TTCC); X13WF (TTATCCCGGGAT GGCACAACATGTAAGAA ATACTAA) and X13WR (AGTTCTCGAGGGATATT AACAACAA CCAGCTTG); C1F (CACACCCGGGATGGCAGCTAAAGCACACT ACG) and C1R (GTAACTCGAGTTAGCAGCATCCTTTTCCATTG TCC); and NS1F (TTATCCCGGGTCCATGCTCATACCCAAGCAG) and NS1R (GTAACTCGAG TTACCCGGGATAATCTGGAAC). XmaI and XhoI sites are underlined. For MS2-tagging, a synthetic gene, flanked by XmaI–XhoI sites and harboring four tandem repeats of the MS2 aptamer in the middle of the RabX13 intron was made (GeneScript, USA). The amplified PCR products or synthetic genes were digested with XmaI and XhoI and ligated into XmaI/XhoI-digested pEhExHA [37], to produce plasmids HA-U1A, HA-RabX13, HA-RabC1, HA-nNS1, HA-gRabX13, and HA-gRabX13-MS2, respectively. The AIG17 construct was described elsewhere [38]. The transformants expressing the above plasmids were established by liposome-mediated transfection of the wild-type HM1:IMSS Cl6 strain as previously described [39].

2.3. *Antibodies, western blotting, immunofluorescence, and recombinant GST–MS2 purification*

Protein lysates, protein analysis and immunolocalizations by confocal microscopy were carried out essentially as described [40]. For nuclear staining, 4′,6-diamino-2-phenylindole (DAPI) was included in the mounting medium. The plasmid coding for the recombinant GST–MS2 protein was a kind gift from Rei Yoshimoto and Mutsuhito Ohno. Purification of the GST–MS2 protein was carried out essentially as described [41].

2.4. *CLIP assays*

Amoeba transformants were exposed to UV light in a Stratalinker® UV Crosslinker 2400 for 30 min. Then immunoprecipitations with anti-HA agarose or MS2–GST–sepharose were carried out essentially as described [40] with modifications. For nuclear extracts, amoebas were lysed with 2% NP-40 in HEPES+ buffer (10 mM HEPES, 0.15 mM $MgCl_2$, 10 mM KCl, proteinase inhibitors E64 7 nM and Complete Mini 1 pill, added with 30 U/mL of RNase inhibitor). Nuclei were pelleted by centrifugation at 12,000 rcf for 10 min, at 4 °C, and washed two times in HEPES+ buffer. Nuclear pellet was suspended in Splicing-PEG Buffer (35 mM KCl, 4 mM $MgCl_2$, 2 mM ATP, 20 mM creatine phosphate, 1.5 mM DTT, and 50 μg/mL creatine kinase) and lysed by 5 freeze and thaw cycles, and lysates were clarified by centrifugation at 20,000 rcf for 20 min, at 4 °C. For unspecific protein bead-binding, extracts were precleared with 50 μL of protein G-sepharose, and anti-HA agarose or MS2–GST–sepharose was blocked with 0.5 mg of yeast tRNA. Anti-HA agarose or MS2–GST–sepharose immunoprecipitates were eluted with 20 mg/mL of HA peptide or elution buffer (20 mM HEPES–KOH pH 7.9, 100 mM KCl, 0.6% Sarkosyl, 10% Glycerol, 0.1% NP-40, 0.1 mM EDTA, and 1 mM DTT). Eluates were split into two. The first set of eluates were incubated with 1 U of RQ1 RNase-free DNase (Promega) in the appropriate buffer at 37 °C for 30 min, then deproteinized with 100 μg Proteinase K/0.1% SDS at 50 °C for 1 h. RNA was extracted using TRIzol reagent (Invitrogen) according to the manufacturer's instructions and samples were analyzed by RT-PCR. The second set of eluates were treated with 5 μg RNase A and RQ1 RNase-free DNase as mentioned above. Enriched proteins were concentrated with 10K Microcon® centrifuge filters and analyzed by SDS-PAGE and MS/MS.

2.5. RT-PCR

The synthesis of cDNA was performed using the SuperScript III First Strand Synthesis System (Invitrogen) according to the manufacturer's instructions. U2 snRNA, U6 snRNA, Cdc2 and actin mRNA molecules were detected with their respective primer sets: U6f (GGATCCACTTCGGTGGAAAT) and U6r (CTT CTCGTATGAGCGTGTCATC); U2f (TAACAGATCTATCACCTTC TCGGCCTTTATG) and U2r (TAACAGATCTTGTTTCCATGCA CATCCTCG); Cdc2f (GCTGTATTACTTGAACTGAAACATCCT) and Cdc2r (TCTTCATCACA AAATTCAAATACTAAA); and Actf (GGGAGACGAAGAAGTTCAAGC) and Actr (TG GATGGGAATA CAGCTCTTG).

2.6. *Protein and MS/MS analyses*

Eluted proteins from three independent experiments were resolved by 4–20% SDS-PAGE. Gels were stained with the Silver Stain MS kit (Wacko). Duplicate lanes of each protein sample were concentrated by allowing the samples to run on the gel for up to approximately 1 cm by electrophoresis. Gels were stained, lanes were excised and peptides were analyzed at the W.M. Keck Biomedical Mass Spectrometry Laboratory at the University of Virginia. Briefly, the solution samples were transferred to a siliconized tube and washed in 200 μL 50% methanol. The gel pieces were dehydrated in acetonitrile, rehydrated in 30 μL of 10 mM dithiothreitol in 0.1 M ammonium bicarbonate and reduced at room temperature for 0.5 h. The DTT solution was removed and the sample alkylated in 30 μL 50 mM iodoacetamide in 0.1 M ammonium bicarbonate at room temperature for 0.5 h. The reagent was removed and the gel pieces were dehydrated in 100 μL acetonitrile. The acetonitrile was removed and the gel pieces were rehydrated in 100 μL 0.1 M ammonium bicarbonate. The pieces were dehydrated in 100 μL acetonitrile, the acetonitrile was removed and the pieces were completely dried by vacuum centrifugation. The gel pieces were rehydrated in 20 ng/μL trypsin in 50 mM ammonium bicarbonate on ice for 30 min. Any excess enzyme solution was removed and 20 μL 50 mM ammonium bicarbonate added. The sample was digested overnight at 37 °C and the peptides formed were extracted from the polyacrylamide in a 100 μL aliquot of 50% acetonitrile/5% formic acid. This extract was evaporated to 15 μL for MS analysis.

The LC–MS system consisted of a Thermo Electron Velos Orbitrap ETD mass spectrometer system with a Protana nanospray ion source interfaced to a self-packed 8 cm × 75 μm id Phenomenex Jupiter 10 μm C18 reversed-phase capillary column. 7 μL of the extract was injected and the peptides were eluted from the column by an acetonitrile/0.1 M acetic acid gradient at a flow rate of 0.5 μL/min over 1.2 h. The nanospray ion source was operated at 2.5 kV. The digest was analyzed using the rapid switching capability of the instrument acquiring a full scan mass

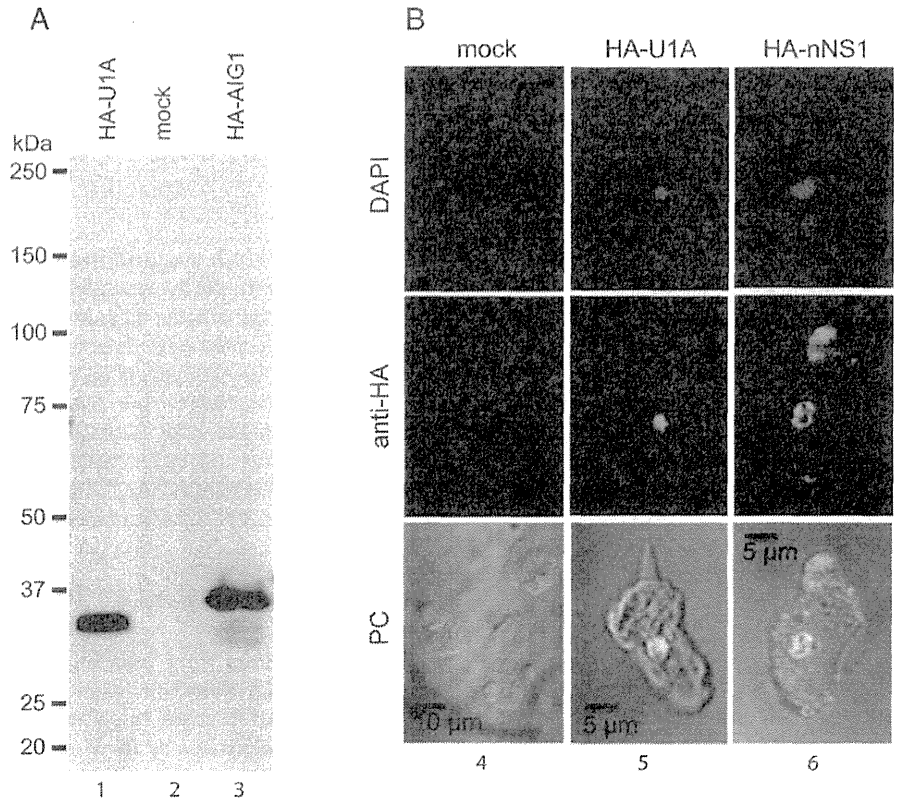


Fig. 1 – HA-U1A expression and nuclear localization in amoeba transformants. (A) Whole cell lysates of HA-U1A, mock and HA-AIG1 (positive control) *Entamoeba* transformants were resolved by 4–20% SDS-PAGE and blotted onto nitrocellulose. Tagged proteins were detected with anti-HA antibodies. The 31 and 34 kDa signals correspond to HA-U1A and HA-AIG1 fusion proteins, respectively. (B) Mock (lane 4), HA-U1A (lane 5), and HA-nNS1 (lanes 6) amoeba transformants were treated for immunofluorescence and observed under a confocal microscope. Green signals correspond to anti-HA Alexa antibodies and blue signals correspond to DAPI-contrasted nuclear DNA. PC, phase contrast images merged with fluorescent signals.

spectrum to determine peptide molecular weights followed by product ion spectra [20] to determine amino acid sequence in sequential scans. This mode of analysis produces approximately 30,000 MS/MS spectra of ions ranging in abundance over several orders of magnitude. Not all MS/MS spectra are derived from peptides.

Trans-Proteomic Pipeline 4.7.0 on MASSyPup was used for protein identification and validation, using Comet as search engine. A concatenated target-decoy database was constructed, using the NCBI protein entries for *E. histolytica* (version 12 June 2014) (ProteomeXchange accession: PXD001080). Results were validated by PeptideProphet/ProteinProphet, using decoy hits to pin down the negative distribution. The identification of the protein was considered significant when at least two non-overlapping peptides of a protein were detected with the probability score of .95 and .99%. Additional protein identification and annotations were carried out with Scaffold 4 (starting with 2110 HA-IP proteins) using AmoebaDB, version 1.7 (http://amoebadb.org/amoeba/). Individual predicted protein sequences were manually analyzed by BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/) against the non-redundant database at NCBI.

3. Results

To search for pre-mRNA processing components that could be also part of the splicing machinery the following experimental approach was used. First we HA-tag cloned the putative *Entamoeba* U1 snRNP component, splicing factor U1A. U1A can also be found in free form, as such interacts with the auxiliary splicing factor PSF (PTB-associated splicing factor), a component of the polypyrimidine-tract binding protein complex, involved in the later stages of pre-mRNA processing [42]. Therefore CLIP of HA-tagged U1A would increase the chances of detecting E complex components as well as B complex and C complex factors. To circumvent mRNA degradation by the numerous *Entamoeba* nucleases [43] during nuclear fractionation, and to ensure detection of RNA processing proteins bound to the RNA for MS/MS analysis, CLIP assays were carried out from nuclear extracts obtained from UV cross-linked amoebae.

Using HA antibodies, HA-U1A expression and intracellular localization in amoeba transfectants were analyzed by western blot (Fig. 1A) and confocal microscopy (Fig. 1B). Whereas the 31-kDa HA-U1A fusion proteins were expressed as well as the

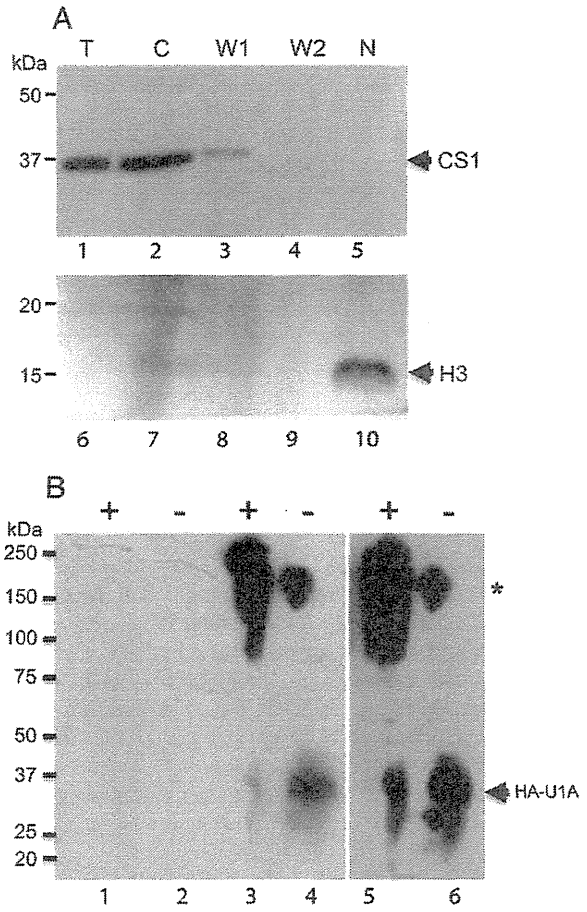


Fig. 2 – Cellular fractionation assessment and UV cross-linking of HA-U1A in enriched nuclear fractions. (A) Protein extracts from total (T, lanes 1 and 6), cytoplasmic (C, lanes 2 and 7), first and second wash (W1, lanes 3 and 8; W2, lanes 4 and 9, respectively) and enriched nuclei (N, lanes 5 and 10) fractions were blotted onto nitrocellulose and probed against the cysteine synthase 1 (CS1) cytoplasmic protein (lanes 1–5) or Histone 3 (lanes 6–10). (B) CLIP (+) or no UV treated (–) nuclear enriched protein extracts of mock (lanes 1 and 2) and HA-U1A (lanes 3 and 4) amoeba transformants were probed with anti-HA antibodies. After development, blots were exposed for 10 s or 1 min (lanes 5 and 6) to discriminate free HA-U1A signals after UV cross-linking, and artifact signals (asterisk).

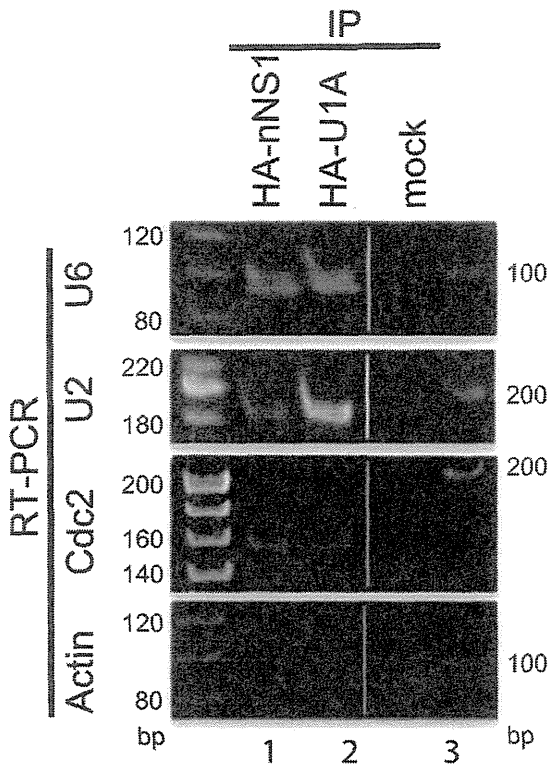


Fig. 3 – HA-U1A interacts with components of the splicing machinery and mRNP particles in vivo. CLIP assays were carried out with HA-nNS1 (lane 1), HA-U1A (lane 2) and mock (lane 3) amoeba transformants. Total RNA was isolated form the immunoprecipitates and was used as template to amplify by RT-PCR the U6 snRNA, the U2 snRNA, and the Cdc2 (intron containing), and Actin (intronless) mRNAs.

cytoplasmic positive control HA-AIG1 proteins of 34 kDa (lanes 1 and 3, respectively), no signals were observed in the extracts from empty vector (mock) transfectants (lane 4). Correspondingly, HA signals (green channel) were detected within the amoebae nuclei of HA-U1A and HA-nNS1 transfectants (lanes 5 and 6, respectively), colocalizing with the DAPI-stained DNA signals (blue channel). Since the N-terminus of NS1 interacts with U2–U6 snRNA dimers inhibiting pre-mRNA splicing [44], HA-nNS1 nuclear localization was expected, although HA-nNS1 was expressed in the cytoplasm as well. Only blue signals were detected in the mock controls (lane 4).

To validate our experimental approach, nuclei enrichment/purification efficiency and HA-U1A cross-reactivity and exposure/availability after UV cross-linking were monitored. Whereas the cytosolic protein cysteine synthase 1 (CS1) was detected in the total, cytosolic and first wash fractions (Fig. 2A, lanes 1–3, respectively), Histone 3 was detected in the nuclear fraction only (lane 5), indicating that no significant cross-contamination occurred during nuclear enrichment/purification. In addition, HA-U1A signals shifted upwards after UV treatment (Fig. 2B, lanes 3 and 5). This suggested HA-U1A–RNA complex formation, RNA-mediated HA-U1A protein–protein association and HA-U1A epitope exposure and availability during CLIP (UV cross-linking and HA-IP) procedures. Longer exposures showed enhanced HA-U1A cross-linked and input signals without significant increase of artifact signals (compare lanes 3 and 4 with lanes 5 and 6).

Next we verified that splicing RNA components and intron-containing pre-mRNAs were associated with CLIP products. To this end, RNA was purified from CLIP eluates of HA-nNS1, HA-U1A and mock-amoeba transfectants. These RNAs were used as templates to amplify by RT-PCR the two components of the spliceosome catalytic core U6 snRNA and U2 snRNA. To explore HA-U1A association with the mRNA

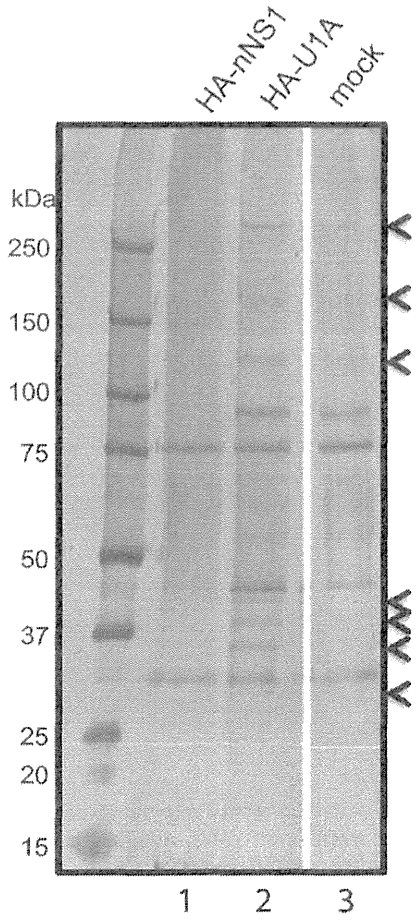


Fig. 4 – HA-U1A immunoprecipitated proteins. CLIP assays were carried out as in Fig. 3. Samples of the immunoprecipitates were loaded onto a 4–20% SDS-PAGE gel and silver-stained. Bands of 280, 200, 125, 48, 40 35, and 28 kDa appearing in the HA-U1A immunoprecipitates (lane 2), above the HA-nNS1 (lane 1) and mock (lane 3) background are indicated by arrowheads.

processing apparatus, the intron containing (CdcB) and intron-less (actin) mRNAs were amplified also. As expected, the HA-U1A and HA-nNS1 precipitates contained U2 snRNA, U6 snRNA as well as Cdc2 mRNA (Fig. 3, top panels of lanes 1 and 2), whereas no actin mRNA was detected in these eluates (lanes 1 and 2, bottom panel). As expected, none of the tested RNA molecules were found in the mock CLIP (Fig. 3, lane 3) and no amplifications were observed in the minus RT controls (Supplemental Fig. 1). Thus far, our results suggest that the *Entamoeba* splicing factor HA-U1A localizes in vivo in the nucleus and is able to interact with RNA molecules from the splicing machinery and with intron containing pre-mRNAs.

Parallel anti-HA CLIPs were analyzed by silver-stained SDS-PGE. The same protein pattern between HA-nNS1 and mock controls showed the light and heavy chains of the antibodies and two bands of 90 and 45 kDa (Fig. 4, lanes 1 and 3). However, in the HA-U1A CLIP seven distinct bands of 280, 200, 125, 48, 40, 35, and 28 kDa were observed above the background (Fig. 4, lane 2). Mock and HA-U1A lanes were analyzed by MS/MS and the resulting protein dataset (2110 proteins) was filtered using the minimal threshold indexes of the least represented high (U5-200K/Prp8) and low (SmD1) molecular weight splicing factors (number of peptides/molecular mass), and eliminating the proteins present in the mock controls (Supplemental Fig. 2). We confirmed and complemented our data with the proteins probed from MS2–sepharose–IP eluates of *Entamoeba* MS2 aptamer-tagged RabX13 intron transformants, compared to untagged gRabX13 and the intron-less RabC1 transformants, both with or without UV cross-linking. Thus 36 splicing proteins corresponding to 32 cognate splicing factors (Table 1) and 50 out of 100 hypothetical mRNP proteins (Table 2) were identified and categorized.

3.1 *Entamoeba pre-mRNA processing associated factors*

Apart from the HA-tagged U1A, the HA-U1A-exclusive selection included 36 *Entamoeba* splicing components, including 13 DExH/D-box RNA helicases required for splicing, 11 splicing-related helicases (Table 1), and 50 hypothetical proteins (Table 2). In keeping with their CLIP recovery, categories were: MS2-specific factors, intron-bound/unspecific (i.e. bound to untagged intron as well) factors, pre-mRNA binding/cotranscriptional (all RNAs tested), pre-mRNA binding (except tagged intron), and intron-less RNA-binding; and according to attributed functions, subcategories were: transcription/translation factors, kinases, membrane/trafficking, and unassigned factors.

Altogether, probed proteins included pre-mRNA processing factors and components of all stages of the splicing reactions: complex A, complex B, complex B* (activated), complex C, splicing step II factors, disassemble components, and factors of the exon junction complex/messenger ribonucleoprotein particles (EJC/mRNP); Fig. 5 compares the splicing factors found here with the previously reported human and yeast spliceosomes. As previously reported for in vitro and in vivo intron-labeled recruited spliceosomes [17–30], our combined approach with MS2-CLIP allowed us to probe 11 additional splicing factors, not detected with the HA-U1A CLIP (Table 1), which include components of the U2 snRNP (U2A′, SF3a120, SF3a60/Prp9, and Prp43), or of the U6 snRNP, U4/U6 · U5 tri-snRNP (LSm2, LSm5, CPR6, 65K/SAD1, and Prp38), or components involved in complex C formation (p68/DDX5, and Abstrakt). In addition, two splicing factors (CPR6, and the Prp19/CDC5L complex core, Prp19), and two splicing-involved (EhDExH13 and EhDEAD3), and one splicing-related (Chain A of Mtr4) DExH/D RNA helicases were recovered with this approach.

Taking advantage of their conservation to human splicing factors, antibodies against Prp8, TIAR and U2AF35 were used to probe cytosolic and nuclear fractions of *E. histolytica* (Fig. 6). Despite Prp8 signal being somewhat lower than expected (250 kDa), this factor was detected in the nuclear faction only. Both TIAR and U2AF35 were detected in the cytosol, but the latter is more abundant in the nuclear fractions. Their apparent molecular weight corresponded to those expected for *Entamoeba*, 35 and 29 kDa, respectively. These data partially validate these *E. histolytica* previously predicted splicing factors. Functional assays are being conducted for

Table 1 – Components of *Entamoeba histolytica* pre-mRNA splicing complexes.

Particle/complex/class	Protein name	Gene ID	Protein access	MW	Splicing factor	Number or peptides[a]							Max % coverage[b]	
						U1A	C1	C1+	X	X+	M	M+	HA	MS2
Sm proteins	Small nuclear ribonucleoprotein Sm D1	EHI_052090	183234262	12	SmD1		1						12	
	Small nuclear ribonucleoprotein	EHI_163710	67476256	13	SmD3	2	2		3				8	31
	Small nuclear ribonucleoprotein F	EHI_060400	67475017	22	SmF		1						7	
U1 snRNP	U1snRNP-specific protein	EHI_050780	67482465	27	U1A, HA-tagged	19	22						15	
	U1 small nuclear ribonucleoprotein subunit	EHI_153670	67482015	63	U1-70K		2						3	
Accessory	RNA-binding protein TIA-1	EHI_056660	67465872	35	TIA-1/TIAR	5	5		1		2		20	7
U2 snRNP	Leucine rich repeat	EHI_167290	67468662	96	U2A′			1		1		2		5
	Splicing factor	EHI_058680	67468502	28	SF3a120		1							3
	Splicing factor 3A subunit 3	EHI_038600	67483146	54	SF3a60/Prp9				1					2
	Splicing factor 3B subunit 1	EHI_049170	183235217	103	SF3b1	3	7		1	1	2	1	6	6
	Splicing factor 3b subunit 3	EHI_048160	67473926	125	SF3b3/DDX42	2		2	1	1	1	1	2	3
U2-related	U2 snRNP auxiliary factor large subunit	EHI_098300	67475980	84	U2AF65	1		1		1	2	3	1	2
	U2 snRNP auxiliary factor	EHI_192500	67476636	29	U2AF35						1	1		10
	EhDExH9	EHI_184530	183233848	110	Prp43			3		1		1		4
	EhDExH13	EHI_090040	67465050	140	Prp43	2		1		1			6	1
	EhDExH7	EHI_096230	67483238	77	Prp43	2	7	1	3		1	1	2	5
	EhDEAD3	EHI_013960	183231239	82	Prp5		3					1	6	3
U5 snRNP	Splicing factor Prp8	EHI_060350	67475030	266	220K/Prp8		2		1		1	2	2	2
	EhDExH10/U5 snRNP-specific 200kd	EHI_045170	67484208	206	200K/Brr2	2		3		4		1	1	2
	EhDExH1	EHI_131080	67480341	206	200K/Brr2	1	2	1					2	2
	U5 small nuclear ribonucleoprotein subunit	EHI_021380	183231926	110	116K/Snu 114		3	2	1	1			4	2
	Pre-mRNA splicing factor	EHI_093960	67484108	101	102K/Prp6	1				3		1	1	3
	EhDEAD4	EHI_021440	67469545	66	U5-100K (DDX3X)/Prp28	2		1					7	
LSm proteins	U6 snRNA-associated Sm-like protein LSm2	EHI_068580	183230265	7	LSm2				2					37
	LSM domain containing protein	EHI_076840	183232658	10	LSm5							1		27
U4/U6 snRNP	Peptidyl–prolyl cis–trans isomerase	EHI_020340	67478366	20	CPR6			2						14

	Peptidyl–prolyl cis–trans isomerase	EHI_125840	67482289	18	CPR6		2			1				8	8
U4/U6.U5 tri-snRNP	Ubiquitin carboxyl-terminal hydrolase domain containing protein	EHI_152110	67484272	52	65K/SAD1			2			1				5
	PRP38 family protein	EHI_000490	67474026	23	Prp38								1		15
Prp19/CDC5L complex	WD domain containing protein	EHI_130870	67478341	52	Prp19	3	5		1					5	6
Prp19/CDC5L complex-related	Regulator of nonsense transcripts	EHI_193520	67472499	108	KIAA0560 (fSAP164)		3							4	
	EhDEAD20	EHI_096390	67483276	62	p68 (DDX5)			1		2					6
Complex B/B[a]	EhDExH4	EHI_033720	67469329	87	Prp2	1	3		1			1		5	2
Complex C	EhDExH8	EHI_077640	67466830	105	Prp22		1			1		1		3	3
	EhDEAD1	EHI_175030	67475258	66	Abstrakt			1		1			1		5
Step II factors	EhDExH5	EHI_122790	67477533	98	Prp16	2	1			1	1		3	3	8
EJC/mRNP	EhDEAD18	EHI_151600	67480889	47	Sub2p/UAP56	2		4	1		2	2	1	4	15
Helicases with other splicing-related functions															
MS2-specific/splicing related	Rad3p DNA repair helicase	EHI_132410	67466685	90	Transcription/ translation	2	6					2		2	5
	Ssl2p-like DNA repair helicase	EHI_077260	67479133	69		2	4				2			4	4
Intron-binding, unspecific	Ssl2p DNA repair helicase	EHI_088430	67467062	75		1							1	5	2
Intron-binding, unspecific	Rvb1p-like DNA helicase	EHI_040360	67471882	48	Chromatin remodeling		2	1						7	4
	Sgs1p recQ family DNA helicase	EHI_023090	67475629	59		1					1			3	3
	Sgs1p recQ family helicase	EHI_028890	67469885	138		1		1	1					1	2
	isw2p helicase	EHI_012470	67479899	98		4								2	
	isw2p helicase	EHI_044890	67483974	112			2	1						3	1
Intron-binding, unspecific	Chain A structure of Mtr4	EHI_125170	67464927	123	Poly-adenylaton/ nuclear exosome complexes	1		1	1	1		1		2	2
Intronless RNA-binding	EhDExH12	EHI_134610	67472639	111		3			1					3	1
Nonsense-mediated mRNA decay and rRNA processing / biogenesis	EhDEAD6	EHI_145050	67463088	72	Intronless-binding		1		1					3	1
	EhDEAD9	EHI_151190	183229616	58	Pre-mRNA-binding/ cotranscriptional	2			1			2		7	5

[a] Number of peptides are listed for each sample: HA-IP (HA) using nuclear extracts from mock and HA-U1A (U1A) amoeba transformants, and MS2-IP (MS2) using nuclear extracts from untreated or UV crosslinked (+) HA-RabC1 (C1), HA-gRabX13 (X) and HA-gRabX13-MS2 (M) amoeba transformants.

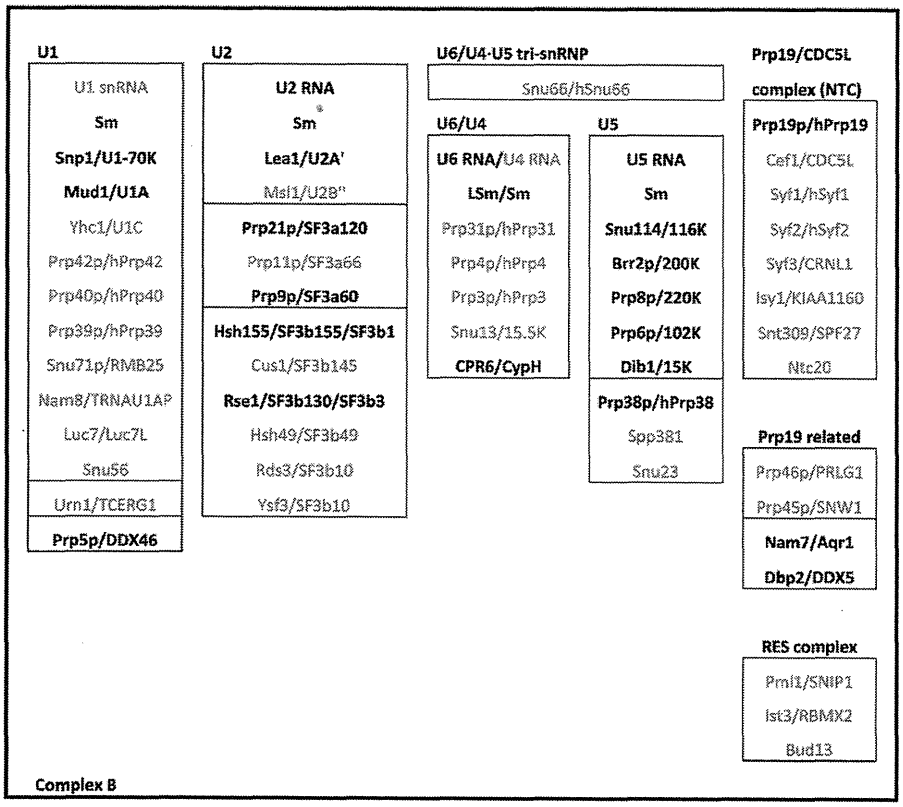
[b] Percent of maximal coverage obtained for each protein in the different IP assays.

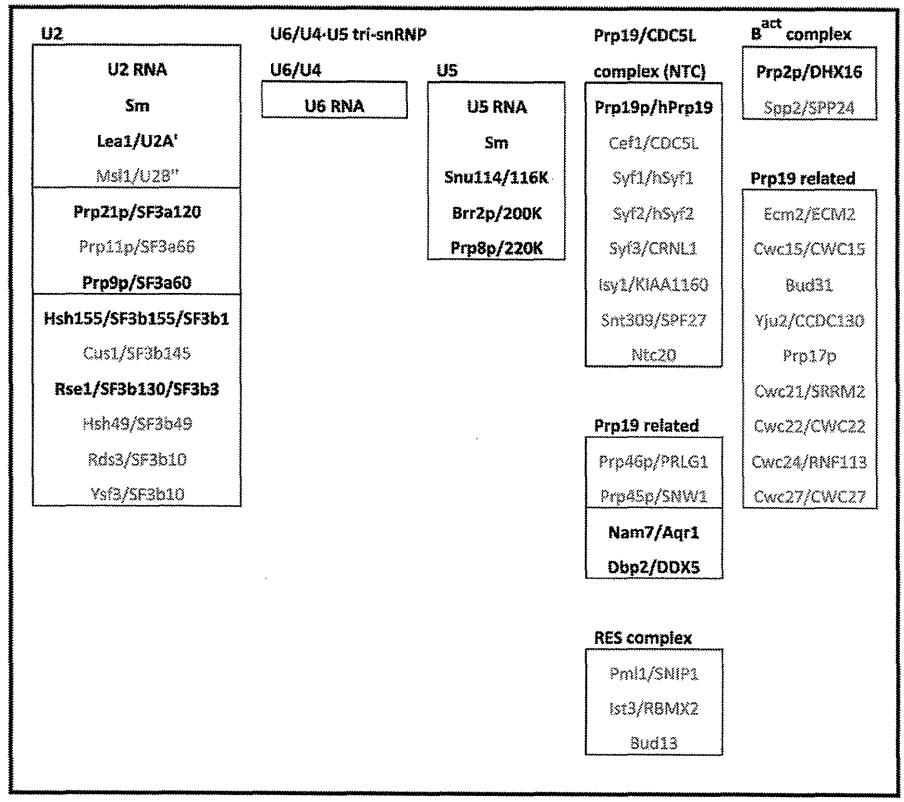
Table 2 – High ranked proteins selected from MS/MS analysis of HA-U1A and gRabX13-MS2 IPP. Data are presented as in . For hypothetical proteins (H labels on the first column) or for proteins identified in other organisms, BLAST scores and accession identifiers are also presented.

					Number of peptides [a]							Max % coverage [b]		BLAST scores				
IP/function	Protein	Gene ID	Prot access	MW	U1A	C1	C1+	X	X+	M	M+	HA	MS2	Total score	Query cover	E value	Ident	Accession
MS2-specific																		
	EhSec7 domain	EHI_048270	67473904	196	3	2				1	1	4	1					
	Ehzinc finger protein	EHI_134790	67472676	59	2	1					2	12	6					
	EhCXXC-rich protein	EHI_129490	183235170	51	1	1				2	2	4	7					
H	KIAA0216, myosin XVIIIA	EHI_098010	67470175	84	2	3				1		6	3	182	66%	6.0E−18	11%	BAA13206.2
H	Transmembrane protein, putative	EHI_180940	67472461	78		2					1	5	5	70.5	25%	2.0E−09	30%	XP_002908461.1
Intron-binding/unspecific																		
	EhRING zinc finger protein	EHI_023310	67480583	71	4	2		1	2			14	7					
H	Similar to zinc finger protein 650	EHI_114110	67462946	151	1	2		1				3	2	81.6	36%	2.0E−12	22%	XP_969420.2
H	Uncharacterized protein LOC101882025	EHI_107540	67463446	94	1	2		1	1			3	3	304	56%	3.0E−88	36%	XP_005164424.1
	EhProtein kinase domain containing	EHI_059040	183235228	154	11	10		2			2	4	4					
	EhProtein kinase domain containing	EHI_062090	67468296	156	10	7			6	2		3	3					
	EhProtein tyrosine kinase domain-containing	EHI_180150	183232515	135	5	3		1	1	4		6	10					
	EhProtein kinase domain containing	EHI_064610	183231977	119	3	2			2	1		7	1					
	EhProtein kinase	EHI_067070	67465327	134	2	1		2	1	2	2	2	4					
	EhPhosphatidylinositol-4,5-bisphosphate 3-kinase	EHI_009890	67483780	120	1	1		1		1		2	3					
H	Protein tyrosine kinase	EHI_065240	183233793	219	4	1			1	2	1	5	1	1911	25%	5.0E−31		
H	WD40 repeat-containing protein	EHI_118100	67481393	98	1	1			2	2	1	2	5	451	36%	2.0E−08	29%	WP_006508829.1
Pre-mRNA-binding/co-transcriptional Transcription/translation factors																		
	EhCXXC-rich protein	EHI_198900	67479193	47	8	12	3		3		2	36	16					
H	Zinc finger domain, LSD1 subclass family	EHI_061870	183235096	235	7	11	2	2	5	3	2	5	9	5	750	83%	3.0E−48	24%	XP_001015310.1
	EhDNA topoisomerase III	EHI_042880	67473767	95		1	1		2	2	1		2	5					
H	cca1p	EHI_177440	183231390	108	1	1	2	1			1		2	2	43.1	26%	0.67	25%	EJS43939.1
	Eh40S ribosomal protein S13	EHI_035460	183231337	17		1		1			4	6	9	29					
Kinases																		
	EhProtein kinase	EHI_092300	183230507	150	11	10		1	1	2		3	7	5					
	EhProtein kinase	EHI_175920	67465319	125	1	2		1	2		1		3	3					
	EhTyrosine kinase	EHI_152420	67483395	200	14	10	1			1	3	1	5	4					
	EhTyrosine kinase	EHI_138750	67468156	124	7	6	3	1	2	1	4		4	6					

	EhTyrosine kinase	EHI_127440	67473072	187	3			1		3	1	1		5	5					
	EhTyrosine kinase	EHI_030120	67476356	155	2	1		2			2			5	5					
	EhProtein tyrosine kinase domain-containing	EHI_101280	183234347	134	1	3	1	1	3	2	1			6	3					
	EhProtein tyrosine kinase domain-containing	EHI_064500	67476998	133	3		1				3			5	3					
	EhReceptor protein kinase	EHI_037140	67465629	129	2		1	1	1		3	3		3	5					
H	Protein kinase domain containing protein	EHI_128800	183236164	145	1	2		2	1	2		2		5	4	187	21%	4.0E−46	39%	XP_004338582.1
H	Putative AGC protein kinase family protein	EHI_055710	67472469	129		2	2					2		4	3	272	36%	4.0E−76	36%	AFW67835.1
Membrane/traffic																				
	EhCysteine surface protein	EHI_116260	183235261	103	11	10	6				2	3		11	11					
	EhCysteine surface protein	EHI_160750	183235324	64	4	1		2		2	1			13	6					
	EhGal/GalNAc lectin heavy subunit	EHI_046650	67469085	135	8	14	2	3			2	1		11	6					
H	Ras-like GTPase	EHI_090940	183234136	93	3	2	2	1	1	1	2			6	3	336	56%	1.0E−103	39%	XP_005164424.1
Unassigned																				
H	Predicted protein	EHI_064100	183235702	155	2			1				1		2	1	114	35%	2.0E−22	21%	XP_001772062.1
H	Hypothetical protein EMIHUDRAFT_240854	EHI_094080	67465773	97	1	1	1	1	1	1		2		3	3	467	19%	3.0E−21	40%	EOD21833.1
H	T17H7.18	EHI_159840	183231093	110		2		1			3			3	3	49.3	13%	0.008	28%	AAD32943.1
Pre-mRNA-binding																				
	EhCleavage stimulation factor	EHI_098370	183231153	65		2	1	1		2				2	6					
	EhNucleotide-binding protein	EHI_047750	67469203	36	3	1	1			1				8	8					
	EhUDP-glucose:glycoprotein glucosyltransferase	EHI_015280	67478161	150	2			1	1					2	2					
H	Piso0_004841	EHI_078270	67472246	86	1		1		1					1	5	54.3	33%	2.0E−04	27%	XP_004194353.1
Intronless RNA-binding																				
	EhActin	EHI_107290	67482879	13	10	8								32						
	EhGTP-binding protein	EHI_014370	67479391	42	7	1								7						
	Eh40S ribosomal protein S19	EHI_198740	67479617	18	1	1								7						
	Ehtyrosine kinase	EHI_017760	183232838	113	1									5						
	DNA-directed RNA polymerase II largest subunit	EHI_017570	183232826	25	1	2								14		36.6	46%	8.9	24%	XP_003072524.1
H	Ehzinc finger domain containing protein	EHI_154180	67468884	117		1		1						2	3					
H	NF-X1 finger transcription factor	EHI_169070	183232471	69		3		1						3	3	341	48%	5.0E−43	36%	EFY87775.1

H: BLAST results are shown for the annotated hypothetical proteins.
IP/function: Functions attributed according to the MS2-IP sample(s) where proteins were found (tagged intron, intron-containing and intron-less mRNA).
[a] Number of peptides are listed for each sample: HA-IP (HA) using nuclear extracts from mock and HA-U1A (U1A) amoeba transformants, and MS2-IP (MS2) using nuclear extracts from untreated or UV crosslinked (+) HA-RabC1 (C1), HA-gRabX13 (X) and HA-gRabX13-MS2 (M) amoeba transformants.
[b] Percent of maximal coverage obtained for each protein in the different IP assays.


U1
U1 snRNA
Sm
Snp1/U1-70K
Mud1/U1A
Yhc1/U1C
Prp42p/hPrp42
Prp40p/hPrp40
Prp39p/hPrp39
Snu71p/RMB25
Nam8/TRNAU1AP
Luc7/Luc7L
Snu56
Urn1/TCERG1
Prp5p/DDX46
U2
U2 RNA
Sm
Lea1/U2A'
Msl1/U2B''
Prp21p/SF3a120
Prp11p/SF3a66
Prp9p/SF3a60
Hsh155/SF3b155/SF3b1
Cus1/SF3b145
Rse1/SF3b130/SF3b3
Hsh49/SF3b49
Rds3/SF3b10
Ysf3/SF3b10
U6/U4-U5 tri-snRNP
Snu66/hSnu66
U6/U4
U6 RNA/U4 RNA
LSm/Sm
Prp31p/hPrp31
Prp4p/hPrp4
Prp3p/hPrp3
Snu13/15.5K
CPR6/CypH
U5
U5 RNA
Sm
Snu114/116K
Brr2p/200K
Prp8p/220K
Prp6p/102K
Dib1/15K
Prp38p/hPrp38
Spp381
Snu23
Prp19/CDC5L complex (NTC)
Prp19p/hPrp19
Cef1/CDC5L
Syf1/hSyf1
Syf2/hSyf2
Syf3/CRNL1
Isy1/KIAA1160
Snt309/SPF27
Ntc20
Prp19 related
Prp46p/PRLG1
Prp45p/SNW1
Nam7/Aqr1
Dbp2/DDX5
RES complex
Pml1/SNIP1
Ist3/RBMX2
Bud13
Complex B


U2
U2 RNA
Sm
Lea1/U2A'
Msl1/U2B''
Prp21p/SF3a120
Prp11p/SF3a66
Prp9p/SF3a60
Hsh155/SF3b155/SF3b1
Cus1/SF3b145
Rse1/SF3b130/SF3b3
Hsh49/SF3b49
Rds3/SF3b10
Ysf3/SF3b10
U6/U4-U5 tri-snRNP
U6/U4
U6 RNA
U5
U5 RNA
Sm
Snu114/116K
Brr2p/200K
Prp8p/220K
Prp19/CDC5L complex (NTC)
Prp19p/hPrp19
Cef1/CDC5L
Syf1/hSyf1
Syf2/hSyf2
Syf3/CRNL1
Isy1/KIAA1160
Snt309/SPF27
Ntc20
Prp19 related
Prp46p/PRLG1
Prp45p/SNW1
Nam7/Aqr1
Dbp2/DDX5
RES complex
Pml1/SNIP1
Ist3/RBMX2
Bud13
B^act complex
Prp2p/DHX16
Spp2/SPP24
Prp19 related
Ecm2/ECM2
Cwc15/CWC15
Bud31
Yju2/CCDC130
Prp17p
Cwc21/SRRM2
Cwc22/CWC22
Cwc24/RNF113
Cwc27/CWC27
Complex B^act

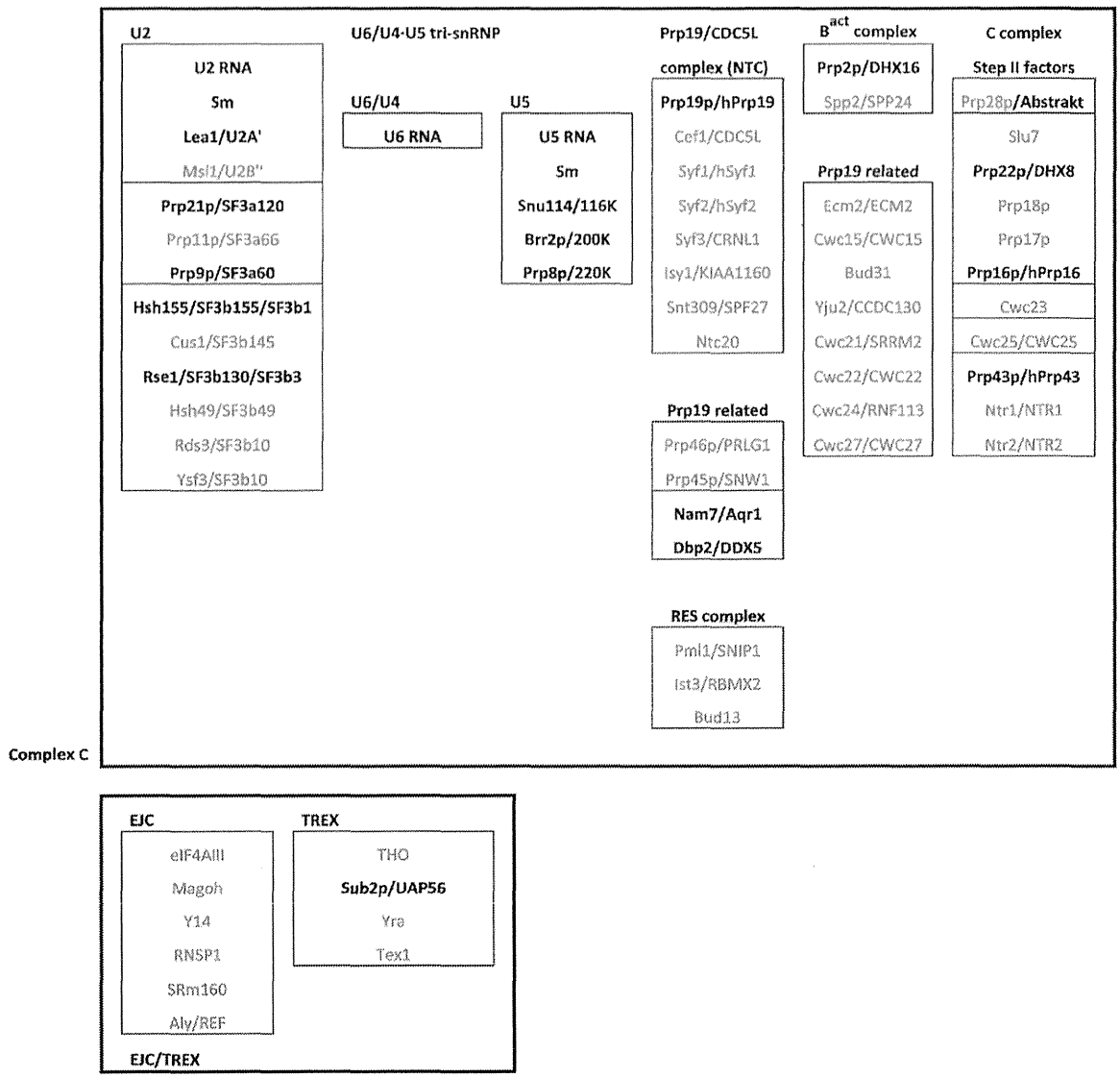


Fig. 5 – Splicing factors of *Entamoeba histolytica*. The compositional dynamics of the yeast and human spliceosomes (Refs. [29,74]) was used as template to compare the *Entamoeba* splicing factors here identified (in bold). Yeast protein factors appear first, followed by the human orthologs, separated by a slash. The *Entamoeba* snRNAs, previously identified are also shown.

thorough characterization of some of these splicing and mRNP factors.

4 Discussion

Consistent with their nuclear localization, our HA-CLIP assays showed that in vivo, the *E. histolytica* nuclear U1A splicing factor associates with the catalytic core snRNA components U2 snRNA and U6 snRNA, as much as the transfected nNS1 influenza virus protein, used here as a control. Possibly, the C-terminus of NS1 is also required for its nuclear-exclusive localization. U1A interacts with intron-containing mRNA molecules, suggesting that the methods used here allowed us to detect mRNP particles. This view is supported by the fact that additional protein signals were revealed in the HA-U1A CLIP compared with those observed with nNS1 or the mock background. Furthermore, additional bands above the background were obtained from MS2-CLIP assays also, with no difference between treatments (not shown). Both cDNA and genomic RabX13 constructs rendered a single 25 kDa protein signal, indicating that the gRabX13 clone product is properly spliced (not shown), and suggesting that the same stands for the gRabx13-MS2 clone products and that its splicing signals were sufficient to recruit components of the splicing machinery.

4.1. *Splicing/mRNP factors*

To gain insights into the *E. histolytica* spliceosome, the MS/MS analysis of the HA- and MS2-CLIP eluates allowed us to probe 100 splicing related proteins. Out of these 100 proteins, only 36