Published in final edited form as:

J Struct Biol. 2019 June 01; 206(3): 267–279. doi:10.1016/j.jsb.2019.03.004.

Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations

Giulia Palermo^{a,*}, Lorenzo Casalino^b, Alessandra Magistrato^e, and J. Andrew McCammon^{b,c,d}

^aDepartment of Bioengineering, Bourns College of Engineering, University of California Riverside, 900 University Avenue, Riverside, CA 92521, United States

^bDepartment of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, United States

^cDepartment of Pharmacology, University of California San Diego, La Jolla, CA 92093, United States

^dNational Biomedical Computation Resource, University of California San Diego, La Jolla, CA 92093, United States

^eConsiglio Nazionale delle Ricerche–Istituto Officina dei Materiali, Democritos National Simulation Center c/o International School for Advanced Studies (SISSA), 34136 Trieste, Italy

Abstract

Noncoding RNA (ncRNA) has a key role in regulating gene expression, mediating fundamental processes and diseases via a variety of yet unknown mechanisms. Here, we review recent applications of conventional and enhanced Molecular Dynamics (MD) simulations methods to address the mechanistic function of large biomolecular systems that are tightly involved in the ncRNA function and that are of key importance in life sciences. This compendium focuses on three biomolecular systems, namely the CRISPR-Cas9 genome editing machinery, group II intron ribozyme and the ribonucleoprotein complex of the spliceosome, which edit and process ncRNA. We show how the application of a novel accelerated MD simulations method has been key in disclosing the conformational transitions underlying RNA binding in the CRISPR-Cas9 complex, suggesting a mechanism for RNA recruitment and clarifying the conformational changes required for attaining genome editing. As well, we discuss the use of mixed quantum-classical MD simulations in deciphering the catalytic mechanism of RNA splicing as operated by group II intron ribozyme, one of the largest ncRNA structures crystallized so far. Finally, we debate the future challenges and opportunities in the field, discussing the recent application of MD simulations for unraveling the functional biophysics of the spliceosome, a multi-mega Dalton complex of proteins and small nuclear RNAs that performs RNA splicing in humans. This showcase of applications highlights the current talent of MD simulations to dissect atomic-level details of complex biomolecular systems instrumental for the design of finely engineered genome editing machines. As well, this review aims at inspiring future investigations of several other ncRNA regulatory

*Corresponding author. giulia.palermo@ucr.edu (G. Palermo).

systems, such as micro and small interfering RNAs, which achieve their function and specificity using RNA-based recognition and targeting strategies.

Keywords

CRISPR-Cas9; RNA splicing; Genome editing; Group II intron; Spliceosome

1. Introduction

Noncoding RNA (ncRNA) has a key role in regulating gene expression, mediating fundamental processes and diseases via a variety of yet unknown mechanisms (Cech and Steitz, 2014). The biological function of ncRNAs has direct effects on transcription, on RNA processing and on translation, as well as on DNA synthesis and chromatin remodeling (Sabin et al., 2013), with repercussions on the onset of cancer and genetic diseases (Anastasiadou et al., 2018; Matsui and Corey, 2017). In these processes, ncRNA mechanistically operates through distinct means, which involve catalysis and a tight interplay with protein complexes, ultimately regulating gene expression. In this multifaceted scenario, we are at the beginning to understand the nature and the broad variety of functions exerted by ncRNA, the latter being the most intriguing “*dark matter*” of biology.

Understanding the mechanisms by which ncRNA molecules carry out their biological functions requires knowledge of their structure and dynamics (Cruz and Westhof, 2009; Sanbonmatsu, 2016). In this respect, extraordinary developments in the field of structural biology – fostered by the recent increase in resolution of cryo-electron microscopy techniques – are providing unprecedented pictures of the macromolecular assemblies involved in ncRNA function (Cate, 2016; Fica and Nagai, 2017; Zhao and Pyle, 2017). In parallel, molecular simulations are attempting to unravel the mechanistic basis of ncRNA molecules with atomic-level detail. Molecular Dynamics (MD) virtually mimic the chemical, physical and thermodynamic conditions, to predict atoms interactions and dynamics following the fundamental laws of Newtonian physics (Karplus and McCammon, 2002; McCammon et al., 1977). This approach uniquely complements structural biology, going beyond overall the architecture, describing conformational landscapes and relating structural transitions to function. After the first MD simulation of RNA (Harvey et al., 1984), we have seen an explosion of applications and the development of novel and more robust MD methodologies for the study of RNA structure and dynamics (Banas et al., 2010; Bergonzo et al., 2016, 2015, 2014; Bottaro et al., 2016; Casalino, 2017; Cheatham and Young, 2000; Chen and García, 2013; Hashem and Auffinger, 2009; Hayatshahi et al., 2017; Perez et al., 2007; Schrod et al., 2015; Tan et al., 2018; Zgarbova et al., 2011). Early computational studies on non-coding RNA have been performed on riboswitches and micro RNA, highlighting the capability of the methodology in capturing the mechanistic aspects that underlie their biological function. A recent review article discussed the applications of MD simulations methods in the broad field of RNA, including also studies on riboswitches and their interactions with ligands, with implications for the design of novel antibiotics targeting riboswitches (Sponer et al., 2018). This contribution further described the efforts of computational biophysics for the study of RNA structure, catalysis and function, indicating

also the drawbacks of the methodology (such as the need of improving the available empirical force fields) and overall highlighting the potential of theoretical models in unraveling structure and dynamics of RNA.

Here, we review the latest applications of conventional and emerging enhanced MD methods to address the mechanistic function of large biomolecular systems that are tightly involved in the ncRNA function and that are of key importance in life sciences. Specifically, we discuss the use of MD simulations methods to unravel the function and improve applications of the RNA-guided CRISPR-Cas9 system (Jinek et al., 2012), which is emerging as the forefront technology for editing and manipulating nucleic acids. Moreover, we showcase the use of molecular simulations to access the catalytic features of ncRNAs and their functional interplay with protein complexes, focusing on group II intron ribozyme (Toor et al., 2009; Marcia and Pyle, 2012; Marcia et al., 2013; Pyle, 2016) and on the spliceosome (Papasaikas and Valcarcel, 2016). Such systems attain the editing of premature messenger RNA (pre-mRNA) by performing RNA splicing, a fundamental biological step prior to translation. The size and complexity of these biomolecular systems represent a challenge for MD simulations, which requires the application of advanced methods overall attempting the characterization of an intricate and fascinating biological function. In the following, we discuss the application of a variety of computational methodologies, including traditional approaches such as classical and quantum (QM)-based MD, as well as emerging enhanced sampling methods. We show how the use of accelerated MD, which improves the configurational sampling and access long time scale biophysical processes, has been key for deciphering the mechanism of RNA binding in the CRISPR-Cas9 complex and the conformational changes of the endonuclease Cas9 toward DNA editing. Our compendium also pinpoints the essential use of QM and free energy methods to characterize the RNA autocatalytic function operated by self-splicing ribozymes. We report the splicing mechanism in group II intron ribozymes, providing the description of the two-metal-aided catalysis in RNA and highlighting the remarkable differences with two-metal-aided protein enzymes. Finally, we debate the challenges and opportunities in the field, discussing the application of MD methodologies for unraveling the mechanistic function of the spliceosome, a multi-mega Dalton assembly of proteins and small nuclear RNA (snRNA), which performs RNA splicing in eukariotes, among which humans (Fica and Nagai, 2017). Overall, our showcase of applications aims to inspire future investigations of several other ncRNA regulatory systems, such as micro RNAs, small interfering RNAs and other small RNAs, which achieve their function and specificity using RNA-based recognition and targeting strategies (Gorski et al., 2017). As well, we hope to highlight the potential of novel computational methodologies for the investigation of large bimolecular complexes involved in ncRNA function. In light of the current extraordinary development of single particle cryo-EM techniques (Cheng et al., 2017; Nogales, 2016), future computational investigations will face the challenge of interpreting structure and dynamics of multi-mega Dalton biological complexes, whose characterization will require the development and application of novel and more advanced simulation methods.

2. Molecular mechanism of the CRISPR-Cas9 system

The RNA-guided CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 system recently emerged as a transformative genome-editing technology (Fig. 1). From the discovery of its RNA-programmable capabilities (Jinek et al., 2012), this system has been rapidly harnessed for efficiently manipulating nucleic acids, including ncRNAs (Yang et al., 2018), overall resulting in a variety of ground-breaking applications across basic biological sciences, medicine and biotechnology (Doudna and Charpentier, 2014; Hsu et al., 2014). In this system, the endonuclease Cas9 can be programmed with single guide RNAs to site-specifically recognize and cleave any DNA sequence bearing a Protospacer Adjacent Motif (PAM) (O'Connell et al., 2014; Sternberg et al., 2014). Upon PAM recognition, the double stranded DNA binds Cas9 by matching the RNA with one strand (i.e., the target strand, TS), while the non-target strand (NTS) is displaced. Both DNA strands are accommodated in the protein and cleaved by means of two nuclease domains (HNH and RuvC for the TS and NTS, respectively), ultimately enabling genome editing. X-ray crystallography and cryo-EM have been critically informative in depicting the structural features of the CRISPR-Cas9 complex (Jiang and Doudna, 2017; Nishimasu and Nureki, 2017). However, the mechanistic basis by which the endonuclease Cas9 associates with RNA to bind and cleave the target nucleic acids has been difficult to reach via experimental techniques. In this scenario, molecular simulations have been employed to characterize the conformational changes underlying nucleic acid binding and processing, contributing in understanding the mechanism of action of this unique genome editing machinery. As discussed below, this effort is providing a fundamental mechanistic framework that contributes to the improvement of this genome editing technology.

2.1. Conformational changes underlying RNA binding

The mechanism of RNA binding is perhaps the most intriguing biophysical process within RNA-binding proteins. Indeed, it can involve complex conformational changes of the protein, which has to cope with the high flexibility of the RNA strand to accommodate its highly charged backbone. The Cas9 protein is a large multi-domain protein, composed of a recognition lobe (REC), which mediates the nucleic acid binding through three regions (REC1–3), and a nuclease lobe (NUC) including the RuvC and HNH catalytic cores and the PAM-interacting (PI) region (Fig. 1) (Anders et al., 2014; Jiang et al., 2015, 2016; Nishimasu et al., 2014).

In this scenario, the structural transitions leading the protein from the “open” apo state to a “closed” conformation in which the enzyme binds RNA are unknown. Mechanistic questions are concerned with understanding how the REC lobe transitions to enable RNA binding. As well, considering that the binding of nucleic acids in proteins is strongly affected by electrostatic interactions, it is unknown how the electro-statics would intervene in the conformational transition of the protein to foster RNA binding. Addressing these fundamental questions represents a remarkable challenge for MD simulations, which are limited to the investigation of short time scale events (i.e., occurring over nano-to-microseconds) and cannot directly access the large amplitude conformational changes underlying RNA binding in CRISPR-Cas9. This investigation therefore required the

combination of biasing methods and enhanced sampling techniques, which extend the time scale limit of MD simulations (Abrams and Bussi, 2014; Miao and McCammon, 2016b). In detail, an initial pathway describing the conformational transition from the apo Cas9 up to RNA binding has been obtained through the application of a Targeted MD (TMD) approach (Karplus and McCammon, 2002), which reduces the root-mean-square deviation between an initial and final target conformations, therefore recovering the conformational transition (Palermo et al., 2017a). However, since TMD adds an external force to drive the conformational change, these simulations only provided a “biased” picture of the conformational transition. To overcome the limitations of this biasing approach, and to enable the unbiased investigation of intermediate states for which structural characterization is absent, extensive unconstrained simulations have been performed employing a novel Gaussian accelerated MD (GaMD) method (Miao et al., 2015). This methodology accelerates the transitions between low-energy states and enables access to long time scale biophysical events occurring on and – in some cases – beyond the millisecond range, also allowing for an accurate derivation of the associated free energy profiles (Miao and McCammon, 2016a). GaMD simulations have been applied over the initial pathway obtained from TMD simulations in multiple replicas and reaching an aggregate sampling of ~20 μ s. As a result, a stable pathway for the conformational transition has been identified (Fig. 1), in which the REC1–3 regions of the protein move in opposite directions relative to each other, leading to the closure of the REC lobe. Remarkably, the observed conformational changes confirm the previous hypotheses based on structural (Jiang et al., 2015) and biophysical (Sternberg et al., 2015) data, as well as on conventional MD (Palermo et al., 2016). Indeed, cross-correlation and principal component analyses over conventional MD runs suggested that highly coupled and specific opposite motions of the REC component would allow for the binding of the nucleic acid substrate. The free energy landscape associated to the conformational transition – i.e., the Potential of Mean Force (PMF) – describes three local minima, corresponding to the crystallographic apo (M1) and RNA-bound (M2) states, as well as to an intermediate state (M3, Fig. 2A). This intermediate conformation is characterized by the exposure toward the solvent of an arginine-rich (R-rich) α -helix. Considering that in the RNA- and DNA-bound X-ray structures the R-rich α -helix contacts the guide RNA (Anders et al., 2014; Jiang et al., 2015, 2016; Nishimasu et al., 2014), this finding has indicated a possible mechanism for RNA recruitment where the electrostatics could play a key role. With the purpose of understanding the role of electrostatics, the ensemble averaged electrostatic potential has been computed over the pathway of the conformational transition (Fig. 2B) (Baker et al., 2001). As a result, upon the exposure of the R-rich α -helix, a positively charged cavity is formed poised to accommodate the nucleic acids backbone (pre-RNA bound state), whereas in the RNA-bound state, the positively charged cavity closes to accommodate the RNA backbone. These computations, together with the available experimental data and the outcomes of GaMD simulations (Jiang et al., 2015; Palermo et al., 2017a), have indicated a mechanism for RNA binding and recruitment in which the relocation of the REC1–3 regions cause the formation of a positively charged cavity for the binding and recruitment of the negatively charged nucleic acid backbone.

2.2. Conformational activation toward DNA cleavage

Upon DNA binding, remarkable structural transitions involve the catalytic HNH domain, which undergoes a conformational change from an inactivated state, in which the catalysis is hampered, to an activated state prone for the cleavage of the TS (Nishimasu et al., 2014). While the inactivated state has been well characterized via X-ray crystallography, high-resolution data for the activated state are missing. Indeed, the most complete X-ray structure of Cas9 from *Streptococcus pyogenes* in complex with the nucleic acids captured a pre-activated state of the system, with the HNH catalytic H840 located ~ 19.4 Å from the cleavage site on the TS (Fig. 3A) (Jiang et al., 2016). Clearly, the activation of the system toward catalysis requires a further conformational change of the HNH domain. Through comparisons with homologous systems and Förster Resonance Energy Transfer (FRET) experiments, Sternberg and colleagues have indicated the conformational features of the activated HNH docked at the cleavage site (Fig. 3B) (Sternberg et al., 2015). However, the detailed conformational rearrangements leading up to the catalytically active Cas9 protein and its structure have not been clarified. Specifically, it is unclear how HNH would dock at the cleavage site and, importantly, how the protein framework would facilitate this process mechanistically.

To understand this conformational change and to identify the structure and dynamics of an activated HNH domain, enhanced sampling simulations using the GaMD method have been performed (Palermo et al., 2018, 2017a). Fig. 3C shows the time evolution of the distance between the catalytic H840 (Ca atom) and the scissile phosphate (at position -3 of the TS) during three simulation replicas of GaMD (warm colors), in comparison with a conventional MD run (gray). During GaMD, the catalytic H840 approaches the scissile phosphate, revealing the tendency for a fast conformational change of HNH in the late step of activation, in agreement with the available experimental data (Dagdas et al., 2017; Shibata et al., 2017; Sternberg et al., 2015). The observed conformational change has been confirmed by a recent continuous MD simulation carried out for ~ 16 μ s, during which the HNH domain approaches the cleavage site and stably docks at the TS (Palermo et al., 2018). Remarkably, the continuous simulation recovers the transition identified by running ~ 400 ns of GaMD, highlighting also the capability of accelerated simulations to capture biomolecular structural transitions occurring over longer time scales (Pierce et al., 2012). In the final configuration, the catalytic H840 approaches the cleavage site, in good agreement with Sternberg's model (Fig. 3B) (Sternberg et al., 2015) and with a recently published cryo-EM structure (EMD-8236, solved at 5.2 Å resolution) (Huai et al., 2017).

The simulations also show that the transition of the HNH domain depends on the structural remodeling of the REC lobe (Fig. 4A). Indeed, an opening of REC3 and a remarkable outward translation of REC2 are observed, facilitating the HNH domain to dock at the cleavage site on the TS. Moreover, REC1 approaches HNH, establishing ionic interactions that drive the conformational change and stabilize the catalytic HNH domain at the cleavage site (Palermo et al., 2018). Remarkably, the recent cryo-EM structure (EMD-8236) also displays an outward displacement of REC2-3, and the approach of REC1 to HNH (Huai et al., 2017). To further understand the role of the inter-dependent motions and the cooperative dynamics of the Cas9 domains during the activation process, dynamic correlation analyses

have been performed (Fig. 4B). In detail, a generalized correlation (GC) method has been employed, measuring the degree of correlation between residues based on their mutual information in consideration of Shannon entropy (Lange and Grubmuller, 2006). The GC matrix provides information on the per-residue correlations, which makes difficult, however, to localize the effective correlations, due to a high noise. To overcome this limitation, a coarse representation of the GC matrix (i.e., a per-domain GC score matrix) has been introduced (Palermo et al., 2017b). The latter accumulates the inter-domain correlations, providing a measure of the per-domain correlations. This analysis revealed that the REC lobe and the HNH domain are highly correlated in motions, indicating a key role of the REC lobe in affecting the conformational dynamics of HNH. The simultaneous conformational changes of the REC1–3 regions, and the highly coupled motions of the HNH domain, have suggested a functional role for the REC lobe in the activation process, which also agreed with the available experimental hypotheses. Based on single molecule FRET experiments, it has been suggested that an opening of REC3 would allow for HNH activation upon “sensing” the formation of a complete RNA:DNA hybrid (Chen et al., 2017). As well, REC2 would “regulate” the conformational transition of HNH, enabling the docking at the cleavage site on the TS. The dynamic interplay observed among the REC1–3 domains and HNH suggests that the three recognition regions ensure the catalytic competence of the CRISPR-Cas9 complex by “sensing” (REC3), “regulating” (REC2) and “locking” (REC1) the catalytic HNH domain. The per-domain correlation matrix also reveals a high dynamic coupling between the two catalytic domains. In this respect, the HNH and RuvC domains have been experimentally shown to cleave the two DNA strands via concerted mechanism that relies on interdependent conformational dynamics (Sternberg et al., 2015). A recent computational study combining quantum-classical MD simulations and accelerated MD has further demonstrated that the formation of a catalytic complex requires the activation of both HNH and RuvC catalytic functionalities, such enabling concerted cleavages of the two DNA strands (Palermo, 2019). Taken together, experiments and computations suggest a tight interdependent function.

3. Catalytic mechanism of RNA cleavage in self-splicing ribozymes

Catalysis is a fundamental property of RNA, which is at the basis of the multiple functions of ncRNA and enables the regulation of gene expression (Ward et al., 2014). Here, we focus on RNA splicing, which is a key biological step operated by ncRNA prior to translation. Splicing is the editing of the premature messenger RNA (pre-mRNA), which is cleared of its non-coding sections (introns), while the coding regions (exons) are ligated, forming the mature messenger RNA (mRNA) filament, subsequently coded in proteins. RNA splicing comprises two sequential trans-esterification reactions, facilitated by two Mg^{2+} ions, as in a two-metal aided catalysis (Steitz and Steitz, 1993). The extreme biological importance of RNA splicing lies in a tight regulation of gene expression, where RNA splicing aberrations are responsible for 13% of genetic diseases, cancer and neurodegeneration in humans (Papasaikas and Valcarcel, 2016). In bacteria and lower eukaryotes, RNA splicing is operated by Group II introns (G2Is), which are widespread mobile ribozymes capable of self-splicing and retrotransposition (i.e., reverse splicing) reactions (Toor et al., 2008, 2009; Marcia et al., 2013). G2Is are believed to share common evolutionary origins, with the spliceosome (Galej

et al., 2018; Lambowitz and Zimmerly, 2011; Peters and Toor, 2015; Pyle, 2016), which performs RNA splicing in humans and higher eukaryotes, sharing a common catalytic site with the U6 snRNA of the spliceosome (Fica et al., 2013; Galej et al., 2018; Hang et al., 2015). In this scenario, understanding the mechanistic basis of RNA splicing is of key importance, as it can support the development of gene modulation tools and novel therapeutic strategies. Multiple structures of the *Oceanobacillus iheyensis* (*O.i.*) G2I in the process of engaging in first-step hydrolysis have been determined through X-ray crystallography (Toor et al., 2008; Zhao and Pyle, 2017). The pre-catalytic state of this G2I was determined by replacing Mg^{2+} with Ca^{2+} ions (4FAQ, Fig. 5A), (Marcia and Pyle, 2012). Building on this structure, the first and rate-determining step of the splicing reaction of the *O.i.* G2I ribozyme, in which the 5'-exon cleavage is mediated by a nucleophilic water molecule, has been recently investigated (Casalino et al., 2016).

The authors employed a hybrid quantum mechanics/molecular mechanics (QM/MM) approach, which is suited for the investigation of chemical reactions in their realistic protein/nucleic acid environment, including the RNA scaffolds (Banas et al., 2009; Palermo et al., 2015a,b). In this approach, the reactive center (i.e., active site) is treated at a QM level of theory, while the rest of the system in explicit solution is treated at the classical MM level (Fig. 5). In order to determine the activation free energy associated to the first step of splicing, QM/MM MD has been combined to free energy methods. The authors employed a thermodynamic integration approach that enables to compute the free energy of a chemical step along a given reaction coordinate (RC, details are reported in the methods section) (Carter et al., 1989). QM/MM MD in combination with free energy methods have been shown to accurately determine the activation free energy of the chemical reaction and, within transition state theory, the associated rate constant with remarkable agreement with the experimental catalytic rates (Garcia-Viloca et al., 2004). As well, electronic structure theory provides a reliable description of the metal ions chemistry (Palermo et al., 2013), which are key for RNA catalysis and are often not adequately described at the classical MM level (Bergonzo et al., 2016; Casalino and Magistrato, 2016; Casalino et al., 2017; Cunha and Bussi, 2017).

In the QM/MM investigations of G2I ribozyme, the free energy profile has been computed as a function of the difference in distance between the breaking ($O3'-P_{RNA}$) and forming ($O_{WAT}-P_{RNA}$) bonds (which is used as a RC). This RC has been shown to be appropriate for the phosphodiester bond cleavage in RNA (De Vivo et al., 2008; Rosta et al., 2011). Indeed, by imposing the constraint only on the bond breaking/formation, this RC does not *a priori* specify the group accepting a proton from the attacking water nucleophile, while also not imposing a constraint on the leaving group, enabling these events to freely occur. As a result of the QM/MM investigations, the computed free energy profile reveals that the chemical step occurs through a dissociative mechanism, in which the dissociation of the leaving group ($O3'$) is reached at the Transition State ($TS^\ddagger$), while the deprotonation (DEP) of the attacking water is subsequent and occurs with proton shuttle to the bulk (Fig. 5B, C). This process shows a Helmholtz free energy (ΔF) of $\sim 18.8 \pm 1.5$ kcal/mol, in line with the experimental catalytic rate of 0.011 min^{-1} corresponding to a ΔG value of ~ 23 kcal/mol (Marcia and Pyle, 2012). This dissociative mechanism distinguishes from the two- Mg^{2+} -ion mechanism observed in enzymes by employing a similar QM/MM MD methodology (De

Vivo et al., 2008; Palermo et al., 2015a,b; Rosta et al., 2011; Sgrignani and Magistrato, 2015). In proteins, an associative mechanism S_N2 is observed (Fig. 6A). The two Mg^{2+} ions aided catalysis occurs via the activation of the nucleophilic water by MgA, which acts as a Lewis acid. Next, the activated nucleophile attacks the scissile phosphate and the associative $TS^\ddagger$ is stabilized by the joint action of the two metals. Conversely, in G2I, the reaction proceeds through a dissociative mechanism S_N1 (Fig. 6B). In this mechanism, MgA acts as a Lewis acid and activates the electrophile (the scissile phosphate group) by increasing its positive charge, rather than the nucleophilic water, contributing also to the stabilization of a dissociative $TS^\ddagger$. The water is converted into an OH^- group only after the formation of the $O_{WAT}-P_{RNA}$ bond has started. As such, the proton release to the bulk occurs with no barrier and without the need of a specific base. This mechanism suggests that the role of catalytic Mg^{2+} ions can remarkably differ in RNA, compared to what reported for most enzymes metabolizing DNA and/or RNA. Based on this comparison, it has been speculated that this cleavage mechanism might be specific for the ribozyme scaffold (Casalino et al., 2016). This hypothesis and the outcomes of the QM/MM investigations of G2I call now for the use of experimental methods, such as kinetic isotope effects experiments (Cassano et al., 2004; Gu et al., 2013), which could unambiguously define the nature of the $TS^\ddagger$ in the catalysis of G2I ribozymes. This would ultimately clarify whether these ribozymes, considered the evolutionary ancestors of the human spliceosome, have tailored a specific mechanism to ensure an specific scaffold-tailored autocatalytic function.

4. Biophysics of the spliceosome

In eukaryotes, pre-mRNA splicing is accomplished by the spliceosome, a multi-mega Dalton machinery composed of five snRNAs (U1, U2, U4, U5, and U6) and associated proteins forming the ribonucleoprotein nsRNPs subunits that, with other complexes like nineteen complex (NTC) and NTC-related (NTR), enzymes and cofactors, accomplish the spliceosomal function (Papasaikas and Valcarcel, 2016). The spliceosome exhibits exceptional conformational flexibility and compositional dynamics (Fica and Nagai, 2017; Hoskins and Moore, 2012; Matera and Wang, 2014). Indeed, it undergoes a continuous structural and compositional remodeling during the splicing cycle, with the scope of properly repositioning the various RNA filaments for cleavage. Over the last five years, thanks to the advancements in resolution of the modern cryo-EM detectors, we assisted at an explosion of structures of the spliceosome along the splicing cycle (Bertram et al., 2017; Cate, 2016; Fica and Nagai, 2017; Fica et al., 2013; Hang et al., 2015; Plaschka et al., 2017; Wan et al., 2016; Yan et al., 2015, 2016, 2017). These structures are paving the way for atomistic simulations aimed at understanding the complex biophysics underlying RNA splicing in humans. However, the complexity of the spliceosomal function and the size of the system, is challenging all-atom MD simulations in explicit solvent.

A recent computational study engaged in describing the functional conformational dynamics of this multi-mega Dalton complex (Casalino et al., 2018), focusing on the structure solved at near-atomic resolution from the yeast *Schizosaccharomyces pombe* (*S. Pombe*) (Hang et al., 2015; Yan et al., 2015), which captures the intron lariat spliceosome (ILS) complex at the late stage of the splicing cycle (Fig. 7A). Multi-micro-second length MD simulations have been performed, reaching an aggregate sampling of $\sim 3.25 \mu s$. While this simulation

time is certainly extensive considering the size of the simulated system (i.e., ~1,000,000 atoms), in order to extract functional insights from the MD trajectories, as well as to understand the subtle interplay between proteins and snRNA, correlation analyses have been carried out. Due to the high noise arising from the correlation matrices, the per-domain correlations have been computed, following the protocol used for understanding the inter-dependent domain motions of the CRISPR-Cas9 complex (Fig. 4B) (Palermo et al., 2017b). As a result, a tight interdependency in the domain motions is observed (Fig. 7B).

Relevant dynamic couplings are observed for the Spp42 protein, which is highly correlated with all the snRNAs (U2, U5, U6) and with the intron lariat (IL), while also displaying high dynamic correlations with most of the other protein components. This result has indicated a key role of the Spp42 protein in modulating the structural plasticity of the ILS. Indeed, by establishing a tight dynamic coupling with the majority of the protein and snRNA components, the Spp42 protein governs their conformational dynamics. This has indicated a protein-directed function for the spliceosome remodeling at a late step of splicing (Yan et al., 2015). In order to better extract the functional role of the proteins and snRNAs from the dynamics of the ILS, principal component analysis and electrostatic calculations have also been performed, in analogy to the investigations on CRISPR-Cas9 (Palermo et al., 2017a,b). The “essential dynamics” extracted from principal component analysis (Amadei et al., 1993) revealed the displacement of the IL/U2 helix, whose unrolling is promoted cooperatively by Cwf19 and Spp42 (Fig. 7C). This finding well characterizes the late step of splicing, when the IL/U2 branch helix, formed before the catalysis to bulge out the branch point adenosine, has to be displaced before the IL linearization (or debranching) and degradation (Garrey et al., 2014; Khalid et al., 2005; Martin et al., 2002; Yoshimoto et al., 2009). As well, these findings propose that the Cwf19 is effectively involved in the displacement of the IL, suggesting a functional role in agreement with previous hypotheses (Garrey et al., 2014). Electrostatic calculations have further revealed that Cwf19 and Spp42 act as “polar tweezers” to trap the negatively charged nucleotides of U2 and IL (Fig. 7D). Importantly, a very recent cryo-EM study of the human post-catalytic spliceosome (P complex) and IL spliceosome has supported the key role of Cwf19 in the IL/U2 displacement and unwinding as revealed by the simulations (Zhang et al., 2019). Overall, based on these computational results and their matching with the available experimental data, a fascinating biophysical process has been suggested, where Spp42 exerts a leading role by tuning and governing the motions of the distinct spliceosomal components, like an “orchestra conductor” of the “gene maturation symphony”. In parallel, the protein Cwf19 cooperates with Spp42 in the displacement of the IL, overall suggesting a protein-directed remodeling in the late step of splicing.

5. Computational methods

Here, we discuss the computational methodologies employed in the studies reported above, to provide a description for a broad audience, including structural biologists and non-experts in theoretical and computational biophysics.

5.1. Molecular Dynamics (MD) simulations

Molecular Dynamics (MD) simulations build on the basic principles of Newtonian physics to numerically integrate the equations of motion for a given set of atomic coordinates, generating a dynamical trajectory of the system. From the first MD simulation in the 70s' (McCammon et al., 1977), ever-increasing applications and development of computational algorithms and architectures have elected this method for capturing the dynamic and thermodynamic properties of biomolecular machineries with atomic level resolution. In classical MD simulations, the atomic and molecular interactions are described through classical potential energy functions – known as force fields – empirically parameterized to reproduce experimental or *ab-initio* data. A typical atomistic potential MD energy function is defined as the sum of different contributions, and is usually composed of bonded terms, describing bond, angle bending, and torsional degrees of freedom (first three terms of Eq. (1)) and non-bonded terms, accounting for van der Waals and electrostatic forces (last term of Eq. (1)):

$$V = \sum_{bonds} K_R (R - R_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \quad (1)$$

$$+ \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j e^2}{4\pi \epsilon_0 R_{ij}} \right]$$

To date, the most commonly used force fields in the simulations of biological systems are OPLS (Jorgensen et al., 1996), AMBER (Duan et al., 2003), GROMOS (Christen et al., 2005) and CHARMM (MacKerell et al., 1998). These force fields have been proven to excellently perform in simulations of proteins and peptides. However, in the case of nucleic acids (and therefore RNA), the description of base-pairing, stacking and base/sugar interactions can be a challenging task. The AMBER force field is currently widely employed for RNA. Early development has been done by Cornell and colleagues (AMBER ff94) (Cornell et al., 1995), followed by the correction of several dihedral parameters (ff98 (Cheatham et al., 1999), ff99 (Wang et al., 2000)). Subsequent improvements include the bsc0 corrections for nucleic acids (Perez et al., 2007), which correct the unbalanced α/γ transitions, and to the $\chi OL3$ corrections for RNA (Banas et al., 2010; Zgarbova et al., 2011), which balances the anti and the high-anti conformation related to the χ angle. A further development focused on the adjustment of the van der Waals parameters, leading to a better description of stacking interactions and base-pairing strength (Chen and García, 2013). Recent advances combined *ab-initio* and empirical methods to improve the electrostatics, van der Waals and torsional parameters, yielding agreement with experimental data for both extended and structured RNAs (Tan et al., 2018). For what concerns MD simulations of protein/RNA complexes (such as CRISPR-Cas9 and the spliceosome), recent studies have indicated the reliability of the ff12SB (Maier et al., 2015) force field for proteins and of the ff99bsc0 $\chi OL3$ for RNA (Krepl et al., 2015; Sponer et al., 2016). Building on this work, the studies discussed in this review have been based on this force field.

5.2. Extending the time scale limit of MD simulations

While powerful, MD simulations have traditionally been limited to very short time-scales, making difficult the description of folding and energetic dynamics of many large proteins, including the assembly of the CRISPR-Cas9 complex. To overcome this limitation, accelerated MD (aMD) methodologies enhance the exploration of the configurational space and extend the time scales accessible to MD simulations (Miao and McCammon, 2016b). aMD is a method originally developed by Hamelberg et al. (2004), which adds a boost potential to the potential energy surface (PES), effectively decreasing the energy barriers and thus accelerating transitions between low-energy states. Recent advances led to the development of a novel method – i.e., a Gaussian aMD (GaMD) (Miao et al., 2015) – which allows robust sampling and simultaneous free energy reweighting, using harmonic functions to construct a boost potential that is adaptively added to the PES. This extends the use of aMD to the simulations of large biomolecular complexes (such as protein-RNA complexes) and big-data analysis, in which a large statistical noise can make prohibitive the derivation of the associated energetics through the reweighting procedure (Miao et al., 2014). Given the purposes of this review, a concise description of the GaMD methodology is given as follows. For a system of N atoms at positions $\vec{r} = \{\vec{r}_1, \dots, \vec{r}_N\}$, when the system potential $V(\vec{r})$ is lower than a threshold energy E , the energy surface is modified by adding a boost potential as:

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}), \quad V(\vec{r}) < E, \quad (2)$$

$$\Delta V(\vec{r}) = \frac{1}{2}k(E - V(\vec{r}))^2, \quad (3)$$

where k is the harmonic force constant (Fig. 8). The two adjustable parameters E and k are determined from cMD runs to collect the maximum ($V_{\max}$), minimum ($V_{\min}$), average (V_{avg}) and standard deviation (σ_V) of the system potential energies. When E is set to the lower bound, k_0 can be calculated as:

$$k_0 = \min(1.0, k'_0) = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \cdot \frac{V_{\max} - V_{\min}}{V_{\max} - V_{\text{avg}}}\right), \quad (4)$$

where σ_0 is a user-specified upper limit (e.g., 10 kBT) for accurate energetic reweighting. Alternatively, when the threshold energy E is set to its upper bound $E = V_{\min} + 1/k$, k_0 is:

$$k_0 = k''_0 \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right) \cdot \frac{V_{\max} - V_{\min}}{V_{\text{avg}} - V_{\min}}, \quad (5)$$

with k_0'' calculated between 0 and 1. Since the boost potential follows a near-Gaussian distribution, this allows for robust sampling and accurate reconstruction of the free energy profiles. GaMD has been shown to capture long time scale conformational events in a variety of biological systems, providing routine access to the millisecond time scale dynamics (Pierce et al., 2012) and accurate estimation of the associated energetics (Chuang et al., 2018; Miao and McCammon, 2016a, 2018; Miao et al., 2014; Sibener et al., 2018; Wang and Chan, 2017). As well, GaMD simulations preserve the characteristic features of DNA and RNA, (Palermo et al., 2017a; Ricci et al., 2019) holding the potential to be applied to large macromolecular systems involved in nucleic acid processing, which are being obtained via cryo-EM. Finally, it is important to note that the recent implementation of the GaMD method in the Amber 18 (Case et al., 2018) code enables to efficiently perform enhanced sampling simulations over graphics processing units (GPUs) cards, obtaining MD trajectories at remarkably *low cost*, with the *high reward* of accessing long time scale events.

5.3. Exploring chemical reactions via QM/MM MD simulations

The exquisite capability of describing bond breaking and formation during chemical reactions requires the application of QM methods. Due to the high computational cost of considering an entire protein in complex with nucleic acids (as well as the surrounding solvent) at a QM level of theory, a mixed QM/MM method can be employed. In this approach, the catalytically active atoms within the active site are treated at a higher level of accuracy (QM level), while the rest of the system and the solvent is treated at the level of a classical force field (Warshel and Levitt, 1976). In the general form of a hybrid QM/MM scheme, the total energy of the system (i.e., given by the Hamiltonian $\mathcal{H}$) contains the Hamiltonians for the quantum ($\mathcal{H}_{QM}$) and classical ($\mathcal{H}_{MM}$) systems and the interaction between the QM and MM regions ($\mathcal{H}_{QM/MM}$):

$$\mathcal{H} = \mathcal{H}_{QM} + \mathcal{H}_{MM} + \mathcal{H}_{QM/MM} \quad (6)$$

where the QM Hamiltonian ($\mathcal{H}_{QM}$) can be based on different QM electronic structure methods, from semiempirical to *ab initio* Hartree-Fock or Density Functional Theory (DFT). The QM/MM scheme can be combined with *ab-initio* MD, enabling to integrate the equation of motion for the electronic degrees of freedom over time by invoking the Born-Oppenheimer approximation (Marx and Hutter, 2009) or by introducing a Newtonian fictitious dynamics for the electronic degrees of freedom (as in the method of Car and Parrinello) (Car and Parrinello, 1985). Traditionally applied on solid-state physics, *ab-initio* MD has been combined with the QM/MM method by Roethlisberger and co-workers, paving the way for the study of the dynamics of chemical reactions in biological systems (Laio et al., 2002a,b). This Hamiltonian QM/MM approach fully accounts for the electrostatic effects of the classical environment by applying the following scheme. To treat the electrostatic interaction of the classical environment with the electronic density, the MM atoms are divided in three layers: the “short range” NN atoms, the “electrostatic coupling” EC atoms and the “long range” ME atoms (Fig. 9).

The “short range” interaction of the QM atoms from the classical force field on the NN atoms is computed through a modified Coulomb potential that goes to $1/r$ for large r and to a finite value for small r . To deal with the interaction between the QM atoms and the second EC shell, dynamically generated electrostatic potential derived charges are computed for the QM atoms. The charges are fitted to reproduce the electrostatic potential due to the QM density on the position of the NN atoms. Then, the electrostatic interaction between QM and EC is computed as the Coulomb interaction between the QM charges and the force field charges of the EC atoms. Finally, the “long range” interaction between the QM and ME atoms is computed through a multipole expansion. The introduction of this multiple layers model enabled to fully take into account the effect of the electrostatic field of the classical environment on the QM electronic charge density. The approach is based on DFT for the description of the QM part, while one can interface any desired force field for the classical part. While being very accurate, this *ab-initio* QM/MM approach is known to require massive computational resources. To overcome this limitation, a novel multi-scale framework is becoming available, which promises to exploit the power of emerging exascale computing resources (Olsen et al., 2019). This recent breakthrough in the QM/MM multiscale methods – called MiMiC – will hopefully reduce the computational cost for more accurate QM/MM simulations.

In *ab-initio* MD, the time-step required for integrating the equations of motion is on the order of a few picoseconds. This enables to describe biochemical phenomena in the picosecond time scale. However, chemical reactions can have rate constants on the time scale of seconds and therefore represent a “rare event” on the accessible time scales of QM/MM simulations. To overcome this problem, methods enabling the enhancement of the sampling of “rare events” are valuable tools. Among others, a thermodynamic integration approach within the “blue moon” ensemble can be applied to capture “rare events” observable “once in a blue moon”. In this approach, to explore a reaction step of the catalytic mechanism, a series of constraints are applied to enable exploration of the configurations along them. Then, the average converged constraint forces are computed and integrated along the given reaction coordinates (RC, ζ), deriving the associated free energy profile. Briefly, in the case a simple distance (or difference in distances) is used as a RC, from the average constraint force F at each point along the RC, the free energy difference between two points ζ_2 and ζ_1 along the RC can be calculated as:

$$F(\zeta_2) - F(\zeta_1) = \int_{\zeta_1}^{\zeta_2} d\zeta' \frac{dF}{d\zeta'} \quad (7)$$

where $F(\zeta_i)$ is the free energy at point ζ_i . Overall, *ab-initio* QM/MM MD in combination with free energy methods has been shown to be solid in the description of catalytic mechanisms in several biological and chemical systems (Brunk et al., 2011), including metal complexes (Adhireksan et al., 2017; Ma et al., 2016), DNA/RNA processing enzymes (De Vivo et al., 2008; Palermo et al., 2015a,b) and the metal-aided catalysis of G2I ribozyme (Casalino et al., 2016). It is important to note that alternative free energy methods – e.g., transition path sampling (Bolhuis et al., 2002) and metadynamics (Laio and Parrinello,

2002) – can also be applied in the framework of *ab-initio* MD simulations to compute the activation free energy of chemical reactions. Finally, we would like to recall that, given the complexity of *ab-initio* MD, a comprehensive understanding has to refer to the original papers and books (Marx and Hutter, 2009).

5.4. Tailoring correlation analysis for multi mega Dalton biomolecular complexes

The analysis of the dynamic correlations is a useful and established tool in computational biophysics. Traditionally, Pearson coefficients (Pc) offer a measure of the collinear correlation between two atoms i and j , measuring whether they tend to move in lockstep (positive Pc) or show opposed motion (negative Pc), with their magnitude indicating the strength of the correlation. Pc for each couple of residues atoms i and j can be computed from:

$$Pc_{ij} = \frac{\langle \Delta r_i(t) \cdot \Delta r_j(t) \rangle}{\left(\langle \Delta r_i(t)^2 \rangle \langle \Delta r_j(t)^2 \rangle \right)^{1/2}} \quad (8)$$

where r_i and r_j are the fluctuation vectors of the atoms i and j , respectively. The angle bracket represents an average over the sampled period. Pc do not capture correlations that are not collinear or occur out of phase with each other. An other popular correlation scheme – i.e., the Generalized Correlation (GC) method – measures the degree of correlation between residues based on their mutual information (Lange and Grubmuller, 2006). This approach captures the each type of dependence in the motion of two atoms, regardless of the direction of motion, therefore capturing non-linear correlations. However, it does not provide information on the direction of the correlations, being a normalized and unsigned measure. These correlation schemes complement each other and can be used to build correlation matrices, describing the correlation for each couple of residues in the macro-molecular system. However, correlation matrices provide information on the per-residue correlations, which makes difficult to localize the effective correlations, due to a high noise. This problem is particularly severe for multi mega Dalton biomolecular systems, hampering the identification of the effective correlations. To overcome this issue, Ricci and co-workers (Ricci et al., 2016) have originally introduced the concept of Correlation Score (Cs), which accumulates the per-residue correlations as follows:

$$Cs_i = \sum_{j \neq i}^N CC_{ij} \quad (9)$$

The Cs_i is a measure of both the number and the intensity of the correlation coefficients (CC_{ij}) displayed by each residue i with residues j . Building on this concept, a coarse matrix describing the per-domain correlations can be built. Indeed, the Cs_i can be computed for each residue i belonging to a specific protein domain, with the residues j belonging to another protein domain of interest. Then, by accumulating the Cs_i over all residues j of each specific domain, as well as upon the normalization, a coarse per-domain matrix (Fig. 4B) or,

alternatively, a histogram diagram (Fig. 7B) can be obtained, providing a measure of the strength of the overall correlation that each domain establishes with the others. This coarse representation has been introduced in the study of CRISPR-Cas9 (Palermo et al., 2016, 2017b) and subsequently applied on the spliceosome (Casalino et al., 2018) and on the estrogen receptor α involved in breast cancer (Pavlin et al., 2018). This representation is indeed particularly useful in the case of big size biological systems, where the motions of the high number of residues and nucleobases give rise to a high noise. It enables to identify the most important correlated motions among protein/nucleic acid domains taking place in the simulations, helping in identifying how specific regions mechanistically intervene in the overall correlation network.

6. Conclusions

Here, we review recent applications of conventional and enhanced MD simulations methods to address the mechanistic function of large biomolecular systems that are tightly involved in the ncRNA function and that are of key importance in life sciences. Our compendium focuses on three biomolecular systems, namely the CRISPR-Cas9 genome editing machinery, group II intron ribozyme and the ribonucleoprotein complex of the spliceosome, which edit and process ncRNA. We report the use of MD simulations in characterizing the biophysics of the RNA-guided CRISPR-Cas9 system, providing a mechanistic framework of its function. We show how the application of a novel accelerated MD simulations method – i.e., a Gaussian accelerated MD (GaMD) (Miao et al., 2015) – has been key in disclosing the conformational transitions from the apo form up to RNA binding (Palermo et al., 2016, 2017a). Building on these simulations, a mechanism for RNA recruitment occurring with formation of a positively charged cavity hosting nucleic acids during binding has been suggested. This work also provided insights into the conformational changes required for the activation of the HNH domain toward cleavage (Palermo et al., 2018, 2017b). These findings have been a reference for experimental studies aimed at fully characterizing the dynamic features of the CRISPR-Cas9 system (Dagdas et al., 2017), and provide a fundamental framework for the optimization of ameliorated Cas9 proteins and tailored single guide RNAs for improved genome editing applications. Our showcase of applications further reports the use of molecular simulations to access the catalytic mechanism of RNA splicing as catalyzed by group II intron ribozyme, one of the largest RNA structures crystallized so far. Mixed quantum-classical MD simulations have revealed an RNA-specific two-metal aided-catalysis, which differs from what observed in proteins, and calls now for experimental validation (Casalino et al., 2016). Finally, we discuss the application of MD for unraveling the mechanistic function of the spliceosome, a multi-mega Dalton ribonucleoprotein complex, which performs RNA splicing in humans (Casalino et al., 2018). Given the size and the complexity of the system, essential dynamics calculations have been flanked by correlation analyses to extract functional insights from MD and to understand the subtle interplay between proteins and snRNA. As a result, the late step of splicing appears to be governed by a protein-directed function, where the Spp42 and Cwf19 proteins cooperatively act to tune the motions of the spliceosomal components and to enable the displacement of the intron lariat before its degradation. While this computational study represents the first effort in understanding the mechanistic biophysics of the spliceosomal

function, the complex structural and compositional remodeling along the splicing cycle remains not completely understood. In this respect, the broad time scales associated to the spliceosome structural remodeling make of this multi mega Dalton complex an ideal case for the application of accelerated MD simulations. Overall, the studies reported in this review article highlight the potential of MD simulations to unravel the mechanistic function of ncRNA regulatory systems, inspiring future investigations of several other biomolecular systems whose function relies on ncRNA. Finally, we would like to remark that the current extraordinary advances of single particle cryo-EM (Cheng et al., 2017; Nogales, 2016) are poised to revolutionize the field, giving access to structures of the spliceosome and of other ncRNA regulatory systems with increasingly realistic size and at high resolution. This revolution in structural biology calls for an interpretation from computational biophysics, with future computational studies focusing on the development and application of novel methodologies able to interpret the long time scale conformational dynamics of multi-mega Dalton biological complexes.

Acknowledgments

The authors thank all scientists that have authored the papers presented in this review. JAM acknowledges support from the NIH (grant: R01-GM031749–36, Theory of Biomolecular Diffusion).

Funding

This review article was supported by start-up funds from the University of California Riverside to GP. Computer time was awarded by The Extreme Science and Engineering Discovery Environment XSEDE (grant: TG-MCB160059).

References

- Abrams C, Bussi G, 2014 Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy* 16, 163–199.
- Adhikarsan Z, Palermo G, Riedel T, Ma Z, Muhammad R, Dyson PJ, Rothlisberger U, Davey CA, 2017 Allosteric cross-talk in chromatin can mediate drug-drug synergy. *Nat. Commun* 8, 14860. [PubMed: 28358030]
- Amadei A, Linssen ABM, Berendsen HJC, 1993 Essential dynamics of proteins. *Prot. Struct. Funct. Genet* 17, 412–425.
- Anastasiadou E, Jacob LS, Slack FJ, 2018 Non-coding RNA networks in cancer. *Nat. Rev. Cancer* 18, 5–18. [PubMed: 29170536]
- Anders C, Niewoehner O, Duerst A, Jinek M, 2014 Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573. [PubMed: 25079318]
- Baker NA, Sept D, Joseph S, Holst MJ, Mccammon JA, 2001 Electrostatics of nanosystems: application to microtubules and the ribosome. *PNAS* 98, 10037–10041. [PubMed: 11517324]
- Banas P, Hollas D, Zgarbova M, Jurecka P, Orozco M, Cheatham TE, Sponer J, Otyepka M, 2010 Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput* 6, 3836–3849.
- Banas P, Jurecka P, Walter NG, Sponer J, Otyepka M, 2009 Theoretical studies of RNA catalysis: hybrid QM/MM methods and their comparison with MD and QM. *Methods* 49, 202–216. [PubMed: 19398008]
- Bergonzo C, Hall KB, Cheatham TE 3rd, 2016 Divalent ion dependent conformational changes in an RNA stem-loop observed by molecular dynamics. *J. Chem. Theory Comput* 12, 3382–3389. [PubMed: 27294370]

- Bergonzo C, Hall KB, Cheatham TE 3rd, 2015 Stem-loop V of Varkud satellite RNA exhibits characteristics of the Mg^{2+} bound structure in the presence of monovalent ions. *J. Phys. Chem. B* 119, 12355–12364. [PubMed: 26328924]
- Bergonzo C, Henriksen NM, Roe DR, Swails JM, Roitberg AE, Cheatham TE 3rd, 2014 Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J. Chem. Theory Comput* 101, 492–499.
- Bertram K, Agafonov DE, Liu WT, Dybkov O, Will CL, Hartmuth K, Urlaub H, Kastner B, Stark H, Luhrmann R, 2017 Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* 542, 318–323. [PubMed: 28076346]
- Bolhuis PG, Chandler D, Dellago C, Geissler PL, 2002 Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem* 53, 291–318. [PubMed: 11972010]
- Bottaro S, Gil-Ley A, Bussi G, 2016 RNA folding pathways in stop motion. *Nucleic Acids Res* 44, 5883–5891. [PubMed: 27091499]
- Brunk E, Ashari N, Athri P, Campomanes P, Franco de Carvalho F, Curchod BFE, Diamantis P, Doemer M, Garrec J, Laktionov A, Micciarelli M, Neri M, Palermo G, Penfold TJ, Vanni S, Tavernelli I, Rothlisberger U, 2011 Pushing the frontiers of first-principles based computer simulations of chemical and biological systems. *Chimia Aarau* 65, 667–671. [PubMed: 22026176]
- Car R, Parrinello M, 1985 Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett* 55, 2471–2474. [PubMed: 10032153]
- Carter EA, Ciccotti G, Hynes JT, Kapral R, 1989 Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett* 156, 472–477.
- Casalino L, Magistrato A, 2016 Structural, dynamical and catalytic interplay between Mg^{2+} ions and RNA. Vices and virtues of atomistic simulations. *Inorg. Chim. Acta* 452, 73–81.
- Casalino L, Palermo G, Abdurakhmonova N, Rothlisberger U, 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. [PubMed: 28001405]
- Casalino L, Palermo G, Rothlisberger U, 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. [PubMed: 27309711]
- Casalino L, 2017 Pre-mRNA splicing: an evolutionary computational journey from ribozymes to spliceosome. SISSA <http://hdl.handle.net/20.500.11767/59221>.
- Casalino L, Palermo G, Spinello A, Rothlisberger U, Magistrato A, 2018 All-atom simulations disentangle the functional dynamics underlying gene maturation in the intron lariat spliceosome. *PNAS* 115, 6584–6589. [PubMed: 29891649]
- Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE 3rd, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Izadi S, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein DJ, Merz KM, Miao Y, Monard G, Nguyen C, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Smith J, Salomon-Ferrer R, Swails J, Walker RC, Wang J, Wei H, Wolf RM, Wu X, Xiao L, York DM, Kollman PA, 2018 AMBER 2018 University of California, San Francisco.
- Cassano AG, Anderson VE, Harris ME, 2004 Understanding the transition states of phosphodiester bond cleavage: insights from heavy atom isotope effects. *Biopolymers* 731, 110–129.
- Cate JHD, 2016 STRUCTURE a big bang in spliceosome structural biology. *Science* 3516280, 1390–1392.
- Cech TR, Steitz JA, 2014 The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 1571, 77–94.
- Cheatham TE 3rd, Cieplak P, Kollman P, 1999 A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn* 164, 845–862.
- Cheatham TE 3rd, Young MA, 2000 Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers* 564, 232–256.
- Chen AA, García AE, 2013 High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *PNAS* 42, 16820–16825.

- Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA, 2017 Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550, 407–410. [PubMed: 28931002]
- Cheng Y, Glaeser RM, Nogales E, 2017 How cryo-EM became so hot. *Cell* 171, 1229–1231. [PubMed: 29195065]
- Christen M, Hunenberger PH, Bakowies D, Baron R, Burgi R, Geerke DP, Heinz TN, Kastenholz MA, Krautler V, Oostenbrink C, Peter C, Trzesniak D, Van Gunsteren WF, 2005 The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem* 2616, 1719–1751.
- Chuang CH, Chiou SJ, Cheng TL, Wang YT, 2018 A molecular dynamics simulation study decodes the Zika virus NS5 methyltransferase bound to SAH and RNA analogue. *Sci. Rep* 81, 6336.
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA, 1995 A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc* 117, 5179–5197.
- Cruz JA, Westhof E, 2009 The dynamic landscapes of RNA architecture. *Cell* 1364, 604–609.
- Cunha RA, Bussi G, 2017 Unraveling Mg^{2+} -RNA binding with atomistic molecular dynamics. *RNA* 235, 628–638.
- Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A, 2017 A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv* 3, eaao0027. [PubMed: 28808686]
- De Vivo M, Dal Peraro M, Klein ML, 2008 Phosphodiester cleavage in ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. *J. Am. Chem. Soc* 13033, 10955–10962.
- Doudna JA, Charpentier E, 2014 Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 3466213, 1258096.
- Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman PA, 2003 A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem* 2416, 1999–2012.
- Fica SM, Nagai K, 2017 Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol* 24, 791–799. [PubMed: 28981077]
- Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA, 2013 RNA catalyses nuclear pre-mRNA splicing. *Nature* 5037475, 229–234.
- Galej WP, Toor N, Newman AJ, Nagai K, 2018 Molecular mechanism and evolution of nuclear pre-mRNA and group II intron splicing: insights from cryo-electron microscopy structures. *Chem. Rev* 1188, 4156–4176.
- Garcia-Viloca M, Gao J, Karplus M, Truhlar DG, 2004 How enzymes work: analysis by modern rate theory and computer simulations. *Science* 3035655, 186–195.
- Garrey SM, Katolik A, Prekeris M, Li X, York K, Bernards S, Fields S, Zhao R, Damha MJ, Hesselberth JR, 2014 A homolog of lariat-debranching enzyme modulates turnover of branched RNA. *RNA* 208, 1337–1348.
- Gorski SA, Vogel J, Doudna JA, 2017 RNA-based recognition and targeting: sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol* 184, 215–228.
- Gu H, Zhang S, Wong KY, Radak BK, Dissanayake T, Kellerman DL, Dai Q, Miyagi M, Anderson VE, York DM, Piccirilli JA, Harris ME, 2013 Experimental and computational analysis of the transition state for ribonuclease A-catalyzed RNA 2'-O-transphosphorylation. *PNAS* 11032, 13002–13007.
- Hayatshahi HS, Roe DR, Galindo-Murillo R, Hall KB, Cheatham TE 3rd, 2017 Computational assessment of potassium and magnesium ion binding to a buried pocket in GTPase-associating center RNA. *J. Phys. Chem. B* 121, 451–462. [PubMed: 27983843]
- Hamelberg D, Mongan J, Mccammon JA, 2004 Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys* 12024, 11919–11929.
- Hang J, Wan R, Yan C, Shi Y, 2015 Structural basis of pre-mRNA splicing. *Science* 3496253, 1191–1198.
- Harvey SC, Prabhakaran R, Mao B, Mccammon JA, 1984 Phenylalanine transfer RNA: molecular dynamics simulation. *Science* 223, 1189–1191. [PubMed: 6560785]

- Hashem Y, Auffinger P, 2009 A short guide for molecular dynamics simulations of RNA systems. *Methods* 473, 187–197.
- Hoskins AA, Moore MJ, 2012 The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem. Sci* 375, 179–188.
- Hsu PD, Lander ES, Zhang F, 2014 Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 1576, 1262–1278.
- Huai C, Li G, Yao RJ, Zhang Y, Cao M, Kong LL, Jia CQ, Yuan H, Chen HY, Lu DR, Huang Q, 2017 Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat. Commun* 8, 1375. [PubMed: 29123204]
- Jiang F, Doudna JA, 2017 CRISPR-Cas9 structures and mechanisms. *Annu. Rev. Biophys* 46, 505–529. [PubMed: 28375731]
- Jiang F, Zhou K, Ma L, Gressel S, Doudna JA, 2015 STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* 3486242, 1477–1481.
- Jiang FG, Taylor DW, Chen JS, Kornfeld JE, Zhou KH, Thompson AJ, Nogales E, Doudna JA, 2016 Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 3516275, 867–871.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E, 2012 A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 3376096, 816–821.
- Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA, 2014 Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 3436176, 1247997.
- Jorgensen WL, Maxwell DS, Tirado-Rives J, 1996 Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc* 118, 11225–11236.
- Karplus M, Mccammon JA, 2002 Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol* 99, 646–652.
- Khalid MF, Damha MJ, Shuman S, Schwer B, 2005 Structure-function analysis of yeast RNA debranching enzyme Dbr1, a manganese-dependent phosphodiesterase. *Nucleic Acids Res* 33, 6349–6360. [PubMed: 16275784]
- Krepl M, Havrila M, Stadlbauer P, Banas P, Otyepka M, Pasulka J, Stefl R, Sponer J, 2015 Can we execute stable microsecond-scale atomistic simulations of protein-RNA complexes? *J. Chem. Theory Comput* 113, 1220–1243.
- Laio A, Parrinello M, 2002 Escaping free-energy minima. *PNAS* 9920, 12562–12566.
- Laio A, Vandevonle J, Rothlisberger U, 2002a D-RESP: dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations. *J. Phys. Chem. B* 106, 7300–7307.
- Laio A, Vandevonle J, Rothlisberger U, 2002b A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys* 1166941, 6941–6947.
- Lambowitz AM, Zimmerly S, 2011 Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol* 3, a003616. [PubMed: 20463000]
- Lange OF, Grubmuller H, 2006 Generalized correlation for biomolecular dynamics. *Prot. Struct. Funct. Bioinf* 624, 1053–1061.
- Ma Z, Palermo G, Adhireksan Z, Murray BS, Von Erlach T, Dyson PJ, Rothlisberger U, Davey CA, 2016 An organometallic compound displays a unique one-stranded intercalation mode that is DNA topology-dependent. *Angew. Chem. Int. Ed* 128, 7441–7444.
- Mackerell AD, Bashford DJ, Bellott M, Dunbrack RL, Evanseck DJJ, Field JM, Fischer S, Gao J, Guo H, Ha S, Joseph-Mccarthy D, Kuchnir L, Kuczera K, Lau KTF, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher EW, Roux BI, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M, 1998 All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Phys. Chem. B* 10218, 3586–3616.
- Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C, 2015 ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput* 118, 3696–3713.

- Marcia M, Pyle AM, 2012 Visualizing group II intron catalysis through the stages of splicing. *Cell* 151, 497–507. [PubMed: 23101623]
- Marcia M, Somarowthu S, Pyle AM, 2013 Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob. DNA* 4, 14. [PubMed: 23634971]
- Martin A, Schneider S, Schwer B, 2002 Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J. Biol. Chem* 277, 17743–17750. [PubMed: 11886864]
- Marx D, Hutter J, 2009 *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* Cambridge University Press, New York.
- Matera AG, Wang Z, 2014 A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol* 152, 108–121.
- Matsui M, Corey DR, 2017 Non-coding RNAs as drug targets. *Nat. Rev. Drug. Discov* 16, 67–179.
- Mccammon JA, Gelin BR, Karplus M, 1977 Dynamics of folded proteins. *Nature* 2675612, 585–590.
- Miao Y, Feher VA, Mccammon JA, 2015 Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory Comput* 118, 3584–3595.
- Miao Y, Mccammon JA, 2016a Graded activation and free energy landscapes of a muscarinic G protein-coupled receptor. *PNAS* 11343, 12162–12167.
- Miao Y, Mccammon JA, 2016b Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol. Simul* 4213, 1046–1055.
- Miao Y, Mccammon JA, 2018 Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *PNAS* 11512, 3036–3041.
- Miao Y, Sinko W, Pierce L, Bucher D, Walker RC, Mccammon JA, 2014 Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *J. Chem. Theory Comput* 107, 2677–2689.
- Nishimasu H, Nureki O, 2017 Structures and mechanisms of CRISPR RNA-guided effector nucleases. *Curr. Opin. Struct. Biol* 43, 68–78. [PubMed: 27912110]
- Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O, 2014 Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 1565, 935–949.
- Nogales E, 2016 The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* 131, 24–27.
- O’Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA, 2014 Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* 516, 263–266. [PubMed: 25274302]
- Olsen JMH, Bolnykh V, Meloni S, Ippoliti E, Bircher MP, Carloni P, Rothlisberger U, 2019 MiMiC: a novel framework for multiscale modeling in computational chemistry. *ChemRxiv* 10.26434/chemrxiv.7635986.v1.
- Palermo G, 2019 Structure and dynamics of the CRISPR-Cas9 catalytic complex. *J. Chem. Inf. Model* 10.1021/acs.jcim.8b00988 (in press).
- Palermo G, Cavalli A, Klein ML, Alfonso-Prieto M, Dal Peraro M, De Vivo M, 2015a Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res* 482, 220–228.
- Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, Doudna JA, Mccammon JA, 2018 Key role of the REC lobe during CRISPR-Cas9 activation by “sensing”, “regulating” and “locking” the catalytic HNH domain. *Q. Rev. Biophys* 51, e9.
- Palermo G, Magistrato A, Riedel T, Von Erlach T, Davey CA, Dyson PJ, Rothlisberger U, 2015b Fighting cancer with transition metal complexes: from naked DNA to protein and chromatin targeting strategies. *ChemMedChem* 11, 1199. [PubMed: 26634638]
- Palermo G, Miao Y, Walker RC, Jinek M, Mccammon JA, 2016 Striking plasticity of CRISPR-Cas9 and key role of non-target DNA, as revealed by molecular simulations. *ACS Cent. Sci* 2, 756–763. [PubMed: 27800559]
- Palermo G, Miao Y, Walker RC, Jinek M, Mccammon JA, 2017a CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *PNAS* 11428, 7260–7265.

- Palermo G, Ricci CG, Fernando A, Rajshekhar B, Jinek M, Rivalta I, Batista VS, Mccammon JA, 2017b PAM-induced allostery activates CRISPR-Cas9. *J. Am. Chem. Soc* 139, 16028–16031. [PubMed: 28764328]
- Palermo G, Stenta M, Cavalli A, Dal Peraro M, De Vivo M, 2013 Molecular simulations highlight the role of metals in catalysis and inhibition of type II topoisomerase. *J. Chem. Theory Comput* 92, 857–862.
- Papasaikas P, Valcarcel J, 2016 The spliceosome: the ultimate RNA chaperone and sculptor. *Trends Biochem. Sci* 411, 33–45.
- Pavlin M, Spinello A, Pennati M, Zaffaroni N, Gobbi S, Bisi A, Colombo G, Magistrato A, 2018 A computational assay of estrogen receptor alpha antagonists reveals the key common structural traits of drugs effectively fighting refractory breast cancers. *Sci. Rep* 81, 649.
- Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE 3rd, Laughton CA, Orozco M, 2007 Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J* 9211, 3817–3829.
- Peters J, Toor N, 2015 Group II intron lariat: structural insights into the spliceosome. *RNA Biol* 12, 913–917. [PubMed: 26121424]
- Pierce LCT, Salomon-Ferrer R, De Oliveira CAF, Mccammon JA, Walker RC, 2012 Routine access to millisecond time scale events with accelerated molecular dynamics. *J. Chem. Theory Comput* 89, 2997–3002.
- Plaschka C, Lin PC, Nagai K, 2017 Structure of a pre-catalytic spliceosome. *Nature* 5467660, 617–621.
- Pyle AM, 2016 Group II intron self-splicing. *Annu. Rev. Biophys* 45, 183–205. [PubMed: 27391926]
- Ricci CG, Chen JS, Miao Y, Jinek M, Doudna JA, Mccammon JA, Palermo G, 2019 Deciphering off-target effects in CRISPR-Cas9 through accelerated molecular dynamics. *ACS Cent. Sci* 10.1021/acscentsci.9b00020 (in press).
- Ricci CG, Silveira RL, Rivalta I, Batista VS, Skaf MS, 2016 Allosteric pathways in the PPAR gamma-RXR alpha nuclear receptor complex. *Sci. Rep* 6, 19940. [PubMed: 26823026]
- Rosta E, Nowotny M, Yang W, Hummer G, 2011 Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc* 13323, 8934–8941.
- Sabin LR, Delas MJ, Hannon GJ, 2013 Dogma derailed: the many influences of RNA on the genome. *Mol. Cell* 495, 783–794.
- Sanbonmatsu KY, 2016 Towards structural classification of long non-coding RNAs. *Biochim. Biophys. Acta, Mol. Cell. Biol. Lipids* 18591, 41–45.
- Schrodt MV, Andrews CT, Elcock AH, 2015 Large-scale analysis of 48 DNA and 48 RNA tetranucleotides studied by 1 μ s explicit-solvent molecular dynamics simulations. *J. Chem. Theory Comput* 1112, 5906–5917.
- Sgrignani I, 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.
- Shibata M, Nishimasu H, Kodera N, Hirano S, Ando T, Uchihashi T, Nureki O, 2017 Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat. Commun* 81, 1430.
- Sibener LV, Fernandes RA, Kolawole EM, Carbone CB, Liu F, Mcaffee D, Birnbaum ME, Yang X, Su LF, Yu W, Dong S, Gee MH, Jude KM, Davis MM, Groves JT, Goddard WA 3rd, Heath JR, Evavold BD, Vale RD, Garcia KC, 2018 Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. *Cell* 1743 (672–687), e627.
- Sponer J, Bussi G, Krepl M, Banas P, Bottaro S, Cuhna RA, Gil-Ley A, Pinamonti G, Poblete S, Jurecka P, Walter NG, Otyepka M, 2018 RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev* 1188, 4177–4338.
- Sponer J, Krepl M, Banas P, Kuhrova P, Zgarbova M, Jurecka P, Havrila M, Otyepka M, 2016 How to understand atomistic molecular dynamics simulations of RNA and protein-RNA complexes? *WIREs RNA* 8.

- Steitz TA, Steitz JA, 1993 A general two-metal-ion mechanism for catalytic RNA. PNAS 9014, 6498–6502.
- Sternberg SH, Lafrance B, Kaplan M, Doudna JA, 2015 Conformational control of DNA target cleavage by CRISPR-Cas9. Nature 527, 110–113. [PubMed: 26524520]
- Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA, 2014 DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62–67. [PubMed: 24476820]
- Tan D, Piana S, Dirks RM, Shaw DE, 2018 RNA force field with accuracy comparable to state-of-the-art protein force fields. PNAS 1157, 1346–1355.
- Toor N, Keating KS, Taylor SD, Pyle AM, 2008 Crystal structure of a self-spliced group II intron. Science 320, 77–82. [PubMed: 18388288]
- Toor N, Keating KS, Pyle AM, 2009 Structural insights into RNA splicing. Curr. Opin. Struct. Biol 19, 260–266.
- Wan R, Yan C, Bai R, Huang G, Shi Y, 2016 Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. Science 3536302, 895–904.
- Wang J, Cieplak P, Kollman PA, 2000 How well does a restrained electrostatic potential RESP model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem 21, 1049–1074.
- Wang YT, Chan YH, 2017 Understanding the molecular basis of agonist/antagonist mechanism of human mu opioid receptor through gaussian accelerated molecular dynamics method. Sci. Rep 71, 7828.
- Ward WL, Plakos K, Derose VJ, 2014 Nucleic acid catalysis: metals, nucleobases, and other cofactors. Chem. Rev 1148, 4318–4342.
- Warshel A, Levitt M, 1976 Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J. Mol. Biol 1032, 227–249.
- Yan C, Hang J, Wan R, Huang M, Wong CC, Shi Y, 2015 Structure of a yeast spliceosome at 3.6-angstrom resolution. Science 3496253, 1182–1191.
- Yan C, Wan R, Bai R, Huang G, Shi Y, 2016 Structure of a yeast activated spliceosome at 3.5 Å resolution. Science 3536302, 904–911.
- Yan CY, Wan RX, Bai R, Huang GXY, Shi YG, 2017 Structure of a yeast step II catalytically activated spliceosome. Science 3556321, 149–155.
- Yang J, Meng X, Pan J, Jiang N, Zhou C, Wu Z, Gong Z, 2018 CRISPR/Cas9-mediated noncoding RNA editing in human cancers. RNA Biol 151, 35–43.
- Yoshimoto R, Kataoka N, Okawa K, Ohno M, 2009 Isolation and characterization of post-splicing lariat-intron complexes. Nucleic Acids Res 373, 891–902.
- Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, Cheatham TE, 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 79, 2886–2902.
- Zhang X, Zhan X, Yan C, Zhang W, Liu D, Lei J, Shi Y, 2019 Structures of the human spliceosomes before and after release of the ligated exon. Cell Res 10.1038/s41422-019-0143-x (in press).
- Zhao C, Pyle AM, 2017 Structural insights into the mechanism of group II intron splicing. Trends Biochem. Sci 426, 470–482.

Fig. 1.

Conformation of the Cas9 protein as apo form (left, (Jinek et al., 2014)) and bound to the RNA (right, (Jiang et al., 2015)). Cas9 is shown in molecular surface, highlighting the protein domains with different colors. The RNA (yellow) is shown as ribbons. The protein is composed by a recognition lobe (REC), which includes the REC1–3 regions, and by a nuclease lobe (NUC), including the HNH and RuvC catalytic domains and the PAM interacting (PI) region. Representative snapshots along the conformational change from the apo Cas9 (“open” conformational state) to the RNA-bound Cas9 (“closed” conformation) are shown. During the conformational transition, the REC1–3 regions collectively move in opposite directions with respect to each other (as indicated using arrows), with the exposure toward the solvent of a R-rich α -helix (magenta) (Palermo et al., 2017a). In its apo form, Cas9 adopts an auto-inhibited conformation, in which the REC and NUC lobes assume an “open” conformational state. Upon RNA binding, a dramatic conformational transition is observed within the REC lobe, which leads to the formation of a “closed” conformational state of the protein (Fig. 1).

Fig. 2.

(A) Potential of mean force (PMF) computed with respect to the E945–D435 distance and to the Root Mean Square Deviation (RMSD) from the X-ray structure of the apo Cas9 (Palermo et al., 2017a). The E945–D435 distance has been employed in Förster Resonance Energy Transfer (FRET) experiments to describe the conformational changes of the protein during RNA binding (Sternberg et al., 2015), allowing the validation of the identified pathway. The PMF reveals three local minima, corresponding to the crystallographic apo (M1) and RNA-bound (M2) states, and to an intermediate state (M3), characterized by the solvent exposure of the arginine residues in the R-rich α -helix. (B) From left to the top right: the electrostatic potential of the Cas9 protein is computed along the conformational transition from the apo protein up to RNA binding (Baker et al., 2001). The positive (blue) and negative (red) potential field strengths are captured at an iso-density of ± 200 kT/e. This shows the formation of a positively charged cavity, upon the exposure of the R-rich α -helix in the pre-RNA bound conformation. Adapted with permission from Palermo et al. (2017a). Copyright 2017 National Academy of Sciences.

Fig. 3.

(A) Conformation of the catalytic HNH domain in the X-ray structure of the pre-activated CRISPR-Cas9 complex (5F9R.pdb) (Jiang et al., 2016). The catalytic residue H840 (magenta) locates at ~19.4 Å from the cleavage site (CS) on the target strand (TS). (B) Close-up view on the HNH domain docked at the cleavage site on the TS in the activated CRISPR-Cas9 complex. (C) Distance of the catalytic residue H840 from the cleavage site (CS) over ~400 ns of conventional MD (cMD, gray) and over three replicas of Gaussian accelerated MD (GaMD, warm colors) (Palermo et al., 2017a). A black bar indicates the value of the distance in the 5F9R X-ray structure (Jiang et al., 2016).

Fig. 4.

(A) Conformation of the three regions of the REC lobe in the activated CRISPR-Cas9 complex (Palermo et al., 2018). REC3 and REC2 show a remarkable outward translation, while REC1 moves toward the HNH domain, suggesting that the REC lobe ensures the catalytic competence of Cas9 by “sensing” (REC3), “regulating” (REC2) and “locking” (REC1) the catalytic HNH domain. Three arrows indicate the conformational change from the pre-activated to the activated state. (B) Correlation analysis including the matrix of the generalized correlations (GC, upper triangle), quantifying the coupled motions among residue pairs, and its coarse representation (lower triangle), accumulating the per-domain correlations (Palermo et al., 2017b). High correlations (green) are observed between the REC-HNH and the HNH-RuvC domains. Adapted with permission from Palermo et al. (2018). Copyright 2017 American Chemical Society. <https://pubs.acs.org/doi/10.1021/jacs.7b05313>. Further permissions related to the material excerpted should be directed to the American Chemical Society.

Fig. 5.

(A) Structure of the *Oceanobacillus iheyensis* (*O.i.*) group IIC intron before splicing (4FAQ) (Marcia and Pyle, 2012). The intron (green) and the exon (blue) are depicted in ribbons. (B) Close-up view of the active site, displaying two catalytic Mg²⁺ ions (orange spheres), the water nucleophile and the surrounding RNA. MD simulations have been performed using the QM/MM approach, treating the active site at a QM (DFT/BLYP) level of theory, while the rest of the system in explicit solution is treated at the classical MM Amber force field. Representative snapshots along the phosphodiester bond cleavage: the reactant (R), transition state (TS[‡]), deprotonation event (DEP) and product (P). (C) Free energy profiles (ΔF , in kcal/mol) for the chemical step, computed via thermodynamic integration, using as a reaction coordinate (RC) the distance between the breaking (O3'-P_{RNA}) and forming (O_{WAT}-P_{RNA}) bonds (Casalino et al., 2016). Adapted with permission from Casalino et al. (2016). Copyright 2016 American Chemical Society. <https://pubs.acs.org/doi/abs/10.1021/jacs.6b01363>. Further permissions related to the material excerpted should be directed to the American Chemical Society.

Fig. 6.

Schematic representation of the two-metal aided mechanism for phosphodiester bond cleavage, as observed in the RNase H protein (A) (Rosta et al., 2011; De Vivo et al., 2008; Palermo et al., 2015a,b) and in group II intron (G2I) ribozyme (B) (Casalino et al., 2016). In the RNase H protein, the reaction proceeds through an associative mechanism S_N2 where the nucleophile is bonded to the phosphorus atom before the leaving group is eliminated. In G2I ribozyme, a dissociative mechanism S_N1 is observed, in which the elimination of the leaving group precedes the nucleophile attacks to an unstable meta-phosphate intermediate. The Reactant (R), Transition State ($TS^\ddagger$) and Product (P) structures are shown for both mechanisms.

Fig. 7.

(A) Intron Lariat Spliceosome from the yeast *Schizosaccharomyces pombe* (*S. Pombe*, 3JB9.pdb) (Hang et al., 2015; Yan et al., 2015). The protein components are shown in molecular surface, while the snRNAs and the Intron Lariat are depicted as cartoons using different colors. (B) The per-domain correlations are accumulated and plotted as a histogram (reporting the per-column correlation scores calculated for each pair of spliceosomal components). Correlations are based on Pearson coefficients (P_c) and are shown using blue ($0 < P_c < 1$) and red ($1 > P_c > 0$). Boxes are used to highlight the correlations of the Spp42 protein with the snRNAs (U2, U5, U6) and with the intron lariat (Casalino et al., 2018). (C) Displacement (highlighted by a black arrow) of the intron lariat/U2 helix, as revealed from principal component analysis of the MD trajectories. The first principal component is computed and plotted on the 3D structure indicating the direction of large-scale motions (shown using arrows). (D) The “polar tweezers” formed by K364, K366, and R388 residues of Cwf19. The electrostatic potential (ESP) is computed and plotted over the molecular surface of the protein components (blue: positive; red: negative) (Baker et al., 2001). Adapted with permission from Casalino et al. (2018). Copyright 2018 National Academy of Sciences.

Fig. 8.

Schematic illustration of the GaMD method (Miao et al., 2015). When the system potential $V(\vec{r})$ is lower than a threshold energy E ($E = V_{max}$), the energy surface is modified by adding a harmonic boost potential that follows Gaussian distribution. The harmonic constant k_0 determines the magnitude of the applied boost potential. By increasing k_0 in the range from 0 to 1, higher boost potential is added to the original potential energy surface.

Fig. 9. Schematic illustration of the electrostatic coupling scheme developed by Rothlisberger and coworkers (Laio et al., 2002a,b), to treat the electrostatic interactions between QM and MM atoms in a mixed QM/MM approach.