A Sponsored Supplement to *Science*

Accelerating the path from structure to function through integrative structural biology solutions

Sponsored by

ThermoFisher SCIENTIFIC

Produced by the *Science*/AAAS Custom Publishing Office

Science

Accelerating the path from structure to function through integrative structural biology solutions

About the cover: Crystal structure of the 80S ribosome from the yeast *Saccharomyces cerevisiae*, which consists of four RNA chains (gray) and 79 different proteins (colored ribbons). Image: Sergey Melnikov, Nicolas Garreau de Loubresse, Adam Ben-Shem, Lasse Jenner, Gulnara Yusupova, Marat Yusupov/Institut de Génétique et de Biologie Moléculaire et Cellulaire, Université de Strasbourg.

This booklet was produced by the *Science*/AAAS Custom Publishing Office and sponsored by Thermo Fisher Scientific.

Editor: Jackie Oberst, Ph.D.; Sean Sanders, Ph.D.
Proofreader/Copyeditor: Bob French
Designer: Amy Hardcastle

ROGER GONCALVES, ASSOCIATE SALES DIRECTOR
Custom Publishing
Europe, Middle East, and India
rgoncalves@science-int.co.uk
+41-43-243-1358

Introductions

Overview

Research articles

White paper

Interview with Dr. Albert Heck

Technology feature

Not so elementary: Deciphering structure–function relationships

Recent advancements in technology have allowed the structures of macromolecules to be deciphered at greater and greater speeds.

Entry-level biochemistry class teaches that the three-dimensional structure of a protein defines not only its size and shape but also its function. But this relationship is anything but simple. Exactly how proteins send signals through their structures, a process known as allostery, is still to be determined. As such, designing drugs to regulate the functions of these proteins remains an arduous goal.

Integrative structural biologists—a diverse collection of scientists from various fields such as cell biology, protein engineering, and computational science—are interested in both structure and function. Their ultimate goal: to build a repository that details these relationships for all macromolecules that reside in the cell. Once this repository has been established, personalized medicine will be closer to becoming a reality.

Much progress has been made toward this goal. The latest version of the Human Protein Atlas, an open-access database created through international collaboration, has used immunohistochemistry and immunofluorescence to illustrate down to the subcellular level the distribution of protein expression in normal and cancer tissues. Similarly, the Protein Data Bank archive (www.rcsb.org/pdb) now holds over 130,000 structures of biological macromolecules based on X-ray crystallography, nuclear magnetic resonance spectroscopy, and electron microscopy (EM).

Looking at the multiple sources of data, one can see that problems in structural biology are often not solved by one technique alone, but require a combination of methods including those mentioned above, as well as structural mass spectrometry and small-angle X-ray scattering (SAXS). The field continues to evolve with an alphabet soup of recent innovations, particularly in cryo-EM, X-ray free-electron laser (XFEL), and fluorescence resonance energy transfer (FRET).

Included in this booklet are articles from the *Science* family of journals, as well as from the booklet sponsor, detailing the analytical tools needed to solve complex challenges in the field. Certainly, significant issues remain, such as using advanced computer modeling techniques to combine the disparate data coming from these methods, or how to deal with the fact that these molecules are continuously in motion.

Recent advancements in technology have allowed the structures of macromolecules to be deciphered at greater and greater speeds. While it once took years to figure out the structure of one protein, it now takes months or even weeks. Yet school continues to be in session for these scientists, as new methods are developed and the challenges of their integration into the growing suite of applications must be overcome. And the introduction of new methodologies does not appear to be slowing down—which is both a blessing and a curse for structural biologists and biochemistry teachers alike.

Jackie Oberst, Ph.D.
Sean Sanders, Ph.D.
Custom Publishing Office
Science/AAAS

Perspectives on integrative structural biology

Advances in biomolecular mass spectrometry (MS) have had a significant impact on the field of structural biology.

Understanding the intricate structures of the proteins in our bodies is key to advancing precision medicine. To do this, it's necessary to look beyond individual proteins and delve into the assembly and structure of protein complexes.

Advances in biomolecular mass spectrometry (MS) have had a significant impact on the field of structural biology. Technology developments in mass analyzers are the driving force behind the growing number of structural biology studies, which are enabled by the increased speed, sensitivity, selectivity, and variety of MS fragmentation techniques. This in turn has led to a plethora of MS methods, particularly at the intact protein and peptide levels , which allow the characterization of biomolecular structures.

At the intact protein level, native MS permits the study of protein assemblies in their native state by analyzing noncovalent protein–protein and protein–ligand complexes. At the peptide level, liquid chromatography (LC)-MS/MS analysis of proteolytic digests provide the amino acid sequence of proteins, allowing protein subunits to be identified from a proteome database. Limited proteolysis and surface labeling techniques such as hydrogen-deuterium exchange MS (HDX-MS) have been employed to monitor conformational changes and characterize protein–protein interfaces. A combination of chemical linking of amino acid residues within a native complex with MS analysis of crosslinked peptides (XL-MS) can determine topological arrangements and also reveal where the protein domains interface.

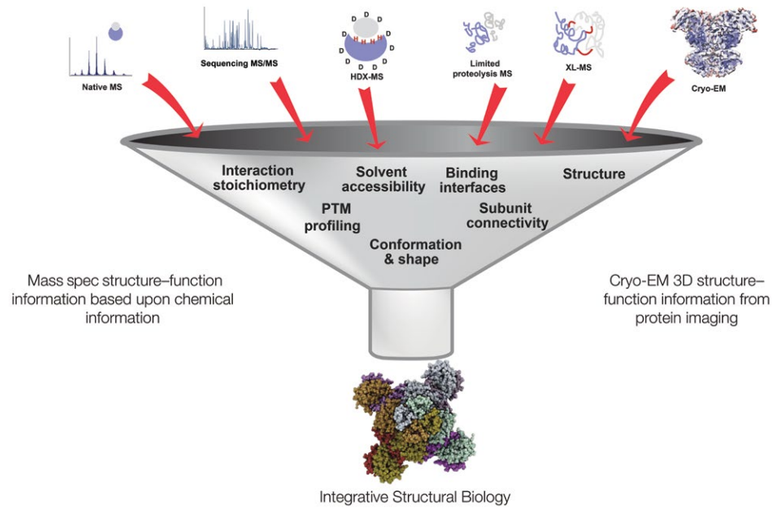
Recently, cryo-electron microscopy (cryo-EM) has emerged as an alternative to traditional techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) imaging. Cryo-EM can directly visualize complete macromolecular complexes instead of just selected parts. As with MS advancements, recent developments in cryo-EM sample preparation, microscope and detector technology, data collection automation, and image processing have made it possible to reproducibly reach near-atomic levels of resolution.

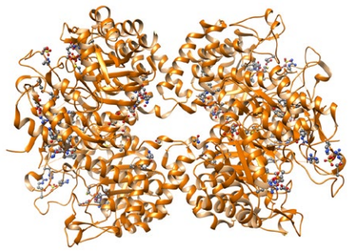
Solving the structure of these large dynamic complexes requires integrating several complementary techniques, such as MS and cryo-EM density maps—an approach known as *integrative structural biology* (see image). One such example uses structural proteomics MS tools to study the stoichiometry of KaiA, KaiB, and KaiC (components of the cyanobacterial circadian clock) and to monitor these well-defined assemblies, followed by structural characterization using single-particle cryo-EM (see page 6).

At Thermo Fisher Scientific, we strive to help our customers deliver the breakthroughs that will translate to real benefits in human health. From our state-of-the-art cryo-EM and Orbitrap MS platforms, to our innovative crosslinking reagents and robust LC systems, we provide integrative structural biologists with the analytical tools they need to solve complex challenges in the field. That's why we're proud to support this booklet and the exciting research contained within.

With continued advancements in both MS and cryo-EM such as we present in this collection, and further applications of these synergistic approaches, integrative structural biology has a bright future in accelerating the knowledge and understanding of even more intricate systems—such as pathways and organelles—along the path from structure to function.

Rosa Viner, Ph.D.
Manager, Integrative Structural Biology Program
Thermo Fisher Scientific





Integrating mass spectrometry in structural biology

Scientists are combining advanced and traditional techniques to understand protein shapes and functions.

By Mike May

A protein's shape plays a fundamental role in its function. As medical biophysicists Mohammad T. Mazhab-Jafari and John L. Rubinstein wrote: "Structural biology strives to construct models, ultimately at atomic resolution, that represent snapshots of biological macromolecules and to describe the ways in which these molecules move" (*1*).

The current dearth of protein structural information reflects the complexity of this challenge. Of the approximately 15,000 protein families, "there are still [about] 5,200 … with unknown structure outside the range of comparative modeling," according to David Baker—a biochemist and director of the Institute for Protein Design at the University of Washington, Seattle—and his colleagues (*2*). Moreover, the behavior of the vast variety of proteins and their rapidly changing conformations depends on the experimental conditions, making it difficult to study them with a single technique.

Over the last few decades, biologists analyzed protein structures using X-ray crystallography, nuclear magnetic resonance (NMR), or electron microscopy (cryo-EM) on samples at cryogenic temperatures. "These are beautiful techniques, because the resolution achieved can be down to the nanometer, angstra, or atomic level," says Albert J. R. Heck, scientific director of the Netherlands Proteomics Centre at Utrecht University. "They provide essential information, but they capture the structure in a frozen state." To unravel protein function, scientists must explore protein dynamics, and that can be done with mass spectrometry (MS).

MS captures a sample's mass-to-charge ratio, which can be used to identify and quantify proteins. By integrating results from different types of MS, scientists can determine protein structures and the mechanisms behind specific functions (see page 28). This process often requires computational tools. The combination of data and models from different experiments reveals how a protein or protein complex works, including the role of binding factors, posttranslational modifications, and interactions with other molecules such as drugs.

Such integrative approaches unveil the basic biology of proteins, and how they can be used. By combining MS with the right set of more conventional techniques, such as EM, researchers can make the most of a method's strong points and offset its weaknesses. For example, scientists in China used MS to confirm cryo-EM data on small nuclear RNA (see page 10). Despite advances in using and combining these techniques, scientists and engineers keep searching for improvements.

MS options

Even though MS can be combined with traditional techniques used in structural biology, one kind of MS is often not enough. "Unfortunately, no MS technique does everything the best," Heck explains. For example, a protein or complex of proteins can be kept in the native state—its typical shape under ordinary biological environmental conditions—and analyzed with MS. With this so-called native MS, says Heck, "the intact weighing of the mass of the protein complex lets us find out which proteins and cofactors are part of it." This method keeps proteins in natural assemblies when delivering them to the detector. Another kind of MS technique, crosslinking MS (XL-MS), can be used to determine which parts of a protein or complex are in contact. "You use a chemical glue to connect two lysine groups in close proximity," Heck explains. "They might be in a single protein or proteins close to each other." Applying this technique to many lysine groups reveals structural constraints, he points out, "because you see which parts of a protein, or which proteins in a group, are in proximity."

XL-MS can also be combined with cryo-EM. Roger Kornberg, a biochemist at Stanford University, and his colleagues combined cryo-EM and crosslinking to study a molecular complex involved in transcribing DNA to RNA (see page 25). In addition, Holger Stark of the Max Planck Institute for Biophysical Chemistry and his colleagues combined cryo-EM and XL-MS to explore the structures involved in splicing RNA (see page 20).

IMAGE: © ERMESS/SHUTTERSTOCK.COM

Scientists can also study the structure of macromolecules with hydrogen-deuterium exchange MS (HDX-MS). Here, the sample is dissolved in heavy water, D_2O. "All of the amide hydrogen on the protein's surface starts to get exchanged for deuterium," Heck notes. "Hydrogens [that are] less accessible—buried somewhere inside the protein structure—are exchanged substantially slower, and this can tell you which parts of the protein are outside and which are inside."

Although scientists developed HDX-MS several decades ago, it could only be used on one small protein at a time. Now, scientists can apply HDX-MS to whole viruses, because of several advances in MS and data processing.

Ups and downs of MS

Although today's scientists can select from a range of MS techniques, that doesn't make structural analysis easy. For one thing, exploring protein structure with MS requires upstream processing, including sample preparation and some form of separation, like liquid chromatography (LC) or capillary electrophoresis. The MS platform also needs to provide high sensitivity. In some samples, scientists search for extremely rare components, such as crosslinked peptides. "There you need nano LC to separate all of the peptides, followed by fast and sensitive MS," says Heck.

Despite some of the challenges of applying MS to protein structure determination, this technology comes with many strengths, such as identifying small binding proteins and protein posttranslational modifications; quantifying the heterogeneity of a sample; determining the ratio of the subunits in a protein complex and how the ratio changes over time or under different conditions; and tracking changes in protein conformations.

Advances in MS technology—both in hardware and software—have turned it into a tool for probing structural biology. "Today's mass spectrometry is so much faster and more sensitive," Heck points out, "and the software to analyze the data is faster, more flexible, and provides smarter algorithms for looking at different sets of large data."

But to dig deep into protein structure, notes Heck, researchers benefit from high-resolution MS technologies, such as time-of-flight platforms or the Orbitrap technology. "You can do it with other machines," he says, "but you really need the fastest and most sensitive mass spectrometer that you can get your hands on."

Tag-team technologies

The conventional techniques used to analyze the structure of biological molecules, like X-ray crystallography, can reveal the locations of components down to the atom. To use this technique, however, the recombinant protein must be crystallized, which is extremely challenging with some proteins, particularly if they are membrane-bound. In those cases, researchers can use cryo-EM to prepare very high-resolution images. But cryo-EM gives you just one image of one specific moment in time (also true for X-ray crystallography).

To study the dynamics of protein structures, today's scientists turn to MS. Although the resolution is lower with MS, the ability to examine temporal changes increases substantially with this technology. Plus, combining the various forms of MS can tease out different aspects of a molecular structure.

So, X-ray crystallography, NMR, or cryo-EM can be combined with one or more forms of MS such that each collects information on some aspect of a protein's structure. In such an integrative approach, scientists must then merge the complementary datasets in some way that produces a unified answer to a specific research question. To do that, they rely on software platforms.

"Today's mass spectrometry is so much faster and more sensitive, and the software to analyze the data is faster, more flexible, and provides smarter algorithms for looking at different sets of large data."

Further on down the road

Currently, scientists must cobble together various methods and techniques, often manually integrating the results to generate the best data. As those steps turn into a more cohesive workflow, integrative structural biology will be applied to an even wider range of questions, including novel functions of particular structures, protein–protein interactions, therapeutic targets, and more. Along the way, this field will uncover new knowledge about how biological systems work, and how they fail. The latter, in particular, will help clinical researchers understand, diagnose, and treat diseases. However, doing that depends on combining areas of expertise—from protein biophysics to drug discovery and beyond—with the right collection of tools for probing and analyzing complicated biological structures, all on a very fine scale. Only then will we have a complete understanding of the very specific ways that a protein's shape determines its function.

REFERENCES

1. M. T. Mazhab-Jafari, J. L. Rubinstein, *Sci. Adv.* **2**, e1600725 (2016).
2. S. Ovchinnikov *et al.*, *Science* **355**, 294–298 (2017).

REPORT

CIRCADIAN RHYTHMS

Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state

Joost Snijder,[1]*† Jan M. Schuller,[2]*‡ Anika Wiegard,[3] Philip Lössl,[1] Nicolas Schmelling,[3] Ilka M. Axmann,[3] Jürgen M. Plitzko,[2] Friedrich Förster,[2,4]§ Albert J. R. Heck[1]§

Cyanobacteria have a robust circadian oscillator, known as the Kai system. Reconstituted from the purified protein components KaiC, KaiB, and KaiA, it can tick autonomously in the presence of adenosine 5′-triphosphate (ATP). The KaiC hexamers enter a natural 24-hour reaction cycle of autophosphorylation and assembly with KaiB and KaiA in numerous diverse forms. We describe the preparation of stoichiometrically well-defined assemblies of KaiCB and KaiCBA, as monitored by native mass spectrometry, allowing for a structural characterization by single-particle cryo–electron microscopy and mass spectrometry. Our data reveal details of the interactions between the Kai proteins and provide a structural basis to understand periodic assembly of the protein oscillator.

Many organisms, from cyanobacteria to animals, have adapted to Earth's day-night cycle with the evolution of an endogenous biological clock. These clocks enable circadian rhythms of gene expression and metabolism with a period close to 24 hours. Many circadian rhythms rely on complex networks of transcription-translation feedback, but simpler posttranslational oscillations have also been described in both cyanobacteria and human red blood cells (*1*). The circadian oscillator of cyanobacteria is composed of three components: the proteins KaiC, KaiB, and KaiA (*2*). This posttranslational oscillator is robust enough to allow reconstitution simply through incubation of purified recombinant KaiC, KaiB, and KaiA in the presence of adenosine 5′-triphosphate (ATP) (*3*). The in vitro oscillator can maintain a stable rhythm for weeks (*4*, *5*), allowing for its detailed study.

The proteins of the Kai system collectively generate a circadian rhythm based on assembly dynamics associated with KaiC autophosphorylation and dephosphorylation (*6*, *7*). KaiC forms a homohexamer consisting of two stacked rings of domains CI and CII, which have adenosine 5′-triphosphatase (ATPase) and kinase activity, respectively (*8*). During the subjective day, the kinase activity of KaiC is stimulated by the binding of KaiA to the intrinsically disordered C-terminal regions of KaiC, resulting in sequential autophosphorylation at Thr^{432} and Ser^{431} of KaiC (*9*). During the subjective night, KaiB interacts with phosphorylated KaiC, forming the KaiCB complex (*8*). Binding of KaiB to KaiC changes the activities of SasA and CikA, which are key signaling proteins of clock-output pathways that modulate transcription (*10*). Moreover, the KaiCB complex exposes an additional KaiA-binding site, sequestering KaiA and thus preventing its productive association with KaiC (*11*). The sequestration of KaiA allows KaiC to readopt its default autodephosphorylation activity, thereby slowly resetting the protein clock to an unphosphorylated state (*4*).

Atomic-level structures of the individual Kai proteins are available (*12*–*14*), but structural information on the KaiCB and KaiCBA complexes is still ambiguous (*15*–*18*). KaiB forms monomers, dimers, and tetramers in solution, with six KaiB monomers binding cooperatively to one KaiC hexamer (*19*). It has been unclear whether KaiB binds to the CI or CII domain of KaiC (*11*, *15*, *19*–*21*). Nuclear magnetic resonance (NMR) spectroscopy studies of engineered and truncated Kai proteins suggested that KaiB binds the KaiC-CI domain and only one subunit of a KaiA dimer (*11*), but it is unclear whether the wild-type, full-length proteins arrange similarly in the KaiCBA complex. Here we use mass spectrometry (MS) and cryo–electron microscopy (cryo-EM) to study the assembly and structures of the full-length clock components to provide a structural basis for the assembly dynamics of the in vitro circadian oscillator.

The standard in vitro Kai oscillator consists of a 2:2:1 molar ratio of KaiC:KaiB:KaiA in the presence of excess MgATP, incubated at 30°C (*3*). We tracked the phosphorylation-dependent assembly of the Kai proteins under these conditions and used native MS to determine the masses and stoichiometries of the formed noncovalent assemblies (*22*). For the in vitro Kai oscillator, we simultaneously detected multiple co-occurring Kai-protein complexes, revealing more than 10 different Kai protein–assembly stoichiometries over the course of 24 hours (Fig. 1A and table S1).

The KaiC starting material had low amounts of phosphorylation. Upon initial mixing, most KaiC therefore existed as a free hexamer, whereas a small fraction formed a complex with one or two KaiA dimers (fig. S1A). These KaiCA complexes have autophosphorylation activity, which led to cooperative formation of phosphorylated $KaiC_6B_6$ complexes through a $KaiC_6B_1$ intermediate (*19*). In our samples, formation of KaiCA and KaiCB complexes peaked at 4 to 8 hours incubation time. The formation of higher-order KaiCBA complexes followed the formation of KaiCB complexes, with a maximum at 8 to 12 hours of incubation followed by a steady decline toward 24 hours. We observed $KaiC_6B_6$ with between one and six KaiA dimers bound. Detailed assignments of peaks and repeated measurements are shown in Fig. 1B and figs. S1 to S3. During the dephosphorylation phase (16 to 24 hours), as KaiCBA complexes disassemble, we detected $KaiA_2B_1$ complexes in the lower-mass region of the spectra (fig. S1B). Thus, the disassembly pathway of the KaiCBA complexes appears not to be simply the reverse of the assembly pathway but rather a distinct route.

We attempted to freeze Kai-protein assembly in specific states, producing particles amenable to more detailed structural characterization. Whereas at 30°C the default activity of KaiC is autodephosphorylation, autophosphorylation is favored at 4°C (*7*, *17*). Therefore, we tested how a lower incubation temperature affected assembly of the complete in vitro oscillator. At 4°C, KaiCBA-complex formation was slower than at 30°C. However, KaiCBA abundance steadily increased, and, even after 24 hours, it did not peak (Fig. 1A). This indicated a possible route for preparation of KaiCBA complexes with full occupancy of the KaiA-binding site. Therefore, we incubated KaiC, KaiB, and KaiA at a 1:3:3 molar ratio at 4°C for one week in the presence of MgATP. We observed near-complete occupancy of the KaiA-binding site, as seen from the predominant formation of $KaiC_6B_6A_{12}$ assemblies (Fig. 1C). The measured mass of this complex was 823.3 ± 0.5 (standard deviation) kDa, compared to a theoretical mass of 821.3 kDa for $KaiC_6B_6A_{12}$ (table S1). Similarly, prolonged incubation of KaiC with KaiB at 4°C resulted in the efficient formation of $KaiC_6B_6$ complexes (measured: 426.9 ± 0.1 kDa; theoretical: 426.4 kDa; table S1). Further experiments revealed that formation of the KaiCB complex is the limiting step for the complete assembly of KaiCBA (fig. S1C).

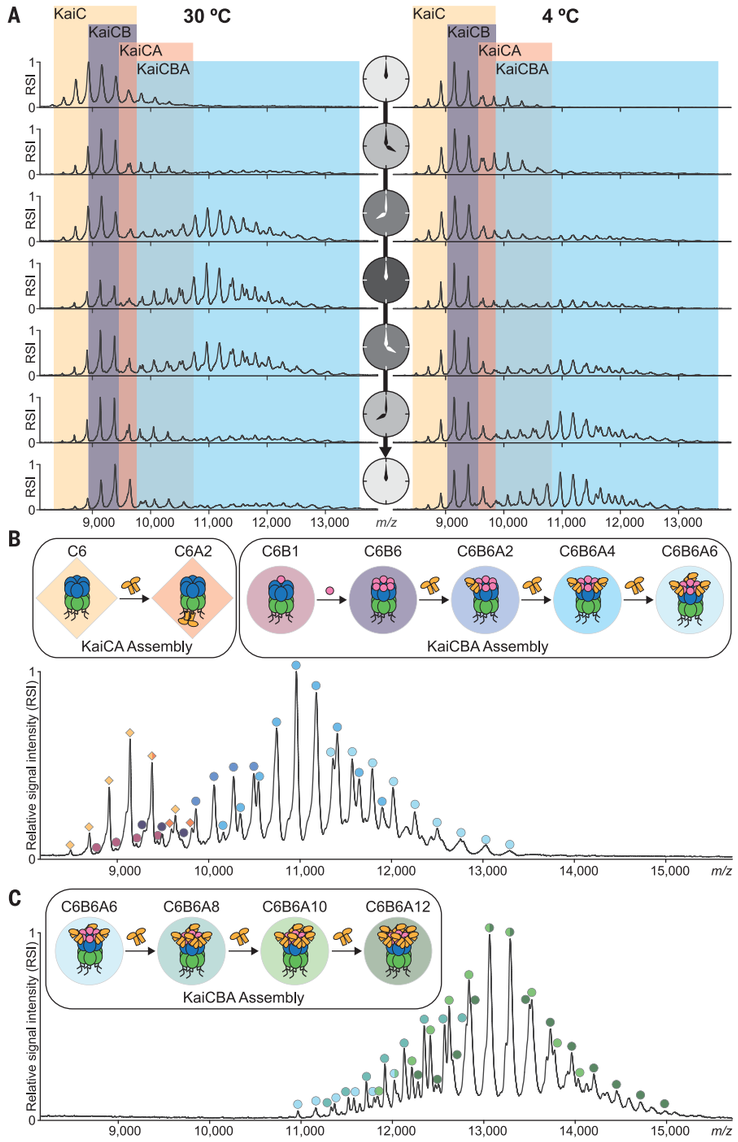


Fig. 1. Monitoring KaiCBA assembly dynamics by native MS. (**A**) Native mass spectra of the in vitro oscillator at 30° or 4°C, as indicated. The relative signal intensity (RSI) is plotted against the mass-to-charge ratio (*m*/*z*). Areas of the spectra corresponding to KaiC, KaiCA, KaiCB, and KaiCBA are indicated. (**B** and **C**) Enlarged mass spectra with full peak annotation. The identified Kai complexes are schematically represented above the spectra (KaiC-CI, green; KaiC-CII, blue; KaiA, yellow; KaiB, pink). The complexes are highlighted with differently colored circles and diamonds that match the symbols used to label the mass spectrum. A detailed explanation of the peak assignment is provided in fig. S2. An overview of all mass assignments is given in table S1. (B) Mass spectra of oscillator at 30°C after 12 hours of incubation. (C) Mixture of KaiCBA containing excess KaiA and KaiB incubated for 1 week at 4°C. These Kai complexes have near-complete occupancy of the KaiA-binding site.

Using the protocol described above, we obtained near-homogeneous $KaiC_6B_6A_{12}$ and $KaiC_6B_6$ assemblies, which we further structurally characterized by single-particle cryo-EM (Fig. 2). Preferred orientations of particles in ice limited the overall resolution of the KaiCB reconstruction to 7 Å, whereas the KaiCBA map was resolved to 4.7 Å (figs. S4 and S5). Superposition of the KaiCB and KaiCBA maps indicated that the KaiCB subcomplex remained essentially invariant in the KaiCBA complex (fig. S4). Both maps clearly show that KaiB binds to the KaiC-CI domain, resolving a controversy in the field (*11*, *19*–*22*). This architecture was further confirmed by cross-linking MS experiments (fig. S6 and table S2). For a molecular interpretation of the cryo-EM maps, we fitted available atomic models of their constituents (Fig. 2C).

The KaiCB structure is composed of three stacked rings (Fig. 2). Fitting of the homohexameric KaiC crystal structure (*12*) showed that the bottom two rings correspond to KaiC, and the upper ring is accordingly assigned to KaiB. A comparison of the map to the various nucleotide-bound states of the KaiC-CI domain indicated that this domain is in an adenosine 5′-diphosphate (ADP)–bound state (fig. S7) (*23*). The nucleotide-binding sites at the CI domain showed an unaccounted-for density, which was hence assigned to a bound nucleotide (Fig. 2D).

The KaiB subunits are arranged in a six-fold symmetrical ring, stacked on the lids of the small KaiC-CI–ATPase subdomains (fig. S8). Isolated KaiB of the cyanobacterium *Synechococcus elongatus* exists in two different folds (*24*). One fold, seen only in KaiB to date, has been observed in protein crystals (*14*, *15*, *25*, *26*). NMR spectroscopy experiments suggested that KaiB switches from the fold observed in crystal structures to a thioredoxin-like fold upon binding to KaiC (*24*). The cryo-EM map of KaiCBA confirmed that KaiC-bound KaiB adopts the thioredoxin-like fold (fig. S9). The observed KaiC-KaiB interface is further supported by hydrogen-deuterium exchange (HDX)–MS experiments (fig. S8, table S3, and data files S1 to S3). The KaiCBA model predicted that KaiC-Ala^{108} is an essential part of the KaiC-KaiB interface. Indeed, by native MS we observed loss of binding upon mutation of Ala^{108} (fig. S10). The position of individual KaiB subunits in the KaiCBA model also suggests possible KaiB-KaiB contacts that could promote cooperativity of KaiB binding to KaiC (fig. S11).

The KaiA protein from *S. elongatus* is composed of an N-terminal pseudoreceiver (PsR) domain and a C-terminal α-helical domain that takes part in homodimerization (*13*). KaiA dimerization is consistent with the $KaiC_6B_6A_{12}$ stoichiometry determined for the fully assembled complex. Fitting of the KaiA dimer structure into the KaiCBA map (Fig. 2) yields excellent colocalization of secondary-structure elements in the map and the model for the C-terminal dimerization domain and is further supported by HDX-MS experiments (fig. S12). KaiB binds to KaiA most prominently with its β_2 strand, which is present in both KaiB folds and comprises the evolutionarily most-conserved residues of the protein (fig. S13). The KaiCBA model predicts that KaiB-Lys^{42}

[1]Biomolecular Mass Spectrometry and Proteomics and Netherlands Proteomics Center, Bijvoet Center for Biomolecular Research, and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH, Utrecht, Netherlands. [2]Max Planck Institute of Biochemistry, Department of Molecular Structural Biology, D-82152 Martinsried, Germany. [3]Institute for Synthetic Microbiology, Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich Heine University Düsseldorf, D-40225 Düsseldorf, Germany. [4]Cryo-electron Microscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH, Utrecht, Netherlands.
*These authors contributed equally to this work. †Present address: Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. ‡Present address: Max Planck Institute of Biochemistry, Department of Structural Cell Biology, D-82152 Martinsried, Germany.
§Corresponding author. Email: a.j.r.heck@uu.nl (A.J.R.H.); f.g.forster@uu.nl (F.F.)

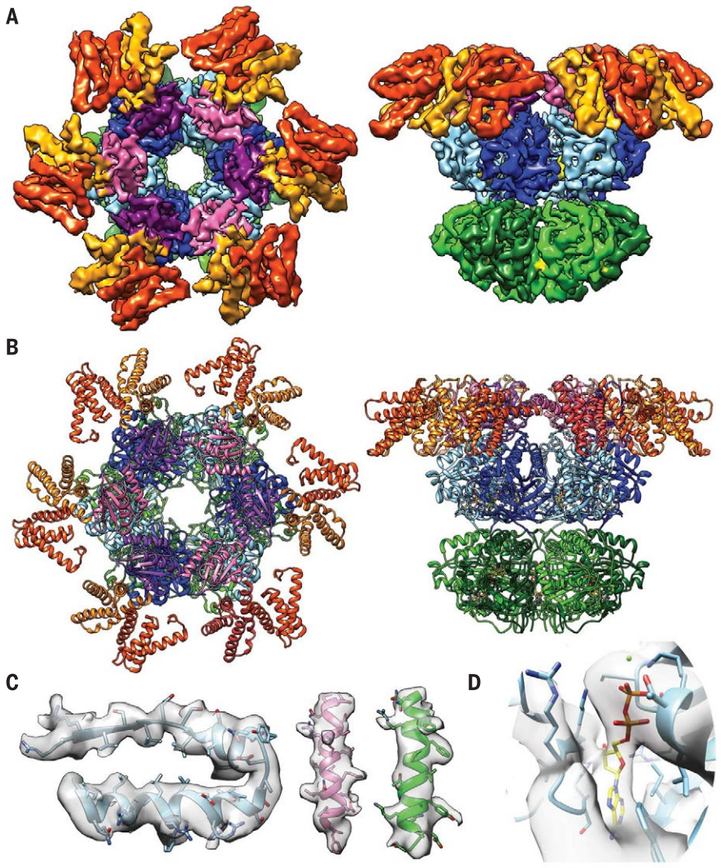


Fig. 2. Cryo-EM map and pseudoatomic model of the KaiCBA complex. (**A**) Top and side view of the three-dimensional (3D) reconstruction of the KaiCBA complex. The CII and CI domains of KaiC are colored in dark green and green and in blue and light blue, respectively. The segmented density corresponding to KaiB is colored alternating in pink and purple, and the individual KaiA homodimers are colored orange and orange-red. (**B**) Top and side view of the model of the KaiCBA complex. Colors are the same as in (A). (**C**) Selected examples of the quality of the map. (**D**) Density in the nucleotide-binding pocket of the KaiC-CI domain superimposed with the nucleotide bound in the KaiC crystal structure bound to ADP (Protein Data Bank 4TLA chain C).

is important for the KaiB-KaiA interaction, which was confirmed by loss of KaiA binding upon mutation of KaiB at this site, as observed in native MS experiments (fig. S10). The presence of KaiB-Lys[42] is also important for the in vivo clock in *S. elongatus* and *Thermosynechococcus*, as *kaiBC* and *psbA* promotor activities become arrhythmic upon mutation of this site (*25*).

For every KaiA dimer in the KaiCBA model, only one monomer is in contact with KaiB. The HDX-MS data also showed signs of asymmetric binding, confirming that the two KaiA protomers in the dimer are distinct in the KaiCBA complex (fig. S14). The density assigned to KaiA in the KaiCBA map does not cover most of the N-terminal PsR domain, indicative of the domain's structural flexibility. Positioning of the PsR domains according to the fitted domain-swapped crystal structure also results in extensive clashes with KaiB (fig. S15). HDX-MS experiments do not indicate that the PsR domain unfolds or becomes disordered (table S3 and data file S1). We therefore suspect that the PsR domain is still folded, but attached with a flexible linker, which would explain the lack of density in the cryo-EM map of the KaiCBA complex. We did observe a small, unassigned KaiA-density segment near the cleft between the homodimeric C-terminal domains. We tentatively assigned this segment to residues 147 to 172, which form the small cross–β sheet and the N-terminal α5 helix in the KaiA crystal structure (fig. S15). Binding of KaiB to the linker

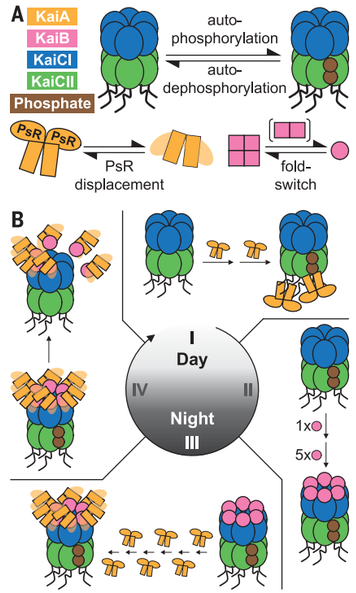


Fig. 3. The structural basis of periodic assembly in the cyanobacterial circadian clock. (**A**) Structural transitions of the individual Kai proteins during the circadian cycle. (**B**) Molecular changes in the KaiCBA oscillator. Stepwise binding of two KaiA dimers triggers KaiC autophosphorylation at Thr^{432} and Ser^{431} (I). These phosphorylation events enable cooperative binding of fold-switched KaiB monomers to the KaiC-CI domain, forming the KaiCB complex (II). KaiCB provides a scaffold for the successive sequestration of KaiA in ternary KaiCBA assemblies, concurring with a rearrangement of the KaiA PsR domains (III). KaiA sequestration promotes KaiC autodephosphorylation, resulting in the regeneration of free KaiC through release of KaiBA subcomplexes (IV).

region of KaiA therefore appears to dissociate the two strands in the dimer, resulting in a large displacement of the PsR domain. The α5 helix of KaiA likely occludes the site to which the flexible C termini of KaiC bind (*27*).

On the basis of these structures and our native MS data, we propose a detailed model for the cyclic phosphorylation-dependent assembly of Kai components in the in vitro oscillator (Fig. 3). Upon mixing the protein components of the in vitro oscillator, unphosphorylated KaiC hexamers bind one or two copies of a KaiA dimer on the C terminus of the KaiC-CII domain (*9*). Binding of the second KaiA dimer stimulates autophosphorylation of KaiC, first at Thr^{432} and then at Ser^{431} (*7*). Serine phosphorylation triggers binding of KaiB in a fold-switched state. Six copies of KaiB bind cooperatively (*19*) to form phosphorylated $KaiC_6B_6$ complexes. The bound KaiB subunits present alternative binding sites for KaiA, away from a phosphorylation-stimulating interaction with the KaiC-CII domain. The KaiA dimer binds

asymmetrically through its linker region to KaiB, resulting in a wide displacement of the PsR domain. As the pool of free KaiA dimers is depleted, KaiC switches back to autodephosphorylation activity. Complete dephosphorylation of KaiC results in dissociation of the KaiCBA complex by loss of $KaiA_2B_1$ subcomplexes, thereby completing one cycle of the oscillator. In cyanobacterial cells, KaiC and KaiB are produced from the same operon and in 10- to 100-fold excess to KaiA (*28*). The high excess of KaiCB over free KaiA could promote efficient sequestration of KaiA in vivo. The model presented here can thus serve as a framework to better understand the circadian clock in cyanobacterial cells.

REFERENCES AND NOTES

1. J. S. O'Neill, A. B. Reddy, *Nature* **469**, 498–503 (2011).
2. M. Ishiura *et al.*, *Science* **281**, 1519–1523 (1998).
3. M. Nakajima *et al.*, *Science* **308**, 414–415 (2005).
4. C. Brettschneider *et al.*, *Mol. Syst. Biol.* **6**, 389 (2010).
5. H. Ito *et al.*, *Nat. Struct. Mol. Biol.* **14**, 1084–1088 (2007).
6. M. Egli, C. H. Johnson, *Curr. Opin. Neurobiol.* **23**, 732–740 (2013).
7. T. Nishiwaki *et al.*, *EMBO J.* **26**, 4029–4037 (2007).
8. X. Qin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14805–14810 (2010).
9. R. Pattanayek *et al.*, *EMBO J.* **25**, 2017–2028 (2006).
10. J. S. Markson, J. R. Piechura, A. M. Puszynska, E. K. O'Shea, *Cell* **155**, 1396–1408 (2013).
11. R. Tseng *et al.*, *J. Mol. Biol.* **426**, 389–402 (2014).
12. R. Pattanayek *et al.*, *Mol. Cell* **15**, 375–388 (2004).
13. S. Ye, I. Vakonakis, T. R. Ioerger, A. C. LiWang, J. C. Sacchettini, *J. Biol. Chem.* **279**, 20511–20518 (2004).
14. R. Murakami *et al.*, *J. Biol. Chem.* **287**, 29506–29515 (2012).
15. S. A. Villarreal *et al.*, *J. Mol. Biol.* **425**, 3311–3324 (2013).
16. S. Akiyama, A. Nohara, K. Ito, Y. Maéda, *Mol. Cell* **29**, 703–716 (2008).
17. R. Pattanayek *et al.*, *PLOS ONE* **6**, e23697 (2011).
18. M. Egli, *J. Biol. Chem.* **289**, 21267–21275 (2014).
19. J. Snijder *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1379–1384 (2014).
20. R. Pattanayek *et al.*, *EMBO J.* **27**, 1767–1778 (2008).
21. Y. G. Chang, R. Tseng, N. W. Kuo, A. LiWang, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16847–16851 (2012).
22. J. Snijder, A. J. Heck, *Annu. Rev. Anal. Chem. (Palo Alto Calif.)* **7**, 43–64 (2014).
23. J. Abe *et al.*, *Science* **349**, 312–316 (2015).
24. Y. G. Chang *et al.*, *Science* **349**, 324–328 (2015).
25. R. Iwase *et al.*, *J. Biol. Chem.* **280**, 43141–43149 (2005).
26. K. Hitomi, T. Oyama, S. Han, A. S. Arvai, E. D. Getzoff, *J. Biol. Chem.* **280**, 19127–19135 (2005).
27. R. Pattanayek, M. Egli, *Biochemistry* **54**, 4575–4578 (2015).
28. Y. Kitayama, H. Iwasaki, T. Nishiwaki, T. Kondo, *EMBO J.* **22**, 2127–2134 (2003).

ACKNOWLEDGMENTS

We thank J. Andres and M. Yazdanyar for help with protein expression, O. Mihalache for help with sample preparation, F. Beck and A. Aufderheide for assistance with image processing, and C. Benda for help with PHENIX software. This work was supported by the Netherlands Organisation for Scientific Research (NWO) Roadmap Initiative Proteins@Work grant 184.032.201 to A.J.R.H., the European Union Seventh Framework Programme ManiFold grant 317371 to A.J.R.H. and P.L., and the German Research Foundation grant GRK1721 to F.F., and grants AX 84/1-3 and EXC 1028 to A.W., N.S., and I.M.A.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/355/6330/1181/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S4
References (*29*–*51*)
Data S1 to S3

10 June 2016; accepted 13 February 2017
10.1126/science.aag3218

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis

Ruixue Wan,[1]* Chuangye Yan,[1]* Rui Bai,[1] Lin Wang,[1] Min Huang,[2] Catherine C. L. Wong,[2] Yigong Shi[1]†

Splicing of precursor messenger RNA is accomplished by a dynamic megacomplex known as the spliceosome. Assembly of a functional spliceosome requires a preassembled U4/U6.U5 tri-snRNP complex, which comprises the U5 small nuclear ribonucleoprotein (snRNP), the U4 and U6 small nuclear RNA (snRNA) duplex, and a number of protein factors. Here we report the three-dimensional structure of a *Saccharomyces cerevisiae* U4/U6.U5 tri-snRNP at an overall resolution of 3.8 angstroms by single-particle electron cryomicroscopy. The local resolution for the core regions of the tri-snRNP reaches 3.0 to 3.5 angstroms, allowing construction of a refined atomic model. Our structure contains U5 snRNA, the extensively base-paired U4/U6 snRNA, and 30 proteins including Prp8 and Snu114, which amount to 8495 amino acids and 263 nucleotides with a combined molecular mass of ~1 megadalton. The catalytic nucleotide U80 from U6 snRNA exists in an inactive conformation, stabilized by its base-pairing interactions with U4 snRNA and protected by Prp3. Pre-messenger RNA is bound in the tri-snRNP through base-pairing interactions with U6 snRNA and loop I of U5 snRNA. This structure, together with that of the spliceosome, reveals the molecular choreography of the snRNAs in the activation process of the spliceosomal ribozyme.

In the precursor mRNA (pre-mRNA) of eukaryotes, the protein-coding sequences, termed exons, are interrupted by noncoding sequences known as introns (*1*, *2*). Pre-mRNA splicing, involving the removal of introns and the ligation of neighboring exons, is carried out by a dynamic, multi-megadalton ribonucleoprotein (RNP) complex known as the spliceosome (*3*). Each splicing cycle entails two sequential transesterification reactions, with the first producing a free 5′-exon and an intron lariat-3′-exon and the second resulting in a freed intron lariat and a joined 5′-exon-3′-exon (*4*). The spliceosome responsible for these two reactions consists of U2 and U5 small nuclear RNPs (snRNPs), U6 small nuclear RNA (snRNA), and a large number of additional proteins (*5*, *6*). Assembly of the catalytically active spliceosome, however, requires a series of concerted steps, with the U4/U6.U5 tri-snRNP playing an indispensable role (*7*–*9*).

According to the prevailing model, U1 snRNP recognizes the 5′ splice site (5′SS) of an intron, and U2 snRNP binds to the branch point sequence and 3′ splice site (3′SS) of the same intron, forming the prespliceosomal complex (the A complex) (*7*). The preassembled tri-snRNP comprises U5 snRNA, the base-paired U4/U6 snRNAs, and more than 30 proteins, including the key factors Prp8 (Spp42 in *Schizosaccharomyces pombe*), Brr2, and Snu114 (Cwf10 in *S. pombe*). Binding of the U4/U6.U5 tri-snRNP to the A complex results in the formation of the precatalytic B complex. Subsequent RNP rearrangement leads to dissociation of the U1 and U4 snRNPs and the recruitment of many additional proteins, producing the activated B complex (B^{act}) and then the catalytically competent B* complex. During this process, the U4/U6 duplex is unwound, allowing U6 snRNA to extensively base-pair with U2 snRNA. The B* complex catalyzes the first transesterification reaction, ending with the C complex, which contains a free 5′-exon and an intron lariat-3′-exon (*7*). The C complex carries out the second transesterification reaction, resulting in the ligation of two exons and formation of the postcatalytic P complex. Release of the ligated exon generates the intron-lariat spliceosomal complex. In the last step, the intron lariat is released, and the protein and RNA components are recycled, a sizable fraction of which reassemble into the U4/U6.U5 tri-snRNP (*7*).

The spliceosome was thought to be a protein-directed metalloenzyme (*10*–*12*), with two catalytic magnesium (Mg^{2+}) ions coordinated by conserved nucleotides in the intramolecular stem loop (ISL) of U6 snRNA (*13*, *14*). These predicted features are observed in the recent electron cryomicroscopy (cryo-EM) structure of a yeast spliceosome at 3.6 Å resolution (*15*, *16*). The spliceosomal catalytic center comprises helix I of the U2/U6 snRNA duplex, ISL of U6 snRNA, loop I of U5 snRNA, and Mg^{2+} ions, all of which are located in a positively charged surface cavity in Prp8 (*15*, *16*). In the U4/U6.U5 tri-snRNP, however, U6 snRNA is thought to exist in an inactive conformation through extensive base-pairing interactions with U4 snRNA (*7*, *17*). Elucidation of the tri-snRNP structure is essential for a mechanistic understanding of spliceosomal assembly and U6 snRNA activation.

The dynamic nature and large size of the spliceosomal complexes have made detailed structural investigation a daunting challenge (*8*). Earlier EM studies of the human and yeast tri-snRNPs, at resolutions of 21 Å or lower, revealed the overall shape and global features (*18*, *19*). More recently, the cryo-EM structure of a U4/U6.U5 tri-snRNP from *Saccharomyces cerevisiae*, determined at 5.9 Å resolution, has allowed positional identification of many proteins and the snRNA components; secondary structural elements are also discernible for many protein components in the tri-snRNP (*20*). Despite these encouraging advances, the relatively low resolution revealed few features of amino acid side chains or nucleotides (*20*), precluding the generation of an atomic model for the tri-snRNP.

Isolation and characterization of the U4/U6.U5 tri-snRNP

Six protein components of the U4/U6.U5 tri-snRNP, each tagged with protein A and calmodulin binding peptide at the C terminus, were individually introduced into *S. cerevisiae* to enable screening of protein expression and pulling down of endogenous tri-snRNP. The best outcome obtained was for Prp6, a protein required for accumulation of the tri-snRNP (*21*). We purified ~260 μg of spliceosomal U4/U6.U5 tri-snRNP from 36 liters of *S. cerevisiae* culture (see the supplementary materials; fig. S1A). The purified tri-snRNP was eluted from gel filtration as a single peak (fig. S1B) and contained three major RNA species (fig. S1C). The lengths of these RNA molecules are consistent with those of U4, U5, and U6 snRNAs from *S. cerevisiae*. The purified tri-snRNP included a large number of proteins (fig. S1D), and the negatively stained sample appeared in EM to contain mostly homogeneous particles (fig. S1E). The EM sample was confirmed by mass spectrometry to include all core components of the U4/U6.U5 tri-snRNP (figs. S1F and S2) (*22*).

The presence of U4, U5, and U6 snRNAs in the purified tri-snRNP sample was confirmed by Northern blots using specific DNA probes (fig. S3). Increasing the exposure time by 50 times also revealed U2 and U1 snRNAs, suggesting the presence of a very small amount of other contaminating complexes. The tri-snRNP particles were disassembled upon incubation with ATP (adenosine triphosphate) but not with ADP (adenosine diphosphate) or the nonhydrolyzable analog AMPPNP (adenosine 5′-(β,γ-imido)triphosphate) (fig. S4). These results suggest a role of ATP hydrolysis–dependent unwinding of the U4/U6 duplex by the ATPase (adenosine triphosphatase)/helicase Brr2. The presence of GDP (guanosine diphosphate), which is known to inhibit the Brr2-activating function of Snu114 (*23*), appears to inhibit the disassembly of tri-snRNP in the presence of ATP (fig. S4A). These findings confirm that the vast majority of the particles observed by means of EM belong to tri-snRNP. To facilitate future structural assignment, we chemically cross-linked the tri-snRNP and performed mass spectrometry analysis on the resulting complex. This analysis uncovered 104 pairs of intermolecular interactions among the protein components of the tri-snRNP (fig. S5).

EM analysis of the U4/U6.U5 tri-snRNP

The U4/U6.U5 tri-snRNP sample was imaged under cryogenic conditions with a K2 direct electron detector mounted on a Titan Krios microscope operating at 300 kV. A total of 3141 micrographs were collected (Fig. 1A and table S1); 635,850 semi-autopicked particles were subjected to particle sorting, reference-free two-dimensional (2D) classification, and 3D classification (Fig. 1B and fig. S6). Using a published protocol (*15*), we manually picked more particles. After two rounds of 3D classification, 207,238 particles were used to produce a cryo-EM map at an average resolution of 4.5 Å (fig. S6). After particle polishing and an additional round of 3D classification, 172,134 particles gave a final reconstruction at an average resolution of 3.8 Å on the basis of the gold-standard Fourier shell correlation (FSC) criterion (Fig. 1, C and D, and fig. S7). Application of soft masks improved the resolutions of local maps to 3.4 to 3.8 Å, with the central regions of the tri-snRNP reaching 3.0 to 3.5 Å (figs. S8 and S9). Throughout the tri-snRNP, most secondary structural elements in the protein components are visible, and ~70% of all amino acids in the core regions of tri-snRNP exhibit discernible side chain features (figs. S10 to S12). Both U4/U6 snRNA duplex and U5 snRNA are well resolved in cryo-EM maps, which, together with prior knowledge of specific base-pairing interactions, allows sequence assignment (fig. S13).

Using a combination of homologous structure docking and de novo assignment, we generated an atomic model for the spliceosomal U4/U6.U5 tri-snRNP that includes most known components (Fig. 2, A and B). The final refined model of the tri-snRNP contains 8495 amino acids from 30 proteins and three snRNA molecules (Fig. 2B and tables S1 to S3), with a combined molecular mass just exceeding 1.0 MD. The full-length U4/U6 and U5 snRNAs contain a total of 486 nucleotides (nt), of which 243 have been tentatively assigned in our structure. Some very weak cryo-EM maps, probably reflecting dynamic components of the tri-snRNP, remain unassigned. Our atomic model includes 18 protein components of U4/U6 snRNP, all 11 core proteins of U5 snRNP, and the tri-snRNP–specific protein Prp6.

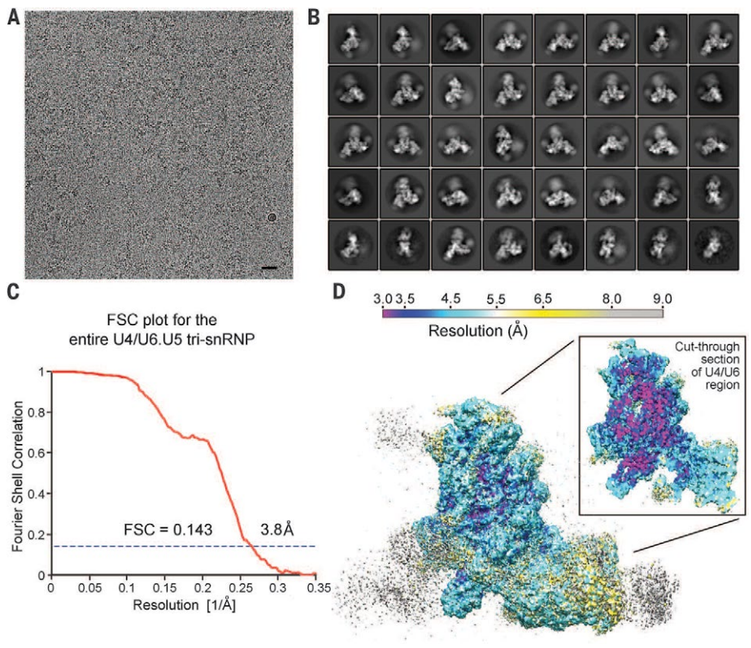


Fig. 1. Cryo-EM analysis of the spliceosomal U4/U6.U5 tri-snRNP from *S. cerevisiae*. (**A**) A representative cryo-EM micrograph of the yeast spliceosomal U4/U6.U5 tri-snRNP. The low image contrast made it challenging to manually pick particles. An entire micrograph is shown. Scale bar, 30 nm. (**B**) Representative 2D class averages of the yeast spliceosomal U4/U6.U5 tri-snRNP. (**C**) The overall resolution is estimated to be 3.81 Å on the basis of the FSC criterion of 0.143. (**D**) An overall view of the cryo-EM maps for the yeast spliceosomal U4/U6.U5 tri-snRNP. The resolution is color-coded for different regions of the complex. A cross section of the tri-snRNP surface view is shown in the inset. The resolution reaches to 3.0 to 3.5 Å for the core regions of the U4/U6.U5 tri-snRNP, including but not limited to the U4/U6 snRNA duplex, protein components of U4 snRNP, much of U5 snRNA, the bulk of Prp8, and Snu114.

Overall structure of U4/U6.U5 tri-snRNP

The spliceosomal U4/U6.U5 tri-snRNP complex has a triangular appearance, with U5 snRNP constituting both the center and two corners of the triangle and spanning the longest dimension of ~315 Å (Fig. 2 and fig. S14). One end of U5 snRNP, comprising Snu114 and the U5 Sm ring, is resolved in the overall cryo-EM maps (Fig. 1D) and exhibits well-defined structural features in the improved local map (figs. S8 and S11). The other end of U5 snRNP, however, appears to be highly flexible (Fig. 1D); this end contains the helicase/ATPase Brr2 and the U4 Sm ring and is linked to the center of tri-snRNP mainly through the Jab1/MPN domain of Prp8 (Fig. 2). This region has a relatively low resolution of 7.9 Å even after the application of a local mask (fig. S8). Relative to its longest side, the tri-snRNP has a height of 210 Å and a thickness of about 150 Å (Fig. 2A). The third corner of the triangle is well resolved in cryo-EM maps (Fig. 1D) and is occupied by the U4/U6 snRNA duplex and its associated proteins (Fig. 2A). Located in close proximity to this corner, the heptameric Sm-like (Lsm) ring and the bound 3′-end of U6 snRNA are poorly resolved in cryo-EM maps, and the intervening components between the Lsm ring and U4/U6 snRNP remain to be structurally identified. Because of their low resolutions, the cryo-EM maps for the Brr2 region and the Lsm ring disallow assignment of side chains or specific interactions; proteins in these regions were fitted into the cryo-EM maps in a rigid-body mode using previously determined crystal structures.

The core of Prp8, comprising residues 749 to 1830, is located at the center of the U4/U6.U5 tri-snRNP (Fig. 2B). The U4/U6 snRNP closely associates with the core of Prp8 through the ferredoxin-like protein Prp3 (*24*) and the Nop domain containing protein Prp31 (*25*). The N-terminal domain (N domain) of Prp8 is mainly responsible for binding U5 snRNA and the only GTPase (guanosine triphosphatase), Snu114, whereas the C-terminal Jab1/MPN domain of Prp8 binds Brr2, and the RNaseH-like domain of Prp8 interacts with Prp3 and the tri-snRNP–specific protein Prp6 (Fig. 2B). The U4-associated Sm ring contacts Brr2 on the opposite side relative to the Jab1/MPN domain of Prp8. The U5 snRNP component Dib1, which has a thioredoxin-like fold (*26*), interacts with an extended loop of Prp31 and bridges the core and the N domain of Prp8.

[1]Ministry of Education Key Laboratory of Protein Science, Tsinghua-Peking Joint Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]National Center for Protein Science Shanghai, Institute of Biochemistry and Cell Biology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.
*These authors contributed equally to this work.
†Corresponding author. E-mail: shi-lab@tsinghua.edu.cn

The U5-associated Sm ring encircles a stretch of U5 snRNA sequence and binds Snu114.

Despite its close proximity to the core of Prp8, the U4/U6 snRNA duplex has few direct interactions with Prp8. The U4/U6 snRNA duplex is surrounded and directly recognized by four proteins: Prp3, Prp6, Prp31, and the globular RNA-binding protein Snu13 (Fig. 2B and fig. S15) (*27*). These interactions are stabilized by the core of Prp8 at the bottom and the β-propeller protein Prp4 at the top; Prp4 directly contacts Prp3, Snu13, and Prp6. Together, these proteins help maintain the inactive conformation of the U4/U6 snRNA duplex in the tri-snRNP; during catalytic activation of the spliceosome, these proteins must be stripped as the U4/U6 snRNA duplex is unwound by the RNA-dependent ATPase Brr2.

Structure of the U4/U6 snRNA duplex

The U4/U6 snRNA duplex and the 5′-stem loop of U4 snRNA both have excellent cryo-EM maps (fig. S13, A to D). The U4/U6 snRNA duplex consists of stem I, which is base-paired by nucleotides 56 to 62 of U6 snRNA and 57 to 63 of U4 snRNA, and stem II between nucleotides 64 and 81 of U6 snRNA and 18 nucleotides at the 5′-end of U4 snRNA (Fig. 3, A and B). Stems I and II are interrupted by the 5′-stem loop of U4 snRNA, which forms a bulged duplex. The catalytic uridine nucleotide 80 (U80) of U6 snRNA (*13*) is sequestered in an inactive conformation in the U4/U6 snRNA duplex by at least two mechanisms. First, U80 of U6 snRNA pairs with A1 of U4 snRNA at the tip of stem II, where the fully extended conformation of the phosphodiester backbone no longer allows coordination of two catalytic Mg^{2+} ions (Fig. 3A, inset). Second, this part of the U4/U6 duplex is bound by Prp3, with the phosphate group of U80 probably forming hydrogen (H) bonds with the side chain of Arg^{399} from Prp3 (fig. S15A).

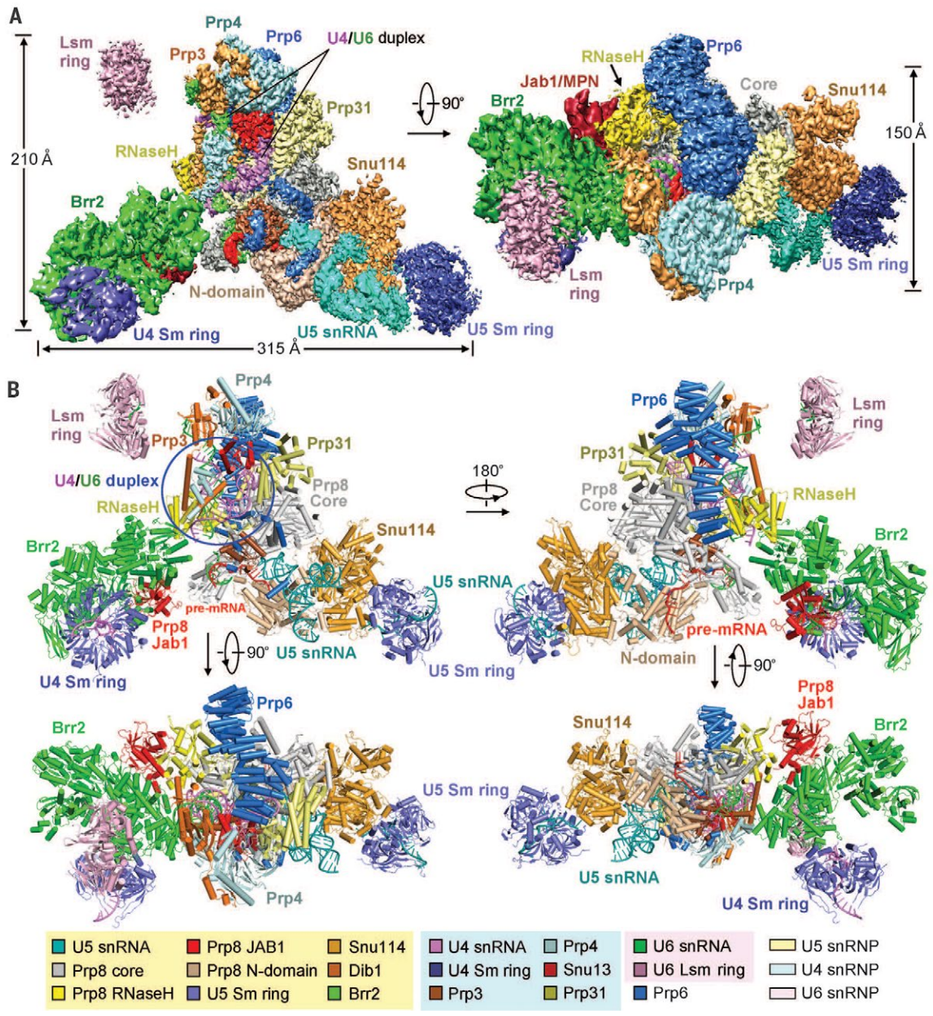


Fig. 2. Structure of the spliceosomal U4/U6.U5 tri-snRNP from *S. cerevisiae*. (**A**) The cryo-EM maps of the yeast U4/U6.U5 tri-snRNP at an overall resolution of 3.81 Å. The cryo-EM maps were generated by Chimera (*54*) and carved using the atomic coordinates. To display cryo-EM maps for all regions of the tri-snRNP, varying contour levels were applied to different regions, with low contour levels for the Brr2 region and the Lsm ring. (**B**) A cartoon of the yeast U4/U6.U5 tri-snRNP complex. The protein and RNA components are color-coded. Four views are shown. This structure includes 30 proteins, three snRNA molecules, and a pre-mRNA molecule, with a combined molecular weight of ~1 MD. The atomic model includes 8495 amino acids; 243 nucleotides from U4, U5, and U6 snRNAs; and 20 nucleotides from pre-mRNA.

Although only 41 nucleotides of U6 snRNA in the U4/U6.U5 tri-snRNP were clearly assigned, they contain most of the catalytically important sequences and are highly conserved from yeast to humans (Fig. 3C). These corresponding nucleotides in U6 snRNA form helix I in the U2/U6 snRNA duplex and the ISL in the activated *S. pombe* spliceosome (Fig. 3, D and E). Superposition of the core of Prp8 with that of Spp42 allows assessment of the relative positioning of U6 snRNA in the tri-snRNP and the *S. pombe* spliceosome. The nucleotides in U6 snRNA of the tri-snRNP are located up to 100 Å away from their corresponding positions in the active *S. pombe* spliceosome (Fig. 3D), with the 5′-stem loop of U4 snRNA placed in between (Fig. 3E). Thus, during Brr2-mediated activation of the spliceosome, U6 snRNA must undergo a dramatic structural rearrangement to arrive at its active conformation.

Recognition of U4 and U6 snRNA

The core of the U4/U6 snRNP is a closely associated complex that consists of U4/U6 snRNA, Prp3, Prp4, Snu13, and Prp31 (Fig. 4A). This complex interacts intimately with the core of Prp8 and the tri-snRNP–specific protein Prp6, forming a compact structure that stands out from the rest of the U4/U6.U5 tri-snRNP (Figs. 2B and 4A). Within this structure, Prp3 only recognizes the U4/U6 duplex, whereas Snu13 and Prp6 mostly bind to the 5′-stem loop of U4 snRNA (Fig. 4B). Prp31, in an extended conformation, interacts with both U4 snRNA and the minor groove in stem I of the U4/U6 snRNA duplex. The U4/U6 snRNA duplex is specifically recognized through a large number of H bonds by amino acids from Prp3, Snu13, Prp31, and Prp6.

Prp3 plays a major role in the recognition of the U4/U6 snRNA duplex (Fig. 4B and fig. S15A). Two extended α-helices in the N-terminal half of Prp3 bind to the major groove and the axial side, respectively, of stem II in the U4/U6 snRNA duplex. In the first helix (residues 238 to 262), six Arg and three Lys residues may donate up to nine direct and water-mediated H bonds to the phosphate backbone in the major groove of stem II (Fig. 4C and fig. S16A). In the second helix (residues 291 to 326), His^{308}, Asn^{312}, and Arg^{315} recognize A11 of U4 snRNA, G71 of U6 snRNA, and C10 of U4 snRNA, respectively, in the minor groove of stem II (fig. S15A, top middle panel). A C-terminal domain of Prp3 directly interacts with the 5′-end of U4 snRNA and the phosphodiester backbone of U6 snRNA around the nucleotide U80 (fig. S15A, top right panel). The doubly methylated 2′-amino group of the 5′-m_3G cap from U4 snRNA may make hydrophobic contacts with the side chains of Glu^{362}, Leu^{363}, and Phe^{391}, whereas Arg^{399} may donate two H bonds to the phosphate group of U80.

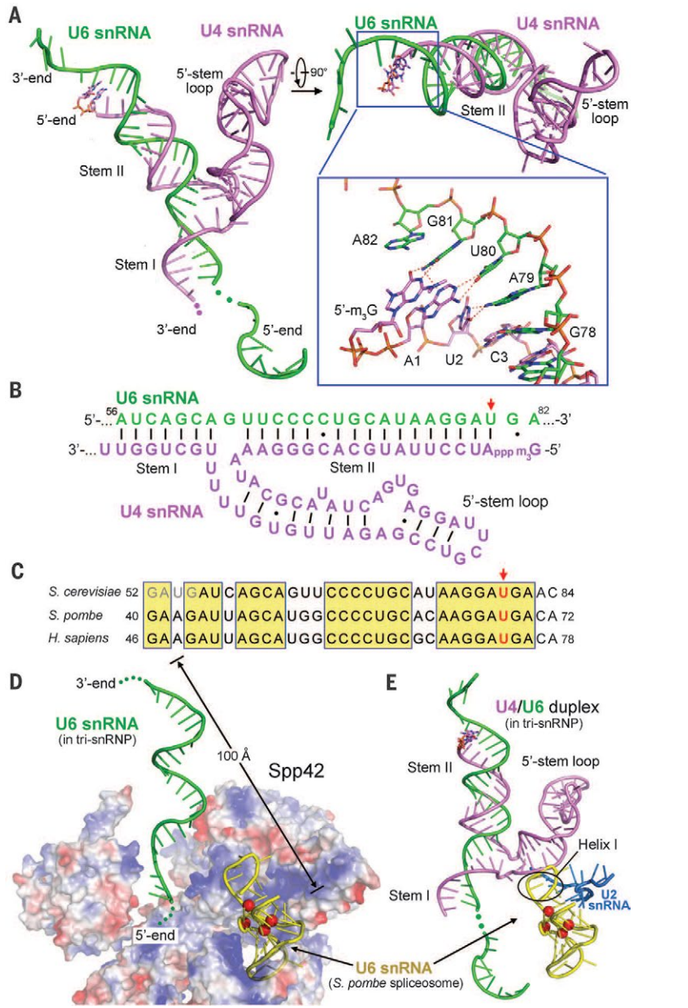


Fig. 3. Structure of the U4 and U6 snRNAs in the spliceosomal U4/U6.U5 tri-snRNP. (**A**) Structure of the U4/U6 snRNA duplex. U4 snRNA (colored violet) and U6 snRNA (green) extensively base-pair with each other to form two stretches of duplexes termed stem I and stem II, which are interrupted by the 5′-stem loop of the U4 snRNA. Base-pairing interactions between four nucleotides at the 5′-end of U4 snRNA and nucleotides G78 to A81 of U6 snRNA are highlighted in the close-up view. (**B**) A schematic diagram of the base-pairing interactions in the U4/U6 snRNA duplex. (**C**) Sequence alignment of U6 snRNA from *S. cerevisiae*, *S. pombe*, and *Homo sapiens*. Only 33 nucleotides are shown from each U6 snRNA. Twenty-nine of these nucleotides have been assigned for U6 snRNA in the cryo-EM maps of tri-snRNP. The catalytic uridine nucleotide (U80 in *S. cerevisiae* and U68 in *S. pombe*) is identified by a red arrow in (B) and (C). (**D**) U6 snRNA undergoes a dramatic conformational switch and a translocation of up to 100 Å during the assembly of a functional spliceosome. The *S. cerevisiae* U4/U6.U5 tri-snRNP is aligned to the *S. pombe* spliceosome on the basis of the core domains of Prp8 and Spp42. The resulting U6 snRNA (green) from U4/U6.U5 tri-snRNP is shown in relation to U6 snRNA (yellow) from the *S. pombe* spliceosome. For visual clarity, only Prp8 is shown by its surface electrostatic potential (red indicates high and blue indicates low surface electrostatic potential). Red spheres indicate the catalytic magnesium ions. (**E**) A close-up view on the U4/U6 snRNA duplex of the tri-snRNP in relation to the activated U6 snRNA in the *S. pombe* spliceosome. A small portion of U2 snRNA (colored marine) from the *S. pombe* spliceosome is shown.

The globular protein Snu13 is wedged between stem II and the 5′-stem loop of the U4/U6 duplex but mainly interacts with the 5′-stem loop (Fig. 4B and fig. S15B). The side chains of Glu^{39}, Lys^{42}, and Arg^{46} contact both the phosphate backbone and specific bases in the 5′-stem loop (Fig. 4D and fig. S16B). Similar to Prp3, Prp31 also adopts an extended conformation. An α-helical domain at the N-terminal half of Prp31 interacts with the tip of the 5′-stem loop of U4 snRNA (Fig. 4B and fig. S15C). A protracted loop (residues 346 to 438) follows the ridge of the 5′-stem loop, goes into the major groove of stem I, and comes out in an extended conformation. In the major groove of stem I, at least two bases of U4 and U6 snRNAs are recognized by the positively charged residues Arg^{367} and Lys^{371} from Prp31 (Fig. 4E and fig. S16C). The extended conformation of Prp31 is sustained through close interactions with other proteins, particularly Prp8 (Fig. 4A).

In our atomic structure, the superhelical protein Prp6 consists of 44 α-helices, of which 36 are organized into 18 tetratricopeptide repeats. Only helices α6 preceding the tetratricopeptide repeats and α39 at the C terminus bind to the stem and tip, respectively, of the 5′-stem loop of U4 snRNA (Fig. 4B and fig. S15D). Arg^{136} and Gln^{140} of α6 may make specific H bonds to U51 and A20/U54, respectively, whereas Lys^{133} and Arg^{143} contact the phosphate backbone (Fig. 4F and fig. S16D). The overall appearance of Prp6 resembles a cup handle, with the N-terminal helices of Prp6 contacting the RNaseH-like domain and the core of Prp8 and the C-terminal domain interacting with Prp4, Snu13, and Prp31 (Fig. 4A).

Structure of U5 snRNP

U5 snRNP adopts an elongated and flexible conformation, with most components assigned in our atomic model (Fig. 5A). One end of U5 snRNP contains a heptameric Sm ring, which is bound to the 3′-end sequences of U5 snRNA. The other end of U5 snRNP is capped by a second Sm ring, which associates with the 3′-end sequences of U4 snRNA. U5 snRNP constitutes two corners of the triangular-shaped U4/U6.U5 tri-snRNP. One corner, consisting of U5 snRNA, Snu114, and the N domain of Prp8, has a well-defined conformation and is connected to the centrally located Prp8 core through multiple interfaces, including that mediated by Dib1. The other corner, comprising mostly Brr2 and the bound Jab1/MPN domain of Prp8, is flexibly attached to the rest of the tri-snRNP and exhibits a dynamic conformation.

U5 snRNA from *S. cerevisiae* has two forms, of which the long form (214 nt) contains 35 extra nucleotides beyond the 3′-end of the short form (179 nt). Nucleotides 28 to 53, 62 to 127, and 163 to 183 are distinguishable in cryo-EM maps (fig. S13) and were explicitly modeled in our structure (Fig. 5B). The U5 snRNA structure consists of loop 1, stem I, stem II, a variable stem loop, and extended sequences at the 3′-end. The core regions of U5 snRNA from *S. cerevisiae* and from *S. pombe* have a very similar structure with respect to both loop I and stems I and II (Fig. 5C). Loop II of the U5 snRNA from *S. pombe* is replaced by a much longer variable stem loop in *S. cerevisiae* (Fig. 5C). Loop III of the U5 snRNA from *S. pombe* is replaced by a predicted internal loop II in U5 snRNA from *S. cerevisiae*. Both internal loop II and stem III of U5 snRNA in our tri-snRNP structure are disordered.

Conformational changes in Prp8

Prp8 in *S. cerevisiae* (Spp42 in *S. pombe*) is a central component of the spliceosome and the splicing reaction. In the U4/U6.U5 tri-snRNP, Prp8 is divided into four distinct structural regions: the N domain (residues 130 to 736), the core (residues 747 to 1830), the RNaseH-like domain (residues 1831 to 2078), and the Jab1/MPN domain (residues 2148 to 2396). The core, also known as the large domain (*28*), consists of the reverse transcriptase palm/finger, thumb/X, linker, and endonuclease-like subdomains. Each of these four regions is able to undergo pronounced rigid-body movement relative to neighboring regions, exemplified by the RNaseH-like domain. The core of Prp8 from the tri-snRNP can be superimposed onto that of Spp42 from the *S. pombe* spliceosome with a root mean square deviation (RMSD) of 3.06 Å between 1099 pairs of aligned Cα atoms (Fig. 6A). With the cores of Prp8 and Spp42 aligned, their RNaseH-like domains adopt different positions and are related to each other by a pseudo–two-fold rotational axis (fig. S17A), yet these two domains can be aligned with an RMSD of 1.24 Å for 225 Cα atoms (Fig. 6B). Similarly, the position of the RNaseH-like domain relative to the core of Spp42 in the *S. pombe* spliceosome is different from that in Prp8 bound to Aar2 (*15*, *28*). The C-terminal Jab1/MPN domain moves freely of the rest of Prp8 and mainly functions to regulate the activity of the ATPase/helicase Brr2 through physical association (*29*).

As with the RNaseH-like domain, the relative position of the Prp8 N domain is different from that of the Spp42 N domain (fig. S17B), yet these two domains can be perfectly aligned with a RMSD of 0.94 Å for 486 Cα atoms (Fig. 6C). Because the catalytic cavity of the *S. pombe* spliceosome is formed between the N domain and the core of Spp42/Prp8 (*15*, *16*), the relative movement between these two domains leads to a misaligned catalytic cavity in Prp8 of the U4/U6.U5 tri-snRNP (Fig. 6D). Such structural distortion generates misalignment of the U5 snRNA between the tri-snRNP and the *S. pombe* spliceosome (Fig. 6E, upper panel). The U5 snRNA in the tri-snRNP can be brought into registry with that in the *S. pombe* spliceosome by a rotation of ~30° (Fig. 6E, lower panel). The N domain of Prp8 associates with the GTPase Snu114 through a highly conserved interface, and alignment of the N domains between Prp8 and Spp42 only produced a slight misalignment between Snu114 and Cwf10 (fig. S18A) that

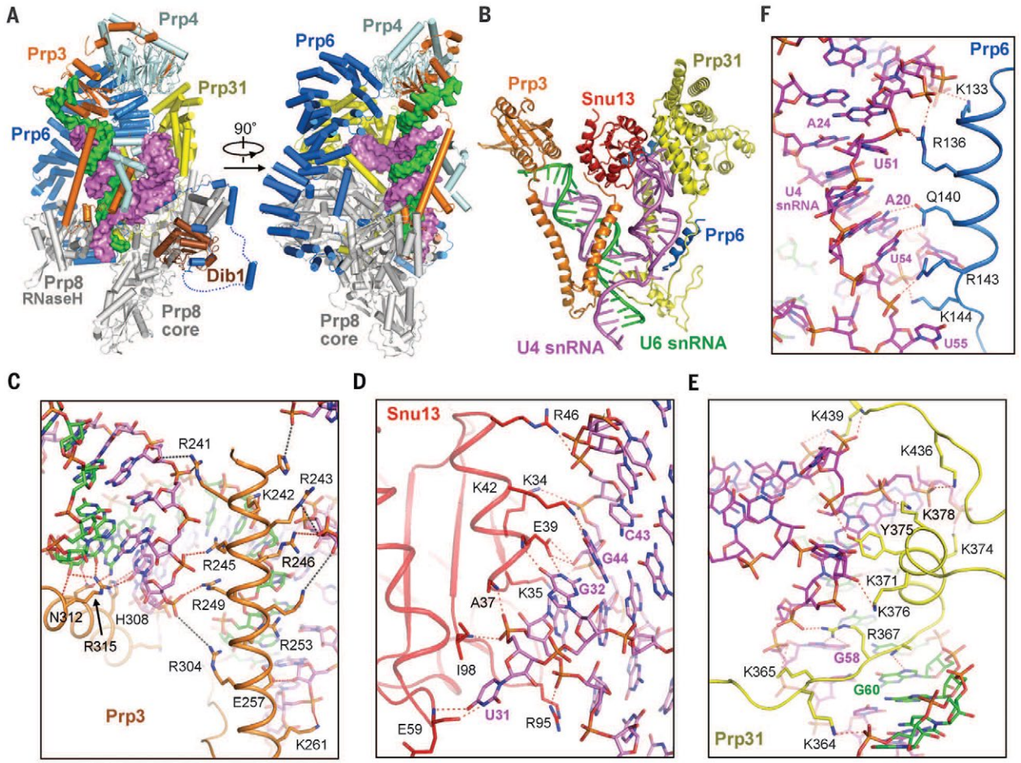


Fig. 4. Recognition of the U4/U6 snRNA duplex by spliceosomal proteins. (**A**) Overall structure of the U4/U6 snRNP core. The U4/U6 snRNA duplex and the intervening 5′-stem loop of U4 snRNA are surrounded by Prp3 (orange), Prp4 (light cyan), Prp6 (marine), Prp8 core (gray), Snu13 (red), Prp31 (yellow), and Dib1 (brown). (**B**) The U4/U6 snRNA duplex is mainly recognized by four proteins: Prp3, Prp6, Snu13, and Prp31. Prp3 interacts only with the U4/U6 snRNA duplex, whereas Prp6 uses two α-helices to contact the 5′-stem loop of U4 snRNA. Both Snu13 and Prp31 mainly associate with U4 snRNA. (**C**) An α-helix from Prp3 binds the major groove of stem II in the U4/U6 snRNA duplex. A few Arg and Lys residues may donate H bonds to the backbone phosphates of RNA duplex. The red and black dashed lines represent direct and water-mediated H bonds, respectively. (**D**) Positively charged amino acids from Snu13 may directly H-bond with backbone phosphates and the bases of U4 snRNA. (**E**) An extended loop from Prp31 interacts with the minor groove of stem I in the U4/U6 snRNA duplex. Arg^{367} may directly recognize the base G58 of U4 snRNA, whereas Lys^{371} probably H-bonds to G63 of U6 snRNA. (**F**) A close-up view of the interactions of an α-helix from Prp6 and the 5′-stem loop of U4 snRNA.

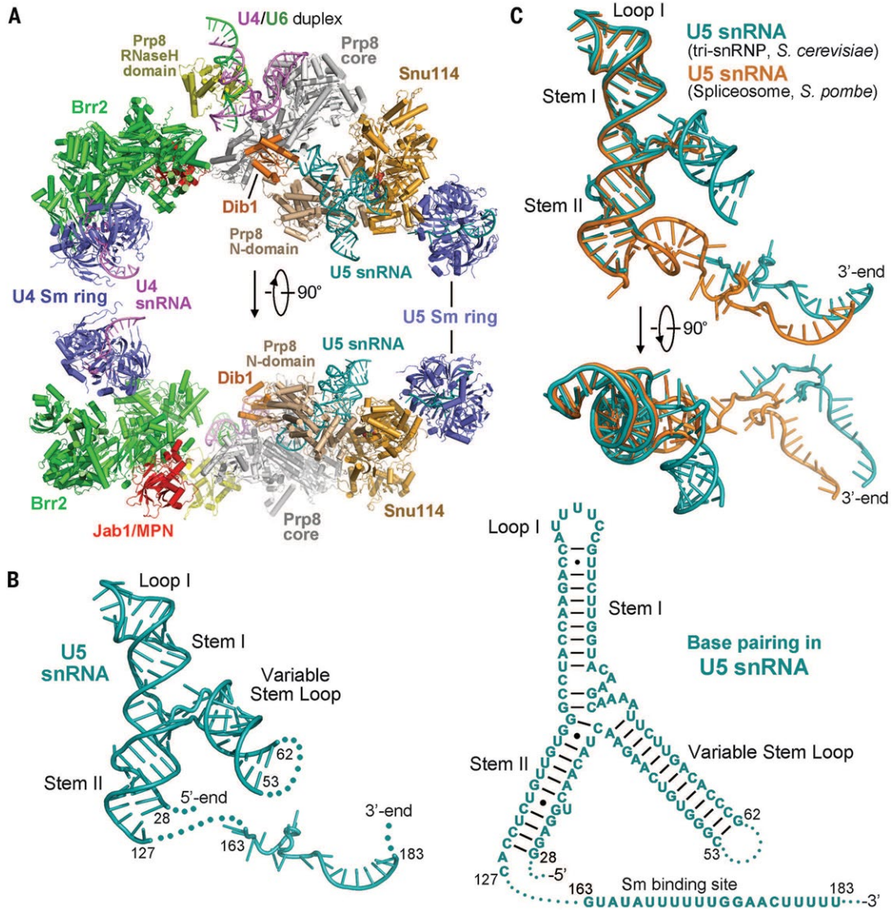


Fig. 5. Structures of U5 snRNP and U5 snRNA in the U4/U6.U5 tri-snRNP. (**A**) Overall structure of the U5 snRNP. The U4/U6 snRNA duplex and the U4 Sm ring are also shown for reference. The centrally located Prp8 is displayed in four colors: tan for the N domain, gray for the core, yellow for the RNaseH-like domain, and red for the Jab1/MPN domain. (**B**) Structure of U5 snRNA. A schematic representation of base-pairing interactions in U5 snRNA is shown in the right panel. (**C**) Structural comparison of U5 snRNAs from the *S. cerevisiae* tri-snRNP and *S. pombe* spliceosome. The two U5 snRNA molecules from U4/U6.U5 tri-snRNP and the *S. pombe* spliceosome are colored cyan and gold, respectively. Two perpendicular views are shown.

could be mitigated by minor rigid-body adjustment (fig. S18B). Compared with the GDP-bound Cwf10, binding of GTP (guanosine triphosphate) in Snu114 appears to induce no apparent conformational changes (fig. S18C).

Recognition of pre-mRNA by tri-snRNP

At the beginning of atomic modeling, we recognized a stretch of cryo-EM density close to Prp8 and loop I of U5 snRNA. The density is characteristic of RNA but cannot be assigned to U4, U5, or U6 snRNA because of topological considerations, suggesting the presence of pre-mRNA. After most components of the U4/U6.U5 tri-snRNP had been assigned, features of this density became clear, with some bulges projecting out from the linear-shaped density (fig. S13, G and H). Consideration of snRNA directionality and local density features only allowed one possible assignment for the pre-mRNA (Fig. 7A). The 5′SS of the intron is base-paired with the ACAGA box of U6 snRNA, whereas three consecutive nucleotides in the preceding 5′-exon sequences are recognized by loop I of U5 snRNA through base-pairing interactions (Fig. 7, B and C, and fig. S19).

Immediately after the 3′-end nucleotide of the 5′-exon, the first two bases of the 5′SS, guanine and uracil, protrude away from the extended pre-mRNA phosphodiester backbone (Fig. 7C and fig. S13H). The distinct configuration of the guanine base is maintained through five candidate H bonds with the side-chain amino group of Lys^{1378} and the main-chain groups of Gly^{1636} and Phe^{1623} in Prp8 (Fig. 7D). The extended conformation of the pre-mRNA sequence is probably a prerequisite for the first-step transesterification reaction involving an adenine nucleotide from the branch point sequence of the intron and two Mg^{2+} ions coordinated by U6 snRNA. The assignment of pre-mRNA also identifies Dib1 as a critical player in the U4/U6.U5 tri-snRNP, because it interacts simultaneously with Prp31, the N domain of Prp8, loop I of U5 snRNA, and 5′SS of pre-mRNA (Fig. 7E). Dib1 also directly contacts residues 1585 to 1598 of Prp8, hereafter termed the 1585 loop, which were found to play an important role in pre-mRNA splicing (*17*).

Implication for pre-mRNA splicing

The structure of pre-mRNA, characterized by its base-pairing interactions with both loop I of U5 snRNA and the ACAGA box of U6 snRNA, is indicative of a productive conformation that is poised for an impending transesterification reaction. To further examine this scenario, we aligned the N domains between Spp42 and Prp8, which brought U5 snRNA from the *S. pombe* spliceosome into registry with that in the tri-snRNP (Fig. 8A). If such an alignment matrix was applied to the entire *S. pombe* spliceosome, it would bring the catalytic Mg^{2+} ions and their coordinating nucleotides into close proximity with the guanine nucleotide at the 5′-end of 5′SS (Fig. 8A, inset). The catalytic metal 1 (M1) (*14*), which is known to stabilize the 3′-OH group of the 3′-end nucleotide of the 5′-exon, is positioned only about 2 Å away from the oxygen atom of the 3′-OH. In addition, the 1585 loop of Prp8 is positioned next to the guanine nucleotide and the catalytic metals. These observations suggest that the conformation of the pre-mRNA bound to tri-snRNP is ready for the first-step transesterification reaction. Our analysis also indicates that at least part of the tri-snRNP, which may include the N domain of Prp8/Spp42, Snu114/Cwf10, and U5 snRNA, already adopts a productive conformation for the splicing reaction. This conclusion is supported by the near-perfect alignment between these corresponding regions from *S. cerevisiae* tri-snRNP and *S. pombe* spliceosome (Fig. 8, B to D).

Discussion

X-ray crystallography on individual components or subcomplexes of the spliceosome has yielded

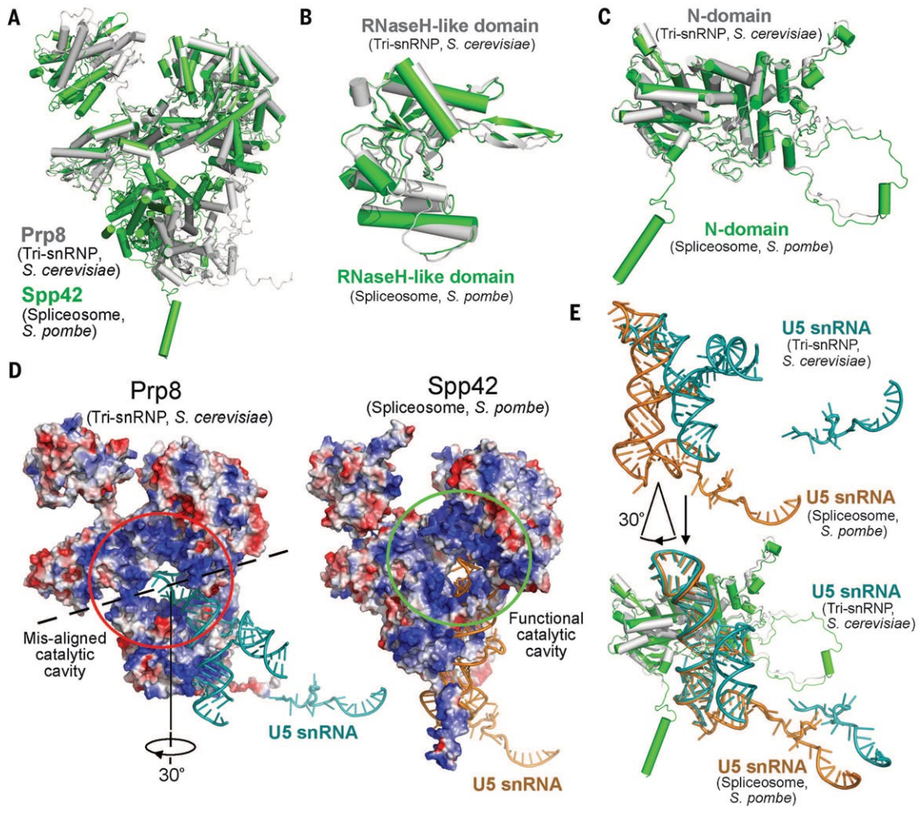


Fig. 6. Structure of Prp8 from tri-snRNP and its comparison with Spp42 from the *S. pombe* spliceosome. (**A**) The structure of Prp8 (gray) is aligned to that of Spp42 (green) on their respective core domains. In contrast to the near-perfect alignment of the core domains, the N and RNaseH-like domains are positioned differently. (**B**) The RNaseH-like and (**C**) N domains from Prp8 and Spp42 exhibit a similar conformation and can be aligned to each other. (**D**) Because of the rotation of the N domain relative to the core, Prp8 has a mis-aligned catalytic cavity (left panel) relative to Spp42 (right panel). (**E**) U5 snRNA of tri-snRNP can be aligned to U5 snRNA of the spliceosome by a rotation of ~30°.

structural information about U1 snRNP (*30*–*32*), U2 snRNP (*33*–*36*), U4 snRNP (*37*), U6 snRNP (*38*, *39*), Prp8 (*28*), and Brr2 (*29*, *40*). EM, on the other hand, has been used to probe the structure of both the human and yeast spliceosomes at various stages of the splicing reaction (*41*–*51*). These structures, mostly at moderate resolutions, have led to identification of global features of the spliceosome. The cryo-EM structure of the *S. cerevisiae* U4/U6.U5 tri-snRNP, at a resolution of 5.9 Å (*20*), allowed assignment of the components and identification of some secondary structural elements but not generation of an atomic model. Recently, we reported the first atomic structure of an intact spliceosome from *S. pombe* at 3.6 Å resolution, which reveals the fine-scale features of the pre-mRNA splicing machinery (*15*). In this study, we report the cryo-EM structure of the *S. cerevisiae* U4/U6.U5 tri-snRNP at an overall resolution of 3.8 Å and present an atomic model for this complex.

Spliceosomal complexes are notorious for their conformational and compositional heterogeneity, which underlies unsuccessful crystallization attempts. Compared to x-ray crystallography, single-particle cryo-EM analysis has the distinctive advantage of 2D and 3D classifications to effectively identify a subgroup of particles that share a similar conformation. In the case of U4/U6.U5 tri-snRNP, conformational heterogeneity is particularly severe, with flexible linkages between the Brr2 region and the tri-snRNP core. In the previous tri-snRNP structure (*20*), about 48% of the 347,241 total particles were used to generate the final cryo-EM map. In our current study, about 57% of the 299,993 total particles were used to generate the final cryo-EM map.

Although the global features of the cryo-EM maps from this study and the previous one (*20*) are similar, our cryo-EM maps reveal atomic details of the U4/U6.U5 tri-snRNP. The "head," "foot," and "arm" described in that report correspond to the three corners in our structure, with the foot including Snu114, U5 snRNA, the U5 Sm ring, and the N domain of Prp8. Because no atomic coordinates were reported for the previous tri-snRNP structure (*20*), we are unable to make a detailed comparison. In this study, we built protein components through either homology modeling or de novo building with most side chains assigned. The U4/U6 snRNA duplex is specifically assigned, and its interactions with surrounding proteins are elucidated.

Similar to the spliceosome (*15*, *16*), the structure of the U4/U6.U5 tri-snRNP reveals rich structural and mechanistic information. For example, Dib1 in *S. cerevisiae* (Dim1 in *S. pombe* and U5-15K in *H. sapiens*) plays an important role in pre-mRNA splicing, indicated by both its central location in the tri-snRNP and its association with U5 snRNA, pre-mRNA, Prp31, and the N domain and 1585 loop of Prp8 (Fig. 7E). Yet the function of Dib1 remains to be determined. This is echoed by Cwf19 in *S. pombe*, which is centrally located in the *S. pombe* spliceosome and interacts with U2 snRNA, U6 snRNA, and the RNaseH-like domain and the core of Spp42 (*15*, *16*), yet its function remains largely unknown. The enigmatic cases of Dib1 and Cwf19 apply to a number of other functionally unknown or uncertain spliceosomal proteins. The structural information provides a framework for functional and biochemical investigations.

An initially unexpected result in this study is the identification of pre-mRNA in the U4/U6.U5 tri-snRNP. The pre-mRNA–loaded tri-snRNP may represent an intermediate that recognizes pre-mRNA but still contains the extensively base-paired U4/U6 duplex. Consistent with this finding, analysis by reverse transcription polymerase chain reaction revealed the presence of *TUB3* pre-mRNA in the cryo-EM sample (fig. S20). Consideration of the tri-snRNP structural features in fact makes this finding unsurprising. First, the ACAGA box of U6 snRNA is exposed in the tri-snRNP and free to recognize the 5′SS of an intron in the pre-mRNA. Second, loop I of U5 snRNA is available in the tri-snRNP to interact with the 3′-end sequences of the 5′-exon in the pre-mRNA. Third, the linker domain of Prp8 is available to bind the

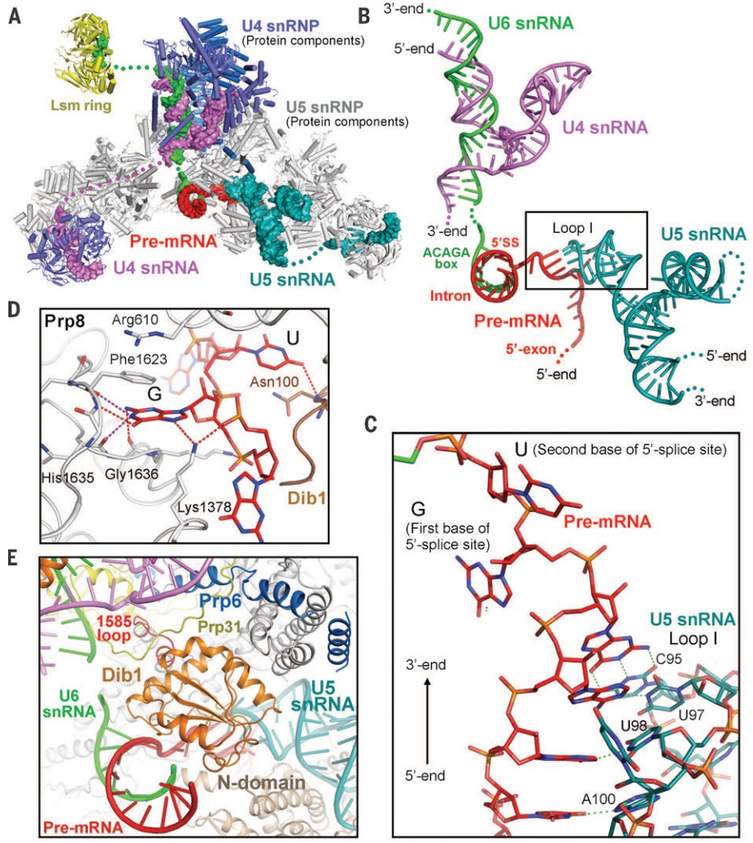


Fig. 7. Pre-mRNA is recognized by U6 snRNA and loop I of U5 snRNA in the U4/U6.U5 tri-snRNP. (**A**) Structure of the U4/U6.U5 tri-snRNP with pre-mRNA bound. All RNA components are shown in surface representation, with U4, U5, and U6 snRNAs colored violet, teal, and green, respectively. Pre-mRNA is highlighted in red. (**B**) The pre-mRNA is located in the center of tri-snRNP and forms duplexes with both U6 snRNA and U5 snRNA. For visual clarity, all protein components are stripped and only the RNA components in the center of tri-snRNA are displayed. The 5′SS of pre-mRNA is base-paired with the ACAGA box of U6 snRNA, whereas the 3′-end sequences of the 5′-exon are recognized by loop I of U5 snRNA. (**C**) A close-up view on the base-pairing interactions between the exon sequences and loop I of U5 snRNA. For base complementarity, the three consecutive nucleotides at the 3′-end of the 5′-exon were modeled as A-A-G (*55*), and they form a duplex with C95-U97-U98 of U5 snRNA. The first two bases of the 5′SS, guanine and uracil, are shown. (**D**) The guanine base of the first nucleotide in the 5′SS is specifically coordinated by amino acids in Prp8. The guanine base is recognized by the side chain of Lys^{1378} and the main chain groups of Phe^{1623} and Gly^{1636} through five putative H bonds. (**E**) Dib1 directly interacts with pre-mRNA, U5 snRNA, and the protein components Prp8 and Prp31. In particular, the N domain and the 1585 loop of Prp8 bind to Dib1.

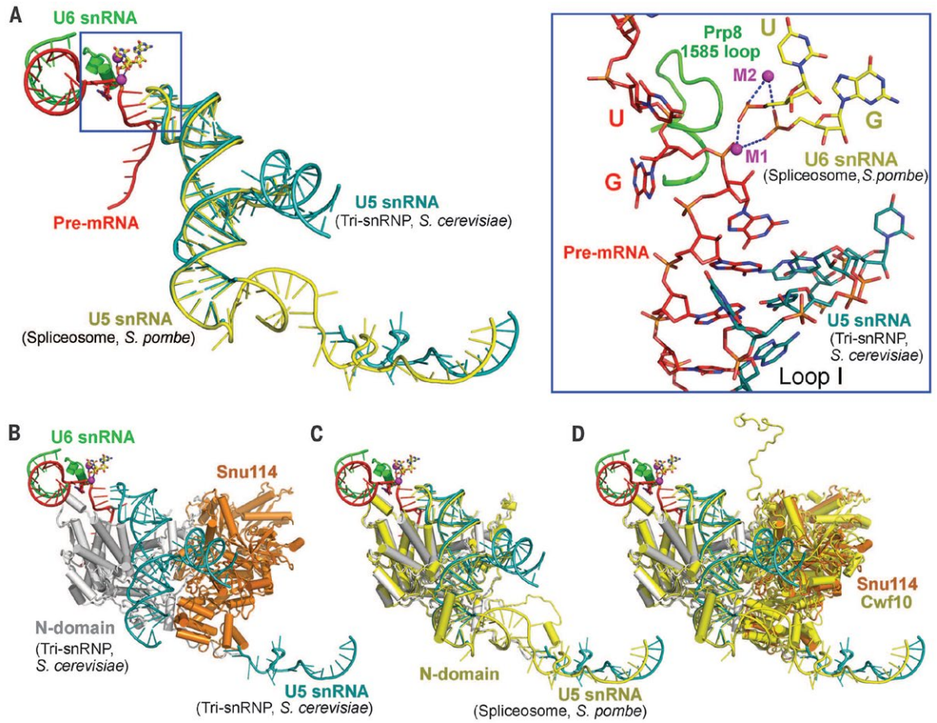


Fig. 8. Pre-mRNA in the U4/U6.U5 tri-snRNP is poised for the splicing reaction. (**A**) Superposition of the U4/U6.U5 tri-snRNP and the *S. pombe* spliceosome brings pre-mRNA in the tri-snRNP into close proximity to the Mg^{2+} ions in the catalytic center of the *S. pombe* spliceosome (*15*). The superposition matrix is the same as that between the N domains of Spp42 and Prp8. This alignment brings the catalytic Mg^{2+} ion (M1) to a distance of about 2 Å from the 3′-OH group of the ribose in the last nucleotide of the 5′-exon (close-up view). (**B**) A subcomplex comprising U5 snRNA, Prp8 N domain, and Snu114 probably constitutes the scaffold for the two-step splicing reaction. An overall structure of this subcomplex is shown here, together with those of pre-mRNA and a short stretch of U6 snRNA. The superimposed Mg^{2+} ions and their coordinating nucleotides from U6 snRNA in the *S. pombe* spliceosome are also shown for reference. (**C**) Superposition of the N domains of Prp8 and Spp42 brings the two U5 snRNA molecules in tri-snRNP and *S. pombe* spliceosome into registry. (**D**) Superposition of the N domains of Prp8 and Spp42 brings Snu114 from tri-snRNP and Cwf10 from *S. pombe* spliceosome into registry.

5′-nucleotide of the 5′SS. Fourth and last, the back side of Prp8 is strongly positively charged (fig. S21A), so it may bind and orient the 5′-exon sequences. Some of the cryo-EM density located in this region appears to be connected to the 5′-end of the pre-mRNA (fig. S21B). Nevertheless, although unlikely, we cannot rule out the possibility that the pre-mRNA–loaded tri-snRNP may represent part of the B complex.

The current understanding of spliceosomal assembly suggests that the spliceosomal A complex, which contains pre-mRNA loaded with U1 and U2 snRNPs, associates with U4/U6.U5 tri-snRNP to form the spliceosomal B complex (fig. S22, red arrows). According to this model, the U4/U6.U5 tri-snRNP is free of pre-mRNA and exists as an inhibitory complex by keeping U6 snRNA in an inactive conformation. Our structural finding suggests additional possibilities (fig. S22, black arrows). In the tri-snRNP, the ACAGA box of U6 snRNA and loop I of U5 snRNA are both free to engage pre-mRNA, and they do so through base-pairing interactions with both the 5′SS of an intron and the 3′-end sequences of the 5′-exon. Thus the U4/U6.U5 tri-snRNP may freely recruit pre-mRNA, independently of U1 snRNP (fig. S22). Our speculative model further predicts that the tri-snRNP loaded with pre-mRNA may directly associate with U2 snRNP and proceed to form a catalytically competent spliceosome (fig. S22). Consistent with this prediction, most protein components of U2 snRNP were identified by mass spectrometry in our sample and exhibited relatively high peptide-spectrum match values, suggesting a reasonable abundance (fig. S23). In contrast, the protein components of U1 snRNP were present with considerably less abundance (fig. S23). Supporting our model, direct recognition of the 5′SS in the pre-mRNA by the U4/U6.U5 tri-snRNP has been previously reported (*52*).

Despite these clues and analyses, our model awaits experimental scrutiny. This speculative model may be inconsistent with some of the reported biochemical data. For example, using an in vitro purification method, inactivation of pre-mRNA binding by U1 snRNP was shown to nearly cripple pre-mRNA binding by all other snRNPs (*53*). However, such studies were performed under highly specific settings and stringent analysis conditions, such as those for detection of snRNA species in stalled splicing reactions, and thus may not fully capture the complex situations in cells.

The molecular choreography of many different components, exemplified by that of Prp8 and Spp42 (Fig. 6) and U6 snRNA (fig. S24), serves to execute the splicing reactions for tens of thousands of distinct pre-mRNA molecules. The near-atomic structures of the *S. pombe* spliceosome (*15*) and the *S. cerevisiae* tri-snRNP provide a principal framework for ultimately elucidating the underlying molecular mechanisms of pre-mRNA splicing.

REFERENCES AND NOTES

1. S. M. Berget, C. Moore, P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171–3175 (1977).
2. L. T. Chow, R. E. Gelinas, T. R. Broker, R. J. Roberts, *Cell* **12**, 1–8 (1977).

3. C. B. Burge, T. Tuschl, P. A. Sharp, "Splicing of precursors to mRNAs by the spliceosomes," in *The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Monograph vol. 37, Cold Spring Harbor Laboratory Press, 1999).
4. M. J. Moore, C. C. Query, P. A. Sharp, "Splicing of precursors to mRNA by the spliceosome," in *The RNA World*, R. F. Gesteland, J. F. Atkins, Eds. (Cold Spring Harbor Monograph vol. 24, Cold Spring Harbor Laboratory Press, 1993).
5. D. A. Wassarman, J. A. Steitz, *Science* **257**, 1918–1925 (1992).
6. C. L. Will, R. Lührmann, *Cold Spring Harbor Perspect. Biol.* **3**, a003707 (2011).
7. M. C. Wahl, C. L. Will, R. Lührmann, *Cell* **136**, 701–718 (2009).
8. W. Chen, M. J. Moore, *Curr. Opin. Struct. Biol.* **24**, 141–149 (2014).
9. A. Hegele *et al.*, *Mol. Cell* **45**, 567–580 (2012).
10. T. A. Steitz, J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498–6502 (1993).
11. E. J. Sontheimer, S. Sun, J. A. Piccirilli, *Nature* **388**, 801–805 (1997).
12. P. M. Gordon, E. J. Sontheimer, J. A. Piccirilli, *RNA* **6**, 199–205 (2000).
13. S. L. Yean, G. Wuenschell, J. Termini, R. J. Lin, *Nature* **408**, 881–884 (2000).
14. S. M. Fica *et al.*, *Nature* **503**, 229–234 (2013).
15. C. Yan *et al.*, *Science* **349**, 1182–1191 (2015).
16. J. Hang, R. Wan, C. Yan, Y. Shi, *Science* **349**, 1191–1198 (2015).
17. W. P. Galej, T. H. Nguyen, A. J. Newman, K. Nagai, *Curr. Opin. Struct. Biol.* **25**, 57–66 (2014).
18. B. Sander *et al.*, *Mol. Cell* **24**, 267–278 (2006).
19. I. Häcker *et al.*, *Nat. Struct. Mol. Biol.* **15**, 1206–1212 (2008).
20. T. H. Nguyen *et al.*, *Nature* **523**, 47–52 (2015).
21. F. Galisson, P. Legrain, *Nucleic Acids Res.* **21**, 1555–1562 (1993).
22. A. Gottschalk *et al.*, *EMBO J.* **18**, 4535–4548 (1999).
23. E. C. Small, S. R. Leggett, A. A. Winans, J. P. Staley, *Mol. Cell* **23**, 389–399 (2006).
24. S. Liu *et al.*, *eLife* **4**, e07320 (2015).
25. S. Liu *et al.*, *Science* **316**, 115–120 (2007).
26. K. Reuter, S. Nottrott, P. Fabrizio, R. Lührmann, R. Ficner, *J. Mol. Biol.* **294**, 515–525 (1999).
27. H. C. Dobbyn *et al.*, *Biochem. Biophys. Res. Commun.* **360**, 857–862 (2007).
28. W. P. Galej, C. Oubridge, A. J. Newman, K. Nagai, *Nature* **493**, 638–643 (2013).
29. S. Mozaffari-Jovin *et al.*, *Science* **341**, 80–84 (2013).
30. G. Weber, S. Trowitzsch, B. Kastner, R. Lührmann, M. C. Wahl, *EMBO J.* **29**, 4172–4184 (2010).
31. D. A. Pomeranz Krummel, C. Oubridge, A. K. Leung, J. Li, K. Nagai, *Nature* **458**, 475–480 (2009).
32. Y. Kondo, C. Oubridge, A. M. van Roon, K. Nagai, *eLife* **4**, e04986 (2015).
33. S. R. Price, P. R. Evans, K. Nagai, *Nature* **394**, 645–650 (1998).
34. E. A. Sickmier *et al.*, *Mol. Cell* **23**, 49–59 (2006).
35. P. C. Lin, R. M. Xu, *EMBO J.* **31**, 1579–1590 (2012).
36. J. L. Jenkins, A. A. Agrawal, A. Gupta, M. R. Green, C. L. Kielkopf, *Nucleic Acids Res.* **41**, 3859–3873 (2013).
37. A. K. Leung, K. Nagai, J. Li, *Nature* **473**, 536–539 (2011).
38. L. Zhou *et al.*, *Nature* **506**, 116–120 (2014).
39. E. J. Montemayor *et al.*, *Nat. Struct. Mol. Biol.* **21**, 544–551 (2014).
40. T. H. Nguyen *et al.*, *Structure* **21**, 910–919 (2013).
41. N. Behzadnia *et al.*, *EMBO J.* **26**, 1737–1748 (2007).
42. D. Boehringer *et al.*, *Nat. Struct. Mol. Biol.* **11**, 463–468 (2004).
43. E. Wolf *et al.*, *EMBO J.* **28**, 2283–2292 (2009).
44. J. Deckert *et al.*, *Mol. Cell. Biol.* **26**, 5528–5543 (2006).
45. S. Bessonov *et al.*, *RNA* **16**, 2384–2403 (2010).
46. M. M. Golas *et al.*, *Mol. Cell* **40**, 927–938 (2010).
47. M. S. Jurica, D. Sousa, M. J. Moore, N. Grigorieff, *Nat. Struct. Mol. Biol.* **11**, 265–269 (2004).
48. J. O. Ilagan, R. J. Chalkley, A. L. Burlingame, M. S. Jurica, *RNA* **19**, 400–412 (2013).
49. P. Fabrizio *et al.*, *Mol. Cell* **36**, 593–608 (2009).
50. M. D. Ohi, L. Ren, J. S. Wall, K. L. Gould, T. Walz, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3195–3200 (2007).
51. W. Chen *et al.*, *RNA* **20**, 308–320 (2014).
52. P. A. Maroney, C. M. Romfo, T. W. Nilsen, *Mol. Cell* **6**, 317–328 (2000).
53. S. W. Ruby, J. Abelson, *Science* **242**, 1028–1035 (1988).
54. E. F. Pettersen *et al.*, *J. Comput. Chem.* **25**, 1605–1612 (2004).
55. E. Bon *et al.*, *Nucleic Acids Res.* **31**, 1121–1135 (2003).

ACKNOWLEDGMENTS

We thank X. Fang for technical assistance and the Tsinghua University Branch of the China National Center for Protein Sciences (Beijing) for access to the EM facility. We also thank the Explorer 100 cluster system of the Tsinghua National Laboratory for Information Science and Technology, the Computing Platform of the China National Center for Protein Sciences, and Lenovo for providing high-performance computing. This work was supported by funds from the Ministry of Science and Technology (grant 2014ZX09507003006) and the National Natural Science Foundation of China (grants 31430020 and 31321062). The atomic coordinates have been deposited in the Protein Data Bank under the accession code 3JCM. The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with the accession codes EMD-6561 for the overall map and EMD-6562 to EMD-6573 for the 12 local maps. The authors declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/351/6272/466/suppl/DC1
Figs. S1 to S25
Tables S1 to S3
References (*56*–*80*)

15 October 2015; accepted 24 December 2015
Published online 7 January 2016
10.1126/science.aad6466

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Molecular architecture of the human U4/U6.U5 tri-snRNP

Dmitry E. Agafonov,[1]* Berthold Kastner,[1]* Olexandr Dybkov,[1]* Romina V. Hofele,[2,3]† Wen-Ti Liu,[4,5] Henning Urlaub,[2,3]‡ Reinhard Lührmann,[1]‡ Holger Stark[4,5]‡

The U4/U6.U5 triple small nuclear ribonucleoprotein (tri-snRNP) is a major spliceosome building block. We obtained a three-dimensional structure of the 1.8-megadalton human tri-snRNP at a resolution of 7 angstroms using single-particle cryo–electron microscopy (cryo-EM). We fit all known high-resolution structures of tri-snRNP components into the EM density map and validated them by protein cross-linking. Our model reveals how the spatial organization of Brr2 RNA helicase prevents premature U4/U6 RNA unwinding in isolated human tri-snRNPs and how the ubiquitin C-terminal hydrolase–like protein Sad1 likely tethers the helicase Brr2 to its preactivation position. Comparison of our model with cryo-EM three-dimensional structures of the *Saccharomyces cerevisiae* tri-snRNP and *Schizosaccharomyces pombe* spliceosome indicates that Brr2 undergoes a marked conformational change during spliceosome activation, and that the scaffolding protein Prp8 is also rearranged to accommodate the spliceosome's catalytic RNA network.

The spliceosome is formed stepwise by recruitment of the U1 and U2 snRNPs (small nuclear ribonucleoproteins) and the U4/U6.U5 tri-snRNP, plus numerous other proteins, to the pre-mRNA (*1*). Initially, U1 and U2 interact with the pre-mRNA's 5′ splice site (SS) and branch site (BS), respectively, generating the A complex. The tri-snRNP then joins, leading to formation of the precatalytic spliceosomal B complex. Subsequent catalytic activation of the spliceosome involves major structural rearrangements of multiple tri-snRNP components (*1*).

The 1.8-MDa tri-snRNP is the largest preformed building block of the human spliceosome. It contains three snRNA molecules (U4, U6, and U5), two heteroheptameric rings of Sm proteins bound to the U4 and U5 snRNAs' 3′-terminal Sm sites, the LSm ring bound to the 3′ end of U6 snRNA, plus 16 additional proteins (*1*) (fig. S1). In the tri-snRNP and B complex, U4 and U6 snRNA are extensively base-paired. During activation, the U4/U6 duplex is disrupted and a highly structured RNA interaction network forms among the U2, U6, and U5 snRNAs and the pre-mRNA, generating the spliceosome's catalytic RNA core (*2*, *3*). Three large U5 proteins—Prp8, the RNA helicase Brr2, and the guanosine triphosphatase (GTPase) Snu114—play key roles during catalytic activation. Prp8 is a major scaffolding protein that interacts with Brr2 and Snu114 (*4*) and all reactive sites of the intron (5′SS, 3′SS, and BS) and is thus located at the heart of the spliceosome's catalytic core (*5*, *6*). Brr2 unwinds the U4/U6 snRNA helices and is the major driving force for catalytic activation (*7*, *8*). However, as Brr2 and its RNA substrate are present in the tri-snRNP and precatalytic B complex, a mechanism must exist to prevent premature dissociation of the U4/U6 helices by Brr2.

Here, we report a 3D cryo–electron microscopy (cryo-EM) structure of the human tri-snRNP at a resolution of 7 Å and resolve its spatial organization with the aid of protein cross-linking. Comparison with the recently reported cryo-EM structure of the yeast tri-snRNP (*9*) reveals unexpected, large differences in the position of the helicase Brr2, including its position relative to its RNA substrate, the U4/U6 duplex. Our model also reveals the nature of tri-snRNP rearrangements that must occur during spliceosome maturation, including a major conformational change within the Prp8 protein, which adopts an open conformation in the human tri-snRNP and a closed one in the *Schizosaccharomyces pombe* spliceosome at late stages of splicing (*10*).

Structure determination and model building

Human tri-snRNPs were affinity-purified from HeLa nuclear extract and prepared for cryo-EM by a modification of the GraFix protocol involving chemical cross-linking of the particles (fig. S1B) (*11*). The 3D structure was determined from ~141,000 particle images after several steps of computational sorting, starting with an initial data set of ~1,150,000 selected particle images (fig. S2). The calculated 3D structure of the tri-snRNP was determined at a final overall resolution of 7 Å with better-resolved parts in the center and somewhat lower-resolution areas in the U4/U6 part of the structure (fig. S3). Overall, the structure is entirely consistent with an earlier, lower-resolution 3D structure (*12*) showing the tri-snRNP as a roughly tetrahedral particle with dimensions of approximately 300 Å × 200 Å × 175 Å (Fig. 1). At this resolution, structured protein domains and double-stranded RNA (dsRNA) elements can be identified clearly, allowing us to fit known x-ray structures or homology models of structured regions of tri-snRNP components into the EM density map (see table S1 for details regarding how proteins were fit into the EM density map). Additionally, we performed chemical protein cross-linking of purified tri-snRNPs together with mass spectrometry (CX-MS) (table S2). These data allowed us to validate the locations of large tri-snRNP proteins and facilitated docking of smaller proteins. Although we could place all snRNAs and structured protein domains in the EM density map in a manner consistent with our protein cross-linking data, ~30% of the calculated stoichiometric mass of human tri-snRNP proteins are very likely intrinsically unstructured regions that could not be localized (table S1).

Structural organization of the U5 Sm core and the U4/U6 snRNP

The helical regions of U4/U6 and U5 snRNA allowed their unambiguous placement in the EM density map (Fig. 1). The U5 Sm core is located at the lower tip of the tri-snRNP, with the 5′-terminal m_3G cap of U5 snRNA positioned close to it, whereas U5 loop 1 is located more centrally and stems 1b and 1c are coaxially stacked (Fig. 1B). The U4/U6 snRNAs are located in the upper, broader region of the human tri-snRNP. Their dsRNA regions are connected by a three-way junction and are located in a deeper, solvent-accessible cleft. The difference in length of U4/U6 stems I and II and the clearly visible three-way junction define the orientation of U4/U6 snRNA in the model and indicate coaxial stacking of stems I and II (Fig. 1). The U4/U6 snRNAs also define the positions of the U4 Sm and U6 LSm protein rings, which are found at two corners in the upper part of the tri-snRNP (Fig. 1B).

The geometry of the U4/U6 snRNA three-way junction allowed us to fit the crystal structures of (i) the U4 snRNA 5′ stem-loop in complex with Snu13 and a large part of the U4/U6 Prp31 protein, (ii) a large part of Prp3 (Prp3-CTF) in complex with U4/U6 stem II and the U6 single-stranded 3′ overhang, (iii) the WD40 domain of the Prp3-associated Prp4 protein, and (iv) the cyclophilin H (CypH) protein into nearby density elements (Fig. 2A). The position of the various U4/U6 proteins was confirmed by protein-protein cross-linking (fig. S4). There is an overall similarity in the organization of U4/U6 proteins in human and yeast tri-snRNPs, with differences in the architectural details of some proteins (fig. S5) (see below).

The Prp6 protein contains 19 tetratrico repeats (TPRs) in its C-terminal region and is required for stable tri-snRNP formation (*13*, *14*). Consistent with this, Prp6 forms a bridge across a deep cleft at the top of the tri-snRNP that connects the U4/U6 and U5 snRNPs (Fig. 2B). This is supported by numerous cross-links, whereby Prp6's N-terminal and C-terminal TPRs exclusively form cross-links to U5 and U4/U6 proteins, respectively. Consistent with intramolecular cross-links between TPR 19

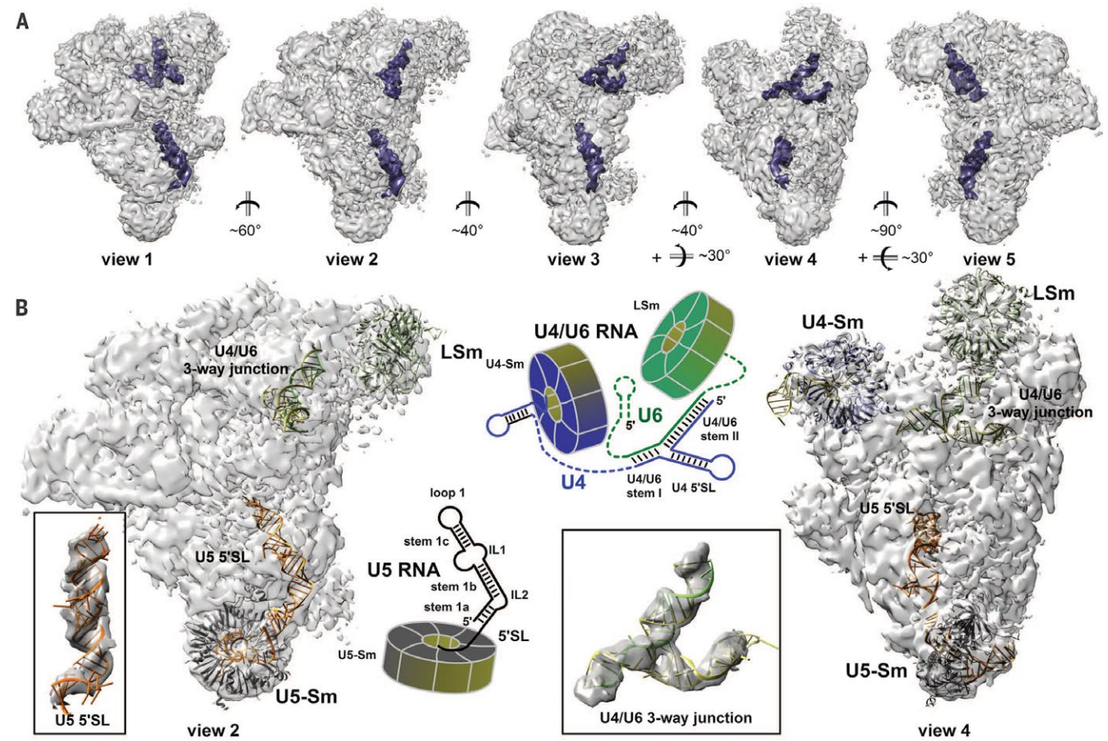


Fig. 1. Three-dimensional cryo-EM structure of the human U4/U6.U5 tri-snRNP and location of U5 and U4/U6 snRNAs and their Sm/LSm cores. (**A**) Different views of the tri-snRNP EM density map with helical high-density elements (blue) representing U5 (in lower region) and U4/U6 snRNA (upper region). (**B**) Position of the U5 Sm, U4 Sm, and U6 LSm cores. A schematic of U5 and U4/U6 snRNA with their Sm/LSm rings is shown. The double-stranded regions of U4/U6 and U5, and their heptameric Sm/LSm rings, are modeled into the cryo-EM map. Insets: RNA elements shown separately.

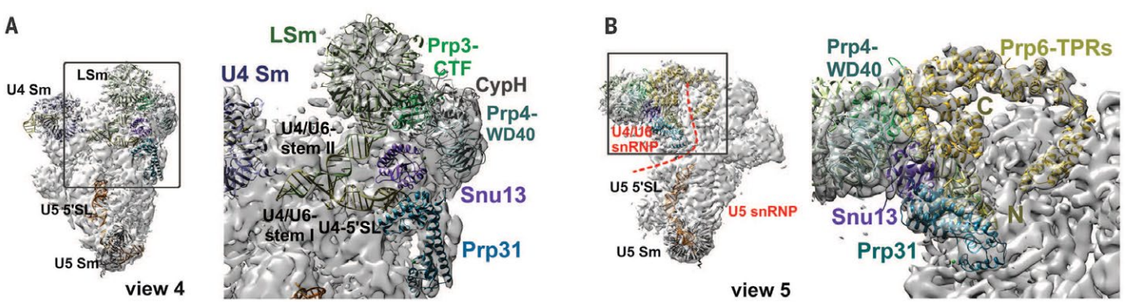


Fig. 2. Structural organization of U4/U6 proteins and Prp6 and their locations in the human tri-snRNP. (**A**) Positions of the U4/U6 proteins and snRNA. Right: Expanded view of boxed region showing the U4/U6 snRNA three-way junction, the crystal structures of Prp31, Snu13, CypH, and the C-terminal fragment (CTF) of Prp3, and a modeled structure of Prp4's WD40 domain, fit into the EM density map. (**B**) Prp6 forms a bridge connecting the U4/U6 and U5 snRNPs. Right: Expanded view of boxed region showing Prp6 TPR repeats and U4/U6 proteins.

[1]Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany. [2]Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany. [3]Bioanalytics Group, Institute for Clinical Chemistry, University Medical Center Göttingen, D-37075 Göttingen, Germany. [4]Department of 3D Electron Cryomicroscopy, Georg-August Universität Göttingen, D-37077 Göttingen, Germany. [5]Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany.
*These authors contributed equally to this work. †Present address: Medimmune, Gaithersburg, MD, USA. ‡Corresponding author. E-mail: reinhard.luehrmann@mpi-bpc.mpg.de (R.L.); hstark1@gwdg.de (H.S.); henning.urlaub@mpibpc.mpg.de (H.U.)

and TPRs 9 to 13 (fig. S4), the C-terminal TPRs fit as a circularly arranged ensemble into a large ringlike density that is connected to U4/U6 proteins (Fig. 2B).

The architecture of Snu114 and Prp8

Aside from its 115-residue N-terminal domain, the 116-kDa Snu114 protein is highly homologous to ribosomal elongation factor EF-2/EF-G (*15*), and we could fit domains D1 to D5 of Snu114 into the lower part of the tri-snRNP, with D1 and D2 located closer to the U5 Sm core and D3 to D5 located more centrally (Fig. 3A). Thus, in the isolated human tri-snRNP, Snu114 adopts a compact form similar to the compact structure of EF-2 (fig. S6) (*16*).

The crystal structure of a large fragment of yeast Prp8 (~110 kDa) containing a reverse transcriptase (RT)–like domain, connected through a linker region to a restriction endonuclease (En)–like domain, fits into a central density element at the base of the upper part of the tri-snRNP; the En domain points outward and is positioned below the U4 Sm core (Fig. 3A). Prp8's C-terminal RNase H (RH)–like domain could be docked into a density element located just above the linker region of the RT/En domain (Fig. 3A), and its orientation was confirmed by cross-linking (tables S1 and S2). The architecture of Prp8's RT/En domain and its position are essentially the same in the human and yeast tri-snRNP models, whereas Prp8's RH domain is rotated by ~180° in yeast relative to human (fig. S7) (*9*).

In the *S. pombe* spliceosome, Prp8's N-terminal 800 amino acids consist of two domains, henceforth termed NTD1 and NTD2, that contain mainly α helices and are separated by a short linker region, termed NTDL (Fig. 3A) (*10*). The larger NTD1 structure fits into a density element in the lower part of the tri-snRNP model and has a substantial interface with Snu114 and also contacts stem1 of U5 snRNA (Fig. 3A) (see below). Consistent with our cross-linking data, the smaller NTD2 is located more toward the U4/U6 three-way junction and interacts with Prp8's RT domain (Fig. 3B and fig. S8A). The crystal structure of Dim1 fits into a density element between NTD1 and NTD2 (Fig. 3B), a position supported by cross-linking (fig. S8A). The overall structure of Prp8's NTD1 and Snu114 is similar in the human tri-snRNP and *S. pombe* spliceosome, including the lasso-like protrusion of NTD1 that interacts with Snu114's D1 domain in a similar manner in both complexes (fig. S8B) (*10*). Guided by multiple cross-links of the RecA2 domain of the Prp28 helicase to Prp8's NTD1 and RT/En domains (fig. S8A), we could fit the crystal structure of the two RecA domains into nearby density elements (Fig. 3C). Prp28, which is not present in isolated *Saccharomyces cerevisiae* tri-snRNPs, exists in an open (inactive) conformation, very similar to its conformation in the crystal structure of isolated Prp28 (*17*, *18*). Finally, we could place the WD40 domain of U5-40K, which is conserved in *S. pombe* but not *S. cerevisiae*, into a density element close to U5's ILS1 (fig. S8C)—a position where it is also found in the *S. pombe* spliceosome (*10*).

Brr2 helicase is found at very different positions in human and yeast tri-snRNPs

The 245-kDa RNA helicase Brr2 contains two tandemly organized helicase cassettes, but only the N-terminal cassette (NC) actively unwinds the U4/U6 duplex during catalytic activation (*19*). The C-terminal Prp8-Jab1 domain binds tightly to Brr2's active NC and regulates its helicase activity (*20*, *21*). The crystal structure of the complete 200-kDa helicase unit of Brr2 bound to Prp8 Jab1 fits very well, as a rigid body, into a major density element in the upper part of the tri-snRNP, near the RT end of the Prp8-RT/En domain, opposite the U4 Sm and U6 LSm rings (Fig. 4A).

Besides this Brr2 NC–Prp8-Jab1 interaction, there appear to be at least two additional density elements connecting the helicase cassettes to other tri-snRNP proteins (Fig. 4, A to C). The N-terminal region of Brr2 contains a noncanonical PWI domain (*22*) and a helical domain (*23*). The PWI domain fits into the density element connecting Brr2's C-terminal cassette (CC) to Snu114 and Sad1 (Fig. 4B), while the N-terminal helical domain (NHD) fits into a density element connecting Brr2's NC to Prp8's RH domain and to the N-terminal–most three TPR repeats of Prp6 (Fig. 4C and table S1). Interestingly, Brr2's NHD is located in front of the RNA binding channel between the RecA2 and helical bundle (HB) domain of Brr2's NC (Fig. 4C), consistent with this element acting like a plug, autoinhibiting Brr2 via substrate recognition (*23*). Brr2's architecture and its connections to the above-mentioned proteins were confirmed by a network of cross-links between Brr2's NHD and NC/CC domains, and between these domains and the Prp8's RH and Jab1 domains, as well as Prp6's N-terminal TPRs (fig. S9), and additionally between Brr2's PWI and CC domains and the Snu114 and Sad1 proteins (fig. S10).

Strikingly, Brr2 is located at radically different positions in the human and yeast tri-snRNP models (Fig. 4D). Human (h)Brr2 (bound to hPrp8 Jab1) is located close to the N-terminal TPR repeats of Prp6 and Prp8's RT end, and its general position in the tri-snRNP is not dependent on the use of a chemical cross-linking reagent during EM sample preparation (fig. S11). In contrast, yeast (y)Brr2 (bound to the yPrp8 Jab1 domain) is found near yPrp8's En domain, which is ~20 nm away from the position of hBrr2. In addition, it is rotated by ~180° around the long axis of the tri-snRNP (Fig. 4D and fig. S12). In the yeast model, yBrr2 appears to be connected to the tri-snRNP primarily via the yPrp8 Jab1 domain (which contacts the tip of yPrp8's EN domain) and the U4 Sm core (Fig. 4D and fig. S12) (*9*). Unfortunately, because of the less well-defined density at the interface between yBrr2 and other tri-snRNP proteins, the locations of yBrr2's N-terminal PWI and helical domains cannot be identified in the yeast structure. Another striking difference is that the yeast U4 Sm core is located at the interface between yBrr2's helicase cassettes, and the central single-stranded region of U4 snRNA, to which Brr2 is thought to dock prior to unwinding the U4/U6 duplex (fig. S1C) (*24*), is positioned at/near the RecA domains of the active NC of yBrr2 (Fig. 4D) (*9*). In contrast, in the human tri-snRNP structure, hBrr2's active NC is located 8 to 10 nm away from its U4/U6 snRNA substrate (Fig. 4D).

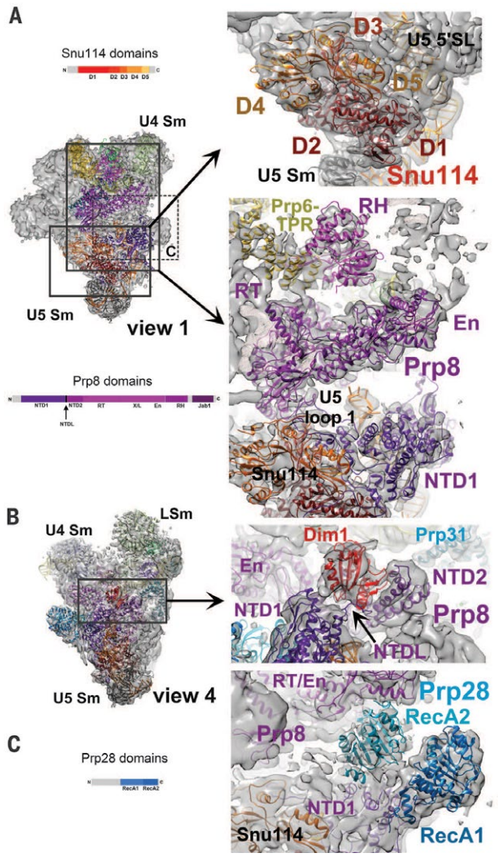


Fig. 3. Structures and positions of Snu114, Prp8, Dim1, and Prp28 in the human tri-snRNP. (**A**) Location and structural organization of Snu114 and Prp8. Left: Organization of Snu114 (domains D1 to D5 homologous to EF-G/EF-2) and Prp8 (NTD1 and NTD2, N-terminal domains 1 and 2; NTDL, NTD linker; RT, reverse transcriptase–like; X/L, linker; En, endonuclease-like; RH, RNase H–like; Jab1, Jab1/MPN-like). Upper right: Fit of Snu114 domains D1 to D5 as a compact structure. Lower right: Fit of Prp8 NTD1, RT/En, and RH domains, with front global clipping to improve Prp8 visibility. (**B**) U5 Dim1 between Prp8's NTD1 and NTD2 domains (extended view at right). (**C**) Structural organization of the RNA helicase Prp28. Left: Domain organization of Prp28. Right: Expanded view of dashed box labeled C in view 1 of (A), showing Prp28's RecA domains fit into two neighboring density elements.

Structural basis for how Sad1 likely tethers Brr2 in a preactivation position

The very different position of Brr2 suggests either that there is a substantial difference in the spatial organization of the yeast and human tri-snRNPs, or potentially that the human and yeast structures represent two different conformational states that are obtained by rearrangements in protein architecture. Although the first possibility cannot be rigorously excluded, we consider it unlikely, as the structures of Brr2 and all other major U5 proteins are evolutionarily highly conserved between yeast and human (*1*, *5*). Instead, differences in the protein composition of the purified human and yeast tri-snRNPs potentially lead to different conformations. That is, in the presence of adenosine triphosphate (ATP), isolated human tri-snRNPs are stable, whereas yeast tri-snRNPs are not (*9*, *25*, *26*). This is likely because the evolutionarily conserved Sad1 protein is stoichiometrically present in purified human tri-snRNPs (*25*) but is lost during purification of the yeast complex (*26*, *27*). Sad1 plays a key role in stabilizing the tri-snRNP, as depletion of Sad1 from yeast cell extracts leads to dissociation of the otherwise stable tri-snRNP in an ATP- and Brr2-dependent manner into a U4/U6 di-snRNP (where U6 and U4 are still base-paired) and U5 snRNP (*28*). Consistent with it contributing to tri-snRNP stability, human Sad1 is located at a strategically important position at the interface between the U4/U6 and U5 snRNPs. The Sad1 UCH domain contacts U4/U6-Prp31 and the Prp8 NTD2 and RT domains, whereas Sad1's Zf-UBP domain has a substantial interface with domains D2, D3, and D4 of Snu114 and is tightly connected to Brr2's PWI domain (Fig. 5, table S1, and fig. S10).

Thus, Sad1 not only potentially acts as a clamp stabilizing the interaction of U4/U6 and U5, it might also help to tether Brr2 in a preactivation position (i.e., away from the U4/U6 duplex) within the human tri-snRNP. This in turn suggests that dissociation of Sad1—as observed during activation of the human B complex (*1*)—might allow Brr2 to undergo a major conformational change that is required for it to interact with its U4/U6 snRNA substrate. Because Sad1 is absent from purified yeast tri-snRNPs, the very different position of Brr2 in the yeast tri-snRNP may therefore represent a conformational state similar to the one that Brr2 normally adopts at a later stage during spliceosome activation. Whereas the yeast cryo-EM model lacks density in the corresponding regions where Sad1 and Brr2 are located in the human tri-snRNP structure, the crystal structures of Sad1 and Brr2 can be docked well onto the surface of the yeast tri-snRNP at the corresponding positions (fig. S13). It will be of interest to determine the 3D structure of the yeast tri-snRNP in the presence of ySad1.

Remodeling of the human tri-snRNP during spliceosome assembly and activation

The spatial architecture of the human tri-snRNP provides important insight into the function of several proteins and also reveals the likely docking site of the tri-snRNP with the spliceosomal A complex during B complex formation. That is, the 3′ end of U6 and Prp8's RH domain, which interact with U2 snRNA to form U2/U6 helix II (fig. S1C) and with the pre-mRNA's 5′SS, respectively, during A complex docking, are located at accessible positions at the "top" of the tri-snRNP (fig. S14A), consistent with the general architecture of the spliceosomal B complex previously revealed by EM (*29*).

The architecture of the human tri-snRNP also indicates that several of its proteins and RNA

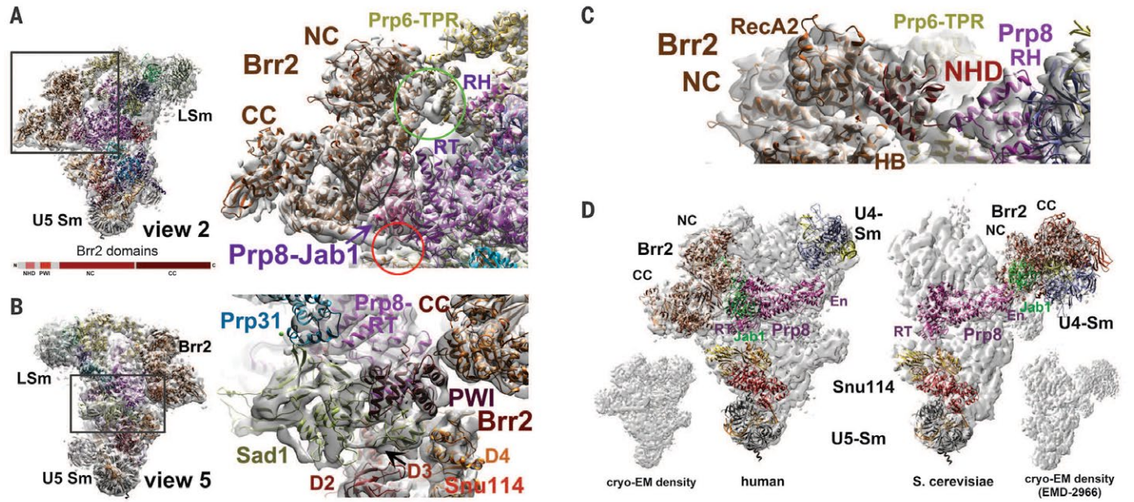


Fig. 4. Structure and location of the RNA helicase Brr2 and Prp8 Jab1 domain. (**A**) Location and structural organization of Brr2. Left: Organization of hBrr2. NHD, N-terminal helical domain; PWI, N-terminal, noncanonical PWI domain; NC/CC, N-terminal/C-terminal helicase cassette. Right: Expanded view showing fit of hBrr2's helicase region in complex with Prp8-Jab1. Circles: Brr2-Jab1 interface (black oval), additional density elements connecting Brr2's NC (green circle) and CC (red circle) to the tri-snRNP [see (B) and (C)]. (**B**) Right: Expanded view showing fit of Brr2's N-terminal PWI domain [red circle in (A)]. (**C**) Expanded view showing fit of Brr2's NHD [green circle in (A)]. (**D**) Brr2 is located at radically different positions in the human and yeast tri-snRNPs, and is found at opposite ends of Prp8's RT/En domain in the two models. The cryo-EM densities of the human (left) and yeast (right) tri-snRNPs (*9*) are shown in corresponding views as insets.

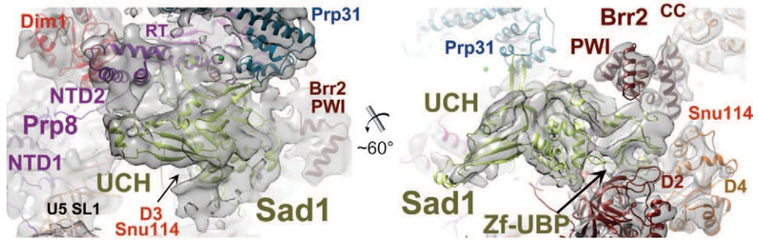


Fig. 5. Sad1 is located in a position bridging U5 and U4/U6 proteins. (**Left**) Fit of Sad1's ubiquitin C-terminal hydrolase (UCH)–like domain (including linker) into a density element that is connected to several U5 proteins and the U4/U6 protein Prp31. (**Right**) Sad1's ubiquitin protease (Zf-UBP)–like domain fits into a neighboring density that is connected to Snu114's D2-D4 domains and Brr2's PWI domain.

elements must undergo major, sequential conformational changes during B complex formation and spliceosome activation. One major rearrangement concerns Prp28, which catalyzes the transfer of the 5′SS from U1 to the ACAGA box of U6 snRNA. As this likely occurs at the Prp8 RH domain (*30*), Prp28 must move from its outward position through the cleft between Brr2 and the U4 Sm domain toward the RH domain (fig. S14A). In fact, the Prp28 "stalk" appears to be intrinsically flexible and undergoes movements within the isolated tri-snRNP consistent with this proposed rearrangement (*31*). For catalytic activation of the spliceosome, Brr2's NC and the U4/U6 duplex must be juxtaposed. This could be achieved by movement of Brr2's helicase domain across the cleft between Brr2 and the U4 Sm core toward the U4/U6 snRNAs (fig. S14A).

Additionally, Prp8 appears to undergo a substantial structural change during spliceosome activation. That is, whereas the overall structure of Prp8's large N-terminal NTD1 domain is similar in the human tri-snRNP and *S. pombe* spliceosome models, the RT/En domain adopts a clearly different position in both complexes (figs. S14B and S15) (*10*). In the tri-snRNP it points upward, whereby the tip of the En domain is ~5 nm away from NTD1, resulting in an open conformation. In contrast, in the *S. pombe* spliceosome, Prp8 adopts a closed conformation where the En domain interacts closely with NTD1 (figs. S14B and S15). As the overall structure of the RT/En domain does not change, Prp8 achieves the closed conformation by a downward movement of the RT/En domain, whereby the pivoting point appears to be located at the interface between the RT and NTD1 domains (figs. S14B and S15A). The position of Prp8's RH domain undergoes a similar downward shift (fig. S14B). This structural change within Prp8 is required to create the pocket into which the rearranged catalytic U2/U6 RNA network and U5 snRNA loop 1 are docked in the *S. pombe* spliceosome (fig. S15, B and C) (*32*). Interestingly, the U5 snRNA loop 1, which also interacts with the 3′ end of the pre-mRNA's 5′ exon in the activated spliceosome (*33*), is already located in the tri-snRNP near Prp8's emerging active-site region, and thus it must not be substantially repositioned (fig. S14B).

The aforementioned rearrangements can only occur when several proteins are displaced concomitantly from their positions in the tri-snRNP. For example, in the tri-snRNP, Dim1 is located in the same area where the center of the U2/U6 catalytic RNA network is found in the *S. pombe* spliceosome (fig. S15, B and C) (*32*). Possibly Dim1 and the RecA2 domain of Prp28, which are both located between Prp8's RT/En and NTD domains (fig. S15B), may stabilize the open conformation of Prp8 in the tri-snRNP. Prp31, Prp3, and Prp4 must also be displaced from the U4 and/or U6 snRNAs. Indeed, except for Prp8, all of these proteins, plus Sad1 and Prp6, are displaced from the spliceosome during activation (*1*). How these multiple rearrangements are orchestrated is currently not clear. Snu114 has been implicated in the activation process (*34*), and if it should undergo a conformational switch from a compact to an elongated state, similar to EF-2/EF-G in the ribosome during translocation (*16*, *35*, *36*), several coordinated movements of other tri-snRNP proteins would result (figs. S6 and S14A). For example, Brr2's PWI domain, which (together with Brr2's NHD) provides major contact points between Brr2 and other U5 proteins as well as Sad1, would likely be destabilized; this could potentially facilitate movement of Brr2 toward U4/U6. The elucidation of the structural dynamics of the various events that take place during spliceosome activation will require numerous cryo-EM "snapshots" of the spliceosome during its multistep assembly pathway.

REFERENCES AND NOTES

1. M. C. Wahl, C. L. Will, R. Lührmann, *Cell* **136**, 701–718 (2009).
2. T. W. Nilsen, in *RNA Structure and Function*, R. Simons, M. Grunberg-Manago, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1998), pp. 279–307.
3. S. M. Fica, M. A. Mefford, J. A. Piccirilli, J. P. Staley, *Nat. Struct. Mol. Biol.* **21**, 464–471 (2014).
4. T. Achsel, K. Ahrens, H. Brahms, S. Teigelkamp, R. Lührmann, *Mol. Cell. Biol.* **18**, 6756–6766 (1998).
5. W. P. Galej, T. H. D. Nguyen, A. J. Newman, K. Nagai, *Curr. Opin. Struct. Biol.* **25**, 57–66 (2014).
6. R. J. Grainger, J. D. Beggs, *RNA* **11**, 533–557 (2005).
7. B. Laggerbauer, T. Achsel, R. Lührmann, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4188–4192 (1998).
8. P. L. Raghunathan, C. Guthrie, *Curr. Biol.* **8**, 847–855 (1998).
9. T. H. D. Nguyen *et al.*, *Nature* **523**, 47–52 (2015).
10. C. Yan *et al.*, *Science* **349**, 1182–1191 (2015).
11. B. Kastner *et al.*, *Nat. Methods* **5**, 53–55 (2008).
12. B. Sander *et al.*, *Mol. Cell* **24**, 267–278 (2006).
13. E. M. Makarov, O. V. Makarova, T. Achsel, R. Lührmann, *J. Mol. Biol.* **298**, 567–575 (2000).
14. F. Galisson, P. Legrain, *Nucleic Acids Res.* **21**, 1555–1562 (1993).
15. P. Fabrizio, B. Laggerbauer, J. Lauber, W. S. Lane, R. Lührmann, *EMBO J.* **16**, 4092–4106 (1997).
16. R. Jørgensen *et al.*, *Nat. Struct. Biol.* **10**, 379–385 (2003).
17. S. Möhlmann *et al.*, *Acta Crystallogr. D* **70**, 1622–1630 (2014).
18. A. Jacewicz, B. Schwer, P. Smith, S. Shuman, *Nucleic Acids Res.* **42**, 12885–12898 (2014).
19. K. F. Santos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17418–17423 (2012).
20. S. Mozaffari-Jovin *et al.*, *Science* **341**, 80–84 (2013).
21. C. Maeder, A. K. Kutach, C. Guthrie, *Nat. Struct. Mol. Biol.* **16**, 42–48 (2009).
22. E. Absmeier *et al.*, *Acta Crystallogr. D* **71**, 762–771 (2015).
23. E. Absmeier *et al.*, *Genes Dev.* **29**, 2576–2587 (2015).
24. S. Mozaffari-Jovin *et al.*, *Genes Dev.* **26**, 2422–2434 (2012).
25. S. E. Behrens, R. Lührmann, *Genes Dev.* **5**, 1439–1452 (1991).
26. S. W. Stevens *et al.*, *RNA* **7**, 1543–1553 (2001).
27. P. Fabrizio, S. Esser, B. Kastner, R. Lührmann, *Science* **264**, 261–265 (1994).
28. Y.-H. Huang, C.-S. Chung, D.-I. Kao, T.-C. Kao, S.-C. Cheng, *Mol. Cell. Biol.* **34**, 210–220 (2014).
29. H. Stark, R. Lührmann, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 435–457 (2006).
30. J. L. Reyes, E. H. Gustafson, H. R. Luo, M. J. Moore, M. M. Konarska, *RNA* **5**, 167–179 (1999).
31. B. Sander, M. M. Golas, R. Lührmann, H. Stark, *Structure* **18**, 667–676 (2010).
32. J. Hang, R. Wan, C. Yan, Y. Shi, *Science* **349**, 1191–1198 (2015).
33. E. J. Sontheimer, J. A. Steitz, *Science* **262**, 1989–1996 (1993).
34. E. C. Small, S. R. Leggett, A. A. Winans, J. P. Staley, *Mol. Cell* **23**, 389–399 (2006).
35. C. M. T. Spahn *et al.*, *EMBO J.* **23**, 1008–1019 (2004).
36. J. Lin, M. G. Gagnon, D. Bulkley, T. A. Steitz, *Cell* **160**, 219–227 (2015).

ACKNOWLEDGMENTS

We thank T. Conrad for HeLa cell production in a bioreactor; H. Kohansal for preparing HeLa cell nuclear extract; I. Öchsner, U. Steuerwald, W. Lendeckel, M. Raabe, and U. Pleßmann for excellent technical assistance; and C. L. Will, K. Hartmuth, N. Fischer, D. Haselbach, M. Wahl, and A. Chari for advice and many helpful discussions. Supported by Deutsche Forschungsgemeinschaft grant SFB 860 (R.L., H.S., and H.U.). EM raw data are available at the Electron Microscopy Pilot Image Archive under the code EMPIAR-10056. The EM map has been deposited in the Electron Microscopy Data Bank with accession code EMD-6581. The atomic coordinates have been deposited in the Protein Data Bank with accession code 3JCR.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/351/6280/1416/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 and S2
Movie S1
References (*37*–*54*)

7 August 2015; accepted 4 February 2016
Published online 18 February 2016
10.1126/science.aad2085

Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex

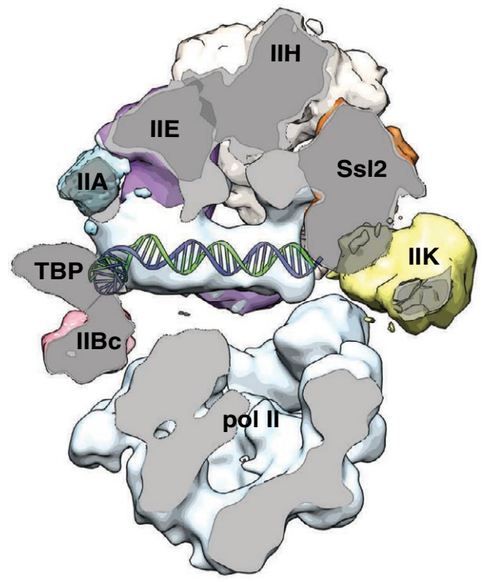
Kenji Murakami, Hans Elmlund, Nir Kalisman, David A. Bushnell, Christopher M. Adams, Maia Azubel, Dominika Elmlund, Yael Levi-Kalisman, Xin Liu, Brian J. Gibbons, Michael Levitt, Roger D. Kornberg*

Introduction: RNA polymerase II (pol II) is capable of RNA synthesis but is unable to recognize a promoter or to initiate transcription. For these essential functions, a set of general transcription factors (GTFs)—termed TFIIB, -D, -E, -F, and -H—is required. The GTFs escort promoter DNA through the stages of recruitment to pol II, unwinding to create a transcription bubble, descent into the pol II cleft, and RNA synthesis to a length of 25 residues and transition to a stable elongating complex. The structural basis for these transactions is largely unknown. Only TFIIB has been solved by means of x-ray diffraction, in a complex with pol II. We report on the structure of a complete set of GTFs, assembled with pol II and promoter DNA in a 32-protein, 1.5 megaDalton "pre-initiation complex" (PIC), as revealed with cryo-electron microscopy (cryo-EM) and chemical cross-linking.

Methods: Three technical advances enabled the structural analysis of the PIC. First, a procedure was established for the preparation of a stable, abundant PIC. Both the homogeneity and functional activity of the purified PIC were demonstrated. Second, an algorithm was developed for alignment of cryo-EM images that requires no prior information (no "search model") and that can distinguish multiple conformational states. Last, a computational method was devised for determining the arrangement of protein subunits and domains within a cryo-EM density map from a pattern of chemical cross-linking.

Results: The density map of the PIC showed a pronounced division in two parts, one pol II and the other the GTFs. Promoter DNA followed a straight path, in contact with the GTFs but well separated from pol II, suspended above the active center cleft. Cross-linking and computational analysis led to a most probable arrangement of the GTFs, with IIB at the upstream end of the pol II cleft, followed by IIF, IIE, and IIH. The Ssl2 helicase subunit of IIH was located at the downstream end of the cleft.

Discussion: A principle of the PIC revealed by this work is the interaction of promoter DNA with the GTFs and not with pol II. The GTFs position the DNA above the pol II cleft, but interaction with pol II can only occur after melting of the DNA to enable bending for entry in the cleft. Contact of the DNA with the Ssl2 helicase in the PIC leads to melting (in the presence of adenosine triphosphatase). Cryo-EM by others, based on sequential assembly and analysis of partial complexes rather than of the complete PIC, did not show a separation between pol II and GTFs and revealed direct DNA–pol II interaction. The discrepancy calls attention to a role of the GTFs in preventing direct DNA-polymerase interaction.



A section through the cryo-EM structure of the complete PIC. Cut surfaces are shown in gray. Locations of densities due to pol II and the GTFs (TFIIA, TFIIB C-terminal domain, TBP subunit of TFIID, TFIIE, and TFIIH, including its helicase subunit Ssl2 and its kinase module TFIIK) are indicated. Density due to DNA is indicated by the superimposed double helix model. TFIIF is not seen in this section.

READ THE FULL ARTICLE ONLINE
http://dx.doi.org/10.1126/science.1238724

Cite this article as K. Murakami *et al.*, *Science* **342**, 1238724 (2013).
DOI: 10.1126/science.1238724

FIGURES IN THE FULL ARTICLE

SUPPLEMENTARY MATERIALS

Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S4 and S5
References (*46*–*63*)
Tables S1 to S3
Movies S1 to S3
Protein Data Bank file for the PIC model

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: kornberg@stanford.edu

Using the most powerful tools in the structural biology toolbox

Taking an integrative approach to solving the field's biggest challenges

Dr. David Schriemer, Professor, University of Calgary and Dr. Rosa Viner, Manager, Integrative Structural Biology program, Thermo Fisher Scientific

We can learn a lot about how proteins function and interact by studying their structure. Equipped with this knowledge, structural biologists can find innovative ways to intervene in disease processes and discover new preventive measures, treatments, and pharmaceutical agents. Understanding molecular mechanisms of diseases with structure–function studies will unravel the fundamentals of life and may consequently lead toward large-scale pharmaceutical solutions for treating disease.

As its name suggests, integrative structural biology involves using a range of analytical techniques to study the architecture of protein systems in intricate detail. Alongside established techniques such as X-ray crystallography, cryo-electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR), an increasingly powerful technique for structural biology is mass spectrometry (MS).

David Schriemer is a protein biochemist at the University of Calgary, who uses MS techniques to answer complex biological questions associated with DNA damage repair and cell division—important cellular mechanisms with cancer-targeting potential. In this article, Schriemer discusses the challenges facing the field and how those looking for answers can make the most of the tools at their disposal.

Powerful labeling techniques

Advances in mass accuracy, resolution, and sensitivity of the latest mass spectrometers are helping structural biologists understand proteins and their complexes at unprecedented levels of detail.

In the simplest experiments, proteins can be examined in their biological state using MS. This approach, referred to as native MS, allows researchers to take an intact protein sample and obtain a mass measurement of the protein or complex to determine its overall size. While this data cannot typically be used to reveal the fine structure of a sample, it can provide a useful overview of its component parts.

However, it's when MS is used in combination with chemical labeling experiments such as crosslinking, covalent labeling, and hydrogen-deuterium exchange (HDX) that the most useful structural insights can be obtained.

Crosslinking experiments, also known as "molecular ruler" measurements, can be extremely powerful for studying protein–ligand interactions, protein–protein interactions, or protein structures to help researchers understand their biological functions. Using a simple bifunctional reagent with a defined length to chemically join components of interacting complexes, protein interactions can be represented by distance restraint low-resolution maps. By multiplexing these measurements using MS, a wealth of information can be obtained to help reveal the proximity of different regions, in turn generating interaction information that is biologically relevant.

Another approach is covalent labeling at reactive side chains using irreversible chemical labels, which can be used to map out a protein system's surfaces, or interaction "footprint." By comparing the mass of a protein labeled with and without a binding partner, the areas that are accessible to the probe can be identified, revealing insights into the protein system's surface structure. Oxidative labeling and photoactive reagents are very useful for this activity.

A third strategy is HDX, which can also provide information on protein structure and function. Accessible backbone amide hydrogens in proteins readily exchange with deuterium in D_2O. Hydrogens present in more tightly folded regions exchange much more slowly than hydrogens in regions exposed to water. Proteins are dynamic structures that move in very functionally relevant ways, and by comparing mass measurements after exchange, structural insights can be gained.

An integrative approach

While many research groups have focused on using and incrementally improving individual analytical techniques to answer specific structural biology questions, Schriemer believes that using these strategies in combination will ultimately prove most effective when it comes to answering the field's biggest questions.

Taking HDX-MS as an example, Schriemer highlights how one approach can guide structural discovery using another. "HDX-MS is a very useful technique for looking at conformational flexibility in a protein system. Likewise, crosslinking is great at defining distances. But there's not a lot of point in taking a precise distance measurement between two points that are in continual motion."

In this way, using HDX-MS experiments to determine which regions are best studied with crosslinking measurements could be a more effective way of planning experiments and getting the most from current technology.

Sample preparation challenges

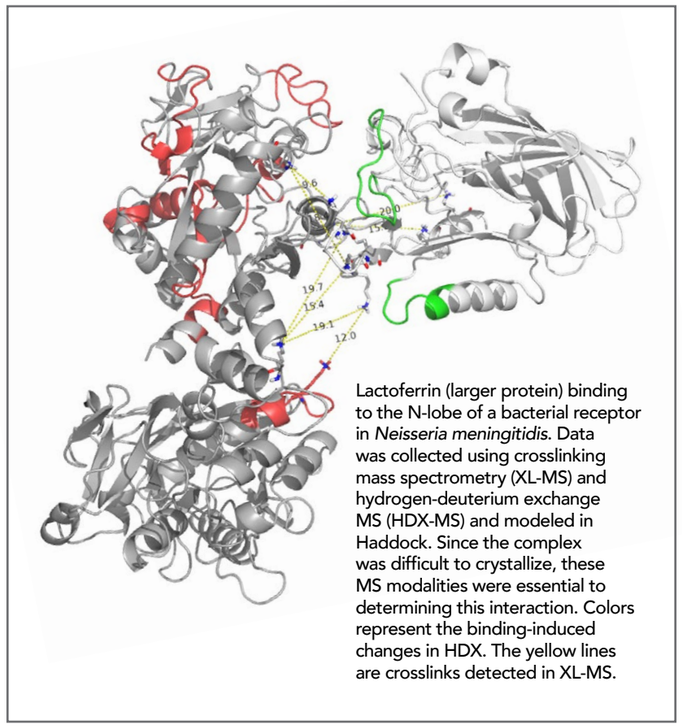
While established techniques such as X-ray crystallography and NMR have proven effective for modeling the structures of smaller systems, they suffer from a number of limitations, with sample preparation being one of the most significant challenges.

"With crystallographic methodologies, your sample needs to be in a solid crystalline form, which is often tricky whatever scale you're working on," says Schriemer. "Creating a diffracting crystal is a trial-and-error process and generally limits the complexity of the protein you can study. Membrane-associated proteins and multiprotein complexes, for instance, are particularly difficult to crystallize."

While Schriemer believes MS approaches are "loosening the shackles," he admits that sample preparation will continue to be an issue for structural biologists, although reagent and sample-providing vendors are improving the capability of these techniques.

Currently, the level of sample preparation associated with reconstituting protein systems is still quite high. "Ideally, we'd mine the information from inside the cell itself, and not have to spend a long time—in some cases years—trying to rebuild it outside the cell."

A combination of approaches, such as installing tags on a particular protein that will enable you to isolate the system, and undertaking crosslinking or labeling before capturing the material for MS analysis, will help to make sample preparation much easier.



Lactoferrin (larger protein) binding to the N-lobe of a bacterial receptor in *Neisseria meningitidis*. Data was collected using crosslinking mass spectrometry (XL-MS) and hydrogen-deuterium exchange MS (HDX-MS) and modeled in Haddock. Since the complex was difficult to crystallize, these MS modalities were essential to determining this interaction. Colors represent the binding-induced changes in HDX. The yellow lines are crosslinks detected in XL-MS.

Innovative labeling strategies

In addition to sample preparation, some of the most important challenges around MS structural biology strategies relate to how we use these tools and interpret the measurements.

For example, crosslinking experiments can yield incredibly useful structural information. But like using precision tools to measure the dimensions of a sponge, we should be cautious about how we interpret the distance measurements obtained using these "molecular rulers."

"Proteins are dynamic structures that bend and flex," says Schriemer. "As a result, the error bars on the distance measurements we obtain are relatively large—though it depends on the crosslinking reagent used and the time required to install it."

While conventional crosslinking reagents tend to be very selective, their slow rate of binding presents a challenge for measurement accuracy and precision. The use of alternatives, such as photoreactive crosslinkers that are activated by UV light, could be one way to improve the utility of crosslinking strategies.

In conjunction with Thermo Fisher Scientific, Schriemer's team is developing fast-acting photochemical labeling reagents that remain inert until activated by light. "Some of the new labeling chemistries we're looking at are based on diazirines, which are orders of magnitude faster than conventional reagents," explains Schriemer. "Photoactivatable reagents show enormous potential in improving the precision of measurements while maintaining the structural integrity of the system you're measuring."

In the meantime, Schriemer believes those engaged in the field need to get better at acknowledging this inherent "fuzziness." "We need to improve how to quantify this error and say, 'I've identified these two sites with this level of confidence, and have this level of confidence on the distance.'"

Integrative analysis workflow

To generate structural models from the MS data, powerful software is essential. For "bottom-up" approaches that involve protein digestion, the proteomics community already has access to software that allows researchers to mine this data and build up a picture of the protein that fits within structural restraints.

But while proteomic peptide identification software is useful for some aspects of the work, Schriemer sees a need to develop other informatics solutions that can integrate and analyze the often disparate datasets produced by all the various MS techniques. His team is collaborating with other groups, including those led by Andrej Šali at the University of California, San Francisco and Alexandre Bonvin at Utrecht University, to develop informatics pipeline that can process the raw data and deliver it into existing structural modeling platforms such as IMP and Haddock.

The result is Mass Spec Studio, a software package capable of identifying the peptides or proteins present in a sample, which localizes the resulting chemical modifications—be it crosslinking, covalent labeling, or HDX. Schriemer hopes this will pave the way toward a single-analysis workflow where raw MS data can be used to quickly generate structural information.

With advanced high-resolution technologies such as cryo-EM poised to make a huge impact on structural biology in the near future, these MS informatics workflows, in combination with innovative labeling techniques, could be a powerful force to unpack the complexity of protein architecture.

The protein clicks in a circadian clock

Detailing the dynamics of molecular structures reveals an ancient timer. A conversation with Dr. Albert Heck on the importance of understanding dynamic protein assembly. **By Mike May**

Life revolves around proteins, from enzymes driving biochemical reactions to antibodies in the immune system army, and beyond. The linear structure of these molecules comes from specific strings of amino acids, often adorned with modifications, such as carbohydrate polymers—but it is the 3-dimensional, tertiary, and quaternary conformation of the protein that is really critical.

While sitting down to talk about the current and future state of structural biology, Albert J. R. Heck, scientific director of the Netherlands Proteomics Centre at Utrecht University, explains how new technologies impact his current work. For proteins, he says, "their shape defines their function." He adds, "We want to know the active structure of a protein, because a protein can malfunction when it does not exhibit the right structure." Using a collection of tools, including mass spectrometry (MS), Heck and his colleagues are unraveling the mechanism behind what he calls "the oldest biological clock on Earth" (*1*).

On the move

To paint a mental picture of the challenge, Heck compares a protein to an image of a person. "We have a global structure, what we look like in a photo," he says, "but then we add action, like moving arms or blinking eyes." Proteins display similar features—static and dynamic aspects. To describe the static structure of a protein, scientists must analyze it on a nanometer scale to get an accurate description of its shape. Researchers must also find ways to describe protein movements. Moreover, most biological mechanisms depend on a group of proteins working together, which makes understanding the movement even more complicated. "We realize more and more that a protein often has multiple dynamic structures," Heck explains. "Rather than just a static picture, we need to see how a protein moves and what it does when it's active."

Consequently, understanding a protein's function poses two key challenges: imaging something on the nanometer scale, and doing so in a way that captures movement. Heck and his team took on those challenges to describe the molecular mechanism that drives a biological clock in prokaryote cyanobacteria.

To get nanometer-scale resolution when studying proteins, scientists can pick from various methods, including electron microscopy, MS, or X-ray crystallography. Although MS doesn't provide more resolution, it does offer some useful advantages. As Heck says, "MS has many different approaches to look at the structure of a protein." The diversity of MS methods gives a researcher many options of mixing and matching the MS techniques that work best for a particular structural study (see page 4). His team used several of these techniques to study the cyanobacteria's clock.

Bacterial biology

Like many other organisms, cyanobacteria keep track of time—or at least know day from night. This internal clock is called a circadian rhythm, which is a roughly 24-hour cycle of a biochemical process—such as a metabolic pathway or gene expression, or a behavior, like sleeping. For cyanobacteria, this rhythm puts the organism in the right place to produce oxygen and make energy through photosynthesis. A few hours before daylight, cyanobacteria rise toward the surface of the water where they harvest energy from the sun, and then they sink back down around sundown. As Heck explains it, "Cyanobacteria know beforehand when it becomes light and when it gets dark."

Heck and his colleagues knew that only a few proteins make this clock work, but they didn't know how. Many circadian rhythms arise from complicated processes of turning genes on to make proteins—all driving a complicated feedback system of transcription and translation. In some cases, though, simpler oscillations of proteins drive the "ticks" of a clock, and that is what happens in cyanobacteria.

To explore the details of the parts and processes of cyanobacteria's clock mechanism, Heck and his team also used a more conventional technology, electron microscopy, and three forms of MS: native MS, crosslinking MS (XL-MS), and hydrogen-deuterium exchange MS (HDX-MS) (see page 4).

All of these techniques helped Heck to explore the proteins in the cyanobacteria's clock. As he explains, "Combining all these approaches, mass spectrometry became really the key technology to understand the structural biology of this clock." MS was essential to figure out the clock's molecular mechanism.

Weighing the mechanism

Fortunately for Heck, that mechanism provides some research gifts. For one thing, it consists of just three proteins: KaiA, KaiB, and KaiC (*2*). Other scientists had already produced structures of each of these proteins at the atomic level, but this data did not tell Heck's team what the complex looked like when the proteins built the clock.

"MS has many different approaches to look at the structure of a protein."
– Albert Heck

In 2005, Japanese biologist Takao Kondo, of the University of Nagoya, reported that this molecular clock can be reconstituted in vitro (*3*). Combining the three proteins and adding adenosine 5′-triphosphate (ATP) turns on the clock, which will run for days up to weeks. The model doesn't make this system easy to study, but it certainly makes it easier.

"These three Kai proteins assemble and disassemble, making movements like a real clock," Heck explains. But there's more. The KaiC proteins also gain and lose phosphate groups. Combined, this transient assembly and phosphorylation, he says, "provide a very complex mixture of relationships of the three proteins as modules of the clock." The proteins can assemble in many ways and phosphorylate and dephosphorylate in different patterns as well.

Using native MS on the oscillator in a dish, Heck's team found that a molar ratio of 6:6:12 of KaiC:KaiB:KaiA makes up the largest structure in the clock's time-cycle; but this was only one of many conformations they discovered. The scientists could freeze the clock and measure the stoichiometry of the co-occurring assemblies—basically the ratio of the molecules to each other—over time. In a 24-hour period, the team found more than a dozen different groupings of the component proteins. In addition, they explored how the Kai proteins bound to each other by using XL-MS. They also applied HDX-MS to see which proteins participated in assemblies. "If KaiA is free, there's a lot of hydrogen access," as Heck describes it. "If KaiA binds to KaiB or KaiC, then less hydrogen is accessible for exchange, especially where the proteins tightly interact."

So, by freezing the process at a point, native MS revealed the stoichiometry of the clock, while crosslinking and H-D exchange MS showed how the proteins connected. In short, the team followed ticks of the clock over time through mass measurements, and the results revealed the 24-hour molecular cycle that drives the rise and fall of cyanobacteria.

Technology teamwork

In addition to using a trio of MS techniques, Heck and his colleagues did even more. They also used advanced cryo-electron microscopy, and the latest detectors, to generate high-resolution images of the protein structures in the various combinations as they stepped through their daily oscillations. As happens in most scientific studies, it took more than one technique to explore and understand the pieces of this bacterial clock and their inner workings.

Moreover, as Heck's research shows us, old and new technologies can work together—and the new ones do not always displace the older ones. More conventional structural biology techniques, like electron microscopy, can confirm or even expand conclusions drawn from more modern approaches, such as advanced combinations of MS methods. The key is combining the right technologies to answer the questions at hand. When it comes to protein structures, it usually takes more than one technology, especially to understand the dynamic side of these molecules.

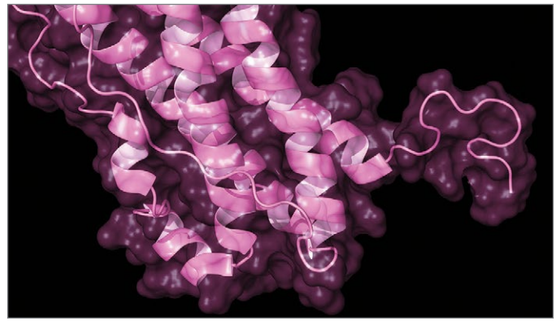
"I will never say that mass spectrometry will take over structural biology, but the techniques are very powerful and complementary to the more conventional techniques," Heck states. "Our study on cyanobacteria's circadian oscillator shows how these technologies come together and give us new biological information."

The cyanobacterial clock is very simple in comparison to other biological examples, but it can boast a long history. "It has survived the dinosaurs and meteorites and climate change," Heck notes. "This clock is very robust, and it's a beautiful piece of biology and nature."

Mike May is a publishing consultant for science and technology.

REFERENCES

1. J. Snijder *et al.*, *Science* **355**, 1181–1184 (2017).
2. M. Ishiura *et al.*, *Science* **281**, 1519–1523 (1998).
3. M. Nakajima *et al.*, *Science* **308**, 414–415 (2005).



Top-Down Proteomics: Turning Protein Mass Spec Upside-Down

Between alternative transcription start sites, alternative splicing, and posttranslational modifications, a given gene may produce dozens of protein variants, each with a different biological activity. Teasing apart those structure-function relationships requires mapping specific variants to their associated biological functions, and the tool of the trade for doing so is mass spectrometry. But not just any mass spec will do. Researchers need a holistic view of protein structure, data that is lost with the popular "bottom-up" proteomics strategy. Powered by today's ultrahigh-resolution, high mass-accuracy mass specs, protein biochemists are increasingly turning bottom-up upside-down. Their new alternative: top-down proteomics. **By Jeffrey M. Perkel**

If you want to know which of a gene's many variants, or "proteoforms," is responsible for a particular biological activity, you need a way to detect that isoform directly. That's easier said than done.

Proteoform analysis is fundamentally a two-part problem. The first part, protein identification, is a simple question of peptide sequencing: matching spectral peaks to a protein's amino acid sequence and thence to the gene that encoded it. This can be complicated if related proteins are present in a sample, because they share identical stretches of amino acid sequences, but in general is relatively straightforward.

Tougher by far is characterization. A given protein may exist in dozens of forms distinguished by just a few daltons, variants that differ in terms of messenger RNA (mRNA) splicing, posttranslational proteolytic processing, and chemical modification. Take histones, for instance. Histone proteins can be heavily modified by methyl, acetyl, and phosphoryl groups, among others, at their N-termini, which in turn can impact chromatin structure and gene expression. In 2009, **University of Pennsylvania** Presidential Associate Professor Benjamin Garcia (then at Princeton University) used a so-called middle-down strategy—in which relatively large protein fragments (bigger than tryptic peptides but smaller than intact proteins) are analyzed and sequenced in the mass spectrometer—and some clever chromatography to resolve and identify 70 proteoforms of human histone H4 and 200 of human histone H3.2.

It isn't clear that every one of those variants has a different biological activity, of course. But the only way to know is to accurately tally them and track their changes under different biological conditions. And therein lies the rub. In bottom-up proteomics, researchers digest their proteoforms to peptides, separate them via liquid chromatography, and deliver them to the mass spectrometer. But as it cleaves the peptide backbone, trypsin also destroys any chance researchers have of understanding how posttranslation modifications are linked. The enzyme can cleave the 15 kilodalton (kDA) histone H3.2 29 times, including more than a dozen sites in the critical N-terminal tail. Using a bottom-up strategy effectively destroys information on how those individual chemical modifications are related, meaning researchers may be able to see that given modifications are present, but are largely blind to their interplay and stoichiometry. They certainly wouldn't be able to determine if, say, two modifications were coincident or mutually exclusive.

In the top-down approach, the histone proteoforms are delivered to the mass spec intact and then sequenced by fragmentation inside the instrument, thereby retaining the critical linkage data. This is a more technically challenging strategy, in that intact proteins are harder to fractionate and fragment than peptides, and much harder to separate by liquid chromatography. Furthermore, it takes relatively high-end instrumentation to resolve such large molecules when they are so similar in size, and special software to do the analysis. Lysine trimethylation, for instance, increases protein mass by 42.0470 Da, while acetylation adds 42.0106 Da, a difference of just 0.0364 Da.

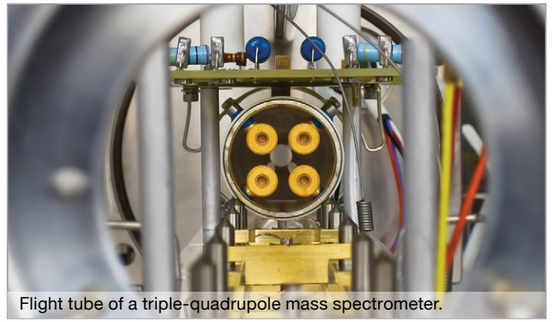
Still, top-down is on the upswing, says Neil Kelleher, the Glass Professor of Life Sciences at **Northwestern University**, founding member of the Consortium for Top Down Proteomics and a top-down evangelist. At the recent annual meeting of the American Society for Mass Spectrometry (ASMS), for example, top-down accounted for "10% to 15%" of the conference, Kelleher says. "A decade ago, it was around 0.1%, or very fringe."

Top of the line

One driver for the growth in top-down is the increasing availability of instrumentation capable of running the experiments. Given the need to distinguish proteins varying by only small chemical changes, top-down researchers typically use high-end, high-resolution instrumentation. Just a few years ago, that mostly meant top-of-the-line Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometers, massive and complicated hardware offering resolution values—and pricing—in the millions. Today, more affordable quadrupole-time-of-flight (qTOF) instruments, such as the **Waters** SYNAPT G2-Si, the **Bruker** maXis II, and the **Thermo Fisher Scientific** benchtop Orbitrap mass spectrometers, have made the technology more accessible.

Still, for some jobs, only an FT-ICR will do. And one of the world's most powerful such systems just went online at

ILLUSTRATION: © PETARG/SHUTTERSTOCK.COM



Flight tube of a triple-quadrupole mass spectrometer.

PHOTO: © JENSGOEPFERT/SHUTTERSTOCK.COM

One driver for the growth in top-down is the increasing availability of instrumentation capable of running the experiments.

the **Pacific Northwest National Laboratory** (PNNL), in Washington state.

FT-ICR mass spectrometers derive their exquisite resolution from the massive cryo-cooled magnets that power them, and as magnetic field strength rises, so too does performance, says Ljiljana Paša-Tolić, lead scientist for mass spectrometry at PNNL's Environmental Molecular Sciences Laboratory user facility. Thus, with a more powerful magnet, "you can think about getting higher resolving power in the same acquisition time, or you can get equal resolving power in a shorter acquisition time."

The PNNL already has several FT-ICR instruments, including systems with 12 and 15 tesla (T) magnets. The new instrument, which went online in mid-March, clocks in at 21 T. With a linear ion trap (Thermo Scientific LTQ-Velos) on the front end and an **Agilent Technologies** magnet on the back, the instrument "occupies almost the whole room; it [weighs] about 24 tons," Paša-Tolić says. The magnet itself requires about 4,000 L of liquid helium to maintain its working temperature of 2.19 Kelvin. (A second 21 T instrument, employing a Bruker magnet, has been installed at the **National High Magnetic Field Laboratory** in Tallahassee, Florida.)

The PNNL instrument went online in mid-March, Paša-Tolić says, and preliminary data were presented at the recent ASMS conference. "We were able to demonstrate resolving power of about 8 million for 12-second transients, which is great," she says. That 12-second analysis time is too slow for the traditional LC-MS workflow, in which proteins flow straight from a liquid chromatography (LC) column into the mass spectrometer (MS), she notes. But even at a more LC-compatible rate, the instrument yields resolutions of about 1 million, she says, and further improvements are in the works. "We have demonstrated a resolving power … an order of magnitude greater than what is attainable with currently available commercial technology."

Among other things, Paša-Tolić hopes to use the 21 T to break the size barrier that bedevils top-down research. Top-down researchers typically struggle to characterize proteins larger than about 50 kDa, though some have used the technique to tackle the posttranslational modifications of 150 kDa biotherapeutic antibodies. But with a more powerful magnet, it may be possible to routinely hit 100 kDa or more, Paša-Tolić says. Indeed, her lab already presented data at ASMS demonstrating "isotopically resolved" analysis of 70 kDa proteins (such as intact bovine serum albumin) at high spectral-acquisition rates.

Paša-Tolić now plans to direct the instrument at secreted fungal enzymes, especially those that degrade lignocellulose. These heavily glycosylated proteins, weighing between 50,000 and 100,000 Da, could advance biofuel development, and Paša-Tolić is developing new reverse-phase chromatography strategies to separate them. "It would be very beneficial to figure out how this pattern of glycosylation relates to function and stability and eventually glycoengineer these enzymes to be more stable and more commercially affordable," she says.

Laser focus

Top-down proteomics is so named because intact proteins are separated and broken down into smaller and smaller pieces in the MS to determine their sequence and modifications. To do that, researchers can apply any of a number of protein fragmentation methods, and the more options available, the better. "You might want to have a lot of fragmentation tools available to really get the most out of the actual experiment," says Andreas Huhmer, proteomics marketing director at Thermo Fisher Scientific. Thermo's new Orbitrap Fusion Lumos, for instance, offers collision-induced dissociation (CID), in which the peptide backbone is broken by collision with a gas molecule, and the related higher-energy collisional dissociation (HCD). It also enables the popular electron-transfer dissociation, which uses a charged donor molecule to induce fragmentation, as well as hybrid methods, such as electron-transfer and higher-energy collision dissociation (EThcD).

Jennifer Brodbelt, the William H. Wade Endowed Professor of Chemistry at the **University of Texas** at Austin, is developing an alternative fragmentation approach. Ultraviolet photodissocation (UVPD) uses ultraviolet laser pulses to cause proteins to shatter along their backbone, producing a ladder of fragments that vary in size by a single amino acid. That's how other fragmentation methods are supposed to work, too, but according to Brodbelt, most tend to fragment more efficiently at protein termini or near charged residues, providing incomplete sequence coverage. UVPD seems to provide relatively uniform coverage across the entire sequence, at least for proteins up to 40 kDa, including the oft-overlooked protein center. "The fragmentation process does not seem to be as charge-modulated as those other methods," she says.

Brodbelt has worked with Thermo Fisher Scientific to implement the technique on Orbitrap instruments. In one recent paper, she applied the method on an Orbitrap Elite to map the linkages in branched poly-ubiquitin chains. The result was a remarkable series of fragment ions, one for each consecutive amino acid of the ubiquitin chain, terminating at the residue to which the ubiquitin moiety is coupled. cont.>

Featured Participants

Agilent Technologies
www.agilent.com

Bruker
www.bruker.com

National High Magnetic Field Laboratory
www.nationalmaglab.org

Northwestern University
www.northwestern.edu

The Ohio State University
www.osu.edu

Pacific Northwest National Laboratory
www.pnnl.gov

The Scripps Research Institute
www.scripps.edu

Thermo Fisher Scientific
www.thermoscientific.com

University of Pennsylvania
www.upenn.edu

University of Texas Austin
www.utexas.edu

Vanderbilt University
www.vanderbilt.edu

Waters Corp.
www.waters.com

By simply counting those ions and watching where they abruptly disappeared, she could determine where the inter-ubiquitin linkages must have occurred.

"You'd see a huge shift, a mass shift when ubiquitin appeared at a particular lysine," Brodbelt explains.

Though still in development, UVPD systems have been installed in several labs. The PNNL 21 T has one. So does John Yates III, the Ernest W. Hahn Professor at the **Scripps Research Institute**, who has mounted the system on an Orbitrap Fusion. Bottom-up proteomics, Yates explains, has long been considered easier than top-down in part because the infrastructure required to do it—the mass spectrometers, the peptide separation methods, and the analytical software—was already mature when the technique was developed. The experiments themselves were thus easier to perform. "For top-down, almost everything has to be reinvented or certainly significantly improved in order to make this whole workflow possible." That, he says, explains his enthusiasm for UVPD. "Hopefully it will get us the kind of fragmentation that we need in order to effectively analyze these things."

From top-down to top-top-down

As top-down adoption grows, so too do the technical developments. One emerging area is what Kelleher calls "top-top-down," or native mass spectrometry. The method allows researchers to examine multiprotein complexes in the MS, and one researcher making significant headway on this approach is Vicki Wysocki, Ohio Eminent Scholar at **Ohio State University**.

Existing fragmentation approaches, such as CID, simply cannot inject enough energy per collision into a protein complex to cause it to fall apart, Wysocki explains. "If you have a very large protein complex …[and] if you are colliding that into argon with a mass of 40, the amount of energy that you can transfer will be fairly small." Rather than falling apart, a protein in such a complex will simply unfold, she says. So, her group developed an alternative approach, surface-induced dissociation (SID), in which complexes are smashed at high speed into an inert fluorocarbon-coated gold surface.

Using SID, Wysocki says, researchers can work out the topology of protein complexes and subcomplexes, teasing them apart to determine, for instance, which protein-protein interfaces are strong and which are weak. Suppose a given complex is a hexamer, she explains—a dimer of trimers. "We will directly see those trimers as products of the SID," she says. In one recent example, Wysocki's team used that approach to work out the stoichiometry of the *Pyrococcus furiosus* RNAse P complex, an RNA-containing tetramer whose structure was previously unresolved.

Waters has been working with Wysocki's group to offer SID capability to selected investigators on the SYNAPT G2 series qTOFs, and Wysocki has received grant funding to implement the method on Orbitrap and FT-ICR instruments as well. She has also developed more elaborate implementations, including a modified qTOF containing two SID cells flanking Waters' ion mobility separation unit, for performing multiple surface collision events. Ion mobility separation, Wysocki explains, "is sort of like a gas-phase electrophoresis," separating ions by size and shape, and it "has been a huge help in all of this work."

Another emerging development is top-down-based mass spectrometric imaging, Paša-Tolić says. Richard Caprioli at **Vanderbilt University**, and Ron Heeren in the Netherlands have both demonstrated laser ablation-based top-down strategies in the past year using FT-ICR mass analyzers, and Paša-Tolić says she would like to apply such strategies to study the soil rhizosphere, for instance, to determine where different secreted enzymes are located. "If you think about the way we do top-down proteomics right now, it clearly is missing spatial information," Paša-Tolić says. "In many instances, this would be extremely useful to have."

As for Kelleher, he sees a bright future for top-down in clinical research. Indeed, it is in the clinic that one of top-down's biggest successes can already be found. The Bruker BioTyper, a simple matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer for identification of bacterial pathogens based on intact protein masses, "has been a smash success …. Arguably one of the best successes of proteomics in clinical medicine," Kelleher says. Now he hopes to apply that same top-down philosophy to clinical biomarker development for complex disease.

With a growing user community, he won't be alone in that work. But Kelleher remains undaunted. "I'm smiling," he says. "Even if people are telling me they've done it better than my group has, I just say, 'okay, great, it's a big sandbox. Come play!'"

Jeffrey M. Perkel is a freelance science writer based in Pocatello, Idaho.

DOI: 10.1126/science.opms.p1500096

thermo scientific

Go Beyond

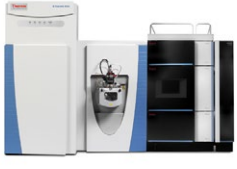
with comprehensive solutions
to characterize biomolecular structures

Meet the challenges of structure-function studies with Thermo Scientific™ Integrative structural biology solutions

The complexities of protein structure-function analyses require a new combined toolset—mass spectrometry (MS) and cryo-electron microscopy (cryo-EM).

Thermo Scientific™ Orbitrap™ MS solutions help enable both peptide and protein-centric strategies that deliver insights into a multitude of biochemical and structural properties.

Armed with superior quality data from proven HRAM Orbitrap systems, you can confidently analyze samples with increasing analytical depth, delivering information to accelerate the path from structure to function.



Thermo Scientific™ Q Exactive™ HF mass spectrometer

Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer

Find out more at **thermofisher.com/MS-structure**

ThermoFisher SCIENTIFIC

For Research Use Only. Not for use in diagnostic procedures.

 AD65065-EN 0717M