A Comparative Analysis of an Orthologous Proteomic Environment in the Yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe

Assen Roguev‡§, Anna Shevchenko§¶, Daniel Schaft‡, Henrik Thomas¶, A. Francis Stewart‡¶, and Andrej Shevchenko¶

The sequential application of protein tagging, affinity purification, and mass spectrometry enables highly accurate charting of proteomic environments by the characterization of stable protein assemblies and the identification of subunits that are shared between two or more protein complexes, termed here “proteomic hyperlinks.” We have charted the proteomic environments surrounding the histone methyltransferase, Set1, in both yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe. Although the composition of these nonessential Set1 complexes is remarkably conserved, they differ with respect to their hyperlinks to their proteomic environments. We speculate that conservation of the core components of protein assemblies and variability of hyperlinks represents a general principle in the molecular organization of eukaryotic proteomes. Molecular & Cellular Proteomics 3:125–132, 2004.

Advances in genomic sequencing, gene manipulation technology, and mass spectrometry have stimulated efforts to decipher the functional organization of eukaryotic proteomes by systematic isolation and characterization of native protein complexes on a genomic scale (1, 2). Proteins can be tagged and purified, along with their interaction partners, under native conditions by immunoaffinity chromatography followed by their identification and characterization of post-translational modifications by mass spectrometry (reviewed in Ref. 3).

Documented protein-protein interactions and the composition of native protein complexes are very valuable resources, although biological interpretation of this knowledge is not straightforward. So far the concordance of the results obtained by two independent (although similar) protein tagging approaches has been rather poor (1, 2, 4). Furthermore, these data are also in poor concordance with a dataset obtained by two-hybrid screening (5) or inferred via various bioinformatic approaches (6, 7). Although the availability of complementary data is always a positive factor, it seems rather unlikely that observed discrepancies and excessive complexity of protein assemblies could be attributed solely to errors in analytical methods.

Bruce Alberts pictured the cell as a “collection of protein machines” (8). Comparison of the composition and linkages of similar machines in phylogenetically diverged organisms could provide insight into their molecular architecture, regulation, and involvement in various intracellular processes. The fission yeast Schizosaccharomyces pombe is an appropriate organism to validate and extend our understanding of the functional organization of the proteome of Saccharomyces cerevisiae. The tandem affinity purification (TAP) procedure (9–11), which has been successfully employed in purifying protein complexes from the budding yeast, also works in the fission yeast (12, 13). In many (although not in all) cases, bioinformatics can be used either to identify orthologous proteins in the two organisms by homology searches and close inspection of aligned full-length sequences, or to limit the selection to a small number of plausible candidates whose sequences share reasonable percentage of identity and/or display similar functional domains. At the same time, the two organisms are quite distant phylogenetically and have remarkably different physiology (14, 15). The completely sequenced genome of S. pombe (16) provides a valuable resource for facile mass spectrometric identification of isolated proteins.

A combination of the TAP method and “shotgun” mass spectrometric sequencing (17) was applied for comparative

© 2004 by The American Society for Biochemistry and Molecular Biology, Inc. This paper is available on line at http://www.mcponline.org
characterization of orthologous complexes involved in splicing (18) and cell cycle regulation (19). A remarkable conservation of the composition of orthologous complexes was reported, and a few novel interactors were discovered. However, the shotgun approach lacked aspects of quantification to determine if novel proteins were present in the stoichiometric or substoichiometric amounts, compared with the conserved core subunits. Furthermore, it is not uncommon that proteins are shared between individual protein complexes (4, 11) (we termed such proteins “proteomic hyperlinks”). If a hyperlink protein was inadvertently used as a bait, a mixture of subunits from two or more protein complexes might be isolated. Therefore, phylogenetic interpretation of the differences identified in proteomic environments could be ambiguous. It would be difficult to distinguish if the core of the complex was altered by adding/removing another subunit, or a new interactor was identified, or a novel association represents a yet unknown hyperlink to another individual complex.

In recent years, we and others successfully applied sequential epitope tagging immunoaffinity chromatography and mass spectrometry (SEAM) to characterize protein complexes and segments of protein interaction networks in the budding yeast (4, 11, 20). Although laborious, the approach enabled us to make clear distinction between individual complexes and to identify relevant proteomics hyperlinks. Using the TAP method, we previously isolated and characterized the Sc_Set1C complex (termed after the set1 gene, whose sequence possesses a characteristic SET domain (21)) that methylates lysine 4 in histone H3 and is implicated in epigenetic regulation (22, 23).

Sc_Set1C is comprised of eight subunits, seven of which pulled down the same eight proteins with similar relative stoichiometry upon TAP tagging and immunoaffinity isolation (22). However, the eighth protein, Swd2, was the notable exception. The pool of proteins co-isolated with Swd2 included the members of Sc_Set1C and nine members of another yeast complex termed CPF for cleavage and polyadenylation factor.

Subsequently, two other groups independently confirmed that Swd2 is a bona fide member of the budding yeast CPF, although its association with Sc_Set1C was not reported (1, 24). Taken together, these data suggest that Swd2 protein is a subunit of two independent complexes, Sc_Set1C and Sc_CPF, and hyperlinks the histone methylation and polyadenylation machinery in the budding yeast. Although both complexes act at the site of active transcription, the significance of this hyperlink remains elusive.

Here we report the application of a comparative proteomic analysis of the proteomic environments of the orthologous Set1 methyltransferases in the budding and fission yeasts.

**Experimental Procedures**

**Epitope Tagging of Genes and Isolation of Protein Complexes—**
The strains used in the study were generated using standard techniques and were isogenic to DB325-P41 (his+, ura4-D18, leu1-32).

Cells were grown in rich medium supplied when necessary with 150 mg/liter G418. Proteins of interest were tagged by in-frame fusion of the 3′-end of their open reading frames with a PCR-generated targeting cassette encoding for the TAP-tag and a selectable marker flanked by long (usually 80 nt) tracts of homology (9, 10, 25). Gene deletions and TAP-cassette integrations were performed by PCR-mediated gene targeting (25). Correct cassette integrations were confirmed by PCR and Western blot analysis.

The procedure for the purification of protein complexes in *S. pombe* was essentially the same as employed previously in *S. cerevisiae*. The breaking and extraction of yeast cells was performed as described by Logie and Peterson (26). TAP purification was performed according to Rigaut *et al.* (9), with the following modifications: 10 ml supernatant collected after 43,000 rpm centrifugation were allowed to bind to 200 μl IgG Sepharose (Amersham Pharmacia Biotech, Piscataway, NJ), equilibrated in buffer E (27) for 2 h at 4 °C using a disposable chromatography column (Bio-Rad, Hercules, CA). Two to three column volumes (the equivalent of 4–6 liters of yeast culture at optical density at 600 nm between 2 and 3) were used per purification. The IgG Sepharose column was washed with 35 ml of buffer E without proteinase inhibitors, followed by 10 ml of the tobacco etch virus (TEV) cleavage buffer. Cleavage with TEV was performed using 10 μl (100 U) rTEV (Life Technologies, Inc., Grand Island, NY) in 1 ml cleavage buffer for 2 h at 16 °C. Calmodulin Sepharose (Stratagene, La Jolla, CA) purification was performed as described (11). Purified proteins were concentrated according to Wessel and Fluge (27).

**Analysis H3-K4 Methylation in Sp_Swd2.1 and Sp_Swd2.2 Deletion Mutants—**Crude cell extracts from exponentially growing cells were prepared by glass bead lysis. Proteins were separated on 15% SDS polyacrylamide gel and blotted onto nitrocellulose membrane following the manufacturer’s instructions. For detecting lysine 4 (K4) methylated histone H3, an antibody recognizing di- and trimethylated forms of K4 was used (Abcam, Cambridge, UK). Secondary antibody was anti-rabbit IgG-HRP conjugate (Amersham Biosciences, Piscataway, NJ). The signals were visualized using the enhanced chemiluminescence system (Amersham Biosciences).

**Identification of Proteins by Mass Spectrometry—**Proteins were separated by electrophoresis using gradient (6–18%) one-dimensional polyacrylamide gels and visualized by staining with Coomassie. Protein bands were excised and in-gel digested with trypsin (Roche Diagnostics, Indianapolis, IN) as described previously (28). Proteins were identified by a combination of matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) peptide mapping and nanoelectrospray tandem mass spectrometry (nanoES MS/MS) sequencing as described (29). Briefly, 1-μl aliquots were withdrawn from the in-gel digests and analyzed on a REFLEX IV mass spectrometer (Bruker Daltonics, Billerica, MA) on AnchorChip™ targets (Bruker Daltonics) as described (30). If no conclusive identification was achieved, gel pieces were extracted with 5% formic acid and acetonitrile. Unseparated mixtures of recovered tryptic peptides were sequenced by nanoES MS/MS on a QSTAR Pulsar quadrupole time-of-flight mass spectrometer (MDS Sciex, Concord, Canada). Database searching with MALDI time-of-flight peptide mass maps and with uninterpreted tandem mass spectra was performed against a database of *S. pombe* proteins using Mascot software (Matrix Science Ltd., London, UK) installed on a local server. Hits with the MOWSE score exceeding 51 (the threshold score suggested by Mascot) were considered significant, but were accepted only upon manual inspection. Borderline hits were additionally verified by nanoES MS/MS.

The list of proteins identified in each pulldown experiment with relevant identification details is presented in the supplemental material.
RESULTS AND DISCUSSION

Protein Tagging, Affinity Purification, and Mass Spectrometric Analysis of Complexes in S. pombe—In total, in the course of this and other projects performed in our laboratories, we attempted to tag 23 genes from the fission yeast S. pombe. The molecular mass of bait proteins was in a broad range from 12 to 129 kDa and pI from 4 to 11. Their codon adaptation index, which represents a relative level of protein expression, was from 0.177 to 0.324, hence suggesting that the selected proteins are lowly abundant. Two out of the 23 genes encoding for tagged proteins are essential. The TAP tag was successfully fused to the C terminus of 22 proteins, and 20 tagged proteins (86%) were subsequently detected by Western blot. Two bait proteins from these 20 did not pull down any detectable interaction partners, although baits were visualized as intense Coomassie stainable bands. Each of the other 18 baits pulled down 2–12 interaction partners. Here we only present the identification of proteins interacting with 12 out of 18 baits that encompass the genes related to S. pombe Set1 and CPF complexes (see supplemental material). Taken together, data from this and other projects suggest that the success rate of protein tagging and isolation of complexes in S. pombe and in S. cerevisiae (4) is similar.

In the budding yeast, TAP of protein complexes was variably accompanied by a co-isolation of a common set of highly abundant background proteins, including housekeeping proteins, metabolic enzymes, and ribosomal proteins (4). We found that persistent background proteins in S. pombe were different from S. cerevisiae, although, once again, highly abundant proteins, including housekeeping proteins and components of the protein synthesis machinery, were mostly observed (Table I).

Sequential rounds of epitope tagging and immunoaffinity isolation were applied to characterize protein complexes from both S. cerevisiae (22) and S. pombe (13). Basically, the bait protein was TAP tagged, integration of the tag and expression of the tagged protein was checked by PCR and Western blot, and the complex was purified from the whole-cell lysate by the two-step affinity chromatography. Then, proteins were separated by gradient SDS-PAGE and visualized by Coomassie staining. All bands detected in the gel (without using mock patterns as controls) were excised and identified by mass spectrometry (Fig. 1). Proteins repeatedly detected in different purifications are listed in Table I and were disregarded as background. The identified proteins that were considered as plausible members of the protein complex were in turn tagged and the purification was repeated. Again, all detectable bands were analyzed by mass spectrometry. No identification has been based on the similarity of gel patterns.

The sequential approach was rather laborious: altogether,

### TABLE I

**Background proteins in the TAP method in S. pombe**

| Gene product | Name | CAI<sup>a</sup> | Molecular mass | pI | Localization |
|--------------|------|----------------|----------------|----|--------------|
| Cut6 | Probable acetyl-CoA carboxylase | 0.373 | 257 | 6.2 | Cytoplasm |
| SPAPB1E7.07 | Putative glutamate synthase | 0.516 | 233 | 6.1 | Cytoplasm |
| Fas1 | Fatty acid synthase, subunit beta | 0.427 | 230 | 6.1 | Cytoplasm |
| Fas2 | Fatty acid synthase, subunit alpha | 0.395 | 202 | 6.0 | Cytoplasm |
| Tit3 | Translation initiation factor eIF3, p110 subunit | 0.378 | 107 | 9.1 | Cytoplasm |
| Pf1 | 6-phosphofructokinase beta subunit | 0.71 | 103 | 6.0 | Cytoplasm |
| Sec21 | Coatamer gamma subunit | 0.297 | 101 | 5.3 | Cytoplasm |
| SPBC16H5.12c | Protein of unknown function | 0.278 | 77 | 5.8 | ? |
| Eif3b | Translation initiation factor eIF-3 subunit | 0.386 | 84 | 5.1 | Cytoplasm |
| Hsp70 | Heat shock protein 70kDa family | 0.789 | 70 | 5.1 | |
| Ded1 | Probable ATP-dependent RNA helicase | 0.523 | 70 | 8.8 | Cytoplasm |
| Sec1 | Heat shock 70 kDa protein, mitochondrial | 0.601 | 73 | 7.0 | Mitochondrion |
| Sks2 | Heat shock protein 70kDa family | 0.802 | 67 | 5.9 | Cytoplasm |
| Eif1-a | Translation elongation factor EF-1 alpha | 0.879 | 50 | 9.3 | Cytoplasm |
| SPBC14C8.02 | Mitochondrial import inner membrane translocase | 0.257 | 49 | 9.5 | Mitochondrion |
| SPAC4H3.01 | Probable DNA-J-like protein | 0.237 | 45 | 6.3 | Cytoplasm |
| Spj1 | dnaJ protein homolog | 0.319 | 42 | 8.5 | Cytoplasm |
| Act1 | Actin | 0.719 | 42 | 5.3 | Cytoplasm |
| Tdh1 | Glyceraldehyde-3-phosphate dehydrogenase | 0.849 | 36 | 6.5 | Cytoplasm |
| Hob3 | RVS161 protein homolog | 0.301 | 30 | 6.7 | Cytoskeleton |
| Gpm1 | Phosphoglycerate mutase | 0.849 | 24 | 7.2 | |
| Other ribosomal proteins<sup>b</sup> | | | | | |
| Mean | | 0.60 ± 0.33 | |

<sup>a</sup> CAI, codon adaptation index.<br>
<sup>b</sup> Background ribosomal proteins may vary. The proteins Rp13–1/2, Rp12, Rp15–1/2, Rp113, Rps3, Rps1–1/2, Rps6, Rps7A/C, Rps9A/B, Rps11A, Rps13, Rps17–1/2, Rps22A, Rps18, Rps8, Rps25A/B, Rp115–2, Rp115, Rp117, Rp120–1/2, Rp121–1/2, Rp111A/B, Rp124, Rp128A/B, Rp125A/B, and Rp136A/B were most typical contaminants.
for three protein complexes characterized in *S. pombe* (Sp_Set1C, Sp_Lid2C, and Sp_CPF), 12 baits were TAP tagged and mass spectrometric identification of 681 bands of co-isolated proteins was performed (Table II), so that each complex was independently purified several times using different subunits as baits (see supplemental material).

Although shotgun methods (17) provide higher throughput compared with band-per-band identification of gel-separated proteins, we see two major arguments in favor of using gels. First, visual inspection of Coomassie stained gel-separated proteins accompanied by densitometry scanning provides semiquantitative estimations of the stoichiometry between the

![Fig. 1. TAP purification of tagged fission yeast homologues of Sc_Swd2. A, Purification of TAP-Sp_Swd2.1; B, TAP-Sp_Swd2.2. All Coomassie detectable bands (52 bands in A (indicated by dots) and 58 bands in B) were identified by mass spectrometry. Sp_Swd2.1 pulled down six members of Sp_Set1 complex, whereas 10 members of CPF complex were identified in association with Sp_Swd2.2 (designated at the panels). A complete list of identified proteins is provided in the supplemental material.](image)

**Table II**

| S. cerevisiae | S. pombe |
|--------------|----------|
| **Protein**  | **ORF**  | **Mr** | **Protein** | **ORF** | **Mr** | **Identity/similarity** |
| Set1         | YHR119w  | 124    | Set1        | SPCC306.04c | 105 | 29/44 |
| Bre2         | YLR015w  | 58     | Ash2        | SPBC13g1.08c | 74  | 27/39 |
| Spp1         | YPL138c  | 41     | Spp1        | SPCC59.05c  | 49  | 40/58 |
| Swd1         | YAR003w  | 49     | Swd1        | SPAC23h3.05c | 45  | 37/55 |
| Swd2         | YKL018w  | 37     | Swd2.1      | SPBC18h10.06c | 40  | 34/53 |
| Swd3         | YBR175w  | 35     | Swd3        | SPBC35.04c  | 43  | 30/49 |
| Shg1         | YBR258c  | 16     | Shg1        | SPAC17g6.09 | 15  | –    |
| Sdc1         | YDR469w  | 19     | Sdc1        | SPCC18.11c  | 12  | 46/75 |
| Cft1         | YDR301w  | 153    | Cft1        | SPBC170.08c | 160 | 25/45 |
| Cft2         | YLR115w  | 96     | Cft2        | SPBC170.15c | 89  | 25/44 |
| Ysh1         | YLR227c  | 88     | Ysh1        | SPAC17g6.16c | 88  | 49/67 |
| Pta1         | YAL043c  | 88     | Pta1        | SPAC4H3.15c | 50  | 23/40 |
| Swd2         | YKL018w  | 37     | Swd2.2      | SPAC82.04  | 38  | 32/55 |
| Pfs2         | YNL317w  | 53     | Pfs2        | SPAC12G12.14c | 58  | 43/61 |
| Glc7         | YER133w  | 36     | Dis2        | SPBC77.02c  | 37  | 89/95 |
| Yhr175w      | 60     | 153    | Ysh1        | SPAC17g6.16c | 88  | 26/48 |
| YDR195w      | 50     | 50     | Ysh1        | SPAC17g6.16c | 88  | –    |
| YOR173c      | 21     | 21     | Ysh1        | SPAC17g6.16c | 88  | –    |
| Lid2         | SPBP19A11.06 | 172  | –    | SPAC74.02c  | 77  | –    |
| Snt2         | SPAC3h1.12c | 129  | –    | –            | –    | –    |
| Jmj3         | SPBC83.07 | 85    | –    | –            | –    | –    |
| Bre2         | YLR015w  | 58     | Ash2        | SPBC13g1.08c | 74  | 27/39 |
| Sdc1         | YDR469w  | 19     | Sdc1        | SPCC18.11c  | 12  | 46/75 |

* Members of Sc_Set1C and Sc_CPF were identified previously (22).

b ORF, open reading frame.

Orthologous Proteomic Environments in Yeasts
subunits and indicates those proteins whose interaction with the core complex may be only transient. Second, gel patterns and spectra of individual bands could be compared directly, thus confidently stipulating the presence and, equally important, the absence of a particular protein in the preparation.

To further characterize the organization of the Set1-anchored proteomic environment in S. pombe, we used the genome of S. pombe to encode for the orthologous protein Sp_Set1, which shares 29% of the full-length sequence identity with its S. cerevisiae homologue. We therefore asked if the fission yeast CPF complex includes the other Swd2 homologue, Sp_Swd2.2, and if it might represent a yet undetected hyperlink between Sp_Set1C and Sp_CPF (Fig. 2). The genome of S. pombe encodes for two relatively distant homologues of Sc_Swd2, namely SPBC18H10.66c (now termed Sp_Swd2.1) and SPAC824.04 (now termed Sp_Swd2.2), sharing 34 and 32% of sequence identity with Sc_Swd2, respectively and sharing 30% of sequence identity between themselves. By tagging Sp_Set1C members, we established that Sp_Swd2.1 is a member of Sp_Set1C (Fig. 2). However, tagging Sp_Swd2.2 (Fig. 1A) or subunits of Sp_CPF (see supplemental materials) did not provide any evidence that it interacts with the Sp_CPF complex.

We therefore asked if the fission yeast CPF complex includes the other Swd2 homologue, Sp_Swd2.2, and if it might represent a yet undetected hyperlink between Sp_Set1C and Sp_CPF. To elucidate the proteomic environment of Sp_Swd2.2, the protein was tagged and used as a bait in the immunopunaffinity purification. It pulled down the Sp_CPF complex, and, importantly, no members of Sp_Set1C complex were detected (Fig. 3). To confirm that Sp_Swd2.2 is a core subunit of the Sp_CPF complex, we further tagged its conserved member Sp_Ysh1. Sp_Ysh1 was chosen because it is a key component of Sc_CPF, and its sequence is remarkably conserved between S. cerevisiae and S. pombe with no other clear homologues shared between these two genomes. Using Sp_Ysh1-TAP as a bait, we pulled down the same subunits of CPF, including Sp_Swd2.2, but no Sp_Swd2.1 was observed in a detectable amount. Taken together, these data suggested Sp_Swd2.2, but not Sp_Swd2.1, is a genuine member of Sp_CPF (Fig. 3).

Thus we demonstrated that although the composition of Set1C is conserved between S. cerevisiae and S. pombe, the hyperlinks of the Set1C complex are different in two ways. First, the fission yeast Sp_Set1C is hyperlinked to another...
complex, Sp_Lid2C, which is not present in the budding yeast (see below). Second, Set1C and CPF are no longer hyper-linked via Swd2-homologous proteins in S. pombe.

The Function of Swd2 Paralogues Is Completely Diverged in S. pombe—We further investigated if Sp_Swd2.1 and Sp_Swd2.2 can substitute each other in Sp_Set1C and Sp_CPF complexes. To this end, we constructed a strain in which the Sp_Swd2.2 gene was deleted and Sp_Swd2.1 was tagged. The mutant strain grew slower compared with the wild-type strain and other strains with TAP-tagged proteins. Although much less protein material was purified, in a pull down experiment with Sp_Swd2.1 we were able to identify all subunits of Sp_Set1C except Sp_Sdc1, which produces only one peptide upon its digestion with trypsin and was undetectable in a heavy mixture with low molecular mass background proteins. At the same time, no subunits of Sp_CPF were detectable (Fig. 4), suggesting that Sp_Swd2.1 cannot substitute for Sp_Swd2.2 in Sp_CPF.

As we demonstrated previously (13), Sp_Swd2.1 is essential for methylation of the K4 residue of histone H3. To test if Sp_Swd2.2 could also be involved in the H3K4 methylation, we created ΔSp_Swd2.1 and ΔSp_Swd2.2 deletion strains. Protein extracts from the deletion strains and from wild type were probed by immunoblotting with an antibody recognizing trimethylated K4 in histone H3. Extracts from wild type and ΔSp_Swd2.1 were used as positive and negative controls.

We therefore concluded that in the fission yeast, duplicated swd2 genes are functionally specialized (33, 34), with...
A Protein Complex in *S. cerevisiae* that Is Orthologous to *Sp_Lid2C*—As was demonstrated above, *Sp_Ash2* and *Sp_Sdc1* hyperlink *Sp_Set1C* and *Sp_Lid2C* complexes (Fig. 2). However, no such link to a complex similar to *Sp_Lid2C* was observed in *S. cerevisiae*, although both *S. pombe* proteins share significant sequence identity with their closest homologues in *S. cerevisiae* (Table II). We therefore attempted to isolate a protein complex orthologous to *Sp_Lid2C* from the budding yeast and to determine if it is hyperlinked to *Sc_Set1C*.

The genome of *S. cerevisiae* encodes for proteins YJR119c and Snt2 (YGL131c), which share 23 and 22% of full-length sequence identity with *Sp_Lid2* and *Sp_Snt2*, respectively, and display a very similar composition of functional domains. We tagged *Sc_Snt2* and YJR119C and isolated their interaction partners, but no proteins homologous to members of *Sp_Lid2C* were identified (data not shown).

Taken together, the data suggested that no protein complex orthologous to *Sp_Lid2C* exists in *S. cerevisiae*, despite the presence of a few reasonable sequence homologues in its genome. We speculate that *Sp_Lid2C* is possibly involved in H3K9 methylation, which does not occur in *S. cerevisiae* but is found in *S. pombe* and in humans (35, 36).

Although genes encoding for subunits of *Sc_Set1C* complex (other than *Swd2*) are nonessential, and certain members of *Sc_Set1C* and *Sp_Set1C* display only marginal sequence similarity (Table II), the overall composition of these complexes is well conserved. However, orthologous *Sp_Set1C* and *Sc_Set1C* complexes function in a differently “wired” proteomic network that comprises conserved (*Sc_CPF* and *Sp_CPF*) and nonconserved (*Sp_Lid2C* and *Sc_Snt2C*) protein assemblies. Conserved Composition and Variable Hyperlinks of Orthologous Complexes—Although *Set1* complexes in *S. cerevisiae* and *S. pombe* are highly conserved, overall their proteomic environment differs substantially.

Our data underscore the value and importance of maximal possible characterizations of the compositions of protein complexes, especially when considering a phylogenetic perspective. Using two entry points, *Sc_Swd1* and *Sc_Swd3* proteins, Gavin *et al.* (1) identified a complex (termed complex #108) with a very similar composition to *Sc_Set1C*. The complex #108 missed two subunits (*Swd2* and *Shg1*) and, consequently, a hyperlink to *Sc_CPF* complex via *Sc_Swd2*. At the same time, complex #108 comprised three other proteins (*yeast.cellzome.com*), whose relation to *Sc_Set1C* was not independently confirmed. Gavin *et al.* also tagged seven out of 20 known subunits of *Sc_CPF* and detected *Sc_Swd2* in all affinity purifications (1). However, *Sc_Swd2* itself was not tagged, and its relation to *Sc_Set1C* was not established. Missed interactions or artificially merged individual protein complexes hampered further comparison of *Set1* proteomic environments and reasonable projection of their organization and function to mammalians.

The human genome encodes for two *Set1*-related genes: KIAA0339 and KIAA1076, sharing 35 and 37% identity to *Sc_Set1* and 55 and 45% identity to *Sp_Set1*. KIAA0339 is engaged in a partially characterized complex comprising at least *hAsh2* and WDR5 (a human homologue of *Sc_Swd3*) (38). This putative *Hs_Set1C* is also involved in H3K4 methylation (38). The entry point to *Hs_Set1C*, protein HCF-1, has no apparent homology to any of the core members of *Sp_Set1C* or *Sc_Set1C* but is associated with human Sin3 histone deacetylase (HDAC). We therefore speculate that the partial purification of *Hs_Set1C* was achieved via a hyperlink protein.

Taken together, our data and other published evidence strongly suggest that although orthologous protein complexes may be remarkably conserved, their proteomic environment and hyperlinks to other complexes are not. Furthermore, we propose that the conservation of the core and variability of links represents a common phenomenon in the molecular organization of eukaryotic proteomes.

CONCLUSION AND PERSPECTIVES

Understanding the phylogenetic variability of proteomic hyperlinks is important for satisfactory extrapolations of global protein-protein interaction patterns, determined in model low-eukaryotic organisms, to higher eukaryotes including humans. To this end, boundaries of individual complexes and their hyperlinks should be accurately determined. The quality of data, the reliability and reproducibility of the discovered protein-protein interactions, as well as the scale and the character of the analysis are all factors important to the merits of proteomics extrapolations.

The comparative analysis of proteomic environments in a multiorganismal perspective offers an intriguing opportunity to extend and complement our understanding of how the evolution of genomes guides the evolution of protein machines. Comparative studies may reach far beyond simple cataloguing of observed differences. Rather, together with advanced bioinformatic approaches, correlations of concerted alterations in sequences of orthologous subunits could highlight functional specializations.

The multiorganismal approach in functional proteomics will likely require biochemical isolation of complexes and identification of cognate proteins beyond the boundaries of known genomes. Although lacking exact protein sequences in a database hampers the identification of proteins, a substantial coverage of yet unknown proteomes might be achieved by sequence-similarity searches (reviewed in Ref. 37). The bottleneck (and the likely focus of further efforts) is in the development of a generic approach for isolating protein complexes from cells and tissues of vertebrate organisms, which might be overcome by advanced genetic engineering methods in the future (39).
REFERENCES
1. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147
2. Ho, Y., Gruler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutillier, K., Yang, L., Wolting, C., Donaldson, I., Schandorf, S., Shevchenko, A., et al. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415, 180–183
3. Seraphin, B. (2002) Identification of transiently interacting proteins and of stable protein complexes. Adv. Protein Chem. 61, 99–117
4. Shevchenko, A., Schaff, D., Roguev, A., Pinjappel, W. M. P., Stewart, A. F., and Shevchenko, A. (2002) Deciphering protein complexes and protein interaction networks by tandem affinity purification and mass spectrometry: Analytical perspective. Mol. Cell Proteomics 1, 204–212
5. Ito, T., Ota, K., Kubota, H., Yamaguchi, Y., Chiba, T., Sakuraba, K., and Ito, H. (2001) The tandem affinity purification (TAP) method: A general procedure of protein complex purification. Methods 24, 219–224
6. Pinjappel, W. W., Schaff, D., Roguev, A., Shevchenko, A., Tekotte, H., Wilm, M., Rigaut, G., Seraphin, B., Aasland, R., and Stewart, A. F. (2001) The S. cerevisiae SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. Proc. R Soc. Lond. B Biol. Sci. 268, 2048–2054
7. Tasto, J. J., Carnahan, R. H., McDonald, W. H., and Gould, K. L. (2001) Vectors and gene targeting modules for tandem affinity purification in Schizosaccharomyces pombe. Yeast 18, 657–662
8. Roguev, A., Schaff, D., Shevchenko, A., Aasland, R., and Stewart, A. F. (2003) High conservation of the Set1/Rad6 axis of histone 3 lysine 4 methyltransferase in budding and fission yeasts. J. Biol. Chem. 278, 8487–8493
9. Sipiccki, M. (2000) What does fission yeast sit on the tree of life? Genom. Biol. 1, RESEARCH0111
10. Eisen, J. A. (2002) Brouhaha over the other yeast. Nature 415, 845–848
11. Wood, V., Guviman, R., Rajandream, M. A., Lyné, M., Lyné, R., Stewart, A., Sguogos, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., et al. (2002) The genome sequence of Schizosaccharomyces pombe. Nature 415, 871–880
12. Link, A. J., Eng, J., Schietz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. B., and Yates, J. R. 3rd. (1999) Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17, 676–682
13. Ohi, M. D., Link, A. J., Ren, L., Jennings, J. L., McDonald, W. H., and Gould, K. L. (2002) Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cre1p, novel pre-mRNA splicing factors, and snRNAs. Mol. Cell Biol. 22, 2011–2024
14. Yoon, H. J., Feoktistova, A., Wolfe, B. A., Jennings, J. L., Link, A. J., and Gould, K. L. (2002) Proteomics analysis identifies new components of the fission and budding yeast anaphase-promoting complexes.Curr. Biol. 12, 2048–2054
15. Seol, J. H., Shevchenko, A., Shevchenko, A., and Deshaies, R. J. (2001) Skp1 forms multiple protein complexes, including RAVE, a regulator of V-ATPase assembly. Nat. Cell Biol. 3, 384–391
16. Nislow, C., Ray, E., and Pilius, L. (1997) SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. Mol. Biol. Cell 8, 2421–2436
17. Roguev, A., Schaff, D., Shevchenko, A., Pinjappel, W. M. W., Wilm, M., Aasland, R., and Stewart, A. F. (2001) The S. cerevisiae Set1 complex contains an Ash2 homolog and methylates histone 3 lysine 4. EMBO J. 20, 7137–7148
18. Krogan, N. J., Dover, J., Khromova, S., Greenblatt, J. F., Schneider, J., Johnston, M., and Shalitafard, A. (2002) COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J. Biol. Chem. 277, 10753–10755
19. Dichti, B., Blank, D., Ohnacker, M., Friedlien, A., Roeder, D., Langen, H., and Keller, W. (2002) A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. Mol. Cell 10, 1139–1150
20. Logie, C., and Peterson, C. L. (1999) Purification and biochemical properties of yeast SHT/SNF complex. Methods Enzymol. 304, 726–741
21. Vessell, D., and Flugge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal. Biochem. 138, 141–143
22. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. 68, 850–859
23. Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M. (1996) Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. U. S. A. 93, 14440–14445
24. Havlis, J., Thomas, H., Sebela, M., and Shevchenko, A. (2003) Fast response proteomics by accelerated in-gel digestion of proteins. Anal. Chem. 75, 1300–1306
25. Aasland, R., Gibson, T. J., and Stewart, A. F. (1995) The PHD finger: Implications for chromatin-mediated transcriptional regulation. Trends Biochem. Sci. 20, 58–59
26. Gildea, J. J., Lopez, R., and Shearn, A. (2000) A screen for new trithorax group genes identified little imaginal discs, the Drosophila melanogaster homologue of human retinoblastoma binding protein 2. Genetics 156, 465–483
27. Hughes, A. L. (1994) The evolution of functionally novel proteins after gene duplication. Proc. R Soc. Lond. B Biol. Sci. 256, 119–124
28. Lynch, M., and Conery, J. S. (2000) The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155
29. Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F., and Allis, C. D. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev. 15, 3288–3295
30. Noma, K., Allis, C. D., and Grewal, S. I. (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science 293, 1150–1155
31. Liska, A., and Shevchenko, A. (2002) Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications. Proteomics 3, 19–28
32. Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N., and Herr, W. (2003) Human Sin3a deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. Genes Dev. 17, 896–911
33. Testa, G., Zhang, Y., Vintersten, K., Benes, V., Pinjappel, W. W., Chambers, I., Smith, A. J., Smith, A. G., and Steward, A. F. (2003) Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. Nat. Biotechnol. 21, 443–447