the same *S. pombe* strain maintained under identical growth conditions as used for the proteomic analyses, followed by hybridization onto *S. pombe* cDNA microarrays (Oliva *et al*, 2005; Zhou *et al*, 2005). Background subtracted hybridization values averaged from three parallel experiments (see Supplementary Data File 2) were used to estimate mRNA abundance. Although it is clear that the hybridization values obtained on cDNA microarrays are influenced by factors other than mRNA abundance (probe length, GC content, etc.), these variations are relatively minor with probes longer than 500 bp as used here (Lyne *et al*, 2003). Similarly, Mata and Bahler (2003) have previously used absolute hybridization signals as approximate measures of mRNA levels in fission yeast.

The comparison of 1367 protein-mRNA pairs for which data were obtained (Supplementary Data File 2) revealed a Spearman rank correlation coefficient ( $r_s$ ) of 0.61 and a Pearson correlation coefficient ( $r_p$ ) of 0.58 (Figure 3A), indicating a substantial correlation between mRNA and protein abundance in fission yeast. The extent of correlation is very similar in



**Figure 3** Correlation of protein and mRNA levels in fission yeast. **(A)** Scatter plot representing the relationship between mRNA and protein (Pombe-ASC). The Pearson correlation coefficient is indicated. **(B)** Protein-mRNA correlation coefficients for proteins belonging to the indicated pathways, protein families, and complexes. Dashed lines indicate 95% confidence intervals (AA=amino acid, UPR=unfolded protein response, TCA=tricarboxylic acid cycle). **(C)** Protein-mRNA ratios for individual members of the indicated pathways, protein families, or complexes. The data are displayed relatively to median centered ratios of the entire data set of 1381 mRNA-protein pairs (black graphs).



budding yeast as determined with the whole-genome TAP tagging data set ( $r_s=0.57$ , Ghaemmaghami *et al*, 2003), and by an independent re-evaluation of additional large-scale budding yeast data sets ( $r_p=0.66$ ; Greenbaum *et al*, 2003).

The mean mRNA intensity of proteins detected in our multidimensional analysis was 2462, whereas for undetected proteins the number was 420 (Supplementary Figure 3). This comparison confirmed the expectation that mass spectrometry-based proteomics has a bias towards detecting proteins encoded by highly expressed mRNAs. However, a significant portion of low-abundance mRNAs may encode proteins that never accumulate in the vegetative state. Consistent with this notion is the demonstration that 1033 genes are induced more than four-fold during nitrogen starvation and meiosis (Mata *et al*, 2002). The actual vegetative translome may therefore be devoid of many of the proteins encoded by such developmentally regulated mRNAs. These proteins may also escape detection by ORF tagging and immunoblotting, thus explaining why the dynamic range of our LC ESI MS/MS analysis was comparable to the whole-genome ORF tagging approach employed in budding yeast (Ghaemmaghami *et al*, 2003; Supplementary Figure 2).

### Functional pathway analysis

Although the overall protein–mRNA correlation is surprisingly high, we wondered whether this correlation is maintained throughout specific functional pathways, protein families, and multisubunit protein complexes. We calculated the Pearson correlation coefficients for several subclasses of protein–mRNA pairs that were highly represented in our data set (see Supplementary Data Files 3 and 4 for individual proteins). Whereas a high coefficient was obtained for kinases ( $r_p=0.80$ ; Figure 3B), the correlation was weak for transporters ( $r_p=0.21$ ) and the unfolded protein response (UPR) pathway ( $r_p=0.12$ ), and moderately strong for glycolytic enzymes ( $r_p=0.36$ ) and transcription factors ( $r_p=0.42$ ). Correlations similar to those observed for all proteins ( $r_p=0.58$ ) were found for the categories amino-acid biosynthesis ( $r_p=0.63$ ), signal transduction ( $r_p=0.61$ ), protein translation ( $r_p=0.5$ ), stress response ( $r_p=0.58$ ), and cell-cycle regulation ( $r_p=0.67$ ; Figure 3B).

For the majority of multisubunit protein complexes, very low or even negative correlation coefficients were obtained (Figure 3B). Previous bioinformatics studies have suggested that a high protein–mRNA correlation (i.e. the higher the mRNA, the higher the protein) as observed here for kinases and cell-cycle components reflects control of protein abundance primarily at the level of mRNA synthesis, whereas poor correlation is indicative of post-transcriptional control (Greenbaum *et al*, 2003). By extension, negative correlations indicate extensive control at the post-transcriptional level (i.e. the higher the mRNA, the lower the protein and vice versa). The subunits of presumed stoichiometric protein complexes such as the 80S ribosome, the 26S proteasome, and the CCT complex would therefore be controlled substantially at the post-transcriptional level.

The poor protein–mRNA correlation for complexes would be expected, if their subunits were coordinately regulated. For example, if all subunits of a protein complex had exactly equal

protein and mRNA levels, say 5.0 units and 1.0 unit, respectively, then all data points would coincide at the very same coordinates of a protein versus mRNA plot ( $x=5$ ;  $y=1$ ; protein–mRNA ratio=5). Consequently, the protein–mRNA correlation would be zero for the subunits of this protein complex.

Indeed, we noticed that the protein and mRNA data points for many protein complexes were not randomly scattered over the entire data map, but tended to cluster together. To comprehensively illustrate this, we determined the protein–mRNA ratio individually for every protein in a given pathway, family, or complex, and compared it to the entire data set. Individual ratios of functional pathway components were used to determine their location and relative distance on the ratio distribution curve of the entire data set of 1367 protein–mRNA pairs. This reference curve indicates the extent and orientation of the deviations of all observed ratios from the median ratio, which was arbitrarily set to 1.0. The partitioning of pathway components along this curve thus informs about the degree to which they cluster around certain protein–mRNA ratios and their distances from the median. The graphical representation of clustering effects was enhanced by displaying the data points for specific pathway components at equal distance laid over the reference curve, thus causing informative phase shifts of the curves.

This analysis revealed strong deviations from the reference curve for several protein complexes, suggesting more consistent protein–mRNA ratios for individual subunits than observed for all proteins. Ribosomal subunits clustered with relatively higher levels of mRNA than protein (Figure 3C; Supplementary Data File 5), whereas the shape of the ratio distribution curve for eIF3, the COPI1 complex, and several other protein complexes (Supplementary Data File 5) indicated clustering around the median ratio (Figure 3C). This differential distribution was even more pronounced for the eight subunits of the CCT complex (Figure 3C). In other words, all eight subunits of the CCT complex displayed highly similar protein–mRNA ratios, and therefore appear to be coordinately regulated at the mRNA and protein levels. Thus, although the protein–mRNA correlations were low for multisubunit protein complexes, clustering of their protein–mRNA ratios around similar values indicated coordinate regulation of complex subunits (Table I). Although this regulation could principally occur at any level, the low protein–mRNA correlation suggests a substantial contribution of post-transcriptional mechanisms (Greenbaum *et al*, 2003). Notably, the UPR pathway showed a similar pattern in correlation and ratio distribution (Figure 3B and C), perhaps suggesting that components of this pathway are also present in stoichiometric amounts.

The reverse scenario, clustering of protein–mRNA ratios around similar values, but relatively high protein–mRNA correlation, was observed for the stress response pathway as well as for glycolysis and amino-acid biosynthesis (Figure 3B and C). This pattern indicated that protein and mRNA expression varied widely among the members of these groups (Table I). This might reflect the fact that proteins involved in hierarchical signal transduction cascades or linear and circular metabolic pathways do not necessarily cooperate in stoichiometric amounts. Rather signal amplification and the specific activities of metabolic enzymes may govern the varying levels of protein required for these functions.

**Table I** Protein–mRNA relationships

		Protein–mRNA ratio	
		Clustering	No clustering
<i>Protein–mRNA correlation</i>	High	Metabolic and signal transduction pathways Non-coordinate expression	Protein families (kinases), cell cycle Transcriptional control
	Low	Multisubunit protein complexes; UPR Coordinate regulation	Protein families (transporters) Post-transcriptional control

Most other pathways and protein families showed a considerable overlap of protein–mRNA ratios with the reference curve, indicating no clustering. Among those were entities with low (transporters; Figure 3B) and high (kinases; Figure 3B) protein–mRNA correlation. For these remaining cases, high protein–mRNA correlations would suggest control primarily at the transcriptional level, whereas low correlations would indicate extensive post-transcriptional control (Table I) (Greenbaum *et al*, 2003).

### Protein and mRNA relationship as a correlate of post-translational modifications

Although no specific enrichment strategies were employed, rigorous interrogation of our peptide data sets obtained by mass spectrometry provided high confidence indications for post-translational modification (PTM) of 53 peptides, which were matched to 51 proteins. A total of 40 proteins contained at least one peptide that was phosphorylated on either serine, threonine, or tyrosine (Table II). The set of phosphoproteins was enriched for protein kinases (15% versus 1.6% in the entire proteome), a finding that is consistent with the known propensity of these enzymes to autophosphorylate and/or be part of kinase cascades. The budding yeast orthologs of eight of these proteins were previously shown to be phosphorylated by methods other than mass spectrometry. In one case, acetyl coenzyme-A carboxylase, the serine phosphorylation site we mapped in fission yeast exactly corresponds to the same position where the budding yeast protein was found to be modified (Ficarro *et al*, 2002).

For another set of 11 proteins, we mapped the precise sites of modification with the diglycine moieties created upon trypsin digestion of ubiquitylated lysines (Table II). Independent evidence for ubiquitylation of the budding yeast ortholog of one of these proteins was provided previously (Peng *et al*, 2003b). Five proteins contained both phosphorylated and ubiquitylated peptides (Table II), a finding that is consistent with the well-established connection between phosphorylation and ubiquitylation (Karin and Ben-Neriah, 2000).

The median abundance of phosphorylated (ASC=3.58) and ubiquitylated (ASC=2.84) proteins was considerably lower than the abundance of all 1465 proteins in the data set (ASC=14.6; Figure 4A). Ubiquitylated proteins also showed a stark dissociation of median mRNA levels, which were relatively high (633 versus 757 in the entire data set), from protein levels, which were very low (2.84 versus 14.6; Figure 4A). This finding indicates that extensive proteolytic control of these proteins through the ubiquitin–proteasome

pathway may be dominant over their relatively high mRNA expression levels. This conclusion was further strengthened by comparing individual protein–mRNA ratios of ubiquitylated proteins to median adjusted ratios for the entire data set. This analysis revealed clustering of ubiquitylated proteins with relatively higher mRNA than protein levels, whereas phosphoproteins showed a distribution largely congruent with the reference curve (Figure 4B).

### Steady-state proteome and transcriptome comparison of *S. pombe* and *S. cerevisiae*

The generation of quantitative fission yeast protein and mRNA data sets and the availability of corresponding data sets for budding yeast enabled the first large-scale comparison of mRNA and protein levels of two eukaryotic organisms. For this, we used the *Cerevisiae*-MS data (Liu *et al*, 2004) with adjustment of spectral counts to the number of tryptic peptides and published mRNA data derived from cDNA microarray analysis of wild-type *S. cerevisiae* grown under conditions comparable to those of our fission yeast strains (Gasch *et al*, 2001). As the raw values of the four data sets were on different scales, they were log-transformed and standardized (see Supplementary information). As a result, each data set contained a continuum of mRNA and protein values ranging from high to low abundance for 445 distinct entities common to all four data sets. A self-organizing map (SOM) algorithm was used to arrange the four data sets into distinct clusters (see Supplementary information). The algorithm was instructed to assemble 16 clusters, because this number achieved good performance in reproducibility (data not shown), average cluster homogeneity (0.81), and separation (−0.048).

The SOM revealed many similarities in the mRNA and protein abundance patterns in the two yeasts, but also marked differences. The most frequent patterns represented roughly equal mRNA and protein levels in both organisms (clusters 3, 4, 7, 9, 10, and 13; Figure 5A). In addition, one pattern was indicative of concordantly low mRNA and high protein abundance in both yeasts (cluster 6), whereas cluster 15 showed the opposite pattern. Among the discordant patterns were those with higher mRNA and protein levels in *S. pombe* (cluster 1), as well as various patterns where either mRNA or protein levels found in one yeast deviated from what was found in the other (clusters 2, 5, 8, 11, 12, 14, and 16).

The clusters were further interrogated for overrepresented *S. pombe* GO terms using the FuncAssociate tool (Berriz *et al*, 2003). In total, seven nonredundant GO attributes were found significantly ( $P \leq 0.0005$ ) overrepresented in the clusters

**Table II** Post-translationally modified peptides

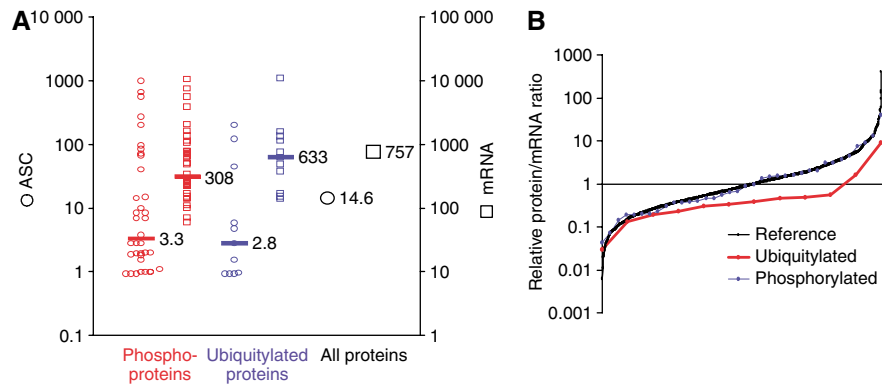
Name	Product	Phosphopeptide	Xcorr	DeltCN	ObsM + H +	SpScore	Ion %
sup45	Translation release factor RF1	K.FHT#EALAEELLES#DQR.F	2.8454	0.1142	1920.1	424.4	0.536
tpi1	Triosephosphate isomerase mitochondrial ribosomal protein subunit	R.RT#IFKES#DEFVADK.T	2.6786	0.1308	1845.7	265.6	0.462
SPAC16E8.10c	S7	K.AKAEKIVAT#ALS#IIQK.E	2.6066	0.1585	1845.1	433.1	0.5
SPBC16G5.07c	Prohibitin	R.FS#RILT#PGVAFLAPIIDK.I	3.3233	0.111	2118.3	474.6	0.441
SPAC18B11.11	GTPase activating protein	K.VLS#EWLT#DLFTIIDD#R.A	3.1806	0.1099	2244.04	273.3	0.406
SPAC23G3.12c	Serine protease	R.Y#VEVCGAKFHNLQYQLAR.Q	2.6886	0.1181	2177.5	643.3	0.412
		K.K@GT#ALVLDKDKGLAVT#S#R.S	3.0538	0.1095	2225.3	809	0.5
SPAC20H4.09	ATP-dependent RNA helicase	R.T#LS#T#DLLLGVLK.R	3.0509	0.1078	1513.53	268	0.545
SPAPB2B4.04c	p-type calcium ATPase	R.T#EGQAT#PLQLRLS#R.V	2.789	0.1696	1809.92	257.4	0.462
SPBC342.02	Glutaminyl-Trna synthetase	R.LMFLPDIKVTLENLDDDS#Y#R.E	2.6182	0.1042	2540.22	574.3	0.447
cek1	Serine/threonine protein kinase Cek1	K.PENLLIS#QNGHLK.L	3.2289	0.1179	1543.77	585.9	0.625
cut6	Acetyl-CoA carboxylase	R.LQS#VSDLSWYVVK.T	3.9381	0.1559	1618.63	1756.7	0.792
ef1a-c	Translation elongation factor EF-1 alpha	K.MVPS#KPMCVEAFTDYAPLGR.F	3.8162	0.4103	2293.29	985.8	0.5
ncs1	Related to neuronal calcium sensor Ncs1	K.NKDGQLTLEEFCEGS#KR.D	2.7358	0.1945	2033.72	335.4	0.406
nog1	GTP binding protein Nog1	R.EGYYDS#DQEIEDADEEEVLEK.A	4.3428	0.1168	2584.66	837.1	0.475
psm3	Mitotic cohesin complex subunit Psm3	K.S#KVALELQSSQLSRQIEFSK.K	3.1029	0.1884	2357.52	699.9	0.447
rps1201	40S ribosomal protein S12	R.QAHLCVLCES#CDQEAYVK.L	2.8358	0.2407	2119.43	519.4	0.412
tef3	Translation elongation factor 3	R.FKLRYKLGNMSEFVK.K	2.6666	0.1287	1940.79	389.6	0.464
tif471	Translation initiation factor eIF4G	R.SGSQVSDQVVEPNSSTLS#PR.N	5.5144	0.1612	2241.52	1063.1	0.55
win1	MAP kinase kinase kinase Win1	R.LSDNELAS#FVK.E	2.7442	0.1567	1302.75	1234.8	0.8
SPCC4B3.11c	BolA domain	K.S#K@AFQGKNTLAQHR.L	2.7209	0.1147	1780.73	296.8	0.462
SPCC16C4.02c	Sequence orphan	K.NLSSATVLS#NLLK.A	2.4108	0.1971	1553.65	521.2	0.462
SPAC4G8.09	Leucine-tRNA ligase	K.VQLSYQKM#S#K.S	2.6622	0.3681	1308.15	645	0.722
SPBC27B12.08	AP-1 accessory protein	K.VVS#LMIELLENLTVNDPK.L	3.3324	0.1322	2178.22	600.8	0.444
SPAC25A8.01c	Fun thirty related protein Fft3	K.KS#QVLDALPKKTR.I	2.6176	0.1062	1563.11	631.2	0.542
SPBC8D2.06	Isoleucine-tRNA ligase	K.NVIVS#GLVMAEDGKKM#SK@R.L	2.9176	0.1584	2273.37	515	0.417
mug28	Meiotically upregulated gene Mug28	K.VHDKENAFAEATGTSILS#S#K.A	2.8545	0.2187	2264.5	582.9	0.421
sec26	Coatamer beta subunit	R.AS#LGEVPILAS#EEQLLK.D	2.6189	0.2339	1956.93	265.6	0.406
SPBC29A3.09c	AAA family ATPase voltage- dependent anion-selective Channel	K.ELEELS#KDQTADQAIS#R.R	3.04	0.2224	2093.99	300.9	0.406
SPAC1635.01	Alpha-glucosidase Mall1	K.Y#ALDKDT#FVK.G	2.7147	0.1355	1360.55	456.1	0.611
mall1	Alpha-glucosidase Mall1	R.TPM#HWDSSPNGGFT#K.A	2.9814	0.2203	1759.18	1212	0.607
ppk22	Serine/threonine protein kinase	-.MARET#EFNDK.S	3.1204	0.2113	1321.71	530	0.722
usol1	Armadillo repeat protein	K.LT#KQLDDIK@NQFGIISSK.N	3.3203	0.1539	2241.64	614.4	0.412
SPBC947.10	Zinc finger protein	K.RAFSEIKNAT#FLNIPER.V	3.1877	0.114	2086.23	748.3	0.5
ppk18	Serine/threonine protein kinase Ppk18	K.QKTELAT#FT#TY#K.E/	2.4263	0.104	1671.67	311.6	0.5
cwf19	Complexed with Cdc5 protein Cwf19	K.QKTELAT#FTT#Y#K.E R.KYGQNYEY#AKQIAK.D	2.4263 2.908	0.104 0.1413	1671.67 1783.43	311.6 394.4	0.5 0.5
erg8	Phosphomevalonate kinase	K.GY#ASTTTLDDKCGTVRVK.S	3.0426	0.1915	1995.3	900.7	0.529
ppk14	Serine/threonine protein kinase	K.SGK@FY#AM#KVLKQEM#IK.R	3.3777	0.1358	2215.44	1034.9	0.562
rgal1	GTPase activating protein	K.NSGAIY#DKNDGTQK.G	2.6184	0.1953	1592.09	492.5	0.462
SPAC11E3.12	Conserved eukaryotic protein	K.IY#GVNTKEKLVDM#EALTQK.K	2.6351	0.115	2388.2	257.5	0.421
SPAC2F3.13c	Queuine tRNA-ribosyltransferase	R.ELVAVILLQLY#VYIKEHGK.E	2.8689	0.103	2396.72	526.9	0.5

Name	Product	Ubiquitylated peptide	Xcorr	DeltCN	ObsM + H +	SpScore	Ion %
ppk14	Serine/threonine protein kinase	K.SGK@FY#AMKVLKQEM#IK.R	3.3777	0.1358	2215.44	1034.9	0.562
SPAC589.10c	Ribosomal-ubiquitin fusion protein	R.TLSDYNIQK@ESTLHLVLR.L	4.396	0.4338	2244.78	1164.9	0.559
kap1	Kinesin-associated protein	K.IGSSATSGSFPVIKSLM#DK@R.S	4.3064	0.2151	2211.84	2044.8	0.408
SPAC23G3.12c	Serine protease	K.K@GT#ALVLDKDKGLAVT#S#R.S	3.0538	0.1095	2225.3	809	0.5
ago1	Argonaute	K.NK@SDGDRNCGNPLPGTIEK.H	2.7063	0.1565	2138.62	282	0.444
		K.LT#KQLDDIK@NQFGIISSK.N	3.3203	0.1539	2241.64	614.4	0.412
SPCC4B3.11c	BolA domain	K.S#K@AFQGKNTLAQHR.L	2.7209	0.1147	1780.73	296.8	0.462
gtp1	GTP binding protein Gtp1	R.LARLPK@SVVISCNM#K.L	2.8006	0.2104	1888.6	551.5	0.5
cct5	CCT-complex epsilon subunit	K.EKFQEM#IK@HVK.D	2.4524	0.2364	1547.71	369.7	0.65
SPCP1E11.11	RNA-binding protein	K.VASKLIVIIK@K.Y	2.2838	0.1112	1326.54	443.2	0.55
rnp24	RNA-binding protein Rnp24	R.FNDAESLQGEDKPNFK@RAR.K	2.6686	0.174	2335.83	296.2	0.444
SPBC8D2.06	Isoleucine-tRNA ligase	K.NVIVS#GLVMAEDGKKM#SK@R.L	2.9176	0.1584	2273.37	515	0.417

# denotes phosphorylated residues, @ denotes ubiquitylated residues, \* denotes oxidated residues.





**Figure 4** Protein and mRNA ratios of post-translationally modified proteins. **(A)** mRNA levels and ASCs for 40 phosphorylated and 11 ubiquitylated proteins as compared to the entire data set. Medians are indicated by the vertical bars. **(B)** Individual protein–mRNA ratios for phosphorylated and ubiquitylated proteins relative to median centered ratios of the entire data set.

(Figure 5A). As noise is a notorious feature of large-scale functional genomics data, the biological significance of these patterns will require further validation by more targeted experiments. However, as not a single GO attribute was enriched in SOM clusters derived from a random data set under identical conditions (data not shown), our data suggest that many pathways and complex subunits are coordinately, albeit not necessarily concordantly regulated in both fission and budding yeasts. For example, 6/13 components of the microtubule cytoskeleton organization GO category present in our data sets were coordinately and concordantly regulated in both yeasts (cluster 10; Figure 5B). In contrast, ATPases and entities involved in chromatin remodelling and intracellular transport were coordinately, but discordantly regulated with mRNA levels being low in budding yeast (cluster 14, Figure 5C–E).

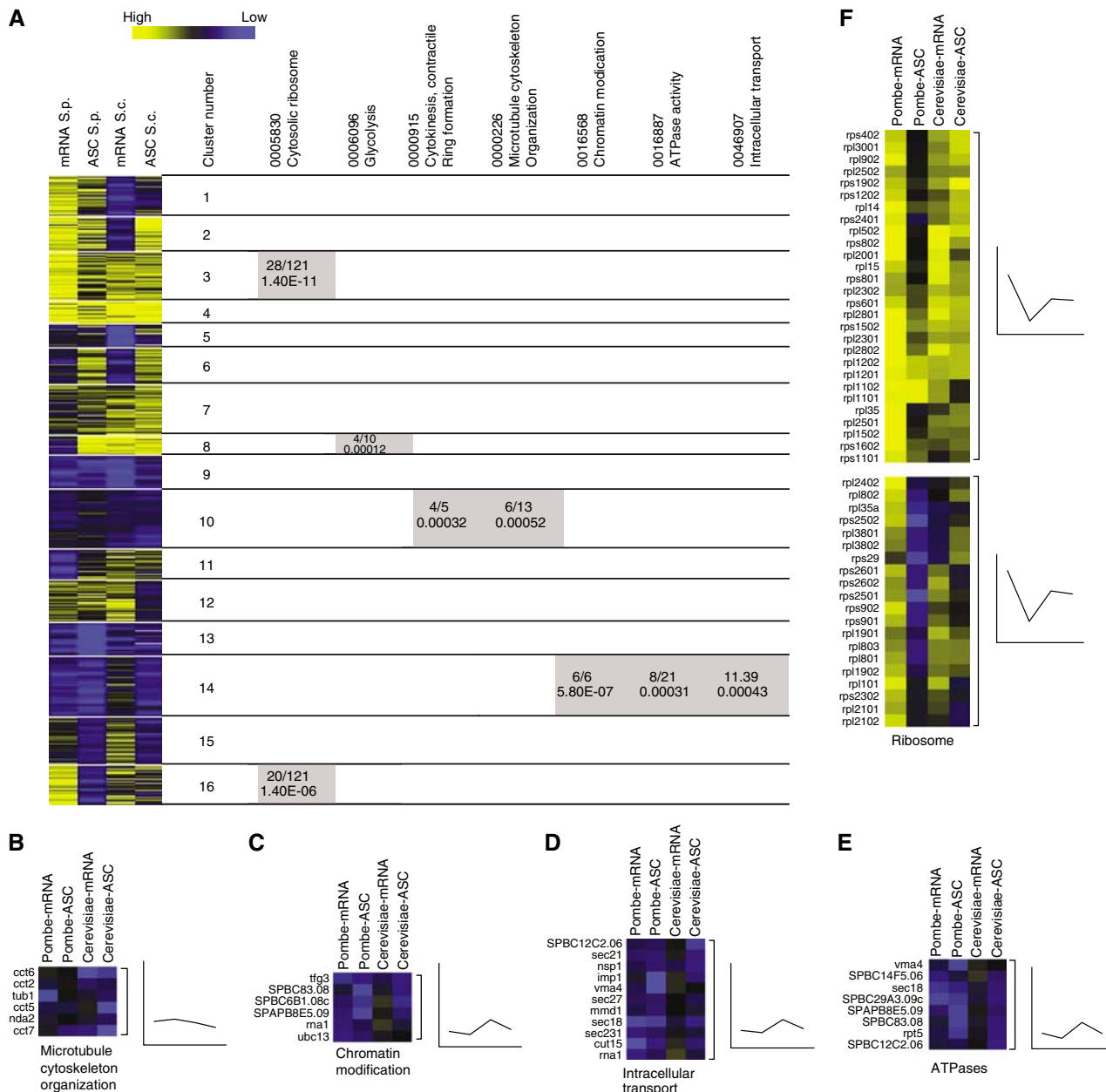
Although 48 out of 121 fission yeast ribosomal subunits present in all data sets were coordinately regulated, they partitioned into two distinct clusters (3 and 16, Figure 5A). Both clusters indicated that fission yeast ribosomal protein mRNAs are typically higher than the subunits they encode (Figures 5F and 3C). Notably, higher mRNA than protein levels were previously reported also in human cells (Ishihama *et al*, 2005). Coordinate post-transcriptional regulation of ribosomal proteins in both budding and fission yeasts was already observed in previous reports (Washburn *et al*, 2003; Bachand *et al*, 2006). Ribosomal proteins are known to be subject to extensive transcriptional and post-transcriptional control as indicated by short mRNA half-lives (Li *et al*, 1999) and extensive translational regulation (Meyuhas, 2000; Bachand *et al*, 2006). Although presumably serving to provide stoichiometric amounts of complex subunits, such control might also ensure the excess availability of individual ribosomal subunits that fulfill extraribosomal functions (Wool, 1996), a repertoire, that may vary from one organism to another.

Overall, our comparison reinvigorates the conclusion gained from previous functional genomics studies that similarities in the control of gene expression in the two yeasts are less pronounced than expected from genome comparisons (Mata *et al*, 2002; Rustici *et al*, 2004; Oliva *et al*, 2005). Only a remarkably small fraction of transcriptomic changes during

cell-cycle progression (Rustici *et al*, 2004; Oliva *et al*, 2005) and sexual differentiation (Mata *et al*, 2002) is shared among the two yeasts. True organism-specific differences are therefore likely to underlie the moderate overall correlation in protein abundance in the two yeasts (Figure 2E and Supplementary Figure 1) as well as the different patterns of mRNA and protein expression revealed here by SOM clustering (Figure 5).

## Conclusions

Shotgun proteomics employing multidimensional prefractionation and tandem mass spectrometry, aided by mathematical modelling of spectral count information, enabled a label-free relative quantitation of ~30% of the theoretical fission yeast proteome corresponding to an estimated 50% of the entire vegetative translome. Whereas Eno101, a subunit of the phosphopyruvate hydratase complex, was revealed as the single most abundant protein, the translation initiation factor eIF4 represents the most abundant protein complex. Highly abundant proteins also included the core set of proteins conserved in metazoans. Among the least abundant proteins observed in this study were *S. pombe*-specific proteins, a series of nonessential proteins, as well as proteins modified by phosphorylation and ubiquitylation. Whereas there was a positive overall correlation between protein and mRNA abundance in fission yeast similar to what was observed in other organisms, simple correlations proved insufficient to assess regulatory patterns of gene expression. Contrasting individual protein–mRNA ratios to the ratio distribution curve representing all entities suggested common schemes of control for subunits of protein complexes, unstable ubiquitylated proteins, and several functional pathways. The first large-scale comparison of mRNA and protein abundance in two related eukaryotic model organisms indicated frequently coordinate, but rarely concordant regulation, an observation that further underscored the marked differences in gene expression in the two yeasts noted previously (Mata *et al*, 2002; Rustici *et al*, 2004; Oliva *et al*, 2005). The data presented should become a valuable resource for the fission yeast community as well as researchers mining comprehensive gene expression data sets for systems biology.



**Figure 5** Comparison of proteome and transcriptome data from *S. pombe* and *S. cerevisiae*. **(A)** Self-organizing map cluster analysis of the fission and budding yeast mRNA and ASC protein data sets. The table on the right shows GO terms overrepresented in the various clusters and the *P*-values of enrichment. Also indicated is the number of genes and proteins with a particular GO attribute enriched in each cluster over the total number with this attribute present in the entire data set. The names of the genes and proteins in the individual clusters are listed in Supplementary Data File 6. **(B–E)** Detailed view of subclusters containing (B) microtubule cytoskeleton organization (GO: 0000226), (C) chromatin modification components (GO: 0016568), (D) components involved in intracellular transport (GO: 0046907), (E) ATPases (GO: 0016887), and (F) ribosomal proteins (GO: 0005830). The graphs next to the heat maps indicate the mean variations in signal intensities.

## Materials and methods

### Preparation of fission yeast cell lysate

*S. pombe* cells (DS 448/2=927 *h-leu-1-32 ura4d-18*) were grown in 50 ml YES to mid-log phase ( $OD_{595}=0.68$ ). Cells were washed in STOP buffer (150 mM NaCl, 10 mM EDTA, 50 mM NaF, 1 mM  $NaN_3$ ) and lysed in 450  $\mu$ l buffer containing 7.7 M urea, 2.2 M thiourea, 0.55% CHAPS, 10 mM Tris (pH 8.5), 200 mM DTT and protease inhibitors by bead lysis in a Fastprep device (Bio 101). The cell homogenate was cleared by centrifugation and the bead lysis was repeated once with the pellet of insoluble debris. The two homogenates were pooled (950  $\mu$ l)

and incubated at room temperature (RT) for 30 min. A volume of 5.2 ml of 99% *N,N*-dimethylacrylamide (Sigma) was added, followed by another incubation at RT for 30 min after which 10  $\mu$ l 2 M DTT was added for 5 min at RT. The homogenate was cleared by centrifugation for 15 min at 14 000 g, resulting in a denatured, reduced, and alkylated sample with a concentration of  $\sim 10$  mg/ml.

### Sample prefractionation

Sample prefractionation by IEF on the ZOOM device (Invitrogen), the multicompartment electrolizer (MCE, Proteome Systems), on immo-

bilized pH gradient (IPG) gel strips, by strong anion exchange (SAX) membrane adsorber spin columns (VivaScience), and by 1D-PAGE was performed as described in detail in Supplementary information.

## LC ESI MS/MS

Trypsin digestion before LC MS analysis as well as protein identification by 1D Nano-LC tandem mass spectrometry and on-line 2-D LC ESI-MS/MS analysis on a Thermo Electron LCQ Deca XP Plus ion trap instrument are described in Supplementary information. This section also contains details on the SEQUEST database searching criteria and the parameters for adjusting the false positive peptide identification rate to 1% as determined by searching a combined forward and reverse *S. pombe* proteome database. Search parameters for the identification of PTMs are also stated in Supplementary information.

## Statistical analysis

Spectral count modelling by likelihood-based goodness-of-fit criteria was performed by negative binomial log-linear regression. The best-fit statistics were obtained for a model considering the number of fully tryptic peptides assuming one miscleavage. This model was used for adjusting spectral counts to protein size.  $r_p$  and  $r_s$  between ASCs and mRNA and budding yeast protein data sets were computed. For SOM cluster analysis, data were preprocessed by log-transforming and subsequent standardization. Each of the variables was standardized by subtracting its mean and dividing by its standard deviation. A full description of all statistical methods is presented in the Supplementary information.

## Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website ([www.nature.com/msb](http://www.nature.com/msb)).

## Acknowledgements

We thank J Leatherwood for help with cDNA microarrays and privileged access to unpublished data, V Wood for access to fission yeast genome data, and K Doud for expert technical assistance. MWS is grateful to DH Wolf (University of Stuttgart) for support. This work was funded by NIH grant GM59780 to DAW and by the NIEHS Center grant ES-00002.

## References

Anderson NL, Matheson AD, Steiner S (2000) Proteomics: applications in basic and applied biology. *Curr Opin Biotechnol* **11**: 408–412

Bachand F, Lackner DH, Bahler J, Silver PA (2006) Autoregulation of ribosome biosynthesis by a translational response in fission yeast. *Mol Cell Biol* **26**: 1731–1742

Bakhtiar R, Tse FL (2000) Biological mass spectrometry: a primer. *Mutagenesis* **15**: 415–430

Berriz GF, King OD, Bryant B, Sander C, Roth FP (2003) Characterizing gene sets with FuncAssociate. *Bioinformatics* **19**: 2502–2504

Chen D, Toone WM, Mata J, Lyne R, Burns G, Kivinen K, Brazma A, Jones N, Bahler J (2003) Global transcriptional responses of fission yeast to environmental stress. *Mol Biol Cell* **14**: 214–229

Doud MK, Schmidt MW, Hines D, Naumann C, Kocourek A, Kashani-Poor N, Zeidler R, Wolf DA (2004) Rapid prefractionation of complex protein lysates with centrifugal membrane adsorber units improves the resolving power of 2D-PAGE-based proteome analysis. *BMC Genomics* **5**: 25

Eng JK, McCormack AL, Yates III JR (1994) An approach to correlate MS/MS data to amino acid sequences in a protein database. *Am Soc Mass Spectrom* **5**: 976–989

Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* **20**: 301–305

Finley D, Tanaka K, Mann C, Feldmann H, Hochstrasser M, Vierstra R, Johnston S, Hampton R, Haber J, McCusker J, Silver P, Frontali L, Thorsness P, Varshavsky A, Byers B, Madura K, Reed SI, Wolf D, Jentsch S, Sommer T, Baumeister W, Goldberg A, Fried V, Rubin DM, Toh-e A (1998) Unified nomenclature for subunits of the *Saccharomyces cerevisiae* proteasome regulatory particle. *Trends Biochem Sci* **23**: 244–245

Gasch AP, Huang M, Metzner S, Botstein D, Elledge SJ, Brown PO (2001) Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol Biol Cell* **12**: 2987–3003

Gebauer F, Hentze MW (2004) Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5**: 827–835

Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA* **100**: 6940–6945

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* **425**: 737–741

Greenbaum D, Colangelo C, Williams K, Gerstein M (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* **4**: 117

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**: 994–999

Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB (2001) molecular evidence for the early colonization of land by fungi and plants. *Science* **293**: 1129–1133

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**: 1265–1272

Karin M, Ben-Neriah Y (2000) Phosphorylation meets ubiquitination: the control of NF- $\kappa$ B activity. *Annu Rev Immunol* **18**: 621–663

Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, Scott MS, Gramolini AO, Morris Q, Hallett MT, Rossant J, Hughes TR, Frey B, Emili A (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **125**: 173–186

Li B, Nierras CR, Warner JR (1999) Transcriptional elements involved in the repression of ribosomal protein synthesis. *Mol Cell Biol* **19**: 5393–5404

Liu H, Sadygov RG, Yates III JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**: 4193–4201

Lyne R, Burns G, Mata J, Penkett CJ, Rustici G, Chen D, Langford C, Vetrie D, Bahler J (2003) Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics* **4**: 27

Marguerat S, Jensen TS, de Lichtenberg U, Wilhelm BT, Jensen LJ, Bahler J (2006) The more the merrier: comparative analysis of microarray studies on cell cycle-regulated genes in fission yeast. *Yeast* **23**: 261–277

Mata J, Bahler J (2003) Correlations between gene expression and gene conservation in fission yeast. *Genome Res* **13**: 2686–2690 [E-pub 2003 Nov 12]

Mata J, Lyne R, Burns G, Bahler J (2002) The transcriptional program of meiosis and sporulation in fission yeast. *Nat Genet* **32**: 143–147

Matsuyama A, Arai R, Yashiroda Y, Shirai A, Kamata A, Sekido S, Kobayashi Y, Hashimoto A, Hamamoto M, Hiraoka Y, Horinouchi S, Yoshida M (2006) ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol* **24**: 841–847, [E-pub 2006 Jun 25]

- Meyuhas O (2000) Synthesis of the translational apparatus is regulated at the translational level. *Eur J Biochem* **267**: 6321–6330
- Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, Skiena S, Futcher B, Leatherwood J (2005) The cell cycle-regulated genes of *Schizosaccharomyces pombe*. *PLoS Biol* **3**: e225
- Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**: 376–386
- Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003a) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2**: 43–50
- Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP (2003b) A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol* **21**: 921–926
- Peng X, Karuturi RK, Miller LD, Lin K, Jia Y, Kondu P, Wang L, Wong LS, Liu ET, Balasubramanian MK, Liu J (2005) Identification of cell cycle-regulated genes in fission yeast. *Mol Biol Cell* **16**: 1026–1042, [E-pub 2004 Dec 22]
- Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**: 1154–1169, [E-pub 2004 Sep 22]
- Rustici G, Mata J, Kivinen K, Lio P, Penkett CJ, Burns G, Hayles J, Brazma A, Nurse P, Bahler J (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat Genet* **36**: 809–817
- Washburn MP, Koller A, Oshiro G, Ulaszek RR, Plouffe D, Deciu C, Winzeler E, Yates III JR (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **100**: 3107–3112
- Washburn MP, Ulaszek R, Deciu C, Schieltz DM, Yates III JR (2002) Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* **74**: 1650–1657
- Washburn MP, Wolters D, Yates III JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242–247
- Wood V (2006) *Schizosaccharomyces pombe* comparative genomics, from sequence to systems. In *Comparative Genomics using fungi as models. Topics in Current Genetics*, Sunnerhagen P and Piskur J (eds) Vol. 15, pp. 233–285. Heidelberg: Springer Verlag
- 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 EJ, 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 MA, Rabinowitsch E, Rutherford K, Rutter S, Saunders D, Seeger K, Sharp S, Skelton J, Simmonds M, Squares R, Squares S, Stevens K, Taylor K, Taylor RG, Tivey A, Walsh S, Warren T, Whitehead S, Woodward J, Volckaert G, Aert R, Robben J, Grymonprez B, Weltjens I, Vanstreels E, Rieger M, Schafer M, Muller-Auer S, Gabel C, Fuchs M, Fritzc C, Holzer E, Moestl D, Hilbert H, Borzym K, Langer I, Beck A, Lehrach H, Reinhardt R, Pohl TM, 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 SJ, Xiang Z, Hunt C, Moore K, Hurst SM, Lucas M, Rochet M, Gaillardin C, Tallada VA, Garzon A, Thode G, Daga RR, Cruzado L, Jimenez J, Sanchez M, del Rey F, Benito J, Dominguez A, Revuelta JL, Moreno S, Armstrong J, Forsburg SL, Cerrutti L, Lowe T, McCombie WR, Paulsen I, Potashkin J, Shpakovski GV, Ussery D, Barrell BG, Nurse P (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**: 871–880
- Wool IG (1996) Extraribosomal functions of ribosomal proteins. *Trends Biochem Sci* **21**: 164–165
- Wu JQ, Pollard TD (2005) Counting cytokinesis proteins globally and locally in fission yeast. *Science* **310**: 310–314
- Yates JR (2000) Mass spectrometry. From genomics to proteomics. *Trends Genet* **16**: 5–8
- Zhou C, Arslan F, Wee S, Krishnan S, Ivanov A, Oliva A, Leatherwood J, Wolf DA (2005) PCI proteins eIF3e and eIF3m define distinct translation initiation factor 3 complexes. *BMC Biol* **3**: 14