Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

# Structural mapping of spliceosomes by electron microscopy

Reinhard Lührmann and Holger Stark

Eukaryotic genes are transcribed as pre-mRNAs which are interrupted by noncoding introns. Selection and accurate removal of introns is an essential step in gene expression that is performed by a large and highly dynamic macromolecular machine, called spliceosome. A major challenge for structural studies of the spliceosome is represented by its large size as well as its dynamic nature. Electron microscopy is an important technique to study the overall shape and architecture of spliceosomes and their components, and to locate subunits and RNA parts. Recent advances in sample preparation of spliceosomes, technical improvements in EM instrumentation, powerful computer hardware, and new image analysis software will be instrumental to push structural studies of spliceosomes to a higher level of resolution.

**Address**
Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Corresponding author: Stark, Holger (hstark1@gwdg.de)

**Current Opinion in Structural Biology** 2009, **19**:96–102

This review comes from a themed issue on
Protein-nucleic acid interactions
Edited by Elena Conti and Daniela Rhodes

0959-440X/$ – see front matter


**DOI** 10.1016/j.sbi.2009.01.001

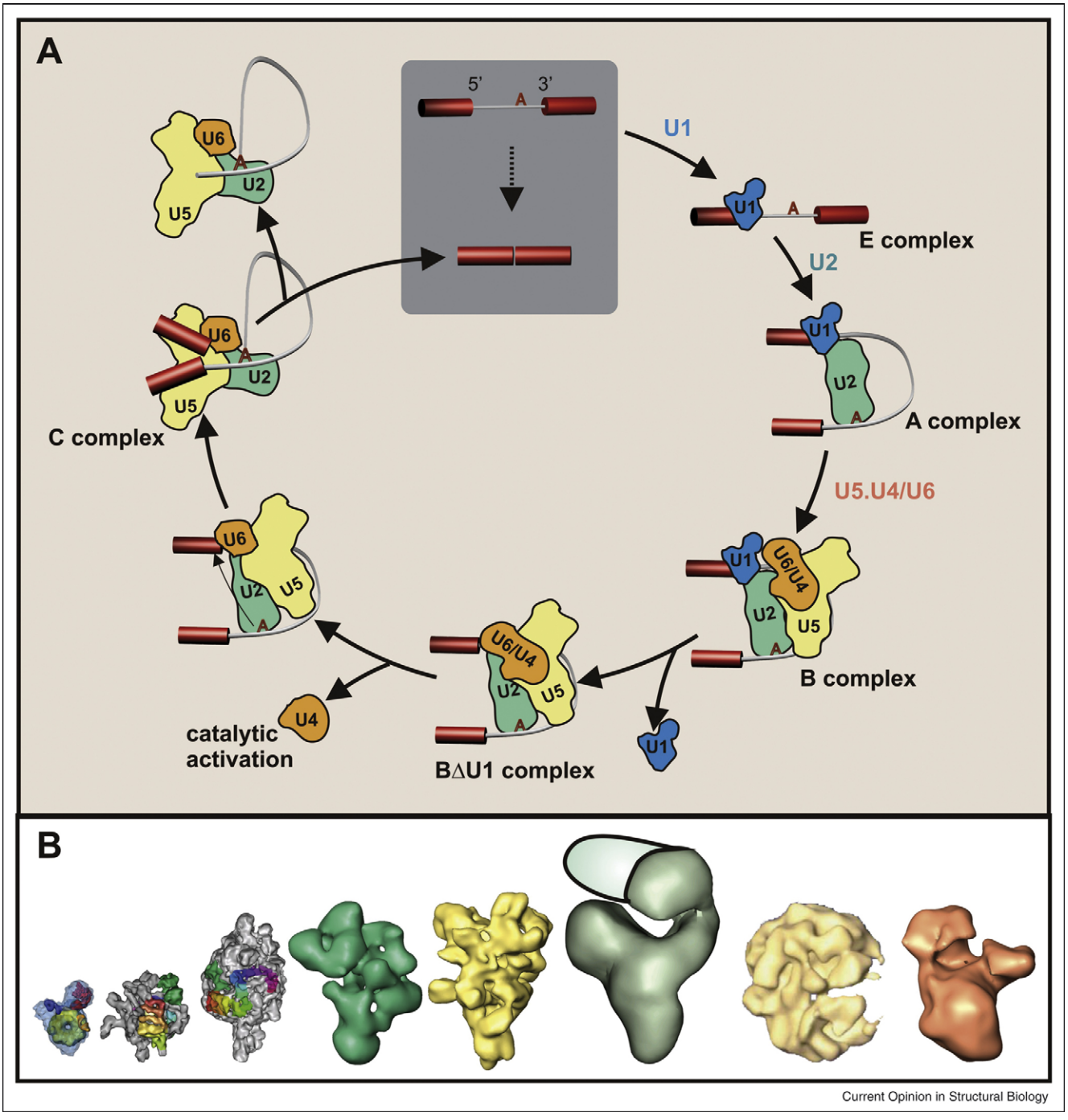
## Introduction

Eukaryotic genes are transcribed as pre-mRNAs which are interrupted by noncoding introns [1]. The post-transcriptional processing step of pre-mRNA splicing removes intronic sequences and ligates together the exonic (coding) sequences, resulting in mature mRNAs. The simple chemical reaction of a two-step transesterification is catalyzed within the spliceosome, a large, multisubunit ribonucleoprotein (RNP) complex [2]. Spliceosome assembly occurs directly onto pre-mRNA-to-be-processed, in a tightly regulated, ordered manner (Figure 1). The initial association of the U1 and U2 snRNP complexes with the pre-mRNA defines the splice sites and allows the further assembly of the U4/U6.U5 tri-snRNP, resulting in a fully assembled spliceosome (termed complex B). Major structural and compositional rearrangements are necessary for complex B to transform into a catalytically active spliceosome (complex C). In addition to changes in protein composition, this catalytic activation requires remarkable rearrangements within the RNA network of the spliceosome [3]. It is this dynamic nature that makes three-dimensional (3D) structure determination of the spliceosome and its components highly challenging [4,5]. Resolution by the current state-of-the-art cryo-EM studies of the snRNPs as well as of the spliceosome is mainly limited by the heterogeneity of the complexes within a sample, due to their dynamic properties as well as to damage to fragile macromolecules during specimen purification and preparation for cryo-EM. Since the resolution of current 3D maps of the spliceosome is too low to provide sufficient information to directly identify individual components, labeling subunits, and identifying the tags in EM provides a powerful way to localize the individual subunits. Different labeling techniques have already been applied to the spliceosomal components and will be discussed in this review. The almost unavoidable problem of heterogeneity of the complexes provides an interesting challenge to our current way of approaching single-particle cryo-EM, which relies on averaging a single 3D structure from the thousands of images taken from a sample. In the future, resolution can be increased by improving sample preparation techniques, EM hardware, and computational analysis with new algorithms. Additionally, we can partially circumvent the heterogeneity problem if we approach it from a different perspective, that is, to obtain information about as many of the variations of the complex (in its composition as well as conformation) as possible within a sample, which could ultimately give us insight into the dynamics of the specific complexes. We will discuss how this can be realistically approached with new technical advances that would allow computational separation of data and simultaneous 3D structure determination of heterogeneous samples.

## 3D structure modeling for the spliceosome

Much progress has been made in recent years in acquiring low (~14–40 Å) resolution EM 3D structures of the spliceosomal snRNPs (e.g. U1, U4/U6.U5, U5, and U4/U6) [6••,7] and spliceosomal complexes (e.g. complex A, BΔU1, and C) [8–10]. Additionally, a structure of the stable protein subcomplex of U2 snRNP, SF3b, was determined at ~10 Å resolution [11], and the structure of the U11/U12 di-snRNP of the minor spliceosome was determined at ~14 Å resolution [12] (Figure 1). Despite this progress, it has only been possible to build a complete 3D model, in which each protein and RNA component is localized, for the relatively small U1 snRNP, for which there was sufficient biochemical and structural data available to support modeling [7]. A similar level of structure interpretation is not yet possible for the larger snRNPs

**Figure 1**




**(A)** Schematic representation of pre-mRNA splicing and the spliceosome assembly pathway. Introns are excised from pre-mRNA by the spliceosome, which is assembled by the stepwise integration of U1, U2, and U4/U6.U5 snRNPs. **(B)** The currently available EM 3D structures of spliceosomes and spliceosomal components. From left to right: U1 snRNP [7], SF3b [11], U11/U12 di-snRNP [12], U5 snRNP [6••], U4/U6.U5 tri-snRNP [6••], BΔU1 spliceosome [9], 'native' spliceosome [22], and the C complex [10].

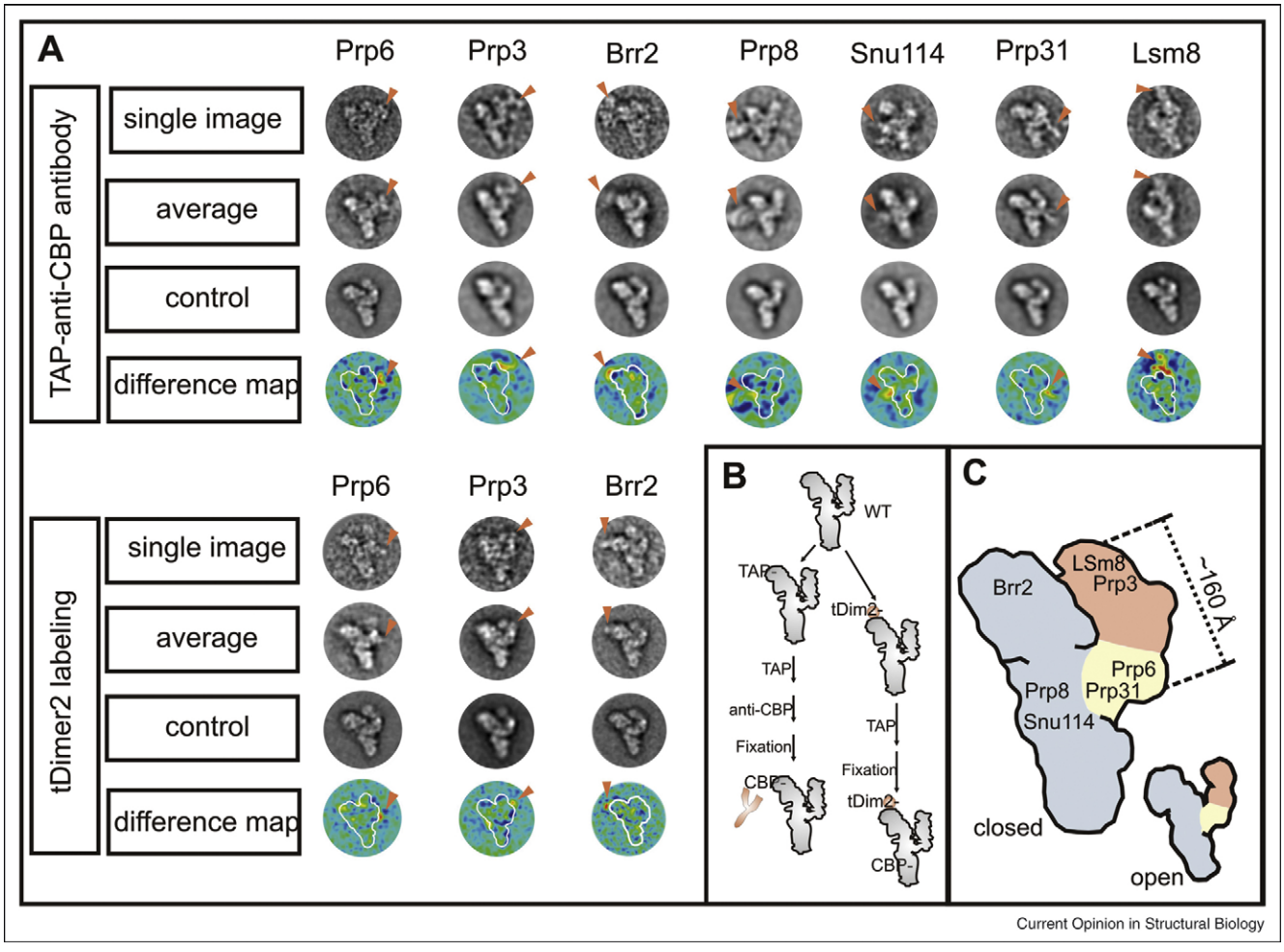
and spliceosomes, because of the increased level of structural complexity that increases with particle size, the limited resolution of the initial structures, and the lack of additional biochemical and structural data. Thus, modeling the 3D arrangement of the individual components, even at low to intermediate resolution levels, within the larger complexes will require higher resolution 3D structures and/or more biochemical and structural information about the localization of the individual components. Here, we will address ways to obtain complete

3D maps through higher resolution structures of dynamic macromolecular complexes, as well as electron microscopic labeling strategies to localize subunits either at the 2D or the 3D level.

## Labeling spliceosomal components — new strategies and current progress

Structure interpretation of cryo-EM 3D maps requires <10 Å resolution to allow secondary structure elements of proteins to be identified directly in the density map. This is beyond the resolution level that has so far been obtained for spliceosomes and snRNP complexes. Subunit labels can be extremely helpful, even in the initial low-resolution phase, to obtain a better understanding of the overall arrangement and subunit organization. Various labeling strategies are currently being tested in different laboratories. With the classical antibody labeling approach, the antibody binding sites can either be detected directly, or gold clusters can be attached to the antibody to increase its visibility in the EM images. A number of protein subunits in various snRNP complexes (of U1, U11/U12, U4/U6.U5, and SF3b) have been identified using this classical labeling technique [6••,7,12]. However, potential disadvantages of this technique are first, the limited binding efficiency of antibodies and second, the fact that antibody binding sometimes destroys the complexes at least partially, making it difficult to locate antibody binding sites accurately.

**Figure 2**




**(A)** EM localization of yeast tri-snRNP proteins by anti-CBP antibody-labeling (upper half) or by genetic tagging with the tDimer2 tag (bottom half). The first row of images shows an individual labeled tri-snRNP complex (single image); the second row, the class average to which the raw images contribute after image processing (average); the third row, a class average of an unlabeled tri-snRNP in the same orientation (control); and the fourth row, the difference maps between the labeled and unlabeled images (average − control = difference map). **(B)** Protein-mapping strategy of tri-snRNPs containing TAP or tDimer2-tagged protein. TAP and tDimer2-tagged tri-snRNP were affinity-purified via the TAP protocol. Particles containing a single TAP tag were incubated with anti-CBP antibody and then subjected to glycerol gradient centrifugation, whereas tDimer2-tagged particles were directly subjected to ultracentrifugation. Proteins were localized via EM by direct visualization of the antibody and by image analysis and class averaging of antibody-bound and/or tDimer2-tagged particles. **(C)** Cartoon model of the yeast tri-snRNP with shaded areas corresponding to U5 (gray), U4/U6 (orange) and the linker region (yellow). The tri-snRNP was found in both a closed and an open conformation, as indicated. (Reprinted with permission from [13••].)

Furthermore, antibodies themselves are quite large and flexible; this further limits the accuracy of the labeling approach. Nevertheless, antibody labeling is still a very powerful method for labeling studies at the level of 2D image analysis (Figure 2).

Antibody labeling can also be combined with genetic labeling studies. Tags such as TAP, HA, or Myc are genetic tags that are usually used for the purification of proteins and macromolecular complexes. Commercial antibodies are available against these tags which makes the antibody labeling approach more reliable because the antibody is not directed against individual subunits of the macromolecule. This strategy was successfully applied in case of the yeast tri-snRNP [13••] (Figure 2) as well as for the mapping of protein subunits in the anaphase promoting complex [14].

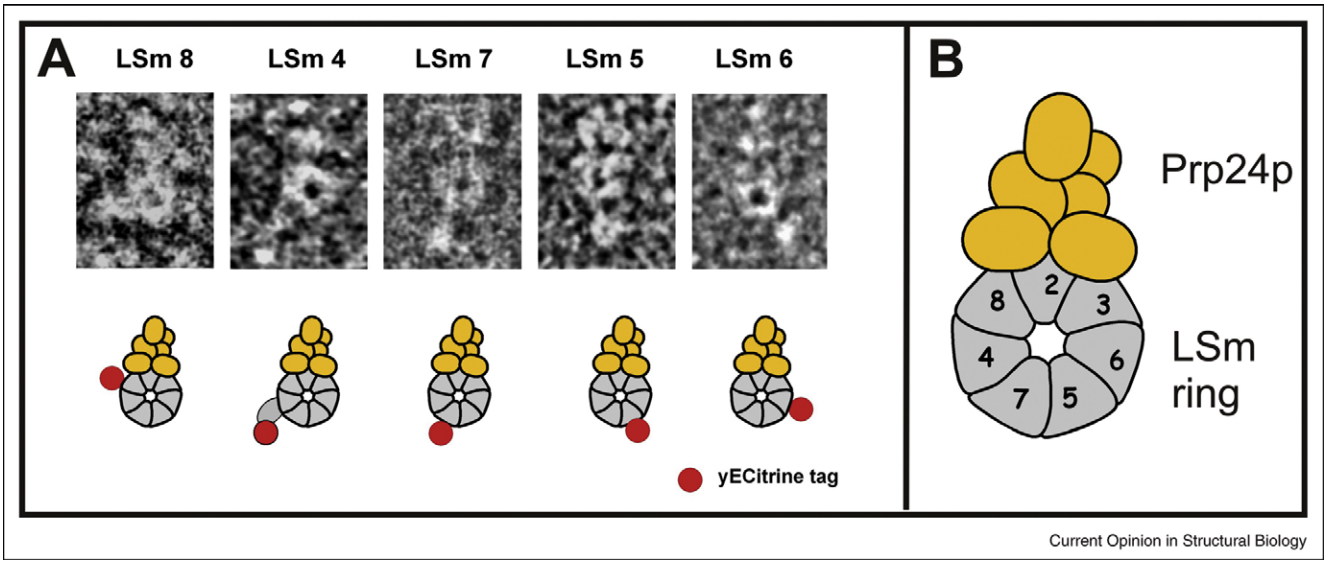
## Labeling yeast U4/U6.U5 tri-snRNP

We have recently started genetic labeling experiments in the yeast U4/U6.U5 tri-snRNP, for which a tDimer2 tag (consisting of two GFP folds) is genetically fused to protein subunits in a haploid yeast strain to assure that every copy of the protein carries the tag [13••]. There are two major advantages of genetic labeling: the genetic tag is smaller than an antibody, allowing higher labeling accuracy, and the tagging efficiency should be 100%. In actual experiments, the tagging efficiency is usually reduced to 70–80%, probably because of endoproteolytic degradation; however, this is still superior to normal antibody labeling efficiency. Tags that protrude from the molecule can be detected already at the level of 2D image analysis (Figure 2). However, owing to its small size with respect to the macromolecular complex, the tag can be masked within the structure at the level of 2D projection images if it is labeling a central protein within U4/U6.U5 tri-snRNP, simply because of significant overlap in projection of the specific protein with other parts of the complex. Once a 3D structure of the tagged complex is available, the tDimer2 tag can very accurately be localized in 3D maps (as demonstrated for the APC/C complex; manuscript in preparation). For the yeast U4/U6.U5 tri-snRNP studies, we corroborated the validity of both labeling approaches in a few cases by combining antibody labeling and genetic tDimer2 labeling for some of the subunits (Brr2, Prp3, Prp6) (Figure 2A and B). Genetic labeling of the U4/U6 proteins Prp3 and LSm8 localized the U4/U6 part within the U4/U6.U5 tri-snRNP; this was identified to be the 'arm' domain that can be found in two different positions (open and closed) with respect to the main body of the complex [13••] (Figure 2C).

## Labeling yeast U6 tri-snRNP

We used a similar genetic approach to label the individual LSm proteins (LSm4-8) within the yeast U6 snRNP, using the yECitrine tag (which contains only one GFP) [15••]. In this particular case, the smaller label was nonetheless still large enough to be seen directly in electron microscopic raw images (Figure 3A) because of the significantly smaller size of U6 snRNP, as compared to the U4/U6.U5 tri-snRNP, and indeed, the label for the individual proteins could be located to distinct positions within the LSm ring. Specifically, the labeling experiments revealed that the

**Figure 3**




EM analysis of yECitrine-labeled U6 snRNPs. Characteristic images of U6 snRNP particles, with a yECitrine tag on LSm8p, LSm4p, LSm7p, LSm5p, or LSm6p, are shown. **(A)** A single yECitrine label can clearly be detected at the level of raw images. **(B)** Schematic model of the yeast U6 snRNP depicting the Prp24 protein (orange) and the LSm ring (gray). The order of LSm proteins within the LSm ring structure was determined according to the labeling results shown in (A). (Reprinted with permission from [15••].)

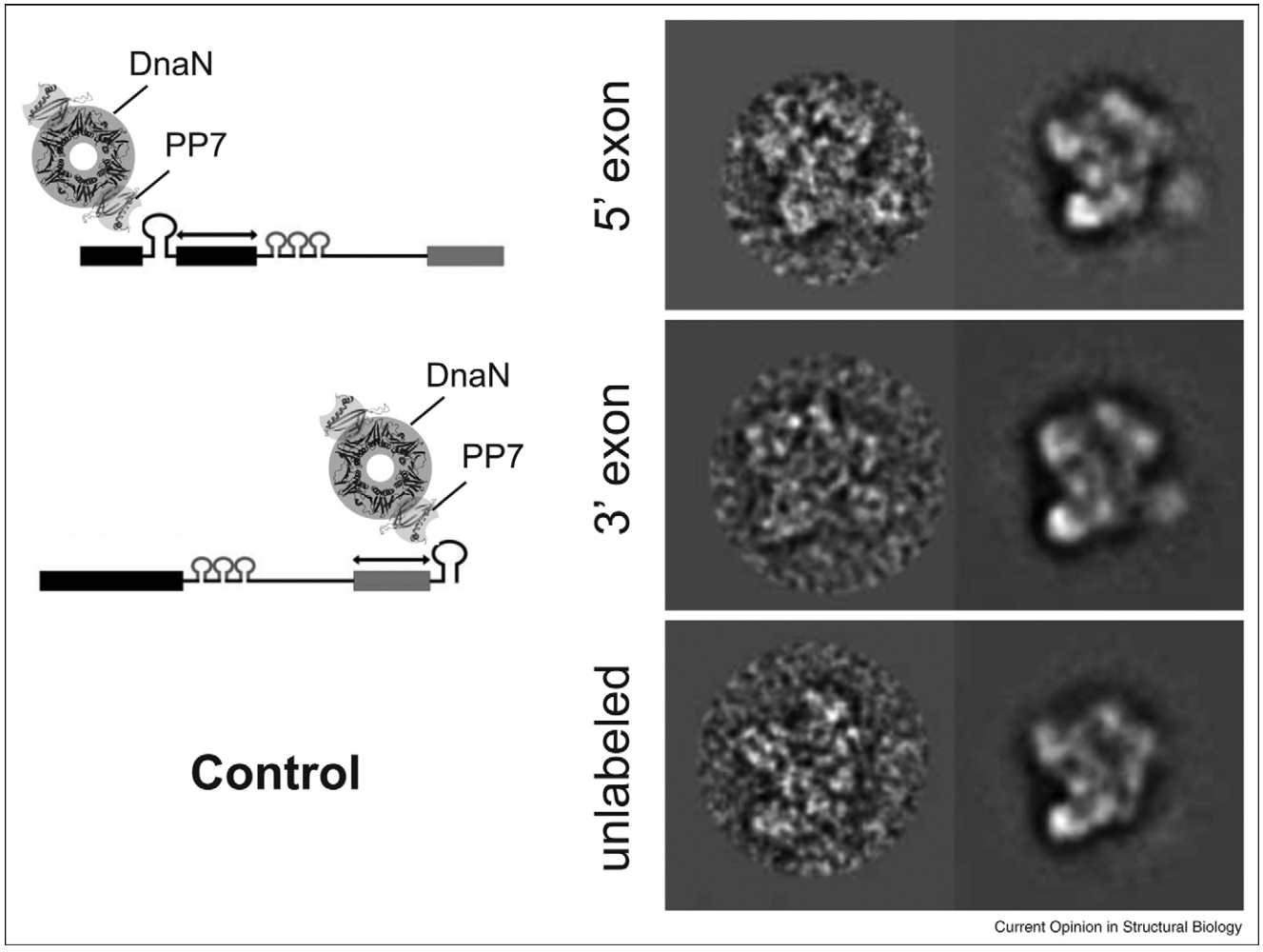
LSm proteins 2, 3 and 8 are localized close to the interface between the LSm ring and the Prp24 protein within the U6 snRNP structure (Figure 3B), in good agreement with biochemical crosslinking data that placed LSm2 near Prp24. Likewise, the genetic labeling allowed the positions of the LSm proteins 4-8 to be determined, and the results verified the proposed order of LSm proteins within the ring structure [16,17]. This proves that yECitrine can indeed be used as a small label to identify distinct positions of proteins in small macromolecular complexes with high accuracy.

## Other genetic labeling strategies

Another type of labeling was recently used to label the pre-mRNA in spliceosomal complex C, namely, that of the protein PP7 (coliphage coat protein) fused to the ring-shaped beta subunit of *Escherichia coli* DNA polymerase [18••]. The PP7 coat protein was bound to a 24-nucleotide RNA hairpin sequence that was introduced at the 5′-exons and 3′-exons of the pre-mRNA, and the ring-like protein subunits could then be located (Figure 4). Image analysis of spliceosomal complexes C labeled at both exons revealed a close proximity of the labeled sites within this complex. Since this new label is highly specific and can already be seen in raw images, it may serve as a tool to label RNA in other spliceosomal complexes as well as in RNP complexes in general.

A further very interesting labeling strategy uses an artificial metallo-fusion protein as a genetic tag [19•]. Like GFP, this is a low molecular weight label (~34 kDa) that cannot always be detected efficiently in a direct manner by electron microscopy and image analysis. However, the metallo-protein has the ability to bind divalent metal ions.

**Figure 4**




The Beta-PP7 label consists of the coliphage coat protein PP7 and the β-subunit of *E. coli* DNA polymerase III (DnaN). This label binds to a 24-nucleotide RNA hairpin sequence that has been inserted into the 5′-exon and 3′-exon of the pre-mRNA. The labeled spliceosomal C complexes reveal an extra density that can be attributed to the Beta-PP7 label at approximately the same position (lower right side of the particles), while there is no extra density in the unlabeled control complex. (Reprinted with permission from [18••].)

The visibility of the label can thus be increased by the ability of the metallo-protein to bind 250 $Cd^{2+}$ ions.

## Combining structural information for 3D modeling

As long as cryo-EM 3D reconstructions of the spliceosome are restricted to low and intermediate resolution, the immediate goal is to build a pseudo-atomic resolution hybrid model of the entire spliceosome. This could be done by combining information from labeling experiments, high-resolution X-ray and NMR structures of individual components or subcomplexes, and crosslinking and biochemical data. A similar approach has recently been taken to obtain computational modeling of the nuclear pore complex [20•,21]. However, for the spliceosome, there is currently not enough structural information available to even approach the 3D modeling level. Considering the complexity, size, and lack of symmetry of the spliceosome, successful modeling studies will certainly require large amounts of data (such as crosslinking, EM labeling) to be collected in the future.

## How to determine high-resolution structures of spliceosomes?

The alternative to obtaining a structural map of the spliceosome by the modeling approach is to directly determine a high-resolution structure of the spliceosome. There are currently several technical limitations to achieve this. Obtaining the amount of high-quality spliceosomes necessary for EM is already technically demanding [8,9,22,23,24•], yet this type of sample preparation would not suffice for X-ray crystallography, which would require an increased concentration (of least two-orders of magnitude) of extremely pure, highly concentrated material. Considering the large degree of structural and conformational variations within a single spliceosomal preparation, there is therefore a low probability of obtaining high-resolution spliceosome crystals in the foreseeable future. However, the sample quality problems that affect X-ray crystallographers likewise affect electron microscopists when trying to determine high-resolution structures. Not surprisingly, therefore, the highest resolution structures determined so far by single-particle cryo-EM were for highly stable and symmetric icosahedral viruses [25,26]. The lack of symmetry and the instability of the spliceosomal complexes thus make it of utmost importance to develop new purification strategies to obtain high-quality samples. Considerable progress has been made in the last few years to improve both the quality of spliceosome purification under low-stringency conditions [24•] and the sample preparation for EM grids [27••].

Sample homogeneity can also be improved by treating the purified spliceosomes with increasing salt concentrations in order to identify a stable and more homogeneous spliceosomal core structure, as has recently been done for a 1 M salt-treated spliceosomal C complex [28••]. Other options are to study snRNPs and spliceosomes in other organisms that have much simpler protein compositions. In contrast to human spliceosomes, spliceosomes isolated from *Saccharomyces cerevisae* are considerably less heterogeneous because of the lack of SR and hnRNP proteins. Furthermore, it should be possible to make use of small molecule inhibitors and/or temperature-sensitive mutants to trap spliceosomes in more defined states of the assembly cycle, which should also improve the homogeneity of purified complexes and thus lead to improved biochemistry and labeling data.

A certain amount of heterogeneity can also be attributed to the electron microscopy sample preparation. A drawback of standard EM sample preparations, such as conventional negative staining, is that the process of binding complexes to carbon foil can negatively affect sample quality. For instance, we have observed that some snRNPs and spliceosomes are easily degraded when they come into contact with carbon support film. To improve EM sample preparation we have recently developed a new procedure, called GraFix [27••], that combines a glycerol gradient centrifugation with a mild chemical fixation before further processing of the samples for microscopy. This procedure greatly enhances the stability of spliceosomes and snRNPs. The mild chemical fixation preserves the structural integrity, which is also important in order to calculate a reliable initial 3D model of a macromolecular complex. GraFix can be used not only with unstained cryo-EM but also with conventional negative-stain EM, where it has the additional advantage of enhancing the image contrast.

Sample heterogeneity remains a limiting factor in any attempt to achieve a high-resolution structure of a spliceosome. Using the GraFix method minimizes the structural degradation during sample preparation that contributes to this heterogeneity. A more critical source of heterogeneity, however, comes from the dynamic nature of the spliceosomes; it is still not possible to isolate spliceosomes in one single, well-defined functional state. All spliceosome preparations imaged by electron microscopy therefore have to be considered to be mixtures of individual complexes in solution that are in a variety of functional states. Since cryo-EM is an in-solution method, the additional purification step that usually happens during crystal growth in X-ray crystallography does not occur. For high-resolution structure determination, therefore, an image dataset needs to be purified computationally in order to separate out molecules that either do not belong to the same 3D structure or that are damaged. We have developed a method that allows this separation to occur at low resolution (25–30 Å), on the basis of the classical 'random-conical-tilt' (RCT) approach [29]. This method includes an additional unsupervised alignment and a weighted averaging scheme on a large set of RCT 3D volumes, followed by multivariate statistical analysis,

which then allows the detection of various conformations of the macromolecule in three dimensions. This method was already successfully applied to determine the structural variations of the human U4/U6.U5 tri-snRNP complex [6••]. Scheres *et al*. have recently developed a maximum-likelihood multi-reference alignment method [30•,31]. Maximum-likelihood methods were introduced into the EM field a decade ago [32•] and have become increasingly popular during the last few years [30•,31,33–35]. Once a reliable initial 3D volume is available for a macromolecular complex, this method can be used to separate the dataset into a maximum of six different states in a nonsupervised manner. These approaches are still limited to a relatively low number of functional states because of practical limitations, such as computer speed and maximum computer memory available. Nonetheless, these new approaches successfully demonstrate that it is possible to separate particle images into subpopulations despite high noise levels in the raw images, making this a promising future strategy. The level of resolution that can ultimately be obtained by single-particle cryo-EM will strongly depend on how accurate these and future algorithms will be in terms of 'purification' of heterogeneous datasets.

An obvious consequence of the strategy of separating data into subpopulations is that considerably more images will be needed for image analysis, since only a small percentage of the entire dataset will contribute to any given 3D structure. Thus, rather than dealing with just several ten thousands of images, we expect the numbers of images in future structural studies to grow rapidly. Using up to several million particle images for a 'high-resolution' structure of a spliceosome is a realistic expectation. This huge amount of data causes considerable practical problems for data collection and processing within an acceptable time frame. Automated data acquisition by state-of-the-art electron microscopes, new developments in image processing software as well as high-performance computing will thus be instrumental for future high-resolution structure determination by single-particle cryo-EM.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest


1. Moore MJ, Query CC, Sharp PA: *The RNA World*. Cold Spring Harbor Lab; 1993:. pp. 303–357.
2. Will CL, Luhrmann R: **Curr Opin Cell Biol** 1997, **9**:320-328.
3. Will CL, Luhrmann R: *Spliceosome Structure and Function*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2006.
4. Jurica MS: **Curr Opin Struct Biol** 2008, **18**:315-320.
5. Stark H, Luhrmann R: **Annu Rev Biophys Biomol Struct** 2006, **35**:435-457.
6. •• Sander B, Golas MM, Makarov EM, Brahms H, Kastner B, Luhrmann R, Stark H: **Mol Cell** 2006, **24**:267-278.
7. Stark H, Dube P, Luhrmann R, Kastner B: **Nature** 2001, **409**:539-542.
8. Behzadnia N, Golas MM, Hartmuth K, Sander B, Kastner B, Deckert J, Dube P, Will CL, Urlaub H, Stark H, Luhrmann R: **EMBO J** 2007, **26**:1737-1748.
9. Boehringer D, Makarov EM, Sander B, Makarova OV, Kastner B, Luhrmann R, Stark H: **Nat Struct Mol Biol** 2004, **11**:463-468.
10. Jurica MS, Sousa D, Moore MJ, Grigorieff N: **Nat Struct Mol Biol** 2004, **11**:265-269.
11. Golas MM, Sander B, Will CL, Lührmann R, Stark H: **Science** 2003, **300**:980-984.
12. Golas MM, Sander B, Will CL, Lührmann R, Stark H: **Mol Cell** 2005, **17**:869-883.
13. •• Hacker I, Sander B, Golas MM, Wolf E, Karagoz E, Kastner B, Stark H, Fabrizio P, Luhrmann R: **Nat Struct Mol Biol** 2008, **15**:1206-1212.
14. Ohi MD, Feoktistova A, Ren L, Yip C, Cheng Y, Chen JS, Yoon HJ, Wall JS, Huang Z, Penczek PA *et al.*: **Mol Cell** 2007, **28**:871-885.
15. •• Karaduman R, Dube P, Stark H, Fabrizio P, Kastner B, Luhrmann R: **RNA** 2008, **14**:2528-2537.
16. Pannone BK, Kim SD, Noe DA, Wolin SL: **Genetics** 2001, **158**:187-196.
17. Zaric B, Chami M, Remigy H, Engel A, Ballmer-Hofer K, Winkler FK, Kambach C: **J Biol Chem** 2005, **280**:16066-16075.
18. •• Alcid EA, Jurica MS: **Nat Struct Mol Biol** 2008, **15**:213-215.
19. • Nishino Y, Yasunaga T, Miyazawa A: **J Electron Microsc (Tokyo)** 2007, **56**:93-101.
20. • Alber F, Forster F, Korkin D, Topf M, Sali A: **Annu Rev Biochem** 2008, **77**:443-477.
21. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J *et al.*: **Nature** 2007, **450**:695-701.
22. Azubel M, Wolf SG, Sperling J, Sperling R: **Mol Cell** 2004, **15**:833-839.
23. Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ: **RNA** 2002, **8**:426-439.
24. • Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, Stark H, Urlaub H, Luhrmann R: **Mol Cell Biol** 2006, **26**:5528-5543.
25. Zhang X, Settembre E, Xu C, Dormitzer PR, Bellamy R *et al.*: **Proc Natl Acad Sci U S A** 2008, **105**:1867-1872.
26. Yu X, Jin L, Zhou ZH: **Nature** 2008, **453**:415-419.
27. •• Kastner B, Fischer N, Golas MM, Sander B, Dube P, Boehringer D, Hartmuth K, Deckert J, Hauer F, Wolf E *et al.*: **Nat Methods** 2008, **5**:53-55.
28. •• Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R: **Nature** 2008, **452**:846-850.
29. Radermacher M: **J Electron Microsc Tech** 1988, **9**:359-394.
30. • Scheres SH, Gao H, Valle M, Herman GT, Eggermont PP *et al.*: **Nat Methods** 2007, **4**:27-29.
31. Scheres SH, Valle M, Carazo JM: **Bioinformatics** 2005, **21(Suppl 2)**:ii243-ii244.
32. • Sigworth FJ: **J Struct Biol** 1998, **122**:328-339.
33. Lee J, Doerschuk PC, Johnson JE: **IEEE Trans Image Process** 2007, **16**:2865-2878.
34. Scheres SH, Nunez-Ramirez R, Gomez-Llorente Y, San Martin C, Eggermont PP, Carazo JM: **Structure** 2007, **15**:1167-1177.
35. Zeng X, Stahlberg H, Grigorieff N: **J Struct Biol** 2007, **160**:362-374.