# Unraveling the Molecular Mechanism of Pre-mRNA Splicing From Multi-Scale Simulations

*Lorenzo Casalino[1] and Alessandra Magistrato[2]\**

*[1] Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, United States, [2] Consiglio Nazionale delle Ricerche–Istituto Officina dei Materiali, International School for Advanced Studies (SISSA), Trieste, Italy*

## INTRODUCTION

The removal of non-coding introns within a precursor messenger RNA (pre-mRNA) transcript is a key step of gene expression and regulation, occurring via two transesterification reactions mediated by at least two $Mg^{2+}$ ions (Kastner et al., 2019). Whereas in lower organisms this process is self-regulated by group II intron ribozymes (G2IRs) performing their own excision from a pre-mRNA strand, in eukaryotes, due to the increased complexity of the genome, these autocatalytic RNAs have evolved into a majestic protein/RNA machinery—the spliceosome (SPL)—composed of hundreds of proteins and five small-nuclear (sn)RNA filaments (Marcia and Pyle, 2012; Yan et al., 2019). The SPL, acting as a protein-directed metallo-ribozyme, promotes the conversion of pre-mRNAs into mature mRNAs. This massive architecture revolves around its central core constituted by Spp42/Prp8 protein (*S. Pombe/S. Cerevisiae* or human, respectively) and a catalytic site fully resembling that of G2IRs (Yan et al., 2019). As the most eminent genome tailor, the SPL undergoes a relentless compositional and conformational remodeling, repetitively assembling and transforming at every splicing cycle into eight distinct complexes (A, B, B$^{act}$, B*, C, C*, P, ILS) to achieve splicing with a single nucleotide precision.

Recent developments in single-particle cryo-EM have led to elucidate a plethora of near-atomic resolution structures of SPL complexes from human and yeast strains, thus allowing decades of biochemical, structural and functional studies to be interpreted. In this context, multiscale simulations can contribute to deciphering the intricacies of the splicing mechanism by assessing the chemical details of the pre-mRNA cleavage, and the role of the extraordinarily convoluted protein/RNA environment in creating the appropriate structural scaffold that finely modulates introns removal (Yan et al., 2019). Nevertheless, the size and the inner complexity of the SPL machinery require a wise use of advanced multiscale simulations to tackle the many different peculiarities of its mechanism, as shown in the following showcased studies.

## CHEMICAL MECHANISM OF PRE-mRNA SPLICING IN PROKARYOTES

The structure of the SPL catalytic site, impressively similar to that of its evolutionary predecessor G2IRs, is well-preserved among the distinct structures that have been solved. A series of crystal structures from *Oceanobacillus iheyensis* captured group IIC intron at sequential stages of the catalytic process, allowing a first structural breakthrough for unraveling the chemical mechanism

of pre-mRNA splicing (Marcia and Pyle, 2012). These crystallographic reconstructions revealed an active site containing a four-metal-ion cluster made of two $Mg^{2+}$ and two $K^+$ ions, the former being catalytically active, while the latter most likely playing a structural role. Building on these structures, classical and hybrid quantum-classical QM/MM simulations enabled the investigation of the first and rate-determining step of the splicing reaction as promoted by G2IRs (Casalino et al., 2016). In particular, this work focused on the water-mediated 5′-exon cleavage mechanism (hydrolytic path). In fact, in G2IRs the hydrolytic catalysis can be as operative as the branching pathway, where, instead, the reaction is started by a conserved bulged adenosine within the branch point sequence (BPS). By using classical and QM(Car–Parrinello)/MM molecular dynamics (MD), with the QM part described at Density Functional Theory (DFT)-BLYP level of theory, and the MM part treated with the AMBER- ff12SB (ff99+bsc0+χOL3) force field (FF) (Pérez et al., 2007; Zgarbová et al., 2011; Maier et al., 2015), in combination with thermodynamic integration to enable the reaction event within the limited time-scale of the QM/MM MD simulations, this study unveiled a novel dissociative two-$Mg^{2+}$-ion mechanism in which the bulk water acts as general base (Casalino et al., 2016).

The two-$Mg^{2+}$-ion motif is a well-established catalytic cofactor shared by many enzymes processing nucleic acids. In these enzymes, the phosphodiester bond hydrolysis is believed to occur according to the Steitz and Steitz's proposal. In its original postulation, confirmed by distinct computational studies, the two $Mg^{2+}$ ions act as Lewis acids activating the nucleophile, stabilizing the leaving group and the transition state (Palermo et al., 2015; Sgrignani and Magistrato, 2015). At variance with this, in G2IRs a dissociative mechanism takes place, with the reactive water detaching from the $Mg^{2+}$ ion and performing the attack on the scissile phosphate while still in its non-deprotonated form. Only after the nucleophilic substitution has started, the catalytic water eventually releases its proton to the bulk water and terminates the reaction. In this mechanism one $Mg^{2+}$ ion activates the scissile phosphate group by making it more electrophilic, while the second $Mg^{2+}$ stabilizes the leaving group. Hence, in this chemical path the role of the two $Mg^{2+}$ ions remarkably differs from that of protein enzymes performing a two-metal-aided catalysis. It is tantalizing to believe that this mechanism may be specific for ribozymes, where the catalytic site is exclusively formed by the RNA sugar–phosphate backbone bearing a lower specificity/efficiency to promote the reaction than that of enzymes. This peculiar mechanism may represent an ancestral version of the two-$Mg^{2+}$-ion catalysis later evolved in enzymes and in protein-directed ribozymes (spliceosome) (Casalino et al., 2016).

## SPLICING MECHANISM MODULATION BY THE PROTEIN ENVIRONMENT

In spite of the large number of cryo-EM structures of the SPL published as of yet, no catalytically competent form has been trapped, thus hampering a study of the chemical mechanism of splicing in eukaryotes. Moreover, the large size and complexity of the SPL pose serious challenges even when attempting to unravel its functional properties. Indeed, the deposited cryo-EM maps usually have a resolution ranging between 3 and 4 Å in the core and even reaching lower values in the peripheral regions of the macromolecular assembly, which often displays structural gaps (Kastner et al., 2019; Plaschka et al., 2019; Yan et al., 2019). For these reasons, in order to perform all-atom simulations of the SPL it is mandatory to find a compromise between system size and accuracy. In the first MD simulation study published to date, based on the first near-atomistic SPL structure solved from yeast S. Pombe capturing the intron lariat spliceosome (ILS) complex (Yan et al., 2015), two explicitly solvated core model-systems containing ∼1,000,000 atoms were built and simulated via multi-replica MD simulations for a cumulative statistics of few microseconds (**Figure 1**). In these simulations the AMBER-ff12SB FF was used for proteins (Maier et al., 2015), whereas ff99+bsc0+χOL3 FF was adopted for RNAs (Pérez et al., 2007; Zgarbová et al., 2011).

Correlation analyses, principal-component analysis (PCA), and electrostatic calculations disentangled the cooperative motions underlying the SPL functional dynamics, unraveling the role of electrostatics in modulating these movements (Casalino et al., 2018). The simulations provided unprecedented insights on the SPL functional plasticity, assigning to Spp42 (Prp8 in human) a central role in finely directing the motions of many distinct SPL components. Metaphorically, the resulting scenario is that of Spp42 as an orchestra conductor of the gene regulation symphony. The essential dynamics extracted from the PCA revealed, consistently with the stage of the splicing cycle investigated, an electrostatically-driven displacement and unrolling of the U2/intron-lariat branch helix co-promoted by Cwf19 (CWF19L2 in human) and Spp42, both involved in the ILS disassembly (Casalino et al., 2018). Strikingly, the implication of Cwf19 in the branch helix unwinding was thereafter corroborated by recent cryo-EM studies on the human SPL (Zhang et al., 2019). Despite the intrinsic limitations of this study due to the large size of the system and the well-known flaws of RNA (Šponer et al., 2018) and $Mg^{2+}$ (Casalino et al., 2017) FFs, this study has opened new avenues for probing this incredible machinery with atomic-level simulations.

## DISCUSSION

A detailed comprehension of the molecular terms of eukaryotic splicing has entailed implications for revolutionary gene modulation therapies and drug discovery studies aimed at fighting the over 200 human diseases associated with splicing defects. Upon the deposition of the first SPL structure from yeast in 2015, many human cryo-EM maps have been solved, thus opening new opportunities to dissect detailed aspects of this machinery (Kastner et al., 2019; Plaschka et al., 2019; Yan et al., 2019). Among the unmet questions that need to be solved
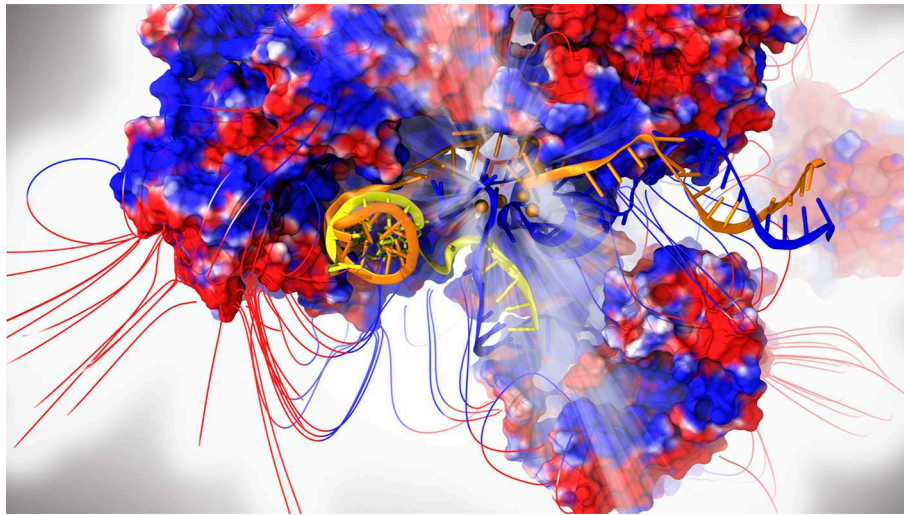
**FIGURE 1 |** The intron lariat spliceosome complex. Proteins are shown with electrostatic surface (blue/red colors for positive/negative charges, respectively) together with the respective field lines. The intron lariat (yellow), U2 (orange) and U6 snRNA are represented as cartoon. $Mg^{2+}$ ions are depicted as orange spheres. The catalytic center is highlighted by light rays.

from an atomic-level perspective, the molecular recognition mechanism by which SPL can recruit key intronic sequences at the 3′ and 5′ splice sites, as well as that of the conserved BPS, stands out. The subtle molecular foundations ensuring the reliable identification of authentic consensus splice sites (constitutive splicing), while simultaneously providing some flexibility in the selection of non-consensus ones (alternative splicing) remain unclear. Deregulated constitutive/alternative splicing is well-known to lead to aberrant mRNA transcripts, which may either induce non-sense mediated decay or result in functionally-altered proteins, deleteriously affecting cells functions. In this context, research efforts have been devoted to understanding the mechanism by which mutations of the splicing factor SF3B1 affect BPS recognition, thus leading to aberrant splicing and to the outbreak of distinct hematological malignancies (Cretu et al., 2018). Splicing modulators hitherto trapped in SF3B1 have been found to target the BPS recognition site, elucidating the structural basis of their inhibition mechanism (Cretu et al., 2018; Zhang et al., 2018). Large-scale genomics studies have recently indicated that splicing abnormalities and cancer onset are strongly entwined. Thus, while eagerly awaiting for more structures to be released in the forthcoming years, we expect SPL to become an increasingly important subject of drug design studies tackling distinct types of cancer.

Although the reported results from all-atom simulations— and all the possible future applications—appear to be very encouraging (Casalino et al., 2018; Palermo et al., 2019), several challenges need to be tackled, starting from the amelioration of current RNA and protein/RNA FFs (Šponer et al., 2018). Moreover, even though we have assisted to a fast development

of computer hardware and software allowing for brute force unbiased MD simulations, biologically relevant time scales remain computationally extremely demanding and out of reach to most computational labs. In this respect, enhanced sampling and free energy methods to study rare events taking place in complex biological contexts call for further improvements (Miao and McCammon, 2016; Valsson et al., 2016). The presence of metals within the catalytic core of the SPL, which in fact makes it a protein-directed metallo-ribozyme, poses serious difficulties for a reliable fully classical prediction of its properties (Vidossich and Magistrato, 2014; Brunk and Rothlisberger, 2015). For this reason, the use of highly parallel QM/MM MD schemes capable of better exploiting large computational infrastructure would be ideal (Bolnykh et al., 2019; Olsen et al., 2019). A timely fashion communication between the QM and MM would in fact allow more efficient QM(DFT)/MM MD calculations, accounting for larger QM regions and longer simulation time than the accustomed ∼100 atoms and ∼100s ps time scale, respectively.

In this scenario, we expect that new methodological advances in computer simulations, modeling and analysis techniques will foster atomic-level studies of the SPL, contributing to an utter comprehension of this fundamental step of gene expression. This will also be of service for a better understanding of the allosteric signaling between distal sites, which occurs via the entangled protein/RNA networks characterizing the SPL, and for the discovery of druggable allosteric sites (Palermo et al., 2017). On a final note, we hope that any related breakthrough might help to elucidate the role of splicing pathways in cancer, concretely opening appealing opportunities for creating therapeutic approaches and innovative gene manipulations tools.

## AUTHOR CONTRIBUTIONS

LC and AM designed research and wrote the paper.

## REFERENCES

Bolnykh, V., Olsen, J. M. H., Meloni, S., Bircher, M. P., Ippoliti, E., Carloni, P., et al. (2019). Extreme scalability of DFT-Based QM/MM MD simulations using MiMiC. *ChemRxiv*. doi: 10.26434/chemrxiv.8067527.v1. [Epub ahead of print].

Brunk, E., and Rothlisberger, U. (2015). Mixed Quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.* 115, 6217–6263. doi: 10.1021/cr500628b

Casalino, L., Palermo, G., Abdurakhmonova, N., Rothlisberger, U., and Magistrato, A. (2017). Development of site-specific Mg(2+)-RNA force field parameters: a dream or reality? Guidelines from combined molecular dynamics and quantum mechanics simulations. *J. Chem. Theory Comput.* 13, 340–352. doi: 10.1021/acs.jctc.6b00905

Casalino, L., Palermo, G., Rothlisberger, U., and Magistrato, A. (2016). Who activates the nucleophile in ribozyme catalysis? an answer from the splicing mechanism of group II introns. *J. Am. Chem. Soc.* 138, 10374–10377. doi: 10.1021/jacs.6b01363

Casalino, L., Palermo, G., Spinello, A., Rothlisberger, U., and Magistrato, A. (2018). All-atom simulations disentangle the functional dynamics underlying gene maturation in the intron lariat spliceosome. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6584–6589. doi: 10.1073/pnas.1802963115

Cretu, C., Agrawal, A. A., Cook, A., Will, C. L., Fekkes, P., Smith, P. G., et al. (2018). Structural basis of splicing modulation by antitumor macrolide compounds. *Mol. Cell.* 70, 265–273 e268. doi: 10.1016/j.molcel.2018.03.011

Kastner, B., Will, C. L., Stark, H., and Luhrmann, R. (2019). Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb. Perspect. Biol.* 2019: a032417. doi: 10.1101/cshperspect.a032417

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255

Marcia, M., and Pyle, A. M. (2012). Visualizing group II intron catalysis through the stages of splicing. *Cell* 151, 497–507. doi: 10.1016/j.cell.2012.09.033

Miao, Y., and McCammon, J. A. (2016). Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol. Simul.* 42, 1046–1055. doi: 10.1080/08927022.2015.1121541

Olsen, J. M. H., Bolnykh, V., Meloni, S., Ippoliti, E., Bircher, M. P., Carloni, P., et al. (2019). MiMiC: a novel framework for multiscale modeling in computational chemistry. *J. Chem. Theory Comput.* 15, 3810–3823. doi: 10.1021/acs.jctc.9b00093

Palermo, G., Casalino, L., Magistrato, A., and Andrew McCammon, J. (2019). Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. *J. Struct. Biol.* 206, 267–279. doi: 10.1016/j.jsb.2019.03.004

Palermo, G., Cavalli, A., Klein, M. L., Alfonso-Prieto, M., Dal Peraro, M., and De Vivo, M. (2015). Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.* 48, 220–228. doi: 10.1021/ar500314j

Palermo, G., Ricci, C. G., Fernando, A., Basak, R., Jinek, M., Rivalta, I., et al. (2017). Protospacer adjacent motif-induced allostery activatesCRISPR-Cas9. *J. Am. Chem. Soc.* 139, 16028–16031. doi: 10.1021/jacs.7b05313

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E. 3rd, Laughton, C. A., et al. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* 92, 3817–3829. doi: 10.1529/biophysj.106.097782

Plaschka, C., Newman, A. J., and Nagai, K. (2019). Structural basis of nuclear pre-mRNA splicing: lessons from yeast. *Cold Spring Harb. Perspect. Biol.* 11:a032391. doi: 10.1101/cshperspect.a032391

Sgrignani, J., and Magistrato, A. (2015). QM/MM MD simulations on the enzymatic pathway of the human flap endonuclease (hFEN1) elucidating common cleavage pathways to RNase H enzymes. *ACS Catal.* 5, 3864–3875. doi: 10.1021/acscatal.5b00178

Šponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R. A., et al. (2018). RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* 118, 4177–4338. doi: 10.1021/acs.chemrev.7b00427

Valsson, O., Tiwary, P., and Parrinello, M. (2016). Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. *Annu. Rev. Phys. Chem.* 67, 159–184. doi: 10.1146/annurev-physchem-040215-112229

Vidossich, P., and Magistrato, A. (2014). QM/MM molecular dynamics studies of metal binding proteins. *Biomolecules* 4, 616–645. doi: 10.3390/biom4030616

Yan, C., Hang, J., Wan, R., Huang, M., Wong, C. C., and Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* 349, 1182–1191. doi: 10.1126/science.aac7629

Yan, C., Wan, R., and Shi, Y. (2019). Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harb. Perspect. Biol.* 11:a032409. doi: 10.1101/cshperspect.a032409

Zgarbová, M., Otyepka, M., Sponer, J., Mládek, A., Banáš, P., Cheatham, T. E. III., et al. (2011). Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* 7, 2886–2902. doi: 10.1021/ct200162x

Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., and Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. *Cell Res.* 28, 307–322. doi: 10.1038/cr.2018.14

Zhang, X., Zhan, X., Yan, C., Zhang, W., Liu, D., Lei, J., et al. (2019). Structures of the human spliceosomes before and after release of the ligated exon. *Cell Res.* 29, 274–285. doi: 10.1038/s41422-019-0143-x