Published in final edited form as:

Methods. 2017 August 01; 125: 55–62. doi:10.1016/j.ymeth.2017.04.023.

Understanding pre-mRNA splicing through crystallography

Sara Espinosa, Lingdi Zhang, Xueni Li, and Rui Zhao*

Department of Biochemistry and Molecular Genetics, University of Colorado Denver Anschutz Medical Campus, 12801 E. 17th Ave. Aurora, CO 80045

Abstract

Crystallography is a powerful tool to determine the atomic structures of proteins and RNAs. X-ray crystallography has been used to determine the structure of many splicing related proteins and RNAs, making major contributions to our understanding of the molecular mechanism and regulation of pre-mRNA splicing. Compared to other structural methods, crystallography has its own advantage in the high-resolution structural information it can provide and the unique biological questions it can answer. In addition, two new crystallographic methods – the serial femtosecond crystallography and 3D electron crystallography – were developed to overcome some of the limitations of traditional X-ray crystallography and broaden the range of biological problems that crystallography can solve. This review discusses the theoretical basis, instrument requirements, troubleshooting, and exciting potential of these crystallographic methods to further our understanding of pre-mRNA splicing, a critical event in gene expression of all eukaryotes.

Theoretical basis and framework of crystallography

X-ray crystallography

Since W. L. Bragg first solved the structure of table salt using X-ray crystallography¹, it has become one of the most powerful tools for determining atomic structures ranging from small molecules to macromolecular complexes as large as the ribosome. Today, the Protein Data Bank has over 120,000 entries of macromolecule crystallographic structures. Of these, almost 900 are related to pre-mRNA splicing or the spliceosome, making significant contributions to our understanding of the mechanism and regulation of pre-mRNA splicing.

Although it remains a daunting challenge to produce large quantities of homogeneous samples of spliceosome for crystallization, X-ray crystallographic studies of components of the spliceosome and their complexes have provided unique functional insight that would have been difficult to obtain using other approaches (Fig. 1). For example, the crystal structure of the U1A RRM (RNA recognition motif) domain bound to RNA revealed the structural basis of how RRM, the most prevalent eukaryotic RNA binding motif, recognizes single stranded RNA². The crystal structures of several other splicing factor-RNA

*Contact author. rui.zhao@ucdenver.edu, 303-724-3269.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

complexes (such as U2AF65 RNA-binding domain bound to RNA³, MSL5/SF1 bound to RNA⁴, Prp24 bound to U6 snRNA⁵) offered molecular details of how these proteins recognize the splicing sites or facilitate U4/U6 snRNA annealing. To date, the largest splicing related crystallographic structure is likely the roughly 250kD human U1 snRNP which revealed the structural basis of 5' splice site (ss) recognition^{6–8}. Long before the recent determination of cryo-EM structure of the entire spliceosome, crystallographic studies of Prp8 (a central component of the U5 snRNP, tri-snRNP, and spliceosome) on its own or with different binding partners (such as Brr2 and Aar2) provided our first glimpse of the potential active site of spliceosome and how it may act as a master regulator of splicing^{9–15}. In addition, crystal structures of the many RNA helicases involved in pre-mRNA splicing illustrated how these helicases recognize their substrates, the mechanism of unwinding, and how their activities may be regulated^{12,13,16–23}.

In this section, we will draw on proteins to illustrate the theoretical framework of conventional X-ray crystallography, which applies to all macromolecules, including RNA, that can be crystallized. Historically, proteins used for crystallization were typically purified from their endogenous source, which often limited the protein that could be studied. Now, most proteins are expressed and purified recombinantly from *E. coli*, yeast, insect cell, or mammalian cell cultures. To identify the crystallization condition, hundreds of conditions of different buffer, pH, salt, and other precipitants are screened with concentrated protein (typically on the order of 10–100 mg/ml) using a hanging or sitting drop method. If successful, the crystals are then placed in a stabilization solution containing a cryo-protectant (such as glycerol) and flash cooled in liquid nitrogen²⁴. Cryo-cooling slows the dispersion of radiation-induced reactive species and significantly reduces secondary radiation damage, even though it does not have much effect on primary radiation damage caused by photoionization and Auger decay^{25,26}. The cryo-cooled crystals are then subjected to X-rays and diffraction data is collected.

A crystal is composed of regular repeats of its smallest three-dimensional unit, the unit cell. When the X-ray encounters a crystal, every encountered electron becomes a new scattering center that further emits X-rays (referred to as Thomson or elastic scattering). These diffracted X-rays interact, either reinforcing or canceling each other. The ultimate consequence of the diffraction from one unit cell is the summation of the diffracted X-ray wave from each electron. In other words, the diffraction is the Fourier transformation of the electron density distribution in the unit cell, which is obviously dependent on the structure of the molecule in the unit cell. This leads to the structure factor equation:

$$\vec{F}_{hkl} = V \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dx dy dz$$

where V is the volume of the unit cell, $\rho(x, y, z)$ is the electron density at (x,y,z) position, and h, k, l are miller indices.

To determine the structure (electron density distribution) from diffraction, the electron density equation, which is the inverse Fourier transformation of the structure factor equation, is used:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| \cos(2\pi(hx + ky + lz + \phi_{hkl}))$$

In this equation, F (the amplitude of the diffracted X-ray) is the square root of intensity, which can be recorded on a detector. The phase information (Φ_{hkl}), however, cannot be determined directly from recorded X-ray diffraction and must be obtained indirectly. This is referred to as the “phase problem” of crystallography.

There are three major approaches to solve the phase problem of crystallography: the Molecular Replacement (MR)²⁷, Multiple Isomorphous Replacement (MIR)²⁸, and the Multiple Anomalous Diffraction (MAD)²⁹ methods. In the MR method, a known structure of a molecule that is homologous to the molecule of interest is required. The higher the homology, the more likely the MR method will succeed. The rotation and translation function search will identify the orientation and position of the known molecular model in the unit cell that best matches the molecule of interest. With this orientation and position information, structure factors, including phases, can be calculated for each diffracted X-ray. These phases, in combination with recorded amplitude information, determine the structure of the molecule of interest through the electron density equation.

For molecules that do not have a homologous structure available, the MIR or MAD method is used to solve the phase problem. The MIR method requires a native crystal and two or more heavy atom containing crystals. Heavy atom containing crystals are generated by soaking crystals with many heavy atom solutions at different times and concentrations to obtain crystals with heavy atom bound without changing the protein conformation. Differences in the diffraction intensities of native and heavy atom containing crystals are used to determine the positions of the heavy atoms, and consequently, the phases for the native crystals. The SIR (Single Isomorphous Replacement) method using one heavy atom containing crystal followed by density modification can also be used to determine the phases. The MAD method uses a crystal containing a heavy atom that can generate significant anomalous signals. Anomalous scattering is wavelength dependent and phase information can be deduced from diffraction data collected at two or more wavelengths. The ease of replacing Met with Se-Met which has significant anomalous scattering in recombinant protein (especially those produced in *E. coli*) made Se-Met MAD method the preferred approach for determining a novel crystal structure without homologous models. A highly useful variation of MAD method is the SAD (Single Anomalous Diffraction) method, which utilizes anomalous diffraction collected from one single wavelength. Phases derived from SAD have significant errors, which must be rectified using density modification approaches (such as averaging or solvent flattening).

In all phasing methods, there are inevitably errors that are either introduced by deriving phases from a homologous model or by inaccuracies arising from experimental measurements in the MIR or MAD methods. Refinement is therefore essential to minimize these errors and obtain the highest quality of atomic models. In the refinement process, a model is built into the initial electron density map, refined to minimize the difference between predicted and observed structure factors (monitored by R_{working} and R_{free} factors

where R_{free} is calculated using 5–10% of the reflections that are excluded from the refinement process) as well as the root mean square deviation between the ideal bond lengths/angles and those in the model. With the refined model, phases and electron density maps can be calculated more accurately. This iterative process continues until the R_{working} and R_{free} factors can no longer be improved.

Compared to other approaches of structural determination (NMR and cryo-EM), X-ray crystallography has its advantages and disadvantages. NMR can obtain solution structure and dynamics of a macromolecule, but it is limited in the size of the molecule that it can work with. Proteins smaller than 20kD are ideal for NMR, and structural determination becomes increasingly difficult as the size of the protein increases, even when a high field NMR instrument is used. Cryo-EM, currently undergoing a resolution revolution, requires the molecule to be sufficiently large. Although there has been a few high-resolution cryo-EM structures of small proteins³⁰, structures of proteins smaller than 200kD are generally difficult to solve using cryo-EM. X-ray crystallography, on the other hand, does not have a molecule size limit. Having the crystal structure of a component remains a valuable resource in interpreting many medium resolution cryo-EM structures of large complexes.

Furthermore, X-ray crystallography can typically generate higher resolution structures than other approaches. For example, Montemayer and colleagues determined the 1.7 Å resolution crystal structure of four RRM domains bound to the central region of U6 snRNA, revealing molecular details of the Prp24-U6 interaction as well as how Prp24 may function as a chaperone for U4/U6 snRNA annealing⁵. Other strengths of X-ray crystallography include its ability to directly visualize electron density (in contrast to NMR structures that are models satisfying distance restraints) and clearly define resolution (in contrast to cryo-EM structures whose resolution is largely a reflection of how well the Fourier shell correlates). In addition, X-ray crystallography can answer some unique biological questions that may be difficult for other structural methods. For example, it is often important to determine the identity and position of cations such as Mg^{2+} that plays an important role in the structure of RNA as well as in the catalytic function of many enzymes including the spliceosome³¹. Mg^{2+} can often be replaced by lanthanide ions³², whose position in the crystal structure can be unambiguously determined by anomalous diffraction. Some other cations, such as K^+ , Ca^{2+} , Mn^{2+} , Cu^{2+} , Fe^{2+} , Co^{2+} , and Ni^{2+} , have detectable anomalous signals at wavelengths used for typical X-ray diffraction experiments and may be located directly by these signals^{33,34}. On the other hand, obtaining large and well diffracting crystals is mainly a trial and error process that is very difficult to predict or control, constituting a major bottleneck.

Serial femtosecond crystallography (SFX)

Several recent developments that strive to overcome intrinsic problems of traditional X-ray crystallography (such as the reliance on large crystals) have generated exciting results that are likely also applicable for crystallographic studies of the pre-mRNA splicing machinery. The first is the development of serial femtosecond crystallography^{35,36} which can utilize micro- or nanocrystals and potentially follow structural changes in real time through time-resolved crystallography. In traditional X-ray diffraction experiments, the X-ray dose needed to generate visible diffraction from very small crystals causes extensive radiation damage

before the diffraction can be recorded. In SFX, millions of micro- or nanocrystals (200 nm to 2 μ m in size) in buffer at room temperature flow through a liquid jet and hit the X-ray free electron laser beam that is perpendicular to the jet stream (Fig. 2). The speed of the flow is such that, at any given time, one crystal is exposed to a femtosecond X-ray free electron laser pulse. The X-ray is over 20-fold more intense than a traditional X-ray beam. However, its pulse time is much shorter than the time scale of most radiation damage processes, achieving “diffraction before destruction”.

Since the vast majority of crystals do not encounter the X-ray, millions of micro- or nanocrystals are needed to generate enough diffraction data. Additionally, the crystal size needs to be relatively homogeneous. Size heterogeneity can create scaling problems, potential detector damage, or liquid jet clogging. Sometimes large crystals can be crushed to produce micro or nanocrystals, but this approach can induce damage and reduce diffraction qualities; therefore, it is not a particularly desirable method. The same principles of growing large crystals are applicable to micro or nanocrystals with some small differences. For example, in order to generate a vast number of small crystals instead of a few large crystals, more instead of less nucleation sites is preferred. Several special methods have also been developed to favor the growth of micro- or nanocrystals, such as the nanocrystalline free interface diffusion method ³⁷, *in vivo* crystallization ^{38–41}, or crystallization using lipidic cubic phase ^{42,43}. Crystals in the order of 5 μ m or larger may be identified by polarized light microscope, while smaller crystals depend on detection and/or confirmation by SONICC (second order non-linear imaging of chiral crystals), X-ray powder diffraction, or transmission electron microscope.

A pixel array detector is typically used to record the diffraction data with a readout speed that matches the repetition rate (120Hz at Linac Coherent Light Source at Stanford) of the X-ray in SFX experiments ⁴⁴. Diffraction data can be indexed using programs such as CrysFEL ⁴⁵, cctbx.xfel ⁴⁶, and nXDS ⁴⁷. Subsequently, intensities are integrated using the Monte Carlo method ^{48,49}. Earlier SFX structures were commonly solved by the Molecular Replacement method; however, a recent proof of principle study demonstrates that SAD method can be applied to determine the structure of lysozyme ⁵⁰. This opens the door for structural determination of novel proteins through phasing methods routinely used for conventional X-ray crystallography, such as the MIR/SID and MAD/SAD methods. In addition, new *ab initio* phasing methods that are only feasible for SFX are being developed, by evaluating the shape transformation ^{35,51–53} or “high-intensity radiation induced phasing” (HI-RIP) ⁵⁴.

SFX has been used to successfully determine a number of crystal structures ³⁶. It is particularly powerful in time resolved crystallography to study the structural changes after a trigger (such as the start of an enzymatic reaction). This technique has yet to be applied to study spliceosome related complexes. Nevertheless, it has the potential to make significant contributions to our understanding of pre-mRNA splicing if the spliceosome or its large sub-complexes can be crystallized but cannot form large crystals. The perspective of using SFX to follow real time structural changes during the splicing reaction in these crystals is particularly exciting.

Electron crystallography

A second recently developed approach to solve structures of proteins that fail to generate large crystals is electron crystallography^{55–58}. Electrons can be regarded as waves; therefore, electron crystallography has the same theoretical bases as X-ray crystallography with some critical differences making it particularly suitable for small 3D crystals. Similar to X-ray crystallography, molecules in a crystal interact with an incoming beam (electrons instead of X-ray in the case of electron crystallography) to generate diffractions. As electrons interact with crystals more strongly than X-rays, useful diffraction can be generated with much smaller and thinner crystals, which would not be able to generate meaningful X-ray diffraction patterns. Electron crystallography has been utilized for 2D crystals with a single layer of molecules, but the electron dose used typically destroys these crystals after one exposure. Consequently, hundreds of crystals are required in order to determine one structure. This approach has not been successful for 3D crystals due to the inherent difficulty associated with the collection and processing of data from hundreds of crystals^{59,60}. Gonen and colleagues developed a new approach called MicroED to carry out electron diffraction of small 3D crystals in a transmission electron microscope^{55,60}. MicroED dramatically reduces electron beam dose in these microscopes to alleviate radiation damage. In addition, radiation damage from electron scattering is much less than that from X-ray scattering⁶¹. Shi et al. found that there is no observable adverse effect on resolution or intensity until the accumulated electron dose reaches $\sim 9 \text{ e}^-/\text{\AA}^2$. The data collection was therefore designed such that the total exposure is under this limit and a complete data set can be collected from a single crystal.

Crystals used for MicroED are much smaller than regular crystals and are barely speckles under a light microscope compared to the crystals used for conventional X-ray crystallography. These crystals are pipetted onto holey carbon grids and frozen in liquid ethane. Although electrons cannot penetrate thick crystals (1.5 μm), excellent diffraction to $<2 \text{ \AA}$ can often be recorded from thin (0.5–1 μm) crystals which still contains millions of unit cells. Previously, data was usually collected as still images at discrete tilt angles. Current methods use a continuous rotation mode and diffraction is recorded on a fast Complementary Metal-Oxide Semiconductor (CMOS) detector. Data processing is typically carried out via standard X-ray crystallographic software including MOSFLM⁶² and XDS⁴⁷.

In most structures solved by MicroED so far, phases were determined using the Molecular Replacement method. Developing new phasing methods would be highly beneficial to solve structures that do not have a homologous structure already known. Possible phasing methods include MIR method using heavy atoms, phasing through single particle cryo-EM maps, or direct phasing when resolution is high (close to 1 $\text{\AA}$). There are over a dozen structures determined by electron crystallography now⁶³, four of which (prion peptides) were determined to 1 $\text{\AA}$ resolution using the direct phasing method⁶⁴. This number is likely to increase exponentially as the technique continues to improve. Considering the highly dynamic nature of the spliceosome, methods such as SFX and MicroED will likely make significant contributions to structural determination of the spliceosome or its large sub-complexes where formation of large crystals may be difficult.

Equipment

X-ray crystallography

The key equipment required for crystallography is an X-ray generator. There are two major approaches of generating X-ray radiation for crystallographic experiments: a rotating anode X-ray generator and a synchrotron. In a rotating anode generator, a cathode filament and rotating anode is placed under high voltage (*e.g.*, 50kV). Electrons are released from the cathode filament, accelerate in vacuum in an X-ray tube, hit the anode, and generate X-ray radiation. The wavelength of the X-ray depends on the material of the anode with which the electrons collide. The most commonly used anode material is copper, which produces X-rays at 1.54 Å. In a synchrotron, electrons circulate in the storage ring generating X-ray beams. Both X-ray generating methods have their advantages and disadvantages. A rotating anode X-ray generator has a small footprint and is easily available in most universities. The major limitation is its fixed wavelength and weaker intensity. One of the major advantages of a synchrotron source is the higher beam intensity which requires much shorter exposure time and often leads to higher diffraction resolution than the rotating anode X-ray generator. The synchrotron X-ray also provides tunable wavelengths essential for structural determination using the multiple wavelength diffraction method. Diffraction data generated from the rotating anode generator or synchrotron is collected on an imaging plate, CCD (charge coupled device) detector, or Hybrid Pixel Array Detector (HPAD). Subsequent structural analyses are carried out on a computer with a number of crystallographic software installed. HKL3000^{65,66}, d*strek⁶⁶, MOSFILM⁶², and XDS⁴⁷ are examples of popular software for data processing. Structural determination and refinement can be carried out in either the Phenix⁶⁷ or CCP4⁶⁸ program suite, and model building is typically performed using program Coot⁶⁹.

Serial femtosecond crystallography

Serial femtosecond crystallography requires ultra fast (10 fs) X-ray free electron laser pulses of extremely high intensity, which can currently be obtained at the Linac Coherent Light Source (LCLS) at Stanford Linear Accelerator Center (SLAC) and a few other synchrotrons in the world (see equipment list).

Electron crystallography

Diffraction data from sample flash frozen on holey grids can be collected on a transmission electron microscope with a CMOS-based detector that has been shown to be beneficial for electron diffraction experiments⁷⁰.

Troubleshooting

Regardless of which crystallographic method is used, the most significant bottleneck in all these methods is obtaining diffracting crystals. The first step toward crystallization is to obtain large quantities of homogeneous sample. This can be a significant challenge for the spliceosome, which assembles on each pre-mRNA by forming the E, A, B, and C complex sequentially prior to catalysis and disassembly. The spliceosome throughout the splicing cycle undergoes numerous compositional and conformational changes, making large

quantities of homogeneous sample difficult to obtain. The spliceosome can be arrested at various stages of the splicing cycle through helicase mutants or by reconstituting the spliceosome using splicing extracts and mutated pre-mRNA substrates ⁷¹. However, quantity and homogeneity of spliceosome purified after arrest (typically through affinity tags) are still far from what is necessary for crystallization. There has not been any reported success of crystallizing the spliceosome, although future exploration of spliceosomes arrested in thermophilic fungi such as *Chaetomium thermophilum* ⁷² could be an approach for obtaining more homogeneous spliceosomes that can potentially be crystallized. Most splicing related crystal structures determined to date contain one to three components, with the human U1 snRNP being the largest complex crystallized. The Nagai group crystallized the human U1 snRNP entirely reconstituted from recombinant components and determined its structure to 5.5 Å resolution ⁶, which enabled the engineering of two sub-complexes of U1 snRNP whose structures were determined to 3.5 and 2.5 Å resolution ⁸. The Wahl and Luhrmann group obtained the crystal structure of human U1 snRNP purified from HeLa cells after treatment with limited proteases in the crystallization drop ⁷. With the development of SFX and MicroED that enables structural determination of much smaller crystals, it is possible that one day we will see crystal structures of much larger spliceosomal-related complexes. Until then, crystal structures of smaller complexes with one or more components will likely continue making significant contributions to our understanding of pre-mRNA splicing.

If the original protein of interest does not crystallize, modifying the protein target is one of the most important approaches to help facilitate crystallization. For example, considering that more compact proteins are generally easier to crystallize, limited proteolysis combined with mass spectrometry is often used to identify flexible ends or loops. These ends or loops can then be removed through standard molecular biology techniques to make the protein more compact, facilitating crystallization. *In situ* proteolysis through addition of a limited amount of protease in the crystallization drop has also been successfully used to crystallize human U1 snRNP purified from HeLa cells ^{7,73}.

If a protein does not crystallize, its homologs from other species can be considered for crystallization. Homologous proteins from thermophilic Archaea have been invaluable for obtaining structures of some difficult proteins due to the increased stability of proteins in these thermophiles. This approach is not widely used for splicing related proteins as Archaea do not undergo pre-mRNA splicing. However, some thermophilic fungi, such as *Chaetomium thermophilum* ⁷², may be helpful in overcoming crystallization problems of splicing proteins. Structures of several *Chaetomium thermophilum* homologs of splicing related proteins, including Prp43 ⁷⁴, Brr2 ⁷⁵, and a SF3b130-SF3b10 dimer ⁷⁶ have been solved.

Modifying the surface of a protein of interest is another approach that can potentially facilitate crystallization. For example, global Lysine methylation ⁷⁷ and surface entropy reduction based mutagenesis ⁷⁸ have been used to successfully rescue proteins that were refractory to crystallization. Proteins fused with MBP mutated based on the principle of surface entropy reduction to facilitate crystallization also have enjoyed significant success ⁷⁹. Software designed to predict protein crystallizability with accuracy around 70–

80% (see references in ⁸⁰) can potentially be harnessed to select protein modifications that should be prioritized among an unlimited number of possibilities.

Crystallizing a protein in complex with its ligand or other protein/DNA/RNA partners is another way to modify the protein target if the protein alone fails to crystallize. For example, co-expression with Aar2 enabled the large fragment of Prp8 (residues 885-2413) to be expressed in large quantities in yeast and be crystallized ¹². Similarly, full-length yeast Brr2 is refractory to crystallization on its own but can be crystallized in complex with the Prp8 Jab1 ⁷⁵ domain ^{22,75}. Other than using a known ligand of a protein, screening approaches are designed to identify small molecules that can act as ligands that bind and stabilize a protein to facilitate crystallization ⁸¹. Yet another interesting way to modify the protein is to synthesize the D-enantiomer of the protein of interest since the racemic mixture of the D- and L-proteins often crystallizes easily ⁸².

Other than modifying the protein target itself as discussed above, many different aspects of the crystallization process have also been explored to potentially facilitate crystallization. For example, the microfluidic technique ⁸³ can sometimes produce crystals that have not been obtained using the traditional hanging drop or sitting drop methods. Nucleation-inducing agents such as porous bioglass ⁸⁴ or nanowrinkle substrates ⁸⁵ can also assist in protein crystallization.

Equipment and supply lists

X-ray crystallography

The most widely used in house X-ray generator is the MicroMax 007 HF from Rigaku, although a sealed tube X-ray generator (such as the MicroMax-003 generator from Rigaku) is also an economical option due to its low maintenance. Most modern X-ray generators are equipped with HPAD detectors for data collection. The major synchrotron facilities in the United States include the Advanced Photon Source (Argonne, Illinois), the Advanced Light Source (Berkeley, California), the National Synchrotron Light Source (Upton, New York), and the Cornell High Energy Synchrotron Source (Ithaca, New York). There are synchrotron facilities throughout the world in over 20 different countries that can be used for macromolecular crystallography experiments, such as the European Synchrotron Facility (Grenoble, France), PETRA III (Hamburg, Germany), Spring-8 (Harima Science Park City, Japan), DIAMOND (Oxfordshire, England), and the Beijing Synchrotron Radiation Facility (Beijing, China). Access to the synchrotron is typically obtained through the submission and approval of a user submitted experimental proposal. Data collection can be done in person or remotely with the help of beamline scientists at the synchrotron.

Although crystallization can be set up manually, several robotic systems can be used, such as Phoenix from Art Robbins Instruments, Mosquito from TTP Labtech, and OryxNano from Douglas Instruments. A recent robot was developed using droplet-based microfluidic technique that only needs 4–8 nL per drop, dramatically reducing the quantity of sample needed for crystallization trials ⁸⁶.

Serial femtosecond crystallography

There are currently three synchrotron facilities capable of producing X-ray free electron laser: Stanford Linear Accelerator Center's Linac Coherent Light Source, the Deutsches Elektronen-Synchrotron's Free-electron laser in Hamburg (FLASH), and RIKEN's Spring-8 Angstrom Compact free electron LAser (SACLA). Access to these facilities is typically obtained through proposal submission, similar to a synchrotron for traditional X-ray crystallographic experiments.

Electron crystallography

Quantifoil 2/2 holey carbon copper EM grid can be used for sample preparation. They can be purchased from Ted Pella, Inc., Electron Microscopy Sciences, or SPI Supplies. The Vitrobot to blot and vitrify the grid in liquid ethane can be obtained from FEI. The cryo-transfer holder for transferring cryo grids to the microscope may be purchased from Gatan, Inc. A Tecnai F20 TEM equipped with a field emission electron source for data collection can be obtained from FEI. An example of the CMOS detector for diffraction data recording is the F416 4kx4k CMOS detector from Tietz Video and Image Processing Systems GmbH.

In summary, X-ray crystallography has made significant contributions to our understanding of the molecular mechanism of pre-mRNA splicing and its regulation. It will continue to complement other structural approaches such as NMR and cryo-EM to provide high-resolution structures of key components of the splicing machinery. New developments such as serial femtosecond crystallography and electron crystallography will likely expand the range of biological questions that crystallography can answer.

Acknowledgments

This work was supported by NIH grants GM114178 (R. Z.) and GM114178-02S2 (S. E.).

References

1. Bragg WL. The structure of some crystals as indicated by their diffraction of X-rays. Proceedings of the royal society A. 1913
2. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*. 1994; 372:432–438. [PubMed: 7984237]
3. Agrawal AA, et al. An extended U2AF(65)-RNA-binding domain recognizes the 3' splice site signal. *Nature communications*. 2016; 7:10950.
4. Jacewicz A, Chico L, Smith P, Schwer B, Shuman S. Structural basis for recognition of intron branchpoint RNA by yeast Msl5 and selective effects of interfacial mutations on splicing of yeast pre-mRNAs. *RNA*. 2015; 21:401–414. [PubMed: 25587180]
5. Montemayor EJ, et al. Core structure of the U6 small nuclear ribonucleoprotein at 1.7-Å resolution. *Nat Struct Mol Biol*. 2014; 21:544–551. [PubMed: 24837192]
6. Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*. 2009; 458:475–480. [PubMed: 19325628]
7. Weber G, Trowitzsch S, Kastner B, Luhrmann R, Wahl MC. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *Embo J*. 2010; 29:4172–4184. [PubMed: 21113136]
8. Kondo Y, Oubridge C, van Roon AM, Nagai K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *Elife*. 2015; 4

9. Zhang L, et al. Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants. *Protein Sci.* 2007; 16:1024–1031. [PubMed: 17473007]
10. Yang K, Zhang L, Xu T, Heroux A, Zhao R. Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc Natl Acad Sci U S A.* 2008; 105:13817–13822. [PubMed: 18779563]
11. Pena V, Liu S, Bujnicki JM, Luhrmann R, Wahl MC. Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. *Mol Cell.* 2007; 25:615–624. [PubMed: 17317632]
12. Nguyen TH, et al. Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure.* 2013; 21:910–919. [PubMed: 23727230]
13. Mozaffari-Jovin S, et al. Inhibition of RNA Helicase Brr2 by the C-Terminal Tail of the Spliceosomal Protein Prp8. *Science.* 2013
14. Mozaffari-Jovin S, et al. The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes Dev.* 2012; 26:2422–2434. [PubMed: 23124066]
15. Galej WP, Oubridge C, Newman AJ, Nagai K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature.* 2013; 493:638–643. [PubMed: 23354046]
16. Zhao R, Shen J, Green MR, MacMorris M, Blumenthal T. Crystal structure of UAP56, a DExD/H-box protein involved in pre-mRNA splicing and mRNA export. *Structure (Camb).* 2004; 12:1373–1381. [PubMed: 15296731]
17. Shi H, Cordin O, Minder CM, Linder P, Xu RM. Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc Natl Acad Sci U S A.* 2004; 101:17628–17633. [PubMed: 15585580]
18. Walbott H, et al. Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *Embo J.* 2010; 29:2194–2204. [PubMed: 20512115]
19. Kudlinzki D, Schmitt A, Christian H, Ficner R. Structural analysis of the C-terminal domain of the spliceosomal helicase Prp22. *Biological chemistry.* 2012; 393:1131–1140. [PubMed: 23096351]
20. Zhang L, et al. Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. *Nat Struct Mol Biol.* 2009; 16:731–739. [PubMed: 19525970]
21. Santos KF, et al. Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. *Proc Natl Acad Sci U S A.* 2012; 109:17418–17423. [PubMed: 23045696]
22. Absmeier E, et al. The large N-terminal region of the Brr2 RNA helicase guides productive spliceosome activation. *Genes Dev.* 2015; 29:2576–2587. [PubMed: 26637280]
23. Jacewicz A, Schwer B, Smith P, Shuman S. Crystal structure, mutational analysis and RNA-dependent ATPase activity of the yeast DEAD-box pre-mRNA splicing factor Prp28. *Nucleic Acids Res.* 2014; 42:12885–12898. [PubMed: 25303995]
24. Hope H. Cryocrystallography of biological macromolecules: a generally applicable method. *Acta crystallographica. Section B, Structural science.* 1988; 44(Pt 1):22–26. [PubMed: 3271102]
25. Ravelli RB, Garman EF. Radiation damage in macromolecular cryocrystallography. *Curr Opin Struct Biol.* 2006; 16:624–629. [PubMed: 16938450]
26. Garman E. 'Cool' crystals: macromolecular cryocrystallography and radiation damage. *Curr Opin Struct Biol.* 2003; 13:545–551. [PubMed: 14568608]
27. Rossmann MG. The molecular replacement method. *Acta crystallographica. Section A, Foundations of crystallography.* 1990; 46(Pt 2):73–82. [PubMed: 2180438]
28. Green DW, Ingram VM, Perutz MF. The structure of Haemoglobin. IV. Sigh determination by the isomorphous replacement method. *Proceedings of the royal society A.* 1954; 225:287–307.
29. Hendrickson, WA., Ogata, CM. Phase Determination from Multiwavelength Anomalous Diffraction Measurements. In: Carters, C., Sweet, R., editors. *Methods in Enzymology.* Vol. 276. Academic Press, Inc; 1997. p. 494–523.
30. Merk A, et al. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell.* 2016; 165:1698–1707. [PubMed: 27238019]

31. Villa T, Pleiss JA, Guthrie C. Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell*. 2002; 109:149–152. [PubMed: 12007401]
32. Dudev T, Chang LY, Lim C. Factors governing the substitution of La³⁺ for Ca²⁺ and Mg²⁺ in metalloproteins: a DFT/CDM study. *Journal of the American Chemical Society*. 2005; 127:4091–4103. [PubMed: 15771547]
33. Pike AC, Garman EF, Krojer T, von Delft F, Carpenter EP. An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr D Struct Biol*. 2016; 72:303–318. [PubMed: 26960118]
34. Ennifar E, Walter P, Dumas P. A crystallographic study of the binding of 13 metal ions to two related RNA duplexes. *Nucleic Acids Res*. 2003; 31:2671–2682. [PubMed: 12736317]
35. Chapman HN, et al. Femtosecond X-ray protein nanocrystallography. *Nature*. 2011; 470:73–77. [PubMed: 21293373]
36. Martin-Garcia JM, Conrad CE, Coe J, Roy-Chowdhury S, Fromme P. Serial femtosecond crystallography: A revolution in structural biology. *Archives of biochemistry and biophysics*. 2016; 602:32–47. [PubMed: 27143509]
37. Kupitz C, et al. Microcrystallization techniques for serial femtosecond crystallography using photosystem II from *Thermosynechococcus elongatus* as a model system. *Philos Trans R Soc Lond B Biol Sci*. 2014; 369:20130316. [PubMed: 24914149]
38. Koopmann R, et al. In vivo protein crystallization opens new routes in structural biology. *Nature methods*. 2012; 9:259–262. [PubMed: 22286384]
39. Redecke L, et al. Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser. *Science*. 2013; 339:227–230. [PubMed: 23196907]
40. Gallat FX, et al. In vivo crystallography at X-ray free-electron lasers: the next generation of structural biology? *Philos Trans R Soc Lond B Biol Sci*. 2014; 369:20130497. [PubMed: 24914164]
41. Ginn HM, et al. Structure of CPV17 polyhedrin determined by the improved analysis of serial femtosecond crystallographic data. *Nature communications*. 2015; 6:6435.
42. Liu W, Ishchenko A, Cherezov V. Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. *Nature protocols*. 2014; 9:2123–2134. [PubMed: 25122522]
43. Ishchenko A, Cherezov V, Liu W. Preparation and Delivery of Protein Microcrystals in Lipidic Cubic Phase for Serial Femtosecond Crystallography. *Journal of visualized experiments: JoVE*. 2016
44. Philipp HT, LJK, Hromalik MS, Tate MW, Gruner SM. Femtosecond radiation experiment detector for x-ray free-electron laser (XFEL) coherent X-ray imaging. *IEEE Trans Nucl Sci*. 2010; 57:3795–3799.
45. White TA, et al. Crystallographic data processing for free-electron laser sources. *Acta Crystallogr D Biol Crystallogr*. 2013; 69:1231–1240. [PubMed: 23793149]
46. Hattne J, et al. Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. *Nature methods*. 2014; 11:545–548. [PubMed: 24633409]
47. Kabsch W. Xds. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:125–132. [PubMed: 20124692]
48. Kirian RA, et al. Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals. *Acta crystallographica. Section A, Foundations of crystallography*. 2011; 67:131–140. [PubMed: 21325716]
49. Hunter MS, et al. Fixed-target protein serial microcrystallography with an x-ray free electron laser. *Scientific reports*. 2014; 4:6026. [PubMed: 25113598]
50. Barends TR, et al. De novo protein crystal structure determination from X-ray free-electron laser data. *Nature*. 2014; 505:244–247. [PubMed: 24270807]
51. Spence JC, et al. Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals. *Optics express*. 2011; 19:2866–2873. [PubMed: 21369108]
52. Kirian RA, et al. Phasing coherently illuminated nanocrystals bounded by partial unit cells. *Philos Trans R Soc Lond B Biol Sci*. 2014; 369:20130331. [PubMed: 24914158]

53. Liu H, Poon BK, Saldin DK, Spence JC, Zwart PH. Three-dimensional single-particle imaging using angular correlations from X-ray laser data. *Acta crystallographica. Section A, Foundations of crystallography*. 2013; 69:365–373. [PubMed: 23778093]
54. Galli L, et al. Towards RIP using free-electron laser SFX data. *Journal of synchrotron radiation*. 2015; 22:249–255. [PubMed: 25723926]
55. Shi D, Nannenga BL, Iadanza MG, Gonen T. Three-dimensional electron crystallography of protein microcrystals. *Elife*. 2013; 2:e01345. [PubMed: 24252878]
56. Doerr A. Electron crystallography goes 3D with MicroED. *Nature methods*. 2014; 11:6–7.
57. Nannenga BL, Gonen T. Protein structure determination by MicroED. *Curr Opin Struct Biol*. 2014; 27:24–31. [PubMed: 24709395]
58. Nannenga BL, Gonen T. MicroED opens a new era for biological structure determination. *Curr Opin Struct Biol*. 2016; 40:128–135. [PubMed: 27701014]
59. Shi D, Lewis MR, Young HS, Stokes DL. Three-dimensional crystals of Ca²⁺-ATPase from sarcoplasmic reticulum: merging electron diffraction tilt series and imaging the (h, k, 0) projection. *J Mol Biol*. 1998; 284:1547–1564. [PubMed: 9878370]
60. Jiang L, Georgieva D, Nederlof I, Liu Z, Abrahams JP. Image processing and lattice determination for three-dimensional nanocrystals. *Microsc Microanal*. 2011; 17:879–885. [PubMed: 22094021]
61. Henderson R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys*. 1995; 28:171–193. [PubMed: 7568675]
62. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*. 2011; 67:271–281. [PubMed: 21460445]
63. Liu S, et al. Atomic resolution structure determination by the cryo-EM method MicroED. *Protein Sci*. 2017; 26:8–15. [PubMed: 27452773]
64. Sawaya MR, et al. Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. *Proc Natl Acad Sci U S A*. 2016; 113:11232–11236. [PubMed: 27647903]
65. Otwinowski, Z., Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, CW., Sweet, RM., editors. *Methods in Enzymology*. Vol. 276. Academic Press; New York: 1997. p. 307–326.
66. Pflugrath JW. The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr*. 1999; 55:1718–1725. [PubMed: 10531521]
67. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:213–221. [PubMed: 20124702]
68. Project, N.C.C. The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst*. 1994; D50:760–763.
69. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:486–501. [PubMed: 20383002]
70. Tani K, et al. Mechanism of aquaporin-4's fast and highly selective water conduction and proton exclusion. *J Mol Biol*. 2009; 389:694–706. [PubMed: 19406128]
71. Fabrizio P, et al. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell*. 2009; 36:593–608. [PubMed: 19941820]
72. Kellner N, et al. Developing genetic tools to exploit *Chaetomium thermophilum* for biochemical analyses of eukaryotic macromolecular assemblies. *Scientific reports*. 2016; 6:20937. [PubMed: 26864114]
73. Wernimont A, Edwards A. In situ proteolysis to generate crystals for structure determination: an update. *PLoS One*. 2009; 4:e5094. [PubMed: 19352432]
74. Taichert MJ, Fourmann JB, Luhrmann R, Ficner R. Structural insights into the mechanism of the DEAH-box RNA helicase Prp43. *Elife*. 2017; 6
75. Absmeier E, Becke C, Wollenhaupt J, Santos KF, Wahl MC. Interplay of cis- and trans-regulatory mechanisms in the spliceosomal RNA helicase Brr2. *Cell Cycle*. 2017; 16:100–112. [PubMed: 27880071]

76. Cretu C, et al. Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. *Mol Cell*. 2016; 64:307–319. [PubMed: 27720643]
77. Walter TS, et al. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure*. 2006; 14:1617–1622. [PubMed: 17098187]
78. Derewenda ZS. Rational protein crystallization by mutational surface engineering. *Structure*. 2004; 12:529–535. [PubMed: 15062076]
79. Moon AF, Mueller GA, Zhong X, Pedersen LC. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci*. 2010; 19:901–913. [PubMed: 20196072]
80. Hu J, et al. TargetCrys: protein crystallization prediction by fusing multi-view features with two-layered SVM. *Amino acids*. 2016; 48:2533–2547. [PubMed: 27299433]
81. Vedadi M, et al. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A*. 2006; 103:15835–15840. [PubMed: 17035505]
82. Kent S, et al. Through the looking glass--a new world of proteins enabled by chemical synthesis. *Journal of peptide science: an official publication of the European Peptide Society*. 2012; 18:428–436. [PubMed: 22674813]
83. Maeki M, Yamaguchi H, Tokeshi M, Miyazaki M. Microfluidic Approaches for Protein Crystal Structure Analysis. *Analytical sciences: the international journal of the Japan Society for Analytical Chemistry*. 2016; 32:3–9. [PubMed: 26753699]
84. Khurshid S, Saridakis E, Govada L, Chayen NE. Porous nucleating agents for protein crystallization. *Nature protocols*. 2014; 9:1621–1633. [PubMed: 24922271]
85. Ghatak AS, Ghatak A. Disordered nanowrinkle substrates for inducing crystallization over a wide range of concentration of protein and precipitant. *Langmuir: the ACS journal of surfaces and colloids*. 2013; 29:4373–4380. [PubMed: 23517340]
86. Zhu Y, et al. Nanoliter-scale protein crystallization and screening with a microfluidic droplet robot. *Scientific reports*. 2014; 4:5046. [PubMed: 24854085]

This review discusses the theoretical basis, instrument requirements, troubleshooting, and exciting potential of X-ray crystallography, serial femtosecond crystallography, and 3D electron crystallography to further our understanding of pre-mRNA splicing, a critical event in gene expression of all eukaryotes.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.
Examples of crystal structures of splicing related complexes with their PDB codes and resolutions listed under each structure.

Fig. 2.
A schematic representation of the serial femtosecond crystallography setup.