Large-Scale Proteomic Analysis of the Human Spliceosome

Juri Rappsilber,[1] Ursula Ryder,[2] Angus I. Lamond,[2] and Matthias Mann[1,3]

[1]*Protein Interaction Laboratory in the Center of Experimental Bioinformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230 Odense M, Denmark;* [2]*Department of Biochemistry, University of Dundee, Dundee DD1 4HN, Scotland, United Kingdom*

In a previous proteomic study of the human spliceosome, we identified 42 spliceosome-associated factors, including 19 novel ones. Using enhanced mass spectrometric tools and improved databases, we now report identification of 311 proteins that copurify with splicing complexes assembled on two separate pre-mRNAs. All known essential human splicing factors were found, and 96 novel proteins were identified, of which 55 contain domains directly linking them to functions in splicing/RNA processing. We also detected 20 proteins related to transcription, which indicates a direct connection between this process and splicing. This investigation provides the most detailed inventory of human spliceosome-associated factors to date, and the data indicate a number of interesting links coordinating splicing with other steps in the gene expression pathway.

Biogenesis of proteins in eukaryotes is a multistep process that involves the concerted action of several complex machineries. Multiprotein complexes containing RNA polymerase II are involved in transcribing genes into pre-messenger RNA. Most human genes contain introns that are removed by splicing, a process orchestrated and catalyzed by the large multiprotein/RNA complex termed the spliceosome. Polyadenylation of the mRNA is also catalyzed by a complex processing machinery before mRNAs are exported to the cytosol, where translation by ribosomes takes place. Although much is known about the individual processes in protein biogenesis, how the separate steps are integrated is much less clear.

The spliceosome is comprised of five small nuclear RNAs (snRNAs)—U1, U2, U4, U5, and U6 snRNA—as well as many protein factors (Staley and Guthrie 1998). Some of these proteins are tightly associated with the snRNAs, forming small nuclear ribonucleoproteins (snRNPs) that are thought to assemble in a stepwise manner onto the pre-mRNA to form the spliceosome. Work over the last decade has elucidated the temporal sequence of recognition of the splice sites by the respective snRNPs and protein factors (Hastings and Krainer 2001). Interestingly, the view of stepwise assembly of the spliceosome has recently been challenged in favor of a more concerted mechanism involving preformed spliceosomes (Stevens et al. 2002). Besides the snRNP subunits, a large number of non-snRNP proteins are known, which perform various functions during the splicing reaction. For example, multiple members of the DEAD-box helicase family are thought to control RNA base-pairing interactions at different stages of spliceosome assembly and catalysis, whereas members of the SR motif family are believed to be link factors promoting protein–protein interactions during spliceosome assembly. In all, ~100 different proteins have been linked to splicing through biochemical and/or genetic evidence (for review, see Will and Lührmann 1997). However, it remains unclear how complete this list might be.

In an alternative systematic approach to the traditional characterization of single splicing factors, the spliceosome can be purified and its components identified collectively using modern proteomic techniques. Initially, heterogeneous nuclear ribonucleoprotein (hnRNP) complexes assembled on mammalian pre-mRNA (Calvio et al. 1995) and subunits of the yeast spliceosome were purified and analyzed by mass spectrometric methods (Neubauer et al. 1997; Gottschalk et al. 1998). Subsequently, our groups performed the first large-scale analysis of a human multiprotein complex on in vitro-assembled spliceosomes (Neubauer et al. 1998) using 2D gel electrophoresis followed by nanoelectrospray (Wilm et al. 1996) mass spectrometric analysis. The relation of many of the newly discovered proteins to splicing was verified by fusing them to the green fluorescent protein, transiently expressing them in human epithelial (HeLa) cells, and showing that they colocalized in vivo with known splicing factors. Further biochemical studies have confirmed a role in splicing for all of the novel proteins analyzed so far in our laboratories, showing the specificity of the spliceosome purification method (Ajuh et al. 2000, 2001; Rappsilber et al. 2001; Lallena et al. 2002).

More recently, other mammalian protein complexes have also been studied by similar methods, combining protein affinity purification with mass spectrometry and database searches (Wigge et al. 1998; Zachariae et al. 1998; Rout et al. 2000; Gavin et al. 2002; Ho et al. 2002). Recently, our groups have reported the identification and analysis of 271 proteins in the human nucleolus, the largest study of an organelle to date (Andersen et al. 2002a; Fox et al. 2002).

Mass spectrometric methods and human sequence databases have continued to improve dramatically in recent years, allowing both increased sensitivity and higher throughput. It is now possible to analyze mixtures of hundreds or even thousands of peptides by liquid chromatography coupled to tan-

[3]Corresponding author.
E-MAIL mann@bmb.sdu.dk; FAX 45 6593 3929.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.473902. Article published online before print in July 2002.

dem mass spectrometry (LC MS/MS; Griffin and Aebersold 2001; Washburn et al. 2001). Improved software and databases containing most human genes—known or putative—are also now available, allowing automated data processing of the large volume of acquired mass spectra.

Building on these advances, we decided to revisit the large-scale analyses of human spliceosome complexes, using these enhanced, state-of-the-art techniques. In the present study, splicing complexes formed on two separate pre-mRNAs were purified, but the resulting proteins were not separated by gel electrophoresis; rather, they were analyzed by automated tandem mass spectrometry of crude peptide mixtures resulting from digest of the entire protein mixture. Using differential mass range pulsing on a quadrupole time-of-flight instrument, a total of 311 proteins were identified. In addition to all the factors reported in our previous spliceosome study and all other essential human splicing factors known, we discovered a further novel 96 proteins about which little or no previous biological information existed. Many of these proteins have a domain structure implicating them directly in splicing/RNA processing. Surprisingly, a number of proteins involved in transcription and cellular regulatory mechanisms copurified with the spliceosome, indicating some form of coupling of these processes to splicing.

RESULTS

Proteomic Analysis of the Human Spliceosome

Preparation of the Spliceosome

A mixture of spliceosomal complexes was assembled on biotinylated, radioactively labeled RNA (see Methods). In contrast to our previous investigation, two standard splicing substrates, adenovirus (AD1) and β-globin (AL4) transcripts, were used in separate experiments. After incubation, which led to the formation of active spliceosomes and assembly intermediates, samples were subjected to gel filtration and affinity selection of the biotinylated pre-mRNA on streptavidin beads.

To identify proteins binding to the beads directly, we performed gel filtration of the nuclear extract without labeled RNA and biotin affinity selection of the same fractions as above. The protein mixture was then applied to a short, one-dimensional sodium dodecyl sulfate (SDS)-poly acrylamide gel electrophoresis (PAGE) gel that allowed removal of SDS, washing, rebuffering, and efficient digestion according to protocols previously described (Shevchenko et al. 1996). The resulting complex peptide mixture was then loaded for chromatographic separation. Based on Coomassie blue staining, we estimate that each substrate sample contained ~6–10 μg of protein in total. This was in agreement with the subsequent mass spectrometric analyses, which indicated an abundance range of ~1–100 fmole per tryptic peptide on the column.

Mass Spectrometric Analysis of the Spliceosome

The peptides bound to the high-performance liquid chromatography (HPLC) column were eluted with an organic gradient into the ion source of a quadrupole Time Of Flight (PE-Sciex) instrument capable of high resolution and mass accuracy. A test run with 10% of the material indicated that the material was sufficient for three LC MS/MS experiments. Therefore, for each of the substrates, we performed three separate, identical chromatographic runs with a third of the material each. The pulsing ability of the QSTAR instrument (Chernushevich 2000) was used to selectively amplify a different region of the peptide mass range for each of the runs (Andersen et al. 2002b). Amplification ranged from a factor of 10 for the lower mass range to a factor of ~4 for the largest mass range. The average resolution and mass accuracy achieved in the six runs were 8000 and 27 ppm, respectively.

Figure 1 shows an overview of the analysis for one of the substrates and all three mass ranges. Figure 1, panels A1, B1, and C1 represent the summed ion currents of all peptides eluting at a specific time in the 100-min gradient. The Figure 1 inset shows the ion current of a particular peptide eluting around the marked position. Peptides typically eluted in 20-sec peaks (full width at half-maximum) from the chromatographic column. Figure 1, panels A2, B2, and C2 show the mass spectrum at the elution time marked in the first row of panels. Several peptides coelute, and up to four are automatically selected for sequencing in order of intensity. The acquisition software was directed to select only peptides in the amplified mass ranges for sequencing (m/z = 350–550 in Fig. 1A, m/z = 550–750 in Fig. 1B, and m/z = 750–1400 in Fig. 1C). The inset of the second row of panels shows the isotopic structure of a peptide ion peak, showing high resolution and unambiguous charge state determination. Figure 1, panels A3, B3, and C3 contain the fragment ion spectra of the peptide ion peak marked with asterisks in the middle panels. The fragments contain amino acid sequence information and were used to determine the identity of the peptides as described below.

Data Analysis and Verification

More than 7000 ion peaks were fragmented. After acquisition, fragment mass lists were generated under script control (`Analyst`, PE-Sciex), added for all six experiments, and submitted to automated database searches using the `Mascot` search engine (Perkins et al. 1999). The peptide sequences retrieved from the database were assembled into protein matches by `Mascot`. The resulting protein list contained more than 1000 entries with a summed peptide score of at least 30. This list was manually verified in the following way: All entries with less than three high-scoring peptides (peptide score >30) were manually inspected and verified. In this manual interpretation, the mass error, the presence of series of fragment ions, the expected prevalence of C-terminus containing (Y-type ions) in the high mass range, and features such as the particular cleavage pattern at proline were all taken into account. As a result of manual verification, a total of 311 proteins were determined with high confidence to be present in our spliceosomal preparation.

The presence of numerous protein isoforms is a practical and recurrent problem in proteomics (Rappsilber and Mann 2002). In determining the list of spliceosome proteins, we used peptides that distinguished between isoforms and/or splice variants when possible. Although information regarding novel splice variants is present in the data, interpretation was beyond the scope of this study.

To generate the complete list of peptides that can be assigned with high confidence to the 311 proteins, we filtered the 9791 peptides first retrieved from the database as follows: First, 4480 peptides that were not the top-ranked sequence for the fragmentation spectrum in question were discarded. A further 1434 peptide matches whose score was <20 were also discarded. Subtracting the peptide matches that did not match to the above list of 311 proteins and removing peptides that matched to several fragmentation spectra (e.g., because of different charge stages selected for sequencing) yielded a final group of 2025 peptides (listed at http://www.pil.sdu.dk).

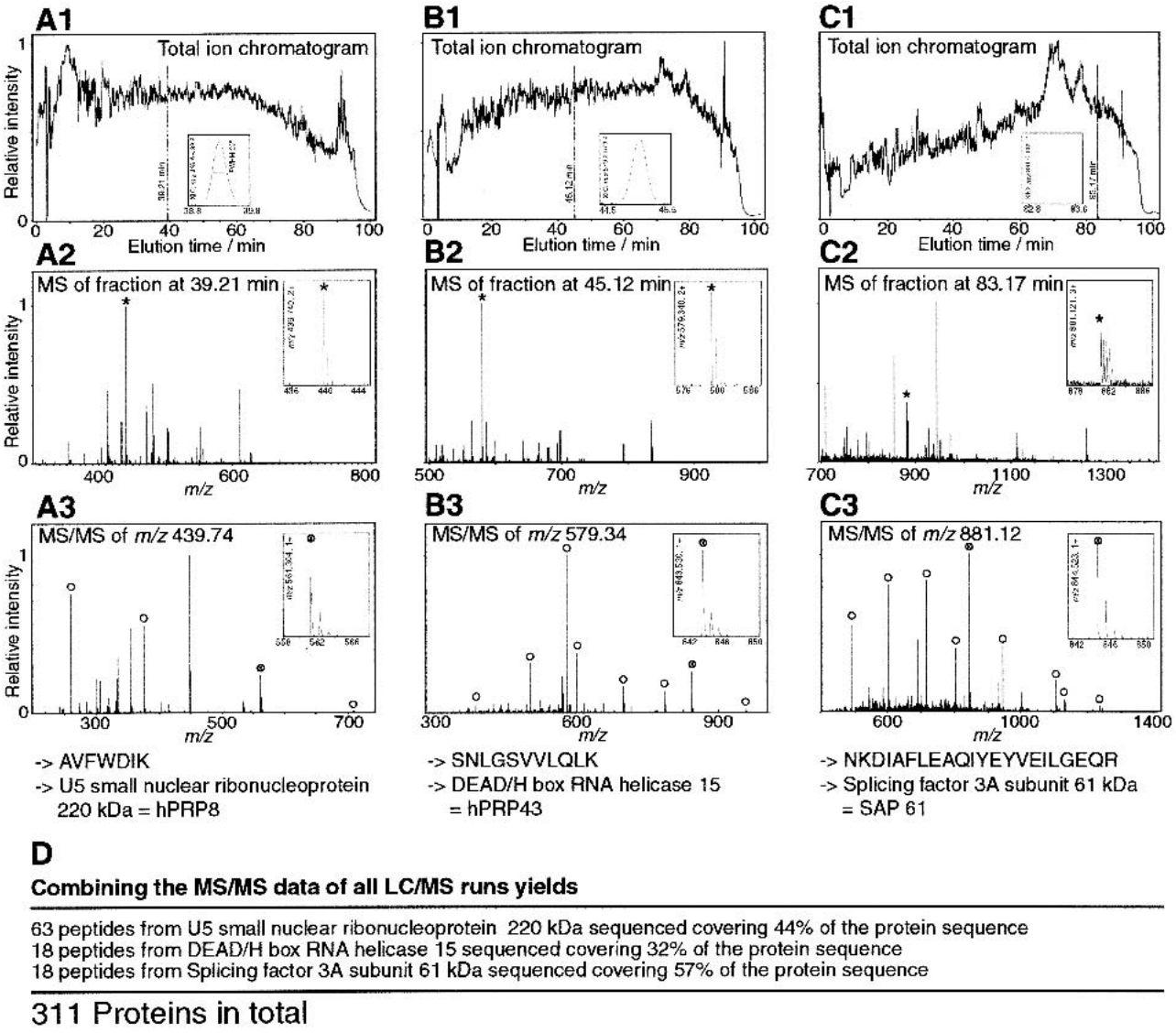

Figure 1 On-line LC MS/MS mass spectrometric analysis of the spliceosome. (*A1*) Sum of the ion intensity measured by mass spectrometry at any point in the chromatogram for the mass range m/z = 350–550 (total ion chromatogram[TIC]), (*B1*) m/z = 550–750, and (*C1*) m/z = 750–1400. (*Insets*) Current for an ion of specific mass (m/z window 0.2) eluting around the marked time in the chromatogram. (*A2*,*B2*,*C2*) Mass spectra obtained at the marked time in the corresponding upper panels. Several peptides coelute. (*Insets*) Zoom of the peak marked with an asterisk. (*A3*,*B3*,*C3*) Fragmentation spectra of the peak marked with an asterisk in the corresponding mass spectra above. The insets show that there was high resolution and the ability to determine the charge state by the differences between isotopic peaks. Open circles mark the peaks corresponding to predicted fragments for the peptide sequence retrieved by the database search. The insets show that there was high-quality data for the fragments, which aided the database search and validation. (*D*) The proteins identified by the peptides shown in *A*–*C* are also independently identified by a large number of other peptide fragmentation spectra, leading to substantial sequence coverage of 32%–57% for these three proteins.

The peptide matches per protein ranged from 63 for the U5 snRNP-specific 220-kD protein to a few proteins that were confidently identified on the basis of only one peptide. The vast majority of proteins, however, were identified on the basis of three or more peptides.

Relative Abundance of the Different Classes of Proteins

For the interpretation of the results in a large-scale study involving hundreds of proteins and very high sensitivity, it is important to obtain some measure of quantification. The mass spectrometric signal for any given peptide is determined by many factors, most importantly its ionizability in electrospray. Therefore, direct quantification of proteins in an LC MS/MS experiment is difficult. However, there is a general correlation between the number of peptides sequenced per protein and the amount of protein present in the mixture. Because larger proteins can give rise to more peptides, we defined a protein abundance index (PAI), which represents the number of peptides identified divided by the number of theoretically observable tryptic peptides. Figure 2 plots the index for the seven different protein classes, which are explained below.

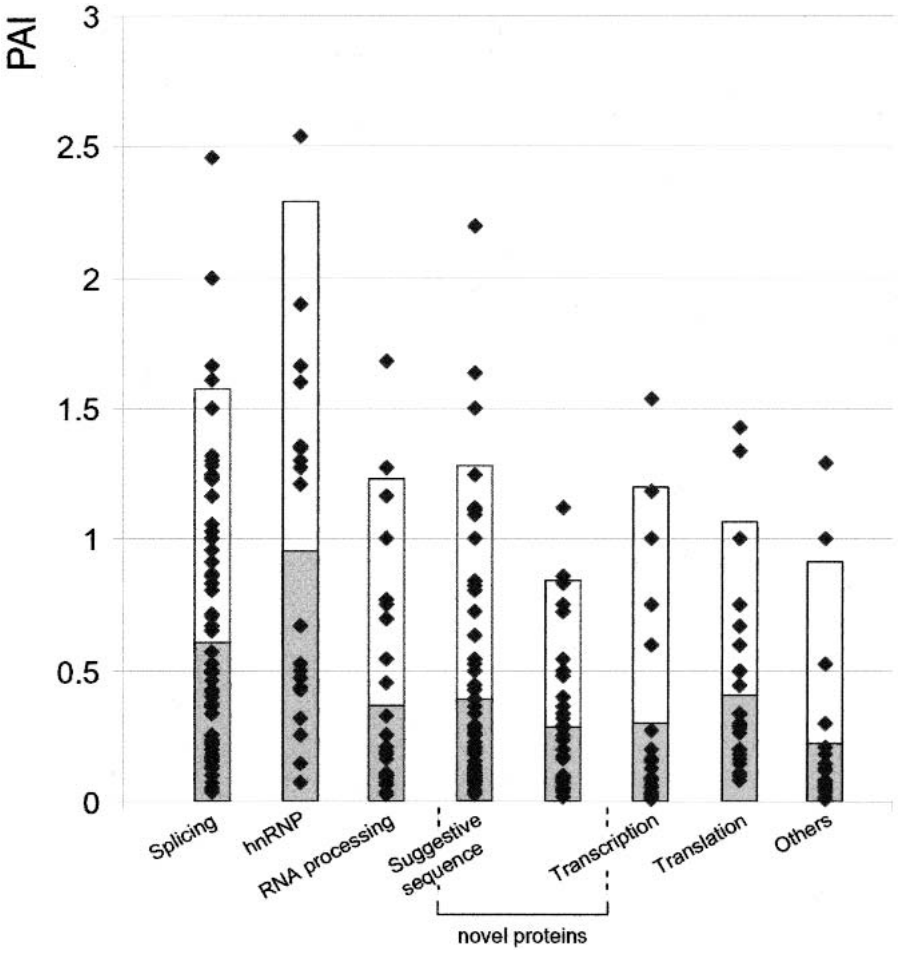

Figure 2 Protein abundance index (PAI) for the identified proteins. Plot of PAI index, which is defined as the number of identified peptides divided by the number of calculated, observable peptides, plotted for the identified proteins in seven different classes. The individual spots represent the PAI for each protein in the category. The bar shaded in gray extents to the average value of the PAI for each category. The white bar encompasses 95% of the proteins in each category.

Categorization of Spliceosome-Associated Factors

We first subtracted 20 proteins (see Methods) from the list of identified proteins, for the following reasons: 16 proteins, mainly keratins, were also identified in the background fraction (beads only). One protein appears in a separate database entry as the C-terminal part of one of these 16 background proteins and was also removed. Finally, one protein that is also abundant in human keratinocytes, like the keratins, and two proteins with a very similar domain structure were also discarded. The remaining 292 proteins identified in this large-scale analysis of the human spliceosome were grouped into eight functional categories (Fig. 3). The low number of cytoskeletal, nuclear matrix, and heat-shock proteins usually highly abundant in less specific protein purifications indicates that our spliceosomal preparation is highly specific.

Known Splicing Factors, hnRNPs, and Other RNA–Processing Proteins

More than 40 percent of the identified proteins have a known function related either directly to splicing or more generally to RNA processing (Fig. 3; Table 1). Encouragingly, all the spliceosomal proteins identified in our previous proteomic characterization were also identified here. All of the core snRNP proteins (Sm proteins) that are present in U1, U2, U4, and U5 snRNPs were identified. These proteins are small and thus difficult to detect by standard 2D gel electrophoresis, having required separate one-dimensional gel analysis in our previous study (Neubauer et al. 1998). Six of the seven Lsm proteins that substitute for Sm proteins in the U6 snRNP (Seraphin 1995) were also detected. The complete protein complement of the U1, U2, and U6 snRNPs was identified, together with all but three proteins from the U5, U4/U6, and U4/U6 · U5 snRNPs (Will and Lührmann 2001, and references therein). The three proteins that were not observed, Lsm5 (9800 D), U5 snRNP 15 kD (16,775 D), and U-snRNP-associated cyclophilin (USA-CyP) (19,196 D), were small and were presumably not identified because they generate few detectable tryptic peptides, because of the low relative protein amount in the sample, and because of the complexity of the peptide mixture resulting in competition of peptides for MS sequencing.

A further 39 non-snRNP protein splicing factors were identified, including early-acting factors such as splicing factor 1 (SF1) (Kramer 1992) as well as late factors such as SLU7 (Frank and Guthrie 1992; Zhou and Reed 1998) and Aly (Zhou et al. 2000), which are required for the second catalytic step of the splicing reaction and for RNA export, respectively.

A total of 20 hnRNP proteins were identified in this analysis. The hnRNP proteins are defined by a common RNA-binding motif (Dreyfuss et al. 1993) and are thought to have diverse functions in RNA protection and processing. Some hnRNPs are known to be splicing factors, such as GRY-RBP/hnRNP Q (Mourelatos et al. 2001), which was also identified here. Furthermore, 27 other RNA-processing-related proteins are listed in Table 1. Of these, a number have functions closely associated with splicing. As an example, we identified the 5′ cap binding proteins (CBP) 20 and CBP 80 (Izaurralde et al. 1994), as well as the cleavage and polyadenylation factor (CPSF) (Keller and Minvielle-Sebastia 1997).

As can be seen from Figure 2, the group of known splicing factors spans a wide range of abundances, with the least abundant factors at the lower detection limit. There are several reasons that the isolated spliceosome proteins are not expected to be observed in stoichiometric amounts. First, some factors such as Sm proteins are present in multiple copies in the spliceosome. Second, to ensure maximum coverage of splicing factors, we have deliberately analyzed a mixture of fully assembled, active spliceosomes and partially assembled spliceosome intermediates that all bound to the pre-mRNA

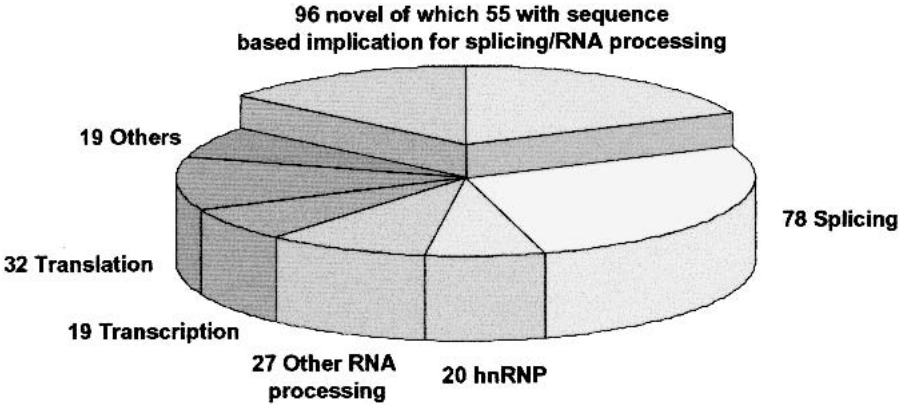

Figure 3 Classification of the identified proteins into eight functional classes.

bait. Some proteins that assemble early are therefore likely to be present in higher levels than factors only assembled at later stages of spliceosome formation. Third, some proteins are less tightly associated with the spliceosome than others, leading to differential losses during purification. Fourth, not all proteins are detected with equal efficiency during MS analysis.

Novel Proteins

In addition to the expected splicing and RNA-processing factors, we discovered a large group of novel proteins with no known function (Table 2). We have also included proteins in this category that were previously observed in large-scale screens or that were cloned because of their homology to a known factor, if no further biological information was available (18 and 5 proteins, respectively). These 96 proteins were submitted for homology and domain searches. Interestingly, this analysis resulted in 55 proteins with sequence similarity to known splicing factors or domains that implicate them in RNA processing.

Listed at the top of Table 2 are 12 proteins with extensive sequence identity to known splicing factors. Two of these are likely orthologs of known yeast splicing factors (*Schizosaccharomyces pombe* Prp4 [Rosenberg et al. 1991] and *Saccharomyces cereviseae* Isy1p [Dix et al. 1999]), and others are similar to snRNP and hnRNP proteins. Because protein identification is performed on the basis of tryptic peptides, in some cases high sequence identity between known and novel proteins could confound the analysis. However, in all cases listed in Table 2, peptides were sequenced that unambiguously identified the novel protein. For example, we identified the hypothetical protein ENSP00000272417 that is identical to U5 snRNP 200 kD in 1627 of 1701 amino acids on the basis of eight peptides that only occur in the novel protein (MLLQSSEGR, TLVEDLFADK, FLYQLHETEK, LLSMAKPVYHAITK, LIILDEIHLLHDDR, MDTDLETMDLDQGGEALAPR, QVLDLEDLVFTQGSHFMANK, and LIILDEIHLLHDDRGPVLEALVAR).

Other groups of novel proteins with evidence linking them to splicing were 8 helicases containing a DEAD- or DEAD/DEAH-box motif (Luking et al. 1998), and 17 proteins containing an RNA-recognition motif (RRM; Shamoo et al. 1995) or other domains typical of splicing factors, such as proline-tryptophan-isoleucine motif (PWI) (Blencowe and Ouzounis 1999) or G patch (Aravind and Koonin 1999).

The novel proteins were further analyzed for localization signals. As expected, none of the proteins scored high for secretory signal sequences. In 21 proteins, a bipartite nuclear localization was apparent. Based on their domain structure, a number of the proteins are predicted to function in signaling or in gene expression. Figure 2 indicates that the novel proteins either with similarity to known splicing factors, or with domains previously associated with RNA-processing factors, show a similar abundance pattern to the known RNA-processing proteins.

Proteins with a Function in Transcription

Table 3 lists proteins with a known function in transcription or translation, that is, the cellular processes that occur upstream and downstream of splicing during gene expression. Among the transcription factors, we found two subunits of RNA polymerase II, members of the histone H2A and H2B families, histone acetylase and deacetylase. Several initiation factors were also present.

Figure 2 indicates that the translation factors detected were generally of lower abundance than the splicing factors. Because many of these factors were close to the threshold of detection with only one or a few peptides sequenced, it is not surprising that only selected members of these complexes appear in the list. This is also the case for some of the other complexes that were observed in the spliceosomal fraction and that are described below.

Ribosomal Proteins and Associated Factors

A number of ribosomal proteins, particularly from the 40S subunit, were identified in the preparation. We furthermore identified several proteins from the signal recognition particle, which binds to the nascent protein chain as well as elongation factor 1.

Proteins with Other Previously Described Functions

Seven proteins with potential regulatory roles were identified. Among these were several signaling-related proteins and cell cycle-associated proteins. The remaining proteins include nucleoprotein TPR, a component of the nuclear pore complex; several cytoskeleton-associated factors; and nucleolin, which is an abundant component of the nucleolus.

DISCUSSION

Mass Spectrometric Analysis of the Spliceosome

In this study we have used enhanced MS technology to characterize the protein composition of human spliceosomes. To ensure maximum coverage of basal human splicing factors, we analyzed a combination of active spliceosomes and intermediate splicing complexes that formed on each of two separate pre-mRNA substrates derived, respectively, from adenovirus (AD1) and β-globin (AL4) transcripts. Spliceosomes assembled in vitro were purified, and the resulting protein mixture was enzymatically degraded to peptides, which were analyzed by liquid chromatography coupled on-line with mass spectrometric sequencing.

Table 1. Proteins With a Known Function in Splicing and RNA Processing

Acc. no.[a]	Name
snRNP core proteins	
SWISS-PROT: Q15357[a]	Sm G
SWISS-PROT: Q15356	Sm F
SWISS-PROT: P08578	Sm E
SWISS-PROT: P13641	Sm D1
SWISS-PROT: P43330	Sm D2
SWISS-PROT: P43331	Sm D3
SWISS-PROT: P14678	Sm B/B′
U1 snRNP	
SWISS-PROT: P09234	U1 snRNP C
SWISS-PROT: P09012	U1 snRNP A
SWISS-PROT: P08621	U1 snRNP 70 kDa
U2 snRNP	
SWISS-PROT: Q15427	SAP 49
SWISS-PROT: Q12874	SAP 61
SWISS-PROT: Q15428	SAP 62
SWISS-PROT: Q15459	SAP 114
SWISS-PROT: Q15393	SAP 130
SWISS-PROT: Q13435	SAP 145
SWISS-PROT: O75533	SAP 155
SWISS-PROT: Q01081	U2AF 35 kDa
SWISS-PROT: P26368	U2AF 65 kDa
SWISS-PROT: P09661	U2 snRNP A′
SWISS-PROT: P08579	U2 snRNP B″
U5 snRNP	
ENSP00000263694	U5 snRNP 40 kDa
ENSP00000261905	U5 snRNP 100 kDa
ENSP00000266079	U5 snRNP 102 kDa
SWISS-PROT: Q15029	U5 snRNP 116 kDa
SWISS-PROT: O75643	U5 snRNP 200 kDa
ENSP00000254706	U5 snRNP 220 kDa
U6 snRNP	
SWISS-PROT: Q9Y333	LSm2
SWISS-PROT: Q9Y4Z1	LSm3
SWISS-PROT: Q9Y4Z0	LSm4
SWISS-PROT: Q9Y4Y8	LSm6
SWISS-PROT: Q9UK45	LSm7
SWISS-PROT: O95777	LSm8
U4/U6 snRNP	
ENSP00000259401	U4/U6 snRNP hPrp4
ENSP00000236015	U4/U6 snRNP hPrp3
ENSP00000291763	U4/U6 snRNP 61 kDa
U4/U6.U5 snRNP	
SWISS-PROT: P55769	U4/U6.snRNP 15.5 kDa
ENSP00000263858	U4/U6.U5 snRNP 65 kDa
ENSP00000256313	SART-1 = U4/U6.U5 snRNP 110 kDa
SR proteins	
SWISS-PROT: Q07955	SF2
SWISS-PROT: Q16629	9G8
SWISS-PROT: Q01130	SC35
SWISS-PROT: P23152	SRp20
SWISS-PROT: Q13242	SRp30C
SWISS-PROT: Q05519	SRp54
SWISS-PROT: Q13247	SRp55
TREMBL: Q8WXA9	Splicing factor, arginine/serine-rich 12
ENSP00000255590	Ser/Arg-related nuclear matrix protein
Other splicing factors	
ENSP00000227503	SF1
ENSP00000235397	SPF27
ENSP00000239010	SPF30
ENSP00000263697	SPF31
TREMBL: O75939; Q96GY6	SPF45
ENSP00000265414	CDC5-related protein
SWISS-PROT: Q13573	SKIP
TREMBL: Q9NZA0	PUF60
TREMBL: O43660	Pleiotropic regulator 1
ENSP00000253363	CC1.3
ENSP00000296702	CA150
SWISS-PROT: Q14562	DEAH-box protein 8
SWISS-PROT: O43143	DEAD/H-box-15
SWISS-PROT: O60231	DEAD/H-box-16
ENSP00000268482	hPRP16
SWISS-PROT: O60508	hPRP17
ENSP00000198939 (+ENSP00000248044)	ERPROT 213-21 (+N-terminal extension of ERPROT)
ENSP00000290341	IGF-II mRNA-binding protein 1
SWISS-PROT: P23246	PTB-associated splicing factor
ENSP00000266611	IK factor
ENSP00000257528	SLU7
ENSP00000216727	poly(A)-binding protein II
ENSP00000293531	KH-type splicing regulatory protein
ENSP00000294623	far upstream element-binding protein 1
ENSP00000227524	nuclear matrix protein NMP200
TREMBL: Q96HB0	HCNP protein
ENSP00000278799	crooked neck-like 1
SWISS-PROT: Q9Y3B4	pre-mRNA branch site protein p14
ENSP00000292123	scaffold attachment factor B
ENSP00000261167	SH3 domain-binding protein SNP70
hnRNP	
ENSP00000257767	GRY-RBP
SWISS-PROT: Q13151	hnRNP A0
SWISS-PROT: P09651	hnRNP A1
SWISS-PROT: P22626	hnRNP A2/hnRNP B1
ENSP00000298069	hnRNP A3
ENSP00000261952	hnRNP AB, isoform a
SWISS-PROT: P07910	hnRNP C
SWISS-PROT: Q14103	hnRNP D
ENSP00000295469	hnRNP D-like
SWISS-PROT: P52597	hnRNP F
SWISS-PROT: P38159	hnRNP G
SWISS-PROT: P31943	hnRNP H
ENSP00000265866	hnRNP H3
SWISS-PROT: P26599	Polypyrimidine tract-binding protein; hnRNP I
ENSP00000297818	hnRNP K
SWISS-PROT: P14866	hnRNP L
SWISS-PROT: P52272	hnRNP M
SWISS-PROT: O43390	hnRNP R
Q00839	hnRNP U
TREMBL: O76022	E1B-55kDa-associated protein 5
RNA processing	
SWISS-PROT: P52298	CBP 20 kDa
SWISS-PROT: Q09161	CBP 80 kDa
SWISS-PROT: P17844	DEAD/H-box-5; RNA helicase p68
SWISS-PROT: P35637	RNA-binding protein FUS
SWISS-PROT: Q01844	RNA-binding protein EWS
SWISS-PROT: Q12906	Interleukin enhancer-binding factor 3
ENSP00000270794	TLS-associated serine-arginine protein 2

(*continued*)

Table 1. (*Continued*)

Acc. no.[a]	Name
RNA processing	
ENSP00000269407	Aly
SWISS-PROT: Q9UBU9	Tap
ENSP00000261600	hHpr1
SWISS-PROT: Q08211	RNA helicase A
ENSP00000264073	ELAV-like protein 1 (Hu-antigen R)
SWISS-PROT: P43243	matrin 3
SWISS-PROT: P55265	Double-stranded RNA-specific adenosine deaminase (DRADA)
ENSP00000300291	CPSF 25 kDa
ENSP00000292476	CPSF 30 kDa
ENSP00000266679	similar to CPSF 68 kDa
SWISS-PROT: Q9UKF6	CPSF 73 kDa
SWISS-PROT: Q9P2I0	CPSF 100 kDa
SWISS-PROT: Q10570	CPSF 160 kDa
ENSP00000227158	cleavage stimulation factor subunit 3
SWISS-PROT: P05455	Lupus La protein: Sjogren syndrome type B antigen
SWISS-PROT: Q06265	Polymyositis/scleroderma autoantigen 1
SWISS-PROT: Q01780	Polymyositis/scleroderma autoantigen 2
SWISS-PROT: Q9Y2L1	Exosome complex exonuclease RRP44
SWISS-PROT: Q9NPD3	Exosome complex exonuclease RRP41
ENSP00000262489	Dhm1-like protein

[a]SWISS-PROT or ENSEMBL accession numbers are given at http://srs.embl-heidelberg.de:8000/srs5/ and http://www.ensembl.org.

A quadrupole time-of-flight instrument provided high resolution and high mass accuracy in the peptide analysis. More than 7000 ion peaks were fragmented in six separate runs. Based on these high-quality data, a total of 311 proteins were identified unambiguously by a combination of automated database search and manual interpretation of peptide fragmentation spectra (Fig. 1). This surprisingly large number of factors is comprised of 125 proteins involved in RNA processing, 71 proteins involved in other, previously described functions, and 96 proteins that have not been functionally described before.

The larger number of proteins found in the present study compared with our previous study (Neubauer et al. 1998) is partly owing to the increased sensitivity of the enhanced proteomics methods now available and partly to the less stringent wash conditions used in this study. The fact that two substrates were used, furthermore, helped to identify additional factors. Human sequence databases have also improved dramatically over the last few years. The previous study used a combination of 2D gel electrophoresis and nanoelectrospray mass spectrometry. The much higher throughput provided by on-line tandem mass spectrometric peptide sequencing combined with automated database searching made it realistic to deal with thousands of peptide fragmentation spectra and even allowed multiple analysis conditions.

Figure 4 shows the calculated positions of the identified proteins in a plot of isoelectric point versus molecular weight (labeled virtual 2D gel). About 40% of the proteins fall outside of the coordinates of a standard 2D gel. For example, the Sm and Lsm proteins are too small, and many other RNA-processing proteins are too basic or too large to be represented on a 2D gel. Note that two proteins, which are outside the box in Figure 4, had migrated anomalously in the previous analysis such that they had been found at positions inside the coordinates of the previously analyzed 2D gels.

We observed a wide variation in the apparent quantity of the spliceosomal proteins (see Fig. 2). The more abundant factors were identified with dozens of sequenced peptides, whereas some of the least abundant factors were identified on the basis of a single peptide. This variation does not only reflect different stoichiometry in the different spliceosomal complexes that were purified, but is also a result of the differential response of the peptides in the analytical method used. To obtain a rough visualization of the abundance of different proteins, we defined a simple protein abundance index (PAI) as the ratio between the sequenced peptides of a protein and the total number of tryptic peptides predicted from the protein sequence (see Methods). Although the PAI in the form presented here is by no means an accurate measure of protein amount, it can be used as a guide for relative classification in abundant and less abundant proteins. For example, the novel proteins G10 protein homolog (EDG-2) and hypothetical protein ENSP00000292314 have a very high index and as such would be excellent candidates for detailed functional studies even though they lack sequence similarities to proteins previously found in splicing/RNA processing. Other proteins with a high index and sequence similarity to known splicing/RNA-processing proteins are the hypothetical proteins similar to U5 snRNP 200 kD, the hypothetical protein similar to U2 snRNP A′, the cyclophilin CGI-124 protein, and the RRM domain-containing Arsenite-resistance protein 2. Among the proteins involved in transcription, Interleukin enhancer-binding factor 2, which binds to the RNA-processing protein Interleukin enhancer-binding factor 3, and the nuclease-sensitive element binding protein 1 also appear to be abundant.

We originally expected that the proteins identified in our previous investigation would have been the most abundant of the much larger number of proteins identified here. However, the average PAI of that group was only moderately higher (0.85 compared with 0.61; data not shown), and many of the previously identified proteins were of low abundance, as indicated by the present analysis. This may reflect the fact that 2D gel electrophoresis with subsequent nanoelectrospray peptide sequencing is also very sensitive for the subgroup of proteins that are readily focused and visualized on the gel.

Proteins associated with the two different substrates were largely identical, especially for the core spliceosomal components. Differences in those components mainly occurred for proteins that were identified with very few peptides, indicating that these were missed in the other purification. However, there were also significant differences in non-core splicing proteins that appear to be unrelated to the analysis and that may have functional significance. As an example, among the clearest differences were the Fuse binding proteins (FBP) 1, 2 and 3, which are unique to the AL4 substrate. FBPs bind to the single-stranded far upstream element (FUSE) upstream of the c-*myc* gene. In addition to its transcriptional role, FBP1 and its closely related siblings FBP2 and FBP3 have been reported to bind RNA and participate in various steps of RNA processing, transport, or catabolism. Interestingly, the insulin growth factor (IGF)-II mRNA-binding protein 3 was also detected exclusively attached to AL4 and is known to recognize c-*myc* and *IGF-II* mRNA, respectively, and to regulate their expression posttranscriptionally. These substrate-specific factors will be the subject of a future investigation. Altogether, 79 factors were unique to the AL4 substrate and 44 to the AD1 substrate.

Table 2. Novel Proteins

Acc. No.[a]	Name	Comments[b]
Novel proteins and proteins with unclear functions with sequence similarities implicating them in splicing/mRNA processing		
ENSP00000295270	Hypothetical protein	Similar to U5 snRNP 200 kDa
ENSP00000272417	CDNA FLJ13778 fis	Similar to U5 snRNP 200 kDa
ENSP00000301345	Hypothetical protein	Similar to U5 snRNP 220 kDa
TREMBL: Q9NUY0	CDNA FLJ11063 fis	Similar to arginine/serine-rich 4
SWISS-PROT: Q13523	Serine/threonine-protein kinase	Ser/Thr protein kinase family, similar to *S. pombe* PRP4
ENSP00000296630	Hypothetical protein	RRM domain, bipartite NLS, similar to arginine/serine-rich 11
ENSP00000266057	CDNA FLJ10998 fis	Similar to RNA lariat debranching enzyme
ENSP00000273541	Hypothetical protein	Similar to Isy 1p, a potential splice factor in yeast
XP_013029	Hypothetical protein	Similar to U2 snRNP A′
ENSP00000286032	Hypothetical protein	Similar to hnRNP A3
ENSP00000301786	Hypothetical protein	Similar to hnRNP U
ENSP000000301784	Hypothetical protein	Similar to hnRNP U
ENSP00000261832	Hypothetical protein DKFZp434E2220	BASIC, basic domain in HLH proteins of MYOD family, PSP, proline-rich domain in spliceosome-associated proteins, zinc finger CCHC, zinc knuckle
ENSP00000244367	CGI-124 protein	Cyclophilin-type peptidyl-prolyl *cis-trans* isomerase
ENSP00000215824	CYP-60	Cyclophilin-type peptidyl-prolyl *cis-trans* isomerase
ENSP00000234288	PPIL3b	Cyclophilin-type peptidyl-prolyl *cis-trans* isomerase
ENSP00000282972	Serologically defined colon cancer antigen 10	Cyclophilin-type peptidyl-prolyl *cis-trans* isomerase, bipartite NLS
SWISS-PROT: Q9UNP9	Cyclophilin E	Cyclophilin-type peptidyl-prolyl *cis-trans* isomerase, RRM domain
ENSP00000261308	KIAA0073 protein	Cyclophilin-type peptidyl-prolyl *cis-trans* isomerase, G-protein beta WD-40 repeats
SWISS-PROT: Q92841	Probable RNA-dependent helicase p72	DEAD/DEAH-box helicase
ENSP00000274514	RNA helicase	DEAD/DEAH-box helicase
ENSP00000242776	Hypothetical protein	Similar to nuclear RNA helicase, DECD variant of DEAD-box helicase family
SWISS-PROT: Q92499	DDX1	DEAD/DEAH-box helicase, SPRY domain
SWISS-PROT: Q9NR30	DDX21	DEAD/DEAH-box helicase, bipartite NLS
SWISS-PROT: Q9UJV9	DEAD-box protein abstract homolog	DEAD/DEAH-box helicase, zinc finger CCHC type
ENSP00000218971	DDX26	DEAD-box, von Willebrand factor type A domain
SWISS-PROT: P38919	Eukaryotic initiation factor 4A-like NUK-34	DEAD-box helicase
ENSP00000297920	Hypothetical protein FLJ11307	Double-stranded RNA-binding domain (DsRBD)
ENSP00000263115	Hypothetical protein	G-patch domain
ENSP00000277477	Far upstream element (FUSE) binding protein 3	KH domain
ENSP00000295749	KIAA 1604 protein	MIF4G, middle domain of eukaryotic initiation factor 4G and MA3 domain, bipartite NLS
ENSP00000298643	PRO1777	PWI domain
SWISS-PROT: Q9Y580	RNA-binding protein 7	RRM domain
SWISS-PROT: O43251	RNA-binding protein 9	RRM domain
ENSP00000295971	Hypothetical protein FLJ20273	RRM domain
ENSP00000266301	KIAA 1649 protein	RRM domain
SWISS-PROT: Q9Y388	Hypothetical protein CGI-79.B	RRM domain
SWISS-PROT: Q02040	B-lymphocyte antigen precursor	RRM domain
ENSP00000262632	Hypothetical 47.4 kDa	RRM domain, ATP/GTP-binding site motif A (P-loop)
ENSP00000293677	Hypothetical protein	RRM domain, Bipartite NLS
SWISS-PROT: Q9BXP5	Arsenite-resistance protein 2	RRM domain, Bipartite NLS
ENSP00000220496	Hypothetical protein FLJ10634	RRM domain, DNAJ heat shock protein, bipartite NLS
TREMBL: O00425	Putative RNA-binding protein KOC	RRM domain, KH domain
ENSP00000262710	KIAA0670 protein	RRM domain, SAP domain
TREMBL: Q96SC6	OTT-MAL	RRM domain, SAP domain
ENSP00000295996	KIAA0332 protein	RRM domain, Surp domain, Bipartite NLS
ENSP00000199814	Hypothetical protein FLJ10290	RRM domain, Zinc finger C-x8-C-x5-C-x3-H type
SWISS-PROT: P98175	RNA-binding protein 10	RRM domain, C2H2 type zinc finger, bipartite NLS
ENSP00000261972 (+ENSP00000261973)	Hypothetical protein S164 (+N-terminal extension: CDNA: FLJ22454 fis, clone HRC09703)	RRM domain, PWI domain, bipartite NLS, Spectrin repeat (ENSP00000261973 encodes the N-terminal extension of ENSP00000261972)
TREMBL: Q9UQ35	RNA-binding protein	RS domain
ENSP00000247001	F23858_1	Surp domain, G-patch domain
ENSP00000299951	Hypothetical protein	U1-like zinc finger, bipartite NLS
ENSP00000281372	HsKin17 protein	C2H2 zinc finger
TREMBL: Q96KR1	Putative Zinc finger protein	C2H2 zinc finger
ENSP00000239893	OPA-interacting protein OIP2	3′ exoribonuclease family
Novel proteins without similarities implicating them in splicing/mRNA processing		
SWISS-PROT: Q9C0J8	WDC146	G-protein beta WD-40 repeats
ENSP00000253952	Hypothetical 34.8 kDa protein	G-protein beta WD-40 repeats

(*continued*)

Table 2. (*Continued*)

Acc. No.[a]	Name	Comments[b]
ENSP00000263222	Hypothetical 57.5 kDa protein	G-protein beta WD-40 repeats
ENSP00000156471	KIAA0560 protein	ATP/GTP-binding site motif A (P-loop)
SWISS-PROT: Q9UH06	Hypothetical 12.4 kDa protein	PHD-finger (C4HC3 zinc finger) belongs to the UPF0123 family of hypothetical proteins
ENSP00000216252	BK223H9	
ENSP00000260210	Hypothetical protein MGC13125	Bipartite NLS, ankyrin similarity
ENSP00000257181	Hypothetical protein FLJ14936	Bipartite NLS, similar to unknowns
ENSP00000290008	Hypothetical protein	Bipartite NLS
SWISS-PROT: Q9NZB2	C9orf10 protein	Bipartite NLS, similar to unknowns
ENSP00000247026	Hypothetical 66.4 kDa protein	Bipartite NLS
ENSP00000236273	GCIP-interacting protein p29	Bipartite NLS, similar to unknowns
ENSP00000292314	Hypothetical protein	Bipartite NLS, similar to unknowns
ENSP00000266923	C21orf70	Bipartite NLS, similar to unknowns
ENSP00000221899	NY-REN-24 antigen	Bipartite NLS, Ezrin/radixin/moesin family; similar to *Drosophila cactin*
SWISS-PROT: Q14331	FRG1 protein (FSHD region gene 1 protein)	Bipartite NLS, Lipocalin-related protein and Bos/Can/Equ allergen domain
SWISS-PROT: P42285	KIAA0052 protein	SKI2 helicase family
ENSP00000221413	CGI-46 protein	DnaB helicase family
ENSP00000222969	G10 protein homolog (EDG-2)	G10 protein family
ENSP00000279839	Adrenal gland protein AD-002	GTP-binding signal recognition particle (SRP54) G-domain
ENSP00000278702	Similar to nuclear mitotic apparatus protein 1	Involucrin repeat, G-protein gamma subunit, DNA gyrase/topoisomerase IV, subunit A, M protein repeat, bZIP (Basic-leucine zipper) transcription factor family
SWISS-PROT: Q92733	Proline-rich protein PRCC	Proline-rich extension
ENSP00000263905	KIAA1461 protein	PWWP domain, Methyl-CpG binding domain
XP_089514	Hypothetical protein	Similar to nucleophosmin
ENSP00000258457	Hypothetical 25.9 kDa protein	Similar to *Xenopus ashwin*
TREMBL: Q8WYA6	Nuclear associated protein	Similar to Bos taurus P14
TREMBL: Q13769	Hypothetical protein	Similarity to intermediate filament b [*Dugesia japonica*]
SWISS-PROT: Q9Y5B6	GC-rich sequence DNA-binding factor homolog	Similar to C-TERMINAL OF GCF/TCF9 and other putative transcription factors
SWISS-PROT: Q9Y224	Hypothetical protein CGI-99	Similarity to putative transcription factors
ENSP00000216038	Hypothetical 55.2 kDa protein	Uncharacterized protein family UPF0027
ENSP00000289509	Hypothetical 80.5 kDa protein	Similar to unknowns
ENSP00000245838	Hypothetical protein LOC57187	Similar to unknowns
ENSP00000289996	Hypothetical protein	Similar to unknowns
ENSP00000252137	DiGeorge syndrome critical region gene DGSI protein	Similar to unknowns
ENSP00000256579	Hypothetical protein FLJ10330	Similar to unknowns
ENSP00000245651	C20orf158 protein	Similar to unknowns
SWISS-PROT: Q9BWJ5	Hypothetical protein MGC3133	Similar to unknowns
ENSP00000272091	Hypothetical protein XP_089191	Similar to unknowns
ENSP00000297526	KIAA1440 protein	Similar to unknowns
ENSP00000271942	Hypothetical protein FLJ21919	Similar to unknowns
TREMBL: Q9BTU2	Hypothetical 31.5 kDa protein	Similar to unknowns
TREMBL: Q8WVN3	Hypothetical protein	Similar to unknowns

[a]SWISS-PROT or ENSEMBL accession numbers are given at http://srs.embl-heidelberg.de:8000 and http://www.ensembl.org.
[b]Domains: RRM: RNA recognition motive; Bipartite NLS: Bipartite Nuclear Localization Signal; SPRY: SP1a/RY anodine receptor SPRY domain; G-patch: named after seven highly conserved glycines; KH: hnRNP K homology domain; PWI: proline-tryptophan-isoleucine motifs; SAP: SAF-A/B, Acinus and PIAS motif; RS: Arginine-Serine repeats; Surp: Suppressor-of-white-apricot splicing regulator domain.

Discussion of Identified Factors

Significantly, all known U1, U2, and U6 proteins were identified in this large-scale study (Table 1). Virtually all of the other known spliceosomal proteins were also observed, which includes the SR proteins that were not detected in our previous study. Five proteins with a described role in U4/U6 and U4/U6 · U5 snRNPs were not identified with either substrate tested. It is possible that these factors are present in the samples but were missed because of low abundance, weak affinity, or other technical reasons affecting detection in this system. Alternatively, it is possible that these proteins are not, in fact, stable components of the spliceosomes formed on the pre-mRNAs analyzed.

A large proportion of the proteins that were detected have a known function in RNA processing. In addition to the known splicing factors, we identified 20 hnRNP proteins, some of which are also implicated in splicing. Likewise, there are several proteins in the category of other RNA-processing proteins that function in splicing.

Table 2 lists 96 novel proteins present in the spliceosomal preparation. At first, this appears to be a surprisingly large number, considering that the spliceosome has been studied intensively for many years. However, we note that a recent analysis of human nucleolar proteins showed that >30% of the factors detected were novel despite more than two hundred years of research into nucleoli (Andersen et al.

Table 3. Proteins Involved in Transcription, Translation, and Other Functions

Acc. no.[a]	Name
SWISS-PROT: P16991	CCAAT-binding transcription factor I subunit A
ENSP00000271939	Interleukin enhancer binding factor 2, 45 kD
TREMBL: O15043	Death associated transcription factor 1
ENSP00000266071	Death associated transcription factor-1 isoform b
SWISS-PROT: P16383	GC-rich sequence DNA binding factor
SWISS-PROT: P78347	general transcription factor II
ENSP00000228251	Cold shock domain protein A
NP_005325	host cell factor CI
SWISS-PROT: P49848	Transcription initiation factor TFIID 70 kD subunit
SWISS-PROT: P12956	ATP-dependent DNA helicase II, 70 kD subunit
ENSP00000283131	SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily a, member 6
SWISS-PROT: P30876	DNA-directed RNA polymerase II 140 kD
SWISS-PROT: P24928	DNA-directed RNA polymerase II largest subunit
SWISS-PROT: P02261	Histone $H2A^{-}$
SWISS-PROT: P20670	H2A histone family member O
SWISS-PROT: Q93080	H2B histone family several members possible
SWISS-PROT: P09429	High-mobility group protein 1
SWISS-PROT: O15347	High mobility group box 4
ENSP00000275182	Histone deacetylase 2
SWISS-PROT: Q16576	Histone acetyltransferase type B subunit 2
SWISS-PROT: P23396	40S ribosomal protein S3
NP_000997	40S ribosomal protein S3A
SWISS-PROT: P12750	40S ribosomal protein S4
SWISS-PROT: P22090	40S ribosomal protein S4Y isoform
SWISS-PROT: P46782	40S ribosomal protein S5
SWISS-PROT: P23821	40S ribosomal protein S7
SWISS-PROT: P09058	40S ribosomal protein S8
SWISS-PROT: P46781	40S ribosomal protein S9
SWISS-PROT: P46783	40S ribosomal protein S10
ENSP00000237131	40S ribosomal protein S12
SWISS-PROT: Q02546	40S ribosomal protein S13
SWISS-PROT: P11174	40S ribosomal protein S15
SWISS-PROT: P39027	40S ribosomal protein S15a
SWISS-PROT: P17008	40S ribosomal protein S16
SWISS-PROT: P08708	40S ribosomal protein S17
SWISS-PROT: P25232	40S ribosomal protein S18
SWISS-PROT: P39019	40S ribosomal protein S19
SWISS-PROT: P25111	40S ribosomal protein S25
SWISS-PROT: P30054	40S ribosomal protein S29
SWISS-PROT: Q05472	40S ribosomal protein S30
SWISS-PROT: P04643	40S ribosomal protein S11
SWISS-PROT: P46777	60S ribosomal protein L5
SWISS-PROT: P35268	60S ribosomal protein L22
SWISS-PROT: P29316	60S ribosomal protein L23a
SWISS-PROT: P12947	60S ribosomal protein L31
TREMBL Q8WT0	Signal recognition particle 9 kD
SWISS-PROT: P09132	Signal recognition particle 19 kD
SWISS-PROT: Q9UHB9	Signal recognition particle 68 kD
TREMBL: Q8WUK2	Signal recognition particle 68 kD isoform
SWISS-PROT: Q76094	Signal recognition particle 72 kD
SWISS-PROT: P04720	Elongation factor 1
SWISS-PROT: P12270	Nucleoprotein TPR
SWISS-PROT: P46940	Ras GTPase-activating-like
ENSP00000268182	protein IQGAP1 (P195)
SWISS-PROT: P52292	Importin alpha-2 subunit
SWISS-PROT: O75909	Cyclin K
SWISS-PROT: P78396	Cyclin A1
SWISS-PROT: P09874	poly(ADP-ribosyl)transferase
SWISS-PROT: O43823	A-kinase anchor protein 8
ENSP00000262971	PIAS1
ENSP00000296215	Smad nuclear-interacting protein 1
ENSP00000234443	Protein kinase, interferon-inducible double stranded RNA dependent activator; protein activator of the interferon-induced protein kinase
ENSP00000300630	Ubiquitin
ENSP00000271238	Phosphatase 2A inhibitor
SWISS-PROT: P19338	Nucleolin
SWISS-PROT: P55081	Microfibrillar-associated protein 1
SWISS-PROT: P11142	Heat shock cognate 71 kD protein

Table 3. (*Continued*)

Acc. no.[a]	Name
ENSP00000286912	Dynein heavy chain
SWISS-PROT: P08670	Vimentin
ENSP00000243115	Tubulin, alpha
ENSP00000259925	Tubulin, beta 5

[a]SWISSPROT or ENSEMBL accession numbers are given. (http://srs.embl-heidelberg.de:8000/srs5/ and www.ensembl.org)

2002a). More than half of the novel spliceosome-associated proteins detected here either showed strong similarity to known splicing factors or had domains such as RRM, DEAD box. and/or PWI that implicate them in RNA processing. Also, a cyclophilin, USA-CyP (Horowitz et al. 2002), has been shown to act in the spliceosome and with six novel proteins that are likely members of the cyclophilin-type-peptidyl-prolyl-*cis*-*trans*-isomerases. Thus this family may play an even larger role in splicing than previously thought.

Interestingly, these novel proteins also show a similar abundance pattern to the known splicing factors (Fig. 2). The fact that these proteins were identified in a spliceosomal preparation, combined with the bioinformatic evidence linking them to splicing, strongly indicates that these proteins are likely to be bona fide splicing factors.

Given the large proportion of proteins implicated in splicing or related RNA-processing activities, it is likely that many of the remaining 42 novel proteins are also involved in these functions. A detailed analysis of all these factors is beyond the scope of this study but will be addressed in future work.

Further studies will be required to assess which of the newly identified spliceosome-associated proteins are directly involved in splicing and which are involved in other activities relating to the synthesis, processing, localization, or transport of nascent mRNA. In this regard, it is interesting that our parallel analysis of host cell factor (HCF), identified here as a spliceosome protein, shows that it is required for splicing in vivo and in vitro (P. Ajuh and A.I. Lamond, in prep.). However, it is likely that not all of the novel spliceosome proteins are directly required for the catalysis of splicing. Rather, we favor the interpretation that the splicing machinery works in the context of a larger series of activities required for the production and cytoplasmic export of mature mRNA. Thus, some of the factors identified may have roles in affecting other related RNA processing, editing, and transport events. Consistent with this idea, the proteins detected include multiple components of the 3′ cleavage and polyadenylation machinery (Minvielle-Sebastia 1999) as well as the double-stranded RNA-specific adenosine deaminase (DRADA) RNA-editing enzyme that is known to associate with a nuclear protein complex (Zhang 2001). The mRNA export machinery was repre-

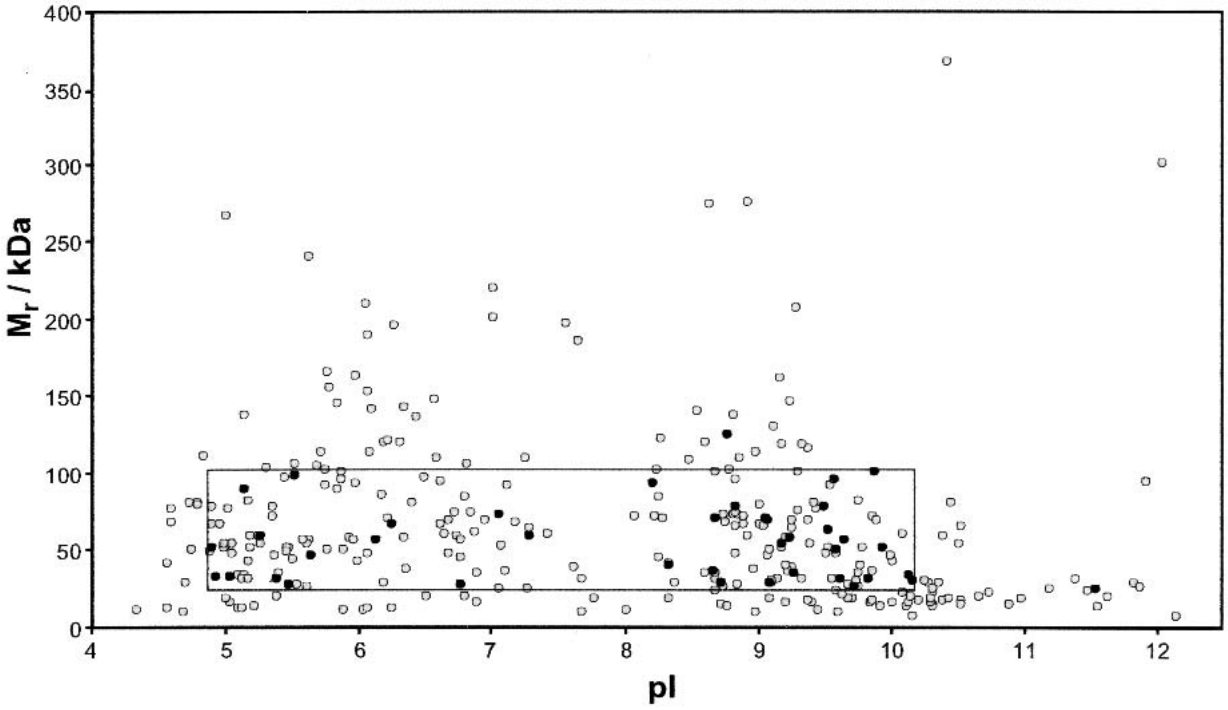

Figure 4 Virtual 2D gel of proteins identified in the spliceosome preparation. The coordinates are the theoretical molecular mass and isoelectric point for each protein. The gray circles represent factors identified in this study, and the black circles represent proteins identified in this and our previous study (Neubauer et al. 1998). The box indicates the coordinate space spanned by our previous investigation using 2D gel electrophoresis.

sented by the proteins Aly (Zhou et al. 2000), Tap (Gruter et al. 1998), hHpr1 (Strasser et al. 2002), and possibly the nuclear pore protein TPR (Bangs et al. 1998; Frosst et al. 2002). Thus, our data provide further support for direct linkage between splicing and mRNA export (Reed and Hurt 2002).

We also identified a number of ribosomal proteins involved in protein translation in the cytosol. At present, we know of no direct evidence linking ribosomal proteins to splicing functions. The ribosomal proteins likely copurified owing to direct binding to the RNA bait, but alternatively may have bound to the mRNA export complex, components of which we have identified here. Thus, the significance of the ribosomal protein data needs to be evaluated cautiously until further studies can be carried out to test their potential link to spliceosomes.

Interestingly, a number of transcription-related proteins including subunits of RNA polymerase II and other transcription factors were identified in this analysis. This finding indicates a tight coupling of transcription with splicing, consistent with recent in vivo and biochemical data indicating such a link (for review, see Bentley 2002). For example, it is already known that CA150, which was identified in our preparation, can bind to RNA polymerase II and SF1 (Goldstrohm et al. 2001). Scaffold attachment factor B can bind to RNA polymerase II and SR proteins (Nayler et al. 1998) and thus also represents a possible direct link between transcription and splicing. In this context, it is interesting to note that Protein inhibitor of activated STAT1 (PIAS1) likewise has the characteristics of a scaffold-attachment factor and has a speckled nuclear localization (Tan et al. 2002), which is typical for splicing factors.

Although it is known that splicing can be tightly regulated, for example, in many instances of alternative splicing (Graveley 2002), less is known about the mechanisms involved in this regulation. In this regard, it is interesting that a number of putative regulatory proteins were found in association with the spliceosome. Three Death-box-containing proteins, one of which is a novel protein, may link the spliceosome to apoptosis. These proteins may, however, have a function similar to hHrp1, a Death-box-containing protein acting in mRNA export. Two other proteins that were found, protein phosphatase II inhibitor, a protein co-immunoprecipitating with SPF30 (J. Rappsilber and M. Mann, unpubl) and poly(ADP-ribosyl)transferase indicate other leads into the regulative mechanisms of the splicing process that will be followed up in future studies. It is also possible that splicing activity in vivo may be regulated during the cell cycle. Consistent with this idea, certain spliceosomal proteins were initially found as cell cycle mutants or have been defined by their homology to cell cycle proteins, for example, the human splicing factor CDC5-like protein (Ajuh et al. 2000). This possible link to the cell cycle may be supported here by the presence of cyclin A1 and K in our spliceosomal preparations. It will be interesting to determine whether either of these factors can act on substrates associated with the spliceosome.

Prospects

We have shown here that the use of enhanced, state-of-the-art proteomic methods facilitate a more detailed characterization of the human spliceosome than was previously possible, as it incorporates both high sensitivity and rapid analysis. This opens up the prospect of detailed proteomic studies addressing the dynamics of the spliceosome, for example, in regulation and in differential splicing, particularly if methods for direct quantification of the proteins can also be used.

Bearing in mind that some of the splicing factors were identified with only one peptide and that we used only two separate substrate RNAs and specific purification conditions, we do not expect the present study to have delivered a final list of spliceosomal proteins. It will be interesting to study alternative purification methods for isolating spliceosomes, including different washing stringencies and different pre-mRNA substrates, to identify even more splicing factors.

There is supporting evidence for functions in splicing for many of the novel factors that we have identified here (see Table 2). For the factors without any domains or sequence identity that links them to splicing, future localization and/or functional studies will be performed to address their putative role in splicing.

The regulatory proteins associated with the spliceosome also prompt multiple new experimental possibilities to study the regulation of splicing both in vivo and in vitro, showing the utility of large-scale proteomic studies as a launch pad for the design of functional studies in molecular cell biology.

METHODS

Purification of the Human Spliceosome

Human complexes were prepared essentially as described, but using less stringent wash conditions (Reed 1990; Calvio et al. 1995; Neubauer et al. 1998). Briefly, a mixture of spliceosomal complexes was assembled on biotinylated, radioactively labeled RNA. Two splicing substrates, adenovirus (AD1) and β-globin (AL4) transcripts, were used in separate experiments. The substrates were each biotin-labeled and incubated under splicing conditions with HeLa nuclear extracts in 1-mL reactions at 30°C for 1 h, forming both active spliceosomes and assembly intermediates. After incubation the samples were immediately loaded onto a 2.5 × 75-cm S-500 gel filtration column, and pooled fractions from the spliceosome peak were affinity-selected on streptavidin beads (Calvio et al. 1995). Proteins bound to the beads were washed three times in wash buffer (100 mM NaCl, 20 mM Tris-HCl at pH 7.5), then eluted in 0.3 mL of elution buffer (2% SDS, 20 mM Tris-HCl at pH 7.5, 20 mM DTT). Eluted proteins were precipitated with 1 mL of methanol together with 12 μg of slipper limpet glycogen carrier and finally resuspended in 50 μL of elution buffer. This procedure was repeated 12 times, and the resulting samples were pooled separately for each of the pre-mRNA substrates. Based on the staining with Coomassie blue, we estimate that each fraction contained ~6–10 μg of protein in total.

For the background control, nuclear extract was incubated without labeled RNA, followed by gel filtration as described above. Beads were mixed with the fractions that corresponded to the ones that contained labeled RNA in the above-described experiment. Beads were washed, and the bound material was eluted as above.

Sample Preparation for LC MS/MS

After purification, the volume of the pooled samples was reduced in vacuo; 15% glycerin, 100 mM dithiothreitol, and Bromophenol blue were added; and the samples were run on a 7.5% SDS-PAGE gel and stained with Coomassie blue. The lightly stained area containing the total, unseparated spliceosomal protein mixture was excised, then the proteins were in-gel reduced, alkylated, and digested using trypsin following described procedures (Shevchenko et al. 1996). Peptides were

extracted using first 70 µL of acetonitrile then 100 µL of 50% acetonitrile/2.5% acetic acid/0.01% heptafluoro butyric acid. Extracts were combined with the respective supernatants and filtered, and the volume was reduced in vacuo to ~25 µL.

LC MS/MS Analysis

Vydac 218MSB3 bulk material (3-µm prototype reversed phase material, a generous gift from Grace Vydac) was packed into pulled fused silica capillaries (PicoTip, New Objective) with a 100-µm ID and an 8-µm tip opening. Particles formed a self-assembled particle frit (SAP-frit) at the tapered end according to the principle of the stone arch bridge (Ishihama et al. 2002). Peptides were loaded using a sample loop. The following gradient was used: buffer A (5% acetic acid/0.02% heptafluoro butyric acid) to buffer B (80% acetonitrile/5% acetic acid/0.02% heptafluoro butyric acid), having the profile: B7% → B15% (0 → 10 min), B15% → B35% (10 → 70 min), B35% → B50% (70 → 80 min), B50% → B80% (80 → 85 min), B80% (90 min). The amount of material was estimated to be sufficient for three analyses using two initial LC MS/MS analyses of 10% of the sample. Subsequently, three identical LC separations were performed with the significant difference that the MS analysis software (`Analyst`, MDS-Sciex) was instructed to select only precursors in a certain mass range (m/z = 350–550, m/z = 550–750, or m/z = 750–1400, respectively) for fragmentation. This was matched by pulsed extraction of fragments, enhancing on m/z = 400, 600, or 800, respectively, as described previously (Andersen et al. 2002b). Tandem mass spectra were acquired for 1.5 sec, and fragmented peptides were excluded from sequencing for 120 sec. The background control was less complex and contained less material and was therefore only run with one LC MS/MS analysis, pulsed in the central region and with a precursor selection window of m/z = 350–1400. Scripts in Analyst created peak lists on the basis of the recorded fragmentation spectra.

Data Analysis

The combined peak lists of all eight runs contained the information on 7019 fragmentation spectra. This list was searched against the International Protein Index (IPI) database (http://www.ebi.ac.uk/IPI/IPIhelp.html) using `Mascot` (Matrix Science) on our in-house server. The most prominently identified peptides were then used to recalibrate the data, and the search was repeated to yield the initial list of identified proteins. All protein entries that were identified with at least three high-scoring peptide-query matches (individual `Mascot` scores above 32) and where the peptides were ranked as the top candidates were accepted as identified. All others were inspected manually as described in Results. In cases of ambiguity, the corresponding fragmentation spectrum was opened in `Inspector` (MDS-Proteomics) and manually interpreted to yield a peptide sequence tag (Mann and Wilm 1994), which was then searched against the IPI database using `PepSea` (MDS-Proteomics). The following proteins were regarded as contaminants on the basis of their occurrence in a blank purification (no biotinylated pre-mRNA added; data not shown): Von Ebner's gland protein (SWISS-PROT: P31025); Lysozyme C precursor (SWISS-PROT: P00695); dermcidin (SWISS-PROT: P81605); NY-REN-6 antigen (ENSP00000255069); trypsin (XP_094996); keratin 1 (SWISS-PROT: P04264); similar to keratin 1 (ENSP00000301445); keratin 2a (SWISS-PROT: P35508); similar to keratin 2a (ENSP00000252247); keratin 5 (ENSP00000252242); keratin, type II cytoskeletal 6F (SWISS-PROT: P48669); keratin 9 (ENSP00000246662); similar to keratin, type I cytoskeletal 10 (SWISS-PROT: P13645); keratin 10 (TREMBL: Q14664); keratin 14 (SWISS-PROT: P02533); keratin 16 (ENSP00000301653). Also, Huntington-interacting protein HYPA/FBP11 (ENSP00000288690) was considered to be a contaminant, because together with NY-REN-6 antigen it is a fragment of formin-binding protein 3 (NP_061255). Other proteins identified here were also classified as contaminants on the following basis: S100 calcium-binding protein A7 (SWISS-PROT: P31151) based on its high expression in keratinocytes (Rasmussen et al. 1992) and the two hypothetical proteins ENSP00000295258 and ENSP00000271816 based on their domain structure, which is very similar to calcium-binding protein A7.

The PAI is here defined as the number of sequenced peptides (fragmentation spectra assigned with significant score and as the top match to an individual identified protein) divided by the number of its calculated, observable peptides. Readily observable tryptic peptides were taken to be those in the mass range 800 to 2400 D. Fragmentation spectra matching the same peptide sequence but with different charge states, modification state, and containing missed cleavage sites were counted separately. For this reason, the index can be >1. The index is an expression describing not only the abundance of the protein in the sample but also its response to the measurement procedure. The latter is a complicated function of the efficiency of digestion, peptide solubility, extraction, ionization, and fragmentation for each protein and its peptides. In the future, more sophisticated versions of the PAI could take an increasing number of such factors into account.

ACKNOWLEDGMENTS

We thank our colleagues in the Protein Interaction Laboratory for fruitful discussions. Jens Andersen helped in devising the analysis strategy, and Leonard Foster developed scripts algorithms that we used here in data handling and especially in parsing the output of Mascot to retrieve the list of identified peptides. Carmen de Hoog is acknowledged for critical reading of the manuscript. Work in M.M.'s laboratory is supported by a generous fund of the Danish National Research Foundation to the Center of Experimental Bioinformatics. A.I.L. is a Wellcome Trust Principle Research Fellow and is funded by a Wellcome Trust Programme grant. J.R. is a Marie Curie Fellow.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J.C., Mann, M., and Lamond, A.I. 2000. Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J.* **19:** 6569–6581.

Ajuh, P., Sleeman, J., Chusainow, J., and Lamond, A.I. 2001. A direct interaction between the carboxyl-terminal region of CDC5L and the WD40 domain of PLRG1 is essential for pre-mRNA splicing. *J. Biol. Chem.* **276:** 42370–42381.

Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. 2002a. Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12:** 1–11.

Andersen, J.S., Rappsilber, J., Ishihama, Y., Wilkins, C., Nigg, E., and Mann, M. 2002b. "LC MS/MS of complex peptide mixtures using the pulsing feature of the hybrid QSTAR-pulsar quadrupole time-of-flight mass spectrometer to increase protein identification." Paper presented at The 50th ASMS Conference on Mass Spectrometry and Allied Topics (Orlando, FL, USA).

Aravind, L. and Koonin, E.V. 1999. G-patch: A new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem. Sci.* **24:** 342–344.

Bangs, P., Burke, B., Powers, C., Craig, R., Purohit, A., and Doxsey, S. 1998. Functional analysis of Tpr: Identification of nuclear pore complex association and nuclear localization domains and a role in mRNA export. *J. Cell Biol.* **143:** 1801–1812.

Bentley, D.L. 2002. The mRNA assembly line: Transcription and processing machines in the same factory. *Curr. Opin. Cell Biol.* **14:** 336–342.

Blencowe, B.J. and Ouzounis, C.A. 1999. The PWI motif: A new protein domain in splicing factors. *Trends Biochem. Sci.* **24:** 179–180.
Calvio, C., Neubauer, G., Mann, M., and Lamond, A.I. 1995. Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. *RNA* **1:** 724–733.
Chernushevich, I.V. 2000. Duty cycle improvement for a quadrupole-time-of-flight mass spectrometer and its use for precursor ion scans. *Eur. J. Mass. Spectrom.* **6:** 471.
Dix, I., Russell, C., Yehuda, S.B., Kupiec, M., and Beggs, J.D. 1999. The identification and characterization of a novel splicing protein, Isy1p, of *Saccharomyces cerevisiae*. *RNA* **5:** 360–368.
Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., and Burd, C.G. 1993. hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62:** 289–321.
Fox, A.H., Lam, Y.W., Leung, A.K., Lyon, C.E., Andersen, J., Mann, M., and Lamond, A. I. 2002. Paraspeckles. A novel nuclear domain. *Curr. Biol.* **12:** 13–25.
Frank, D. and Guthrie, C. 1992. An essential splicing factor, SLU7, mediates 3′ splice site choice in yeast. *Genes & Dev.* **6:** 2112–2124.
Frosst, P., Guan, T., Subauste, C., Hahn, K., and Gerace, L. 2002. Tpr is localized within the nuclear basket of the pore complex and has a role in nuclear protein export. *J. Cell Biol.* **156:** 617–630.
Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415:** 141–147.
Goldstrohm, A.C., Albrecht, T.R., Sune, C., Bedford, M.T., and Garcia-Blanco, M.A. 2001. The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* **21:** 7617–7628.
Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer, G., Colot, H.V., Mann, M., Seraphin, B., Rosbash, M., Luhrmann, R., et al. 1998. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA* **4:** 374–393.
Graveley, B.R. 2002. Sex, Agility, and the regulation of alternative splicing. *Cell* **109:** 409–412.
Griffin, T.J. and Aebersold, R. 2001. Advances in proteome analysis by mass spectrometry. *J. Biol. Chem.* **276:** 45497–45500.
Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K., and Izaurralde, E. 1998. TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol. Cell* **1:** 649–659.
Hastings, M.L. and Krainer, A.R. 2001. Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **13:** 302–309.
Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. 2002. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415:** 180–183.
Horowitz, D.S., Lee, E.J., Mabon, S.A., and Misteli, T. 2002. A cyclophilin functions in pre-mRNA splicing. *EMBO J.* **21:** 470–480.
Ishihama, Y., Rappsilber, J., Andersen, J.S., and Mann, M. 2002. "Microcolumns with self-assembled particle (SAP) frits for proteomics." Paper presented at the 15th International Symposium on Microscale Separations and Analysis (HPCE) (Stockholm, Sweden) *J Chromatogr* (in press).
Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I.W. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78:** 657–668.
Keller, W. and Minvielle-Sebastia, L. 1997. A comparison of mammalian and yeast pre-mRNA 3′-end processing. *Curr. Opin. Cell Biol.* **9:** 329–336.
Kramer, A. 1992. Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. *Mol. Cell. Biol.* **12:** 4545–4552.
Lallena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I., and Valcarcel, J. 2002. Splicing regulation at the second catalytic step by sex-lethal involves 3′ splice site recognition by SPF45. *Cell* **109:** 285–296.
Luking, A., Stahl, U., and Schmidt, U. 1998. The protein family of RNA helicases. *Crit. Rev. Biochem. Mol. Biol.* **33:** 259–296.
Mann, M. and Wilm, M. 1994. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66:** 4390–4399.
Minvielle-Sebastia, L. and Keller, W. 1999. mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr Opin Cell Biol* **11:** 352–357.
Mourelatos, Z., Abel, L., Yong, J., Kataoka, N., and Dreyfuss, G. 2001. SMN interacts with a novel family of hnRNP and spliceosomal proteins. *EMBO J.* **20:** 5443–5452.
Nayler, O., Stratling, W., Bourquin, J.P., Stagljar, I., Lindemann, L., Jasper, H., Hartmann, A.M., Fackelmayer, F.O., Ullrich, A., and Stamm, S. 1998. SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. *Nucleic Acids Res.* **26:** 3542–3549.
Neubauer, G., Gottschalk, A., Fabrizio, P., Séraphin, B., Lührmann, R., and Mann, M. 1997. Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci.* **94:** 385–390.
Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. 1998. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* **20:** 46–50.
Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20:** 3551–3567.
Rappsilber, J. and Mann, M. 2002. What does it mean to identify a protein in proteomics? *Trends Biochem. Sci.* **27:** 74–78.
Rappsilber, J., Ajuh, P., Lamond, A.I., and Mann, M. 2001. SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome. *J. Biol. Chem.* **276:** 31142–31150.
Rasmussen, H.H., van Damme, J., Puype, M., Gesser, B., Celis, J.E., and Vandekerckhove, J. 1992. Microsequences of 145 proteins recorded in the two-dimensional gel protein database of normal human epidermal keratinocytes. *Electrophoresis* **13:** 960–969.
Reed, R. 1990. Protein composition of mammalian spliceosomes assembled in vitro. *Proc. Natl. Acad. Sci.* **87:** 8031–8035.
Reed, R. and Hurt, E. 2002. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108:** 523–531.
Rosenberg, G.H., Alahari, S.K., and Kaufer, N.F. 1991. prp4 from *Schizosaccharomyces pombe*, a mutant deficient in pre-mRNA splicing isolated using genes containing artificial introns. *Mol. Gen. Genet.* **226:** 305–309.
Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B.T. 2000. The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J. Cell Biol.* **148:** 635–651.
Seraphin, B. 1995. Sm and Sm-like proteins belong to a large family: Identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* **14:** 2089–2098.
Shamoo, Y., Abdul-Manan, N., and Williams, K.R. 1995. Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res.* **23:** 725–728.
Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68:** 850–858.
Staley, J.P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92:** 315–326.
Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. 2002. Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell* **9:** 31–44.
Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R., et al. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417:** 304–308.
Tan, J.A., Hall, S.H., Hamil, K.G., Grossman, G., Petrusz, P., and French, F.S. 2002. Protein inhibitors of activated STAT resemble scaffold attachment factors and function as interacting nuclear receptor coregulators. *J. Biol. Chem.* **277:** 16993–17001.
Washburn, M.P., Wolters, D., and Yates III, J.R. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19:** 242–247.
Wigge, P.A., Jensen, O.N., Holmes, S., Soues, S., Mann, M., and Kilmartin, J.V. 1998. Analysis of the *Saccharomyces* spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.* **141:** 967–977.
Will, C.L. and Lührmann, R. 1997. Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9:** 320–328.
———. 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13:** 290–301.
Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L.,

Fotsis, T., and Mann, M. 1996. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379:** 466–469.
Zachariae, W., Shevchenko, A., Andrews, P.D., Ciosk, R., Galova, M., Stark, M.J., Mann, M., and Nasmyth, K. 1998. Mass spectrometric analysis of the anaphase-promoting complex from yeast: Identification of a subunit related to cullins. *Science* **279:** 1216–1219.
Zhang, Z. and Carmichael, G.G. 2001. The fate of dsRNA in the nucleus: A p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* **106:** 465–475.
Zhou, Z. and Reed, R. 1998. Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *EMBO J.* **17:** 2095–2106.
Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407:** 401–405.

WEB SITE REFERENCES

http://srs.embl-heidelberg.de:8000/;access to SWISSPROT.
http://www.ebi.ac.uk/IPI/IPIhelp.html; International Protein Index (IPI) database.
http://www.ensembl.org;access to ENSEMBL.
http://www.pil.sdu.dk; complete list of peptides.

Received May 29 2002; accepted in revised form June 14 2002.