Published in final edited form as:

J Chem Theory Comput. 2023 April 11; 19(7): 1945–1964. doi:10.1021/acs.jctc.2c01313.

Machines on Genes through the Computational Microscope

Souvik Sinha[1], **Chinmai Pindi**[1], **Mohd Ahsan**[1], **Pablo R. Arantes**[1], **Giulia Palermo**[1,2]

[1]Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States

[2]Department of Chemistry, University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States

Abstract

Macromolecular machines acting on genes are at the core of life's fundamental processes, including DNA replication and repair, gene transcription and regulation, chromatin packaging, RNA splicing, and genome editing. Here, we report the increasing role of computational biophysics in characterizing the mechanisms of "machines on genes", focusing on innovative applications of computational methods and their integration with structural and biophysical experiments. We showcase how state-of-the-art computational methods, including classical and ab-initio molecular dynamics to enhanced sampling techniques, and coarse-grained approaches are used for understanding and exploring gene machines for real-world applications. As this review unfolds, advanced computational methods describe the biophysical function that is unseen through experimental techniques, accomplishing the power of the *"computational microscope",* an expression coined by Klaus Schulten to highlight the extraordinary capability of computer simulations. Pushing the frontiers of computational biophysics towards a pragmatic representation of large multi-megadalton biomolecular complexes is instrumental in bridging the gap between experimentally obtained macroscopic observables and the molecular principles playing at the microscopic level. This understanding will help harness molecular machines for medical, pharmaceutical, and biotechnological purposes.

Graphical Abstract

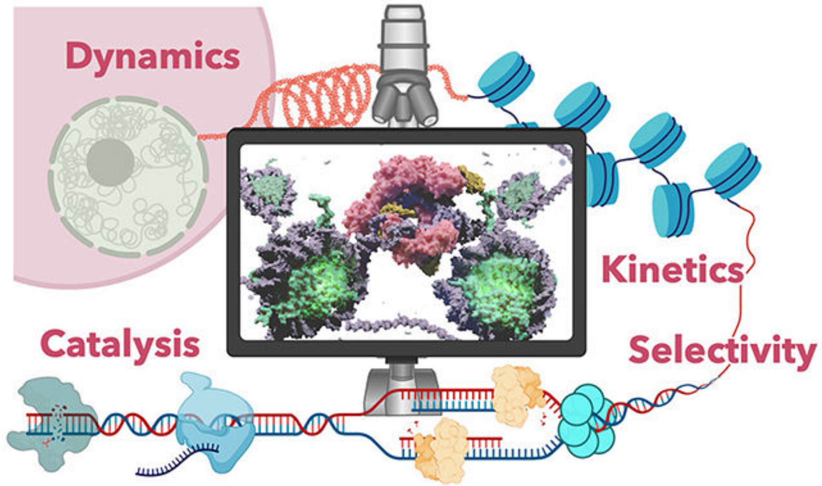
Correspondence: Giulia Palermo (giulia.palermo@ucr.edu).

Author Contribution

S.S., C.P., M.A and P.R.A performed literature search and wrote the paper. GP conceived this review article, supervised the authors and included critical edits to the final manuscript.

Competing Interests

The authors declare no competing interests.



From prokaryotes to eukaryotes, the secret of life lies in strings of five letters – A, T, U, G, C – composing DNA and RNA. How life shapes itself depends on how the information encoded in those strings, i.e., nucleic acids, is expressed and regulated by macromolecular machines. Nucleic acid processing and maintenance are determined by the diverse interplay between large macromolecules and genetic components. This is found across the central dogma, transferring the genetic information from DNA to RNA to proteins. Processes including DNA replication, chromatin packaging, transcription, translation, and RNA splicing involve multiple protein/nucleic acid complexes acting as “machines on genes” with profound biological implications[1]. A clear perspective on their regulatory mechanisms requires an integral knowledge of such molecular assemblies and their associated interactions.

This review article summarizes the mechanisms of action of large macromolecular machineries acting on genes, focusing on the increasing role of computational methods and in their integration with structural and biophysical experimental approaches. Non-stop development of High-Performance Computing (HPC), alongside more powerful algorithms and computational methods, are poised to increase the power of the *“computational microscope”*. This expression, coined by Klaus Schulten, highlights the extraordinary capability of computer simulations in capturing biomolecular mechanisms that are difficult to reach via state-of-the-art experimental techniques[2]. Well-advanced, yet diverse biophysical approaches, such as X-ray crystallography, NMR, SAXS, single-particle cryo-electron microscopy (cryo-EM), and single-molecule spectroscopy, aided significantly in characterizing structure and function of larger protein-nucleic acid assemblies. This offers enormous opportunities toward understanding biophysical function, and provides challenging problems to tackle through molecular dynamics (MD) simulations. Indeed, future MD simulations will have to tackle increasingly large biomolecular systems and access their functional timescales, which may go beyond micro- and milliseconds[3,4].

Here, we review some of the most exciting computational studies, that are pushing the frontiers of computational biophysics in the attempt to characterize structure and dynamics of large macromolecules acting on genes. Our review article offers a computational journey through chromatin dynamics, replication and repair, transcription and translation, RNA splicing, as well as recent applications in the field of genome editing. We discuss the use of multiscale approaches, ranging from quantum-classical (QM/MM) simulations[5] to characterize enzymatic reactions, to all-atom (AA) and coarse-grained (CG) simulations[6] along with the advanced sampling algorithms[7,8] to study long-timescale conformational transitions, aided to understand the regulatory dynamics of biological processes. As this review unfolds, computational methods describe biophysical function that is unseen through experimental techniques, accomplishing the power of Klaus Schulten's "*computational microscope"*. This highlights the significance of computer simulations in shifting the current paradigms in biology, fostering innovations and technological advances. While we summarize some of the most compelling biophysical studies, we strongly suggest referring to the original papers for an in-depth description of computational methods, which goes beyond the scope of the present article. It is also noteworthy to mention that, though we did cite some of the most interesting contributions, readers are encouraged to look further into the literature for a comprehensive knowledge of this field in constant evolution.

Chromatin Dynamics and Nucleosome Function

The majority of DNA in the human genome is packed in chromatin structures, which consist of arrays of nucleosome core particles (NCP) wrapping-up DNA in varying orders of compactness[9,10]. The regulation of gene functions, such as transcription, replication, repair, and recombination, is significantly influenced by the mode and degree of DNA unwrapping from the NCPs. The NCP embodies histone proteins as an octamer swathed by ~147 bp of DNA. Four histone proteins (H3, H4, H2A, and H2B) dimerize to form two heterodimers (H3–H4 and H2A–H2B), four of each heterodimer assemble to set up a histone octamer ensuring 2-fold symmetry (Figure 1A). Recent studies confirmed the presence of significant chromatin dynamics with chromatin fibers enduring diverse lengths and compactness[11,12]. Dynamics of chromatin fibers are a function of inter- and intra-nucleosomal movements[13], such as breathing, twisting, sliding, assembly, and disassembly, also noted to be susceptible to the DNA sequence. It is thereby crucial to explore these movements to understand chromatin dynamics. Computational studies have demonstrated the robustness of all-atom MD simulations to explore nucleosome dynamics, and coarse-grained methods are going beyond single NCPs, describing the dynamics of chromatin fibers[14–39].

After the publication of the first X-ray structure of the NCP by Luger and co-workers[40], dozens of MD simulation studies have been reported. Due to the computational limitations, the early simulation studies could only assess the DNA base pair parameters, histone dynamics, and the associated solvent interactions. Recent technological advancements in computational resources enabled MD simulations to report elastic properties of the DNA, flexibility as a function of DNA sequence and explain DNA curvature and their role in stabilizing the NCP[25,27,28]. Subsequently, several studies using MD simulations characterized the histone dynamics and protein-DNA interactions, accounting for nucleosomal stability[26,29–35]. Microsecond (μs) long MD simulations of the nucleosome showed that it inherits major motions, including DNA tail mobility, linker DNA breathing, DNA twisting, and inter-gyre breathing (Figure 1B)[24,36]. These key motions are instrumental in explaining the intra-nucleosomal movements which subsequently translate into nucleosomal unwrapping and chromatin dynamics. Indeed, these nucleosomal dynamics bring deformations in DNA, resulting in the formation of DNA loops and twists, which are communicated throughout the nucleosome with the cooperation of histone tails. Notably, these µs-long simulations were also able to address the role of DNA sequences and NCP modifications on the intra-nucleosomal dynamics.

Histone tails are important sites for post-translational modifications (PTM), leading to active or inactive regions of chromatin. For instance, the acylation of the H3 or H4 histone tail marks actively transcribed regions of DNA. On the other hand, methylation at positions K9 and K27 of the H3 histone flags the inactive chromatin region, while methylation at K4 of H3 leads to active chromatin regions[41]. These modifications are crucial for epigenetic regulation. The dynamics of histone tails and the consequences of PTM have been extensively investigated by all-atom MD simulations[37,38]. A study by Ishida et al. revealed that the interaction of the highly dynamic histone tails with the DNA holds a crucial role in stabilizing the NCP structure[38]. Moreover, all-atom MD simulations in the absence of histone tails and at high salt concentrations reported the spontaneous opening of the nucleosome[39]. They also reported that the deletion of 20 residues from the H4 histone tail weakens the staking between two nucleosomes. Cojocaru and coworkers collected ~24 µs of all-atom MD simulations, showing that the histone tails of H3 and H2A regulate the nucleosomal opening and closure[16]. The authors suggested that the opening and closure of the NCP are regulated by the displacements and changes in the compactions between histone tails. They also showed that the removal of histone tails increases large-amplitude breathing motions in the nucleosome[16]. These findings were in corroboration with previous FRET experiments that showed large amplitude nucleosome opening upon deletion of either of the two tails[42,43].

Rosanna et al. reported multi-µs MD simulations of a di-nucleosomal complex and highlighted the impact of lysine acylation[31]. The authors reported that lysine acylation increases disorder in the histone tails, which restricts the histone tails to form crucial inter-nucleosomal interactions, hence destabilizing the stacking between the nucleosomes. Ever-increasing advancements in the structural biology of the NCP are opening new opportunities for computational studies. Based on a cryo-EM structure of the 30-nm chromatin fiber, Woods and colleagues performed all-atom MD simulations of poly-nucleosome arrays to determine the role of linker histones on fiber structures and dynamics, a study that can be considered at the frontiers of the field[17]. This suggested that linker histones inhibit DNA flexibility and stabilize repeating tetra-nucleosomal units, giving rise to increased chromatin compaction (Figure 1C).

Several biophysical studies suggested that allosteric communications are critically important in NCP functionality. Indeed, histone modifications and the binding of transcription factors can impressively remodel the structure and dynamics of nucleosomes at a distance. Using all-atom MD simulations, Bowerman & Wereszczynski reported the presence of histone-mediated allostery in the NCP[19]. The authors have shown that histone variants, such as H2A.Zm and macroH2A, substantially alter the allosteric signaling in the NCP, producing structural remodeling. Following this study, Tan et al. studied the DNA-mediated allostery in the nucleosome by employing coarse-grained MD simulations[20]. The authors showed that the binding of transcription factors (Sox2 and Oct4) at the DNA modifies the NCP structural conformation and modulates the cooperativity with other transcription factors at distant sites (Figure 1D). The preferential binding sites of these transcription factors on the nucleosome are later confirmed by biochemical studies and high-resolution NMR[43,44]. Furthermore, the allosteric regulation of the NCP was also leveraged to design anticancer drugs to synergistically target the NCP[21]. This innovative idea contributed to an important paradigm change in biology. Indeed, anticancer drugs have been traditionally designed to target double-stranded DNA. However, the structure of the DNA in chromatin is highly packed, and the NCP induces a critical bend in DNA. Hence, the binding mechanism of drugs can remarkably change when binding nucleosomal DNA vs. naked DNA[22]. Computational methods have been used to study the selectivity of anticancer drugs at the level of nucleosomal DNA, reporting striking differences in analogy to the binding modes in naked DNA[23]. In collaboration with X-ray crystallography and in-vivo assays, these studies have shown that chromosomal DNA and proteins are valuable targets to design anti-cancer drugs.

Replication and Repair Machineries

During cell division, each daughter cell must receive an identical copy of the parental DNA. The process of replication involves multiple enzymes assigned for specific tasks, like parental DNA strand unzipping by helicases to form a ‘replication fork’, elongation and proofreading by DNA polymerases, ligases to fix nicks in daughter strands, etc. (Figure 2A) This entire complex is known as DNA replisome. We will briefly discuss how computational methods based on molecular simulations contributed to our understanding of the replisome machinery’s regulations.

During the initiation phase, hexameric helicases sit around the replication fork, consume ATP at the protomeric interface, and finally hydrolysis of the ATP drives DNA unzipping. The unzipping includes the translocation of single-stranded DNA (ssDNA) through the central pore of the helicase enzyme, subsequently driving away the other strand (Figure 2A). A key question in the field is how the chemical energy from ATP binding and hydrolysis is causing translocation and strand separation. In 2009, Liu et al. simulated the translocation process using a coarse-grained model of the protein-DNA complex with Langevin dynamics, constructing a free-energy surface for the protein-DNA electrostatic interaction[45]. The authors reported that the electrostatic energy of the system produces the directionality of the DNA translocation process. More recently, using a hybrid coarse-grained approach[46], Jin et al. suggested that ATP provides stability at the inter-monomeric interfaces of the hexameric ring-shaped helicases, inhibiting helicase translocation[47]. However, hydrolysis of those interfacial ATP molecules is shown to significantly lower the energy barrier for subunit dissociation, while transient DNA–protein, and protein–protein interactions further aid the translocation process. Ma & Schulten employed the ‘string’ method[48], an advanced path sampling technique, in combination with an enhanced sampling protocol (viz., ‘milestoning’ method[49]) – to characterize the thermodynamics and kinetics of messenger RNA (mRNA) translocation by an ATPase helicase motor[50]. This work provided a general mechanism for sequential ATP binding across the hexameric helicase ring and subsequent translation events (Figure 2B). The authors reported that a complete action phase involves the binding of ATP to an empty interface, the release of previously hydrolyzed "old" ADP, hydrolysis of ATP to new "ADP" and final translocation. Surprisingly, the release of the "old" product is the rate-limiting step of this whole process. By combining molecular simulation and experiment, Colizzi et al. have addressed a long-standing question of how entwined DNA double-strands get unzipped by helicases. The simulations suggested an asymmetric flipping-out probability of paired nucleobases that can strongly influence the unwinding efficiency of DNA helicases toward highly dynamic nucleobases, such as pyrimidines, on the displaced strand. This interplay between asymmetric nucleobase dynamics and helicase unwinding is further confirmed by biochemical and fluorescence-based experiments[51].

During the elongation phase, DNA polymerases are the key players in new strand synthesis, proofreading, and performing the repair of DNA at any time necessary. Trans-phosphorylation of nucleotide triphosphates is the central reaction for adding a base to the new DNA strand (i.e., the elongation process). Roston et al.[53] have extensively simulated DNA polymerase η using a mixed quantum mechanics/molecular mechanics (QM/MM) model, treating the enzyme's catalytic center at a QM(DFTB3) level of theory[54], while describing the remaining system at the classical force field level. The authors also systematically surveyed the reactive space of 10 previously proposed mechanisms using multi-walker metadynamics[55–58]. They found that the smallest activation barrier occurs when Mg^{2+}-coordinated water deprotonates the nucleophilic 3′-OH, in concert with the phosphoryl transfer. The presence of a third Mg^{2+} in the active site also lowers the activation barrier for the water-as-base mechanism.

Notably, polymerases are some of the most computationally studied enzymes through QM/MM methods. Several reports are dedicated to unraveling the metal-aided catalytic mechanism and its role in the close-open transition of the active sites before and after catalytic reactions take place[59–66]. Using bioinformatics analysis and QM/MM simulations, Genna et al. reported a self-activated mechanism (SAM) that allows the formation of a closed-loop catalytic cycle for the efficient polymerization reaction[67]. They observed a structurally conserved H-bond in the Michaelis complex, formed by the nucleophilic 3′-OH and the β-phosphate of the incoming nucleotide, critical for the initiation of the SAM. The latter is triggered by in situ nucleophile formation via 3′-OH deprotonation, nucleotide addition, and finally DNA translocation for nucleic acid polymerization.

DNA is susceptible to damage, resulting in mismatched base pairs, chemically modified nucleosides, strand breaks, etc. Polymerases play a crucial role in scrutinizing such errors during replication and the associated repair processes. During base excision repair, Pol-β first binds DNA and the 2-deoxyribonucleoside 5-triphosphate (dNTP) to form an open ternary complex and then transitions to a closed ternary complex competent for catalysis. After the reaction, Pol-β transitions back to its open enzyme form, and releases pyrophosphate (PPi) and ions. However, once Pol-β binds and forms a closed ternary complex with oxidized guanine such as 8-oxo-dGTP, it rapidly transitions to the inactive form. Kim et al. applied Transition Path Sampling (TPS) to explore the energetics of these closed-to-open transitions[68]. The authors have shown that the open states are energetically disfavored, with respect to the closed state when Pol-β binds deoxyguanosine monophosphate (dGMP). On the other hand, in the presence of oxidized guanine (i.e., 8-oxo-dGTP), open and closed states are energetically comparable, due to the loss of key protein-nucleic acid interactions in the active site. On a similar note, a few other independent investigations also confirmed that the correct complementary dNTP induces a closed catalytically competent state, at odds with complexes of either fidelity-compromised mutants or wild-type protein with non-complementary nucleotides[69,70].

An essential step to achieving high fidelity during the elongation process is the removal of misincorporated nucleotides. For this reason, replicative polymerases keep switching back and forth between DNA synthesis mode at the polymerase (*pol*) active site and excision mode to cut the misincorporated nucleotides at the exonuclease (*exo*) active site. To ensure facile transfer of the DNA primer between the spatially separated *pol* and *exo* sites, the DNA polymerase holoenzyme reorganizes along a well-defined conformational path. Using a transition path optimization protocol, which combined a partial nudged elastic band (PNEB)[71,72] and all-atom MD simulations, followed by detecting the intermediates through Markov state modeling, Dodd et al. addressed the *pol*–to–*exo* switching mechanism and the related kinetics[52]. This study revealed that, for the mismatched primer end to reach the exonuclease active site, three nucleotides must unpair at the primer-template junction and extend toward the *exo* subunit. The authors estimated a stepwise melting of the first three nucleotides from the *pol* active site on a microsecond timescale (Figure 2C). The departure of the terminal base from the *pol* active site results in a stalled polymerase state. Furthermore, the Pol III holoenzyme exploits the natural motion of the DNA inside the Pol-III/β-clamp central cavity to backslide and rotate, which results in the complete removal of the primer-template junction from the *pol* active site. Base unpairing at the second and third position at the end of the primer displays the highest energetic cost along the path, bringing the overall timescale for the *pol*–to–*exo* transition into the millisecond range. Once the third base is out, the closure of the *pol* and *exo* subunits allows the primer to insert into the exonuclease active site in a catalytically competent orientation.

At this point, though the high fidelity of the DNA polymerases is well-explained, how the ‘degree of fidelity’ varies extensively for different enzymes is yet to be understood. To understand such differential fidelity in polymerase enzymes, Geronimo et al. have compared the dynamics of high-fidelity Pol β (~1 error in 60,000 bases) and low-fidelity Pol η (~1 error in 900 bases) employing equilibrium MD and free energy simulations[73]. They reported that the degree of fidelity depends on whether the nucleophile at the reaction center is optimally aligned to incorporate the correct or incorrect nucleotide. In the case of Pol β, local structural distortion at the reaction center in the presence of mismatched base-pair makes the nucleophilic addition unfavorable; thereby promoting high fidelity. On the other hand, Pol η accommodates more flexible base pair shape complementarity at the catalytic site which makes it more tolerant of mismatched base inclusions.

RNA Splicing and Molecular Mechanism of the Spliceosome

The production of mature coding messenger RNA (mRNA) is essential for gene expression and regulation[74]. The coding regions of nascent gene transcripts, the exons, must be excised from the intervening noncoding regions, the introns, to produce the mature mRNA[75] (Figure 3A). Intron splicing from premature mRNA (pre-mRNA) by the spliceosome (SPL) molecular machine is a step of the utmost importance[75,76]. Indeed, splicing defects are linked to more than 200 human diseases, including cancer and neurodegeneration, which highlights the urgent need for in-depth knowledge of the structural and functional cycle of the SPL.

The SPL consists of approximately 150 proteins and five small nuclear RNAs (snRNAs), namely U1, U2, U4, U5, and U6, which form the ribonucleoprotein complexes (snRNPs) through a network of extensive interactions. During each splicing cycle, the SPL, fueled by the hydrolysis of ATP, undergoes a series of conformational changes to form eight major intermediate complexes (A, B, B^{act}, B*, C, C*, P, and ILS)[74]. RNA structural dynamics is an area of intense research for molecular simulations[77]. Recent advances in the structural biology of the SPL[78] and its ancestors[79] triggered new computational studies aimed at characterizing the mechanistic function of large RNA structures[80]. Based on the cryo-EM structure of the intron lariat (i.e., spliced by-products of intron RNA) spliceosome (ILS) complex from the yeast *Schizosaccharomyces pombe,* one million atoms biomolecular systems were simulated by Casalino et al. at all-atom resolution[81]. The study reported hundreds-of-nanoseconds dynamics of the ILS, revealing a key dynamical role of the proteins involved in the spliceosome disassembly at a late stage of the splicing cycle (Figure 3B). Correlation analyses have unveiled the leading role of Spp42 protein (human protein – Prp8), in directing the SPL maturation by governing the motions of the SPL components, owing to its unique multipronged architecture. Cross-correlation matrices and essential dynamics reported a cooperative function between the Cwf19/Spp42 proteins and the IL/U2 snRNA double helix (also named branch helix), proving the functional role of Cwf19 in the IL/U2 branch helix unwinding. These computational findings were subsequently corroborated by a cryo-EM study of the human SPL, revealing a critical role of Cwf19 and Spp42 in the IL/U2 helix unwinding[82].

A further study of the SPL C complex from yeast *Saccharomyces cerevisiae* reported that the communication channels between distal subunits are crucial for SPL functional transition towards the intermediate C* complex[85]. The essential dynamics and community network analyses confirmed the predominant role of Prp8 in the motion of the C complex. Cooperative motions, promoted by the Clf1 and Prp8 proteins, trigger the twisting and repositioning of the IL/U2 branch helix, promoting the transition toward the C* complex. These studies pinpoint a critical role of the Prp8 protein in the modulation of the SPL dynamics by mediating the signal exchange between functional regions of the system.

The SPL catalyzes the chemical “snip & stitch” splicing process in two sequential transesterification reactions (Figure 3C)[74]. Analogous to its self-splicing ancestors, group II intron (GR2I) ribozymes catalyze intron removal in two transesterification steps coordinated by two Mg^{2+} ions. Hybrid QM/MM simulations have been exploited to gain unique, atomic-level insights into the reaction mechanism of splicing[86]. The first computational attempt to study the mechanism of pre-mRNA splicing focused on the first splicing step of GR2I from *Oceanobacillus iheyensis*[87]. QM/MM MD simulations based on density functional theory (DFT), combined with thermodynamic integration, disclosed a dissociative mechanism, in which the catalytic water releases its proton to the bulk water during the nucleophilic attack on the scissile phosphate (Figure 3C). The role of metals was shown to be paramount for the chemical step, as Mg^{2+} ions reduce the negative charge on the electrophilic center (i.e., the scissile phosphate) at the transition state, thereby favoring dissociative catalysis. Additional studies provided further details on the role of metals, discussing the thermodynamics towards the selectivity of multiple ions, including divalent and monovalent ions[88]. Kumar & Satpati revealed that the divalent cation binding sites are very selective towards specific divalent cations; while the monovalent cation binding sites are less selective between monovalent ions. In this respect, it is important to note that GR2I from *Oceanobacillus iheyensis* is among the largest RNAs of known structure, presenting 24 Mg^{2+} binding sites, which also offered the opportunity to assess the reliability of different point charge force field models for Mg^{2+}.[89] A computational study by Huang and co-workers has been based on an SPL structure from the yeast *S. cerevisiae*, trapping the C complex immediately after the branching step, i.e., formation of an intron lariat after the first splicing reaction takes place at the activated B* intermediate[90]. As observed for GR2I, this work confirmed a dissociative pathway for the intron-lariat formation, where the nucleophilic proton is transferred from the O2′ of the branch point adenosine (BPA) (evolutionary conserved SPL recognition site within an intron) to the O3′ atom at the 3′- end of the 5′-exon through the scissile phosphate. The second transesterification reaction was also studied, based on a cryo-EM structure from *S. cerevisiae*[91]. This structure shows a favorable catalytic site in terms of Mg^{2+} ions' positions and their coordination with phosphates of the catalytic core. QM/MM MD simulations showed that the second step of catalysis (exon-ligation) takes place via an associative two-Mg^{2+} ion mechanism, featuring a scissile-phosphate-mediated proton transfer from the nucleophile to the leaving group. The mechanism followed a stepwise pathway via the formation of a metastable phosphorane-like intermediate. Overall, while the first splicing step, in both GR2I and SPL, follows a dissociative mechanism, the second step agrees well with the general associative two-metal-ion-based catalytic mechanism[92–94].

Splicing defects are critical in the onset of different types of cancers. Indeed, cancer proliferation, invasion, and angiogenesis can arise from mutations in splicing factors, or through dysregulation of splicing factor expression levels[95]. Splicing dysregulation creates a vulnerability in cancer cells, that can impart sensitivity to splicing inhibitors. Small molecule splicing modulators (SMs) supply a potential selective therapeutic approach to counteract cancers[96–98]. Cryo-EM and crystallographic structures revealed that pladienolide B and E7107, a pladienolide derivative, bind at the BPA binding pocket of human multiprotein splicing factor SF3B[99,100]. Models of the SF3B1 protein, a member of the SF3B complex, bound to E7107, spliceostatin A, and herboxidiene SMs have been studied using MD simulations to assess the impact of selected SM on the dynamics of the SF3B1 protein[101]. This study reported that each SM differently affects the plasticity of the SF3B1 complex, locking it into an open state, which prevents binding of pre-mRNA, in line with structural studies[100]. These findings provide valuable information that could be used for the rational design and discovery of drugs to fight cancers associated with dysregulated splicing.

Transcription Machineries

Transcription machineries are large protein/nucleic acid complexes that copy information from a segment of DNA into an RNA sequence. The RNA polymerase enzymes, aided by several other transcription factors, unwind DNA, add RNA nucleotides complementary to the template DNA to the growing RNA strand, and ultimately release the RNA strand from the DNA template (Figure 4A). We report the contribution of molecular simulations in clarifying the biophysical functions of selected transcription machineries.

The basic transcriptional machinery includes RNA polymerase II (Pol II) along with a series of general transcription factors (TFIIA, TFIIB, TFIID, TFIIF, TFIIS, TFIIE, and TFIIH), which assemble into a multi-megadalton complex on the promoter DNA to form a preinitiation complex (PIC). The PIC is key to accurately selecting the transcription site, causing promoter melting and subsequently placing the DNA at the RNA Pol II active site for transcription. Despite several experimentally reported structures of Pol II and subcomplexes of PIC, the complete assembly of PIC was missing. Recently, Yan et al. reported a complete model of the human PIC at an atomic resolution and studied the relevant functional dynamics using all-atom MD simulations[102]. The authors built a complete model of the human apo-TFIIH, which is critical for the opening of the promoter DNA and formation of the nascent transcription bubble, by modeling the missing domains separately from structural homologs and finally docked every piece together into the existing cryo-EM map of the human apo-TFIIH. To model TFIIH holo-PIC (core-PIC–TFIIH–DNA), the authors used a cascade MD flexible fitting (cMDFF) approach[103], where the all-atom apo-TFIIH model was sequentially fitted into a series of cryo-EM density maps of the yeast PIC (computationally blurred derivatives of the original map with lower resolution). The final structure was further refined through flexible fitting into the human PIC density. Furthermore, extensive MD simulations of these million atoms complexes elucidated the global dynamics and the concerted motions in the PIC machinery, and how they act together to remodel the DNA substrate and facilitate the promoter opening. This observation is in line with earlier experiments on yeast PIC which showed that PIC-induced DNA distortions in the polymerase cleft is a general mechanism contributing to the promoter opening[104]. Yan et al.[102] also hypothesized that the TFIIH utilizes a ratcheting mechanism to induce negative supercoiling at the target site, which ultimately leads to a strain-induced base-flipping action and results in the formation of the transcription bubble.

Following the formation of the transcription bubble, at the elongation stage, the RNA nucleotides diffuse into the Pol II active site through consecutive nucleotide addition cycles (NAC). In this process, Pol II translocates exactly by one base along the template DNA, on completion of the nucleotide addition reaction at every NAC. Using all-atom MD simulation and Markov State Modelling (MSM)[105], Silva et al. suggested that the translocation of Pol II happens by harnessing the thermal fluctuations of the bridge helix (BH) between bent and straight conformations (Figure 4B)[106]. The BH is a key component in RNA polymerases that locates near the DNA binding channel. The authors have shown that the fluctuations of the BH facilitate the translocation of the upstream RNA:DNA hybrid through direct contact between the BH and i+1 base pair of the upstream RNA:DNA hybrid. In the next step, the transition nucleotide (TN) is translocated to become part of the RNA:DNA hybrid; which again is facilitated by the BH thermal fluctuation. Later, an incoming NTP stabilizes the post-translocation state. MSM helped generate the millisecond dynamics of the translocation process while also identifying additional metastable states that were not previously captured through structural studies.

MSM was also employed by Da and colleagues to investigate the Pol II “backtracking” mechanism, which is essential for proofreading (Figure 4C)[107]. RNA Pol II is very efficient in detecting and cleaving misincorporated nucleotides via “backtracking”, where Pol II traces back to the pre-translocation point for dislodging the RNA 3’-end mismatched nucleotide from the polymerase active site. Da et al. generated the pathways connecting the post-translocation to the backtracked states, as well as its reverse route using a Climber Algorithm[108], which is a morphing method used to generate intermediates along an intended transition. Those intermediates were further subjected to several rounds of unbiased MD, and an aggregated ensemble of 48 μs was used for the MSM construction. They reported a stepwise backtracking process occurring as follows: a sub-μs scale fast-fraying motion at the RNA 3’-end is followed by the flipping of the DNA strand to stack with the bridge-helix, while the reverse translocation of the upstream RNA:DNA hybrid takes place around ~100 μs (rate-limiting step) to reach the final backtracked register. To resume the transcription process, the backtracked RNA 3’-end nucleotides must be cleaved through a TFIIS-catalyzed cleavage reaction. This study also explains the experimentally[109] reported a slower backtracking rate in comparison to the forward translocation process that is due to the stacking interaction of the frayed RNA 3’-end nucleotide with polymerase aromatic residues that slow down the extended backtracking motion. Furthermore, the authors validated these computational observations through site-directed mutagenesis studies and transcript cleavage assays that confirm the role of bridge-helix residues in detecting the mismatched base pairing and promoting subsequent fraying motion of the RNA 3’ end.

Using all-atom and coarse-grained MD simulations[110,111] in conjunction with Umbrella Sampling free-energy simulations[112], Eun et al. suggested that moving from a solvent-exposed apo-state to the Pol II bound state, TFIIS attains a specific closed conformation that is catalytically relevant. After the catalysis, TFIIS switches back to an open form, moving away from the catalytic site to progress with further NACs[113].

Dynamics of the Ribosome

Ribosomes are large macromolecular complexes responsible for protein synthesis by a process called translation. The translation is a multi-step process involving messenger RNA (mRNA), transfer RNAs (tRNAs), and many other proteins and translation factors. It can broadly be divided into three steps: 1. initiation, 2. Elongation, and 3. termination (Figure 5A). Recent advancements in X-ray crystallography and cryo-EM techniques led to a wealth of ribosomal structural information unveiling critical functional states of the ribosomal complexes[114,115]. Structural data analysis along with biochemical and genetic studies showed that the ribosome acquires a specific conformation at every stage of polypeptide chain synthesis and exchanges allosteric signals within several functional sites during its activity. Despite the advancements in computational resources, MD simulations of ribosomal complexes remain challenging, owing to the large size and varying timescales involved in the ribosome functional events. Nonetheless, multiscale modeling techniques from all-atom to structure-based and coarse-grained modeling approaches have been used to study these large ribosomal complexes.

The first all-atom MD simulation of the entire 70S ribosome (~2.6 million atoms) was performed by Sanbonmatsu and coworkers at the Advanced Simulation and Computing (ASCI) Q machine at Los Alamos National Laboratory to understand the cognate tRNA (i.e., tRNA that complements codon sequence of the mRNA) selection into the peptidyl transferase center (PTC), i.e., the decoding step of the protein synthesis pipeline[116]. The authors demonstrated that the flexibility of the acceptor stem and the anticodon arm of the tRNA, are crucial for the accommodation of a cognate tRNA at the PTC. During the initiation step, the start codon AUG is recognized by a pre-initiation complex, initiator tRNA, and several initiation factors (IFs) while scanning the mRNA. Lind et al. focused on the initiation mechanism, utilizing all-atom MD to investigate the role of IFs-eIF1 and eIF1A in the codon recognition process[117]. The study reported that the presence of the IFs on the pre-initiation complex increases the energetic penalty for non-cognate codon binding and thereby enhances the fidelity of the cognate start codon recognition.

During the elongation, aminoacyl-tRNAs in the form of ternary complexes, which include tRNA, elongation factor (translational GTPase), and a GTP molecule, are transported to the ribosome (Figure 5A). Energy factor also plays a role in fidelity checking during the elongation process. All-atom MD revealed that in the presence of tRNA, universally conserved nucleotides A1492 and A1493 of 30S ribosomal subunit opt a flipped-out conformation that causes a complete exclusion of water molecules from the minor groove sites of the codon-anticodon base pairs[118]. This results in an energetically unfavorable complex upon non-cognate base pair binding, as there is no water to compensate for the missing hydrogen bond interactions between mismatched tRNA-mRNA base pairs. Umbrella sampling simulations of the flipping event by Zeng et al. revealed that the A1492/A1493 flipped-out conformation is more favorable for cognate base pairs than for near-cognate, increasing the fidelity of the decoding step[119]. These conformational features of the 30S ribosomal subunit are of great relevance for designing antibacterial agents. Among them, aminoglycosides bind to the 30S subunit and lock the nucleotides in the flipped-out conformation, promoting the binding of near-cognate or wrong tRNAs due to reduced fidelity. This triggers a miscoding effect in bacterial protein expression. Panecka et al. performed all-atom MD simulations of the ribosome bound to paramomycin (an aminoglycoside) and the ribosomal protein S12, in the presence of resistance-inducing mutations[120]. They observed an increased flipping rate when compared to the wild type, suggesting that the miscoding effects of paramomycin are evaded by the mutations. Several other simulation studies also contributed to elucidating the mechanism of antibiotics in the ribosome[121–124].

During the elongation process, the peptide nascent chain (NC) attached to the tRNA located in the P-site (peptidyl site) transfers to the newly added cognate tRNA at the A-site (aminoacyl site, 3’ end to the P-site) forming a peptide bond between NC and new amino acid. Later, the mRNA is pulled through the ribosome by one codon exactly. As a consequence, the A-site tRNA with the elongated NC translocates to the P-site and the deacylated tRNA in the P-site translocates to the E-site (exit site, 5’ end to the P-site) (Figure 5A). Whitford et al. carried out all-atom MD simulations of the pre-translocation state of a 70S ribosome (~ 2.1 million atoms) and calculated the effective diffusion coefficients for the tRNA displacement, describing the rotational movements in the small subunit of the ribosome[125]. Together with previously reported experimental translocation rates, these diffusion coefficients provided an upper estimate of the free energy barriers associated with translocation[126,127].

Later, a study by Bock and colleagues employed extensive simulations to estimate the transition rates between the intermediate states of the inter-subunit rotations and motions during spontaneous tRNA translocation (Figure 5B)[128]. The authors have shown that the tRNA is actively pulled by the L1 stalk from P-site to E-site which occurs in sub-microsecond timescales. This study identified the key structural transitions and rotations accompanying the tRNA translocation, also estimating the associated transition rates. Nguyen et al. used all-atom structure-based models[129] to investigate the relation between ribosomal 30S subunit motions and tRNA translocation (Figure 5C)[130]. The translocation process involves the tRNA hybrid state formation, where tRNAs at A- and P-site move relative to the large 50S subunit (Figure 5C *middle panel*) forming A/P and P/E states from classical A/A and P/P states (i.e., before translocation). The authors suggested that the ribosome samples a chimeric ap/P–pe/E intermediate during this transition, where the 30S head is rotated at ~18°, implying a strained state of the 30S unit, which relaxes as the system reaches the post-translocation states (i.e., P/P and E/E).

The translation process terminates in a series of steps after an mRNA stop codon located in the A-site. First, a release factor (RF) recognizes a stop codon followed by the hydrolysis of the peptidyl-tRNA bond. Then, the separation of the small and large subunits of the ribosome leads to termination. In eukaryotic systems, the stop codons and their count remain elusive. Lind and co-authors performed all-atom MD simulations of a reduced spherical model and calculated the relative binding free energies of several codons when bound to a variety of RFs[131]. In a very recent study, Pelin et al. utilized coarse-grained simulations and network models to reveal a potential allosteric communication pathway along the ribosomal tunnel during translation[132]. They reported the existence of an inherently coupled nucleotide network that can detect the sequence to stop the translation even in the absence of co-factors.

In summary, the past decade witnessed significant progress in bridging the gap between ribosome structure and dynamics through computational techniques. In our view, the realm of ribosomal dynamics and its associated free energy landscape remains to be fully explored. This will require an in-depth representation of the ribosomal large-scale conformational changes while pushing the frontiers of molecular simulations, linking the evolutionary principles of translation with the energy landscape at its various sub-steps.

Gene Editing Systems

Modulation and editing of genetic information in a guided fashion are essential for controlling gene functions. The recently discovered CRISPR-Cas technology is being extensively used for facile genome editing and offers tremendous promises for the treatment of genetic disorders, to development of diagnostic tools, improve genetically modified crops, and a myriad of innovations across life sciences[133–135]. In parallel to the crucial contributions from structural and biochemical studies, molecular simulations contributed to understanding the large-scale dynamics and the local fluctuations driving the biophysical function of gene editing complexes. Here, we briefly review the role of state-of-the-art computations in the CRISPR-Cas revolution[136,137].

The CRISPR-Cas9 genome editing tool is based on the endonuclease enzyme Cas9, which associates with a CRISPR RNA (crRNA), used as a guide for matching DNA sequences. The recognition of a short protospacer adjacent motif (PAM: a specific nucleotide sequence adjacent to the target DNA site) is the primary step for the activation of the Cas9 protein. The endonuclease Cas9 holds two catalytic domains, HNH and RuvC. The HNH nuclease degrades the DNA target strand (TS, i.e., the DNA strand complementary to the guide RNA sequence), while RuvC cleaves the other strand, i.e., the non-target strand (NTS) (Figure 6A). The first all-atom multi-microsecond simulations of CRISPR-Cas9 reported a “striking plasticity” of the HNH domain that was indicated to be crucial for its rapid conformational transition leading to TS cleavage[138], which was also observed through single-molecule experiments[139,140]. In a subsequent study, ~20 μs of unconstrained Gaussian accelerated MD (GaMD) simulations[141] described the pre-catalytic to the catalytic transition of HNH, showing the critical role of the interactions between the L2 loop (joins the nuclease domains) and the RNA:DNA hybrid[142]. This study also explained the conformational transition of Cas9 from its apo form to the RNA-bound form (Figure 6B). During the conformational transition, an arginine-rich helix exposes to the solvent to recruit the RNA, while the recognition lobe (REC1–3: the three domains that recognize and accommodate the RNA-DNA hybrid) subsequently opens off to accommodate the incoming RNA[142].

An intriguing aspect of the CRISPR-Cas9 biomolecular function is characterized by its allosteric signaling that controls the biochemical information transfer to initiate DNA cleavages[143]. Our group used all-atom MD simulations and network models derived from graph theory to reveal how allostery intervenes at different steps of the CRISPR-Cas9 function: 1. affecting DNA recognition, 2. mediating the cleavage, and 3. interfering with the off-target activity. We showed that the binding of the PAM recognition sequence triggers coupled motions within the protein framework, resulting in a shift in the conformational dynamics of the spatially distant HNH and RuvC nucleases. Calculation of the allosteric pathways revealed that the dynamic “cross-talk” between the HNH and RuvC catalytic domains flows through the interconnecting L1/L2 loops, regarded as the “signal transducers”. By combining MD simulations with NMR relaxation experiments, we also provided a path for the allosteric transmission, describing the information transfer between the DNA binding region (i.e., the REC lobe) and the sites of DNA cleavages (i.e., HNH and RuvC)[144]. This allosteric pathway was further investigated by mutating critical residues (K810A, K848A, and K855A) that enhance the specificity of Cas9 against off-target effects[145]. These mutations were found to interrupt the main allosteric connections between REC–HNH–RuvC, suggesting a link between the allosteric regulation and the system’s specificity.

The chemical mechanism of DNA cleavage by the Cas9 endonuclease is a crucial step for genome editing. The HNH and RuvC nuclease domains of Cas9 perform DNA cleavages through the aid of metal ions. While HNH uses a single Mg^{2+} ion, RuvC utilizes two Mg^{2+} ions (Figure 6C)[146,147]. Our group has investigated the catalytic mechanism of phosphodiester bond cleavage by employing first-principles (i.e., ab-initio) MD simulations and a QM/MM approach.

One of the first X-ray structures of the RuvC catalytic site was obtained in the presence of Mn^{2+}, reporting the two ions well-coordinated by three carboxylates (D10, D986, and E762), also approached by a nearby histidine residue (H983)[151,152]. Ab-initio QM/MM MD simulations of this catalytic site in the presence of catalytically competent Mg^{2+} ions revealed a conformational rearrangement of H983, which was shown to act as a catalytic base for the water nucleophile (Figure 6C, *left*)[147]. By further combining free energy methods (i.e., a "blue moon ensemble" approach[153] along with thermodynamic integration), we established a catalytic mechanism by which H983 activates an associative S_N2 chemical mechanism, leading to the breaking of the phosphodiester bond. More recently, our group employed ab-initio QM/MM simulations to characterize the HNH catalysis (Figure 6C, *right*)[146]. We reported that a nucleophilic water is activated by the catalytic base H840, which subsequently lodges the attack on the scissile phosphate to break the phosphodiester bond, similar to the RuvC mechanism.

In HNH, a single Mg^{2+} ion is coordinated by D839 and N863, forming a catalytic triad with H840. This arrangement of the active site, which was observed in recent high-resolution cryo-EM structures[154,155], and the associated catalytic mechanism, clarifies a long-lasting open question on the mechanism of TS cleavage, which was initially modeled based on the homology with the endonuclease T4 endonuclease VII[156]. Interestingly, ab-initio QM/MM simulations also described the protonation of the O3' leaving group, which completes the catalysis. We have shown that one water molecule coordinating the Mg^{2+} ion shuttles a proton from K866 to the DNA O3' to form the final product. That observation was further supported through Constant pH MD simulations, indicating a pK_a of ~10 for K866, which makes it suitable for the protonation of the O3' leaving group (holding a pK_a of ~16). By further analyzing the polarization of the catalytic center, we showed that K866 polarizes the scissile phosphate, making it a better electrophile, and thereby creating a favorable electronic environment for catalysis. This critical role of K866 was further confirmed by DNA cleavage experiments, showing a remarkable reduction in enzymatic activity when K866 is mutated to alanine. In summary, ab-initio QM/MM simulations offered an in-depth understanding of the mechanisms of DNA cleavage in Cas9, which is valuable to increase the enzyme's catalytic efficiency and its metal-dependent function.

Finally, it is important to note that, despite its high intrinsic accuracy in producing targeted DNA breaks, Cas9 can cleave DNA sequences with defective complementarity to the guide RNA, leading to undesired off-target effects. Biophysical studies, employing single-molecule and kinetic experiments, showed that base-pair mismatches drastically affect Cas9 binding and catalysis. Computational investigations used the GaMD method to investigate the long-timescale conformational changes induced by base pair mismatches within the RNA:DNA hybrid of CRISPR-Cas9[150]. These simulations showed that mismatched pairs located at distal sites concerning the PAM recognition sequence induce an extended opening of the RNA:DNA hybrid (Figure 6D). This results in newly established interactions between the unwound RNA:DNA heteroduplex and the L2 loop of the HNH domain, ultimately restricting the conformational mobility of the HNH domain. Notably, a recent experimental study employing X-ray crystallography corroborated the observation of the RNA:DNA hybrid opening in the presence of three PAM distal mismatches[157]. Molecular simulations have also shown that base pair mismatches located at an upstream position within the RNA:DNA gets dynamically incorporated into the hybrid structure and has little impact on Cas9 interactions[157]. These observations were also consistent with the recently reported X-ray structures of Cas9 bound to several off-target substrates. Together, these insights provide a dynamical and structural rationale for the off-target activity of Cas9 and contribute to the rational design of guide RNAs and off-target prediction algorithms.

Ultimately, computational studies on the CRISPR-Cas9 system open the way for the use of molecular simulations to characterize a broad variety of emerging gene editing systems. Among them, the Cas12a system is highly promising for rapid nucleic acid detection. This genome editing system shares similarities as well as remarkable differences with Cas9 since a single catalytic domain, i.e., RuvC, is responsible for the cleavage of both DNA strands. It is unclear how the enzyme’s conformational changes would interplay with the nucleic acids to attain double-stranded DNA breaks. Multi-μs MD simulation revealed that DNA binding induces coupled dynamics of the REC2 and RuvC that prime the conformational transition of TS towards the catalytic site[158]. However, enhanced sampling simulations have recently revealed that an α-helical lid, located within the RuvC domain, also aids in the traversal of the TS by anchoring the crRNA:TS hybrid[159].

Best Practices to Model Protein/Nucleic Acid Complexes

The quality of computational measurements does not only depend on the robustness of the conformational sampling but also on the physically meaningful sampling of the conformational space; the latter depends on the accuracy of the potential energy function and associated parameters (i.e., force field) used for describing the biomolecular system. In comparison to the well-known protein force fields (e.g., OPLS[160], AMBER[161] GROMOS[162], and CHARMM[163]), there has been a recent considerable improvement in nucleic acid force field parameters, due to the rapidly increasing interest in gene machines at an atomic-level resolution.

Cornell et al.[164] developed the earlier major RNA force field parameter set ff94, which is the core of the AMBER force fields for RNAs. This model was further modified several times for torsional refinements; ff94 and its basic variants (ff98[165] and ff99[166], collectively termed as ff9X force fields) were refined to correct the overstabilization of the γ = trans backbone states, leading to the bsc0 refinement. In the AMBER simulation package, the latest available parameter set for RNA, (ff99bsc0)χ_{OL3}[167], is a refinement of the glycosidic torsion, χ, that describes the rotation of the bond linking the base to the sugar moiety. This is the most widely used force field for RNA simulations. Additional efforts have also been made to develop accurate RNA parameters for standard nucleotides and naturally occurring modifications compatible with CHARMM36 additive force field[168,169].

In the context of the AMBER force field, two sets of DNA parameters were proposed: (i) parmbsc1[170], which builds on bsc0 with modifications of the sugar pucker, the χ glycosidic torsion, the ε and ζ dihedrals, and (ii) OL15[171], combining the parm99, bsc0,

χ_{OL4}, $\varepsilon/\zeta_{OL1,}$ and $\beta_{OL1.}$ In-depth studies[172] of bsc1 and OL15 reported that both parameters increase the accuracy in representing averaged structures of DNA, compared to the earlier AMBER models; reporting an agreement within the sub-1 Å range in comparison to the average NMR structure. These parameter sets performed better than the CHARMM36 force field for DNA[173]. Recent developments[174] also focused on polarizable force fields, explicitly representing the electronic polarization induced by the surrounding electric field. Polarizable models for DNA have been able to stabilize both A- and B-forms of DNA on the microsecond timescale; while employing a polarizable force field to larger protein-nucleic acid complexes remains of difficult application due to the timescale limit. These models promise to significantly contribute to studies of protein/nucleic acid complexes, although their current scalability on modern HPC architectures is not computationally economical for multi-mega Dalton complexes.

Another key aspect for accurately modeling protein-nucleic acid complexes is the proper representation of ions (more importantly multivalent ions), as ions play a crucial role in stabilizing the folded structures of RNA/DNA, rearranging enzyme catalytic cores, catalyzing nucleic acid cleavage reactions, etc. Yoo et al.[175] reparametrized van der Waals interactions of monovalent (Li^+, Na^+, K^+) and divalent Mg^{+2} cations using biologically relevant anions (acetate and phosphate) as models, with the aim of simulating highly charged biomolecular complexes. Li and Merz[176] developed a 12–6–4 LJ-type nonbonded model for a total of 16 divalent metal ions by adding a $1/r^4$ term to the 12–6 model to represent ion-induced dipole interactions. Li et al.[177] also parameterized the Lennard-Jones terms for divalent ions compatible with Particle Mesh Ewald (PME) based simulations with different water models. In a recent work by Nierzwicki et al.[146], the Li and Merz model for Mg^{+2} ions in a classical simulation environment has shown excellent agreement with QM/MM simulation. Panteva et al.[178] fine-tuned the pairwise parameters for Mg^{2+}, Mn^{2+}, Zn^{2+}, and Cd^{2+} binding with dimethyl phosphate, adenosine, and guanosine to reproduce experimental site-specific binding free energies. Their ‘m12−6−4’ parameter set has satisfactorily reproduced metal ion migration to crystallographic binding sites of ribozymes. In the plethora[179–184] of several other efforts made to make reliable descriptions of ion-based interactions in biomolecular systems, careful consideration is needed in choosing a specific parameter set for the specific system under investigation[185].

Accurate modeling at the atomistic level provides unprecedented opportunities to track the dynamics of machines on genes, enabling an understanding of the functional regulations from the molecular mechanics (MM) principles. Nevertheless, studying catalytic reactions, which are at the core of these macromolecules’ functionality, is beyond the capabilities of MM. Describing the reaction mechanisms is possible by combining quantum mechanics (QM) and by integrating the electronic degrees of freedom over time. However, the holistic representation of biological complexes at a QM level is computationally intangible, with representation from the first quantum mechanical principles (i.e., ab-initio) only limited to a few hundred atoms. This problem is conventionally tackled by a hybrid QM/MM approach[186], where the enzyme active sites are treated at a QM level of theory, whereas the rest of the system is described at the classical MM level[187]. The accuracy and feasibility of performing reliable QM/MM simulations depend on a judicious choice of the theory describing the electronic structure, deciding on the QM region, and subsequently on the treatment of the QM/MM boundary. While different levels of theory can describe electronic structure at varying accuracy, a good trade-off between accuracy and computational cost is reached with density functional theory (DFT)[188]. DFT-based QM/MM studies generally enable to study QM regions of hundreds of atoms, offering sampling is in the range of 10 – 100ps[5]. Though the choice of the QM region depends on the experimental knowledge of the reaction (i.e., the crucial residues, water molecules and ions are involved, etc.), it is important to find the optimal QM to MM ratio[5]. Indeed, increasing the QM region would result in diminishing sampling of the relevant phase space, in turn decreasing the accuracy of the QM calculation. The sampling issue can be solved (in part) through the application of rare-event sampling methods like the 'blue moon ensemble'[153], 'metadynamics'[189], etc. Another extremely crucial aspect to minimize errors is the definition of the QM/MM boundary. It is generally advisable to define the QM/MM boundary as far as possible from the QM core[5], while also not expanding too much the QM sample space. The QM/MM boundary should preferably be placed where the densities of QM and MM atoms are less overlapping, e.g., a non-bonded boundary. However, if a QM/MM boundary intersects a covalent bond, boundary pseudopotentials, or capping hydrogens, are employed to saturate the atoms' valence. Once the QM region is properly designed as advised above, the dynamics of the electronic structure can be investigated by integrating the equation of motion through the Born–Oppenheimer (BOMD)[190] or Car–Parrinello molecular dynamics (CPMD)[191]. In BOMD, wave function is optimized at each time step of the dynamics, which makes it computationally demanding for a large number of QM atoms. Instead, in CPMD, the electronic degrees of freedom are treated as 'fictitious' dynamic variables, propagated on the BO surface (i.e., the ground state) during time integration, which does not need for a wave function optimization at each step of the simulation as in BOMD. CPMD is therefore preferred for large biomolecular systems, speeding up the dynamics considerably. BOMD remains generally used to properly equilibrate the electronic structure prior to CPMD simulations. Despite the extensive use of QM/MM simulations, extensive research is ongoing to increase scalability, accessing larger timescales and expanding the QM representation.

Conclusions and Perspectives

Despite several challenges, including (but not limited to) sampling the biologically relevant functional timescales for large biomolecular complexes, multiscale simulation approaches have contributed significantly to the understanding of gene processing machines at an unprecedented resolution. This review article summarizes prominent computational contributions aimed at characterizing the dynamics and mechanisms of machines on genes. We discussed nucleosome dynamics and its role in DNA packaging, and subsequently the dynamics of helicases and DNA polymerases in the pursuit of the replication and repair mechanisms. We further outlined how molecular simulations have been critical to understand the dynamics of RNA polymerases and transcription factors crucial for transcription machinery, while also reporting on the dynamics of ribosomal complexes crucial for protein synthesis. We also reported studies on spliceosome dynamics, and we dig into the chemical mechanism of RNA splicing, which is at the basis of mRNA maturation. Finally, we discussed how molecular simulations contributed to understanding the molecular mechanism of CRISPR-Cas gene editing tools. These studies employed a broad spectrum of computational methods spanning from classical and ab-initio molecular dynamics to enhanced sampling techniques and coarse-grained methods to access wide-ranging spatiotemporal scales.

With ever-increasing successes, the biggest challenges of MD simulations remain the accuracy of force fields and the high computational demand, which is significant for large biomolecular complexes. Inaccuracies in the force fields become more apparent as the simulation reaches larger timescales and/or with the system's increasing complexity. Nearly, 40+ years of careful refinement of empirical/additive force fields (FF)[192] led to increasingly realistic parameters, including advancements in polarizable models like Drude[193] and AMOEBA,[194] which promise to significantly contribute to studies of protein/nucleic acid complexes, although their current scalability on modern HPC architectures is not computationally economical for multi-mega Dalton complexes. Furthermore, a constant disparity between biological and simulation timescales is prevalent for large biomolecular systems. Simulations of millisecond timescale events might require years of continuous runs, depending on the size of the system and the computational architecture. Molecular simulations of such complex systems could also be impacted by the lack of convergence of relevant features. Studies have shown that the distributions representing specific properties of large biomolecules do not converge on currently measurable timescales[4]. This shows the requirement of carefully assessing the convergence of the system's properties while carrying out and interpreting large-scale simulations.

Overall, innovative applications summarized here reveal the instrumental role of computational methods in filling the knowledge gap between experimentally obtained macroscopic observables and the molecular guiding principles occurring at the microscopic level. In the forthcoming future, a cross-disciplinary effort, including advancements in deep learning-based modeling and enhanced sampling strategies, advanced GPU-enabled implementation of sampling methods, dedicated HPC architectures for MD simulation, will be critical to characterize these intricate macromolecular machines at a near-realistic spatiotemporal scale. We are hopeful of an exciting time ahead where a well-structured synergy between computational and experimental efforts will expand precise gene editing applications for novel therapeutics.

Acknowledgments

This material is based upon work supported by the National Institute of Health (Grant No. R01GM141329, to GP) and the National Science Foundation (Grant No. CHE-1905374 and CHE-2144823, to GP). Part of this work used Expanse at the San Diego Supercomputing Center through allocation MCB160059 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. Computer time was also provided by National Energy Research Scientific Computing Center (NERSC) under Grant No. M3807.

References

(1). Von Hippel PH From "Simple" DNA-Protein Interactions to the Macromolecular Machines of Gene Expression. Annu. Rev. Biophys. Biomol. Struc 2007, 36, 79–105.

(2). Lee EH; Hsin J; Sotomayor M; Comellas G; Schulten K Discovery Through the Computational Microscope. Structure 2009, 17 (10), 1295–1306. [PubMed: 19836330]

(3). Perilla JR; Goh BC; Cassidy CK; Liu B; Bernardi RC; Rudack T; Yu H; Wu Z; Schulten K Molecular Dynamics Simulations of Large Macromolecular Complexes. Curr. Opin. Struct. Biol 2015, 31, 64–74. [PubMed: 25845770]

(4). Gupta C; Sarkar D; Tieleman DP; Singharoy A The Ugly, Bad, and Good Stories of Large-Scale Biomolecular Simulations. Curr. Opin. Struct. Biol 2022, 73, 102338. [PubMed: 35245737]

(5). Brunk E; Rothlisberger U Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. Chem. Rev 2015, 115 (12), 6217–6263. [PubMed: 25880693]

(6). Jin J; Pak AJ; Durumeric AEP; Loose TD; Voth GA Bottom-up Coarse-Graining: Principles and Perspectives. J. Chem. Theory Comput 2022, 18 (10), 5759–5791. [PubMed: 36070494]

(7). Hénin J; Lelièvre T; Shirts MR; Valsson O; Delemotte L Enhanced Sampling Methods for Molecular Dynamics Simulations. Living J. Comput. Mol. Sci 2022, 4 (1), 1583.

(8). Wang Y; Lamim Ribeiro JM; Tiwary P Machine Learning Approaches for Analyzing and Enhancing Molecular Dynamics Simulations. Curr. Opin. Struct. Biol 2020, 61, 139–145. [PubMed: 31972477]

(9). Ou HD; Phan S; Deerinck TJ; Thor A; Ellisman MH; O'Shea CC ChromEMT: Visualizing 3D Chromatin Structure and Compaction in Interphase and Mitotic Cells. Science 2017, 357 (6349), eaag0025. [PubMed: 28751582]

(10). Ohno M; Ando T; Priest DG; Kumar V; Yoshida Y; Taniguchi Y Sub-Nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs. Cell 2019, 176 (3), 520–534.e25. [PubMed: 30661750]

(11). Robinson PJJ; Fairall L; Huynh VAT; Rhodes D EM Measurements Define the Dimensions of the "30-Nm" Chromatin Fiber: Evidence for a Compact, Interdigitated Structure. Proc. Natl. Acad. Sci. U. S. A 2006, 103 (17), 6506–6511. [PubMed: 16617109]

(12). Song F; Chen P; Sun D; Wang M; Dong L; Liang D; Xu RM; Zhu P; Li G Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. Science 2014, 344 (6182), 376–380. [PubMed: 24763583]

(13). Baldi S; Korber P; Becker PB Beads on a String—Nucleosome Array Arrangements and Folding of the Chromatin Fiber. Nat. Struct. Mol. Biol 2020, 27 (2), 109–118. [PubMed: 32042149]

(14). Kono H; Sakuraba S; Ishida H Free Energy Profiles for Unwrapping the Outer Superhelical Turn of Nucleosomal DNA. PLOS Comput. Biol 2018, 14 (3), e1006024. [PubMed: 29505570]

(15). Öztürk MA; Pachov GV; Wade RC; Cojocaru V Conformational Selection and Dynamic Adaptation upon Linker Histone Binding to the Nucleosome. Nucleic Acids Res 2016, 44 (14), 6599–6613. [PubMed: 27270081]

(16). Huertas J; Schöler HR; Cojocaru V Histone Tails Cooperate to Control the Breathing of Genomic Nucleosomes. PLOS Comput. Biol 2021, 17 (6), e1009013. [PubMed: 34081696]

(17). Woods DC; Rodríguez-Ropero F; Wereszczynski J The Dynamic Influence of Linker Histone Saturation within the Poly-Nucleosome Array. J. Mol. Biol 2021, 433 (10), 166902. [PubMed: 33667509]

(18). Huertas J; MacCarthy CM; Schöler HR; Cojocaru V Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding. Biophys. J 2020, 118 (9), 2280–2296. [PubMed: 32027821]

(19). Bowerman S; Wereszczynski J Effects of MacroH2A and H2A.Z on Nucleosome Dynamics as Elucidated by Molecular Dynamics Simulations. Biophys. J 2016, 110 (2), 327–337. [PubMed: 26789756]

(20). Tan C; Takada S Nucleosome Allostery in Pioneer Transcription Factor Binding. Proc. Natl. Acad. Sci. U. S. A 2020, 117 (34), 20586–20596. [PubMed: 32778600]

(21). Adhireksan Z; Palermo G; Riedel T; Ma Z; Muhammad R; Rothlisberger U; Dyson PJ; Davey CA Allosteric Cross-Talk in Chromatin Can Mediate Drug-Drug Synergy. Nat. Commun 2017, 8 (1), 1–11. [PubMed: 28232747]

(22). Palermo G; Magistrato A; Riedel T; von Erlach T; Davey CA; Dyson PJ; Rothlisberger U Fighting Cancer with Transition Metal Complexes: From Naked DNA to Protein and Chromatin Targeting Strategies. ChemMedChem 2016, 11 (12), 1199–1210. [PubMed: 26634638]

(23). Ma Z; Palermo G; Adhireksan Z; Murray BS; von Erlach T; Dyson PJ; Rothlisberger U; Davey CA An Organometallic Compound Which Exhibits a DNA Topology-Dependent One-Stranded Intercalation Mode. Angew. Chemie Int. Ed 2016, 55 (26), 7441–7444.

(24). Huertas J; Cojocaru V Breaths, Twists, and Turns of Atomistic Nucleosomes. J. Mol. Biol 2021, 433 (6), 166744. [PubMed: 33309853]

(25). Marin-Gonzalez A; Vilhena JG; Moreno-Herrero F; Perez R DNA Crookedness Regulates DNA Mechanical Properties at Short Length Scales. Phys. Rev. Lett 2019, 122 (4), 048102. [PubMed: 30768347]

(26). Elbahnsi A; Retureau R; Baaden M; Hartmann B; Oguey C Holding the Nucleosome Together: A Quantitative Description of the DNA-Histone Interface in Solution. J. Chem. Theory Comput 2018, 14 (2), 1045–1058. [PubMed: 29262675]

(27). Garai A; Saurabh S; Lansac Y; Maiti PK DNA Elasticity from Short DNA to Nucleosomal DNA. J. Phys. Chem. B 2015, 119 (34), 11146–11156. [PubMed: 26134918]

(28). Sun R; Li Z; Bishop TC TMB Library of Nucleosome Simulations. J. Chem. Inf. Model 2019, 59 (10), 4289–4299. [PubMed: 31490686]

(29). Erler J; Zhang R; Petridis L; Cheng X; Smith JC; Langowski J The Role of Histone Tails in the Nucleosome: A Computational Study. Biophys. J 2014, 107 (12), 2911–2922. [PubMed: 25517156]

(30). Zheng Y; Cui Q The Histone H3 N-Terminal Tail: A Computational Analysis of the Free Energy Landscape and Kinetics. Phys. Chem. Chem. Phys 2015, 17 (20), 13689–13698. [PubMed: 25942635]

(31). Collepardo-Guevara R; Portella G; Vendruscolo M; Frenkel D; Schlick T; Orozco M Chromatin Unfolding by Epigenetic Modifications Explained by Dramatic Impairment of Internucleosome Interactions: A Multiscale Computational Study. J. Am. Chem. Soc 2015, 137 (32), 10205–10215. [PubMed: 26192632]

(32). du Preez LL; Patterton HG The Effect of Epigenetic Modifications on the Secondary Structures and Possible Binding Positions of the N-Terminal Tail of Histone H3 in the Nucleosome: A Computational Study. J. Mol. Model 2017, 23 (4), 1–15. [PubMed: 27921184]

(33). Saikusa K; Fuchigami S; Takahashi K; Asano Y; Nagadoi A; Tachiwana H; Kurumizaka H; Ikeguchi M; Nishimura Y; Akashi S Gas-Phase Structure of the Histone Multimers Characterized by Ion Mobility Mass Spectrometry and Molecular Dynamics Simulation. Anal. Chem 2013, 85 (8), 4165–4171. [PubMed: 23485128]

(34). Winogradoff D; Echeverria I; Potoyan DA; Papoian GA The Acetylation Landscape of the H4 Histone Tail: Disentangling the Interplay between the Specific and Cumulative Effects. J. Am. Chem. Soc 2015, 137 (19), 6245–6253. [PubMed: 25905561]

(35). Gatchalian J; Wang X; Ikebe J; Cox KL; Tencer AH; Zhang Y; Burge NL; Di L; Gibson MD; Musselman CA; Poirier MG; Kono H; Hayes JJ; Kutateladze TG Accessibility of the Histone H3 Tail in the Nucleosome for Binding of Paired Readers. Nat. Commun 2017, 8 (1), 1–10. [PubMed: 28232747]

(36). Shaytan AK; Armeev GA; Goncearenco A; Zhurkin VB; Landsman D; Panchenko AR Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions. J. Mol. Biol 2016, 428 (1), 221–237. [PubMed: 26699921]

(37). Zhang R; Erler J; Langowski J Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-Nucleosome Interaction. Biophys. J 2017, 112 (3), 450–459. [PubMed: 27931745]

(38). Ishida H; Kono H H4 Tails Potentially Produce the Diversity in the Orientation of Two Nucleosomes. Biophys. J 2017, 113 (5), 978–990. [PubMed: 28877499]

(39). Saurabh S; Glaser MA; Lansac Y; Maiti PK Atomistic Simulation of Stacked Nucleosome Core Particles: Tail Bridging, the H4 Tail, and Effect of Hydrophobic Forces. J. Phys. Chem. B 2016, 120 (12), 3048–3060. [PubMed: 26931280]

(40). Luger K; Mäder AW; Richmond RK; Sargent DF; Richmond TJ Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. Nature 1997, 389 (6648), 251–260. [PubMed: 9305837]

(41). Alvarez-Venegas R; Avramova Z Methylation Patterns of Histone H3 Lys 4, Lys 9 and Lys 27 in Transcriptionally Active and Inactive Arabidopsis Genes and in Atx1 Mutants. Nucleic Acids Res 2005, 33 (16), 5199–5207. [PubMed: 16157865]

(42). Nurse NP; Jimenez-Useche I; Smith IT; Yuan C Clipping of Flexible Tails of Histones H3 and H4 Affects the Structure and Dynamics of the Nucleosome. Biophys. J 2013, 104 (5), 1081–1088. [PubMed: 23473491]

(43). Gadea FCM; Nikolova EN Nucleosome Topology and DNA Sequence Modulate the Engagement of Pioneer Factors SOX2 and OCT4. bioRxiv 2022, 2022.01.18.476780.

(44). Michael AK; Grand RS; Isbel L; Cavadini S; Kozicka Z; Kempf G; Bunker RD; Schenk AD; Graff-Meyer A; Pathare GR; Weiss J; Matsumoto S; Burger L; Schübeler D; Thomä NH Mechanisms of OCT4-SOX2 Motif Readout on Nucleosomes. Science 2020, 368 (6498), 1460–1465. [PubMed: 32327602]

(45). Liu H; Shi Y; Chen XS; Warshel A Simulating the Electrostatic Guidance of the Vectorial Translocations in Hexameric Helicases and Translocases. Proc. Natl. Acad. Sci. U. S. A 2009, 106 (18), 7449–7454. [PubMed: 19383795]

(46). Lu W; Bueno C; Schafer NP; Moller J; Jin S; Chen X; Chen M; Gu X; Davtyan A; de Pablo JJ; Wolynes PG OpenAWSEM with Open3SPN2: A Fast, Flexible, and Accessible Framework for Large-Scale Coarse-Grained Biomolecular Simulations. PLOS Comput. Biol 2021, 17 (2), e1008308. [PubMed: 33577557]

(47). Jin S; Bueno C; Lu W; Wang Q; Chen M; Chen X; Wolynes PG; Gao Y Computationally Exploring the Mechanism of Bacteriophage T7 Gp4 Helicase Translocating along SsDNA. Proc. Natl. Acad. Sci. U. S. A 2022, 119 (32), e2202239119. [PubMed: 35914145]

(48). Pan AC; Sezer D; Roux B Finding Transition Pathways Using the String Method with Swarms of Trajectories. J. Phys. Chem. B 2008, 112 (11), 3432–3440. [PubMed: 18290641]

(49). Faradjian AK; Elber R Computing Time Scales from Reaction Coordinates by Milestoning. J. Chem. Phys 2004, 120 (23), 10880–10889. [PubMed: 15268118]

(50). Ma W; Schulten K Mechanism of Substrate Translocation by a Ring-Shaped ATPase Motor at Millisecond Resolution. J. Am. Chem. Soc 2015, 137 (8), 3031–3040. [PubMed: 25646698]

(51). Colizzi F; Perez-Gonzalez C; Fritzen R; Levy Y; White MF; Carlos Penedo J; Bussi G Asymmetric Base-Pair Opening Drives Helicase Unwinding Dynamics. Proc. Natl. Acad. Sci. U. S. A 2019, 116 (45), 22471–22477. [PubMed: 31628254]

(52). Dodd T; Botto M; Paul F; Fernandez-Leiro R; Lamers MH; Ivanov I Polymerization and Editing Modes of a High-Fidelity DNA Polymerase Are Linked by a Well-Defined Path. Nat. Commun 2020, 11 (1), 1–11. [PubMed: 31911652]

(53). Roston D; Demapan D; Cui Q Extensive Free-Energy Simulations Identify Water as the Base in Nucleotide Addition by DNA Polymerase. Proc. Natl. Acad. Sci. U. S. A 2019, 116 (50), 25048–25056. [PubMed: 31757846]

(54). Lu X; Gaus M; Elstner M; Cui Q Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications. J. Phys. Chem. B 2015, 119 (3), 1062–1082. [PubMed: 25178644]

(55). de Alba Ortíz AP; Vreede J; Ensing B Sequence Dependence of Transient Hoogsteen Base Pairing in DNA. PLOS Comput. Biol 2022, 18 (5), e1010113. [PubMed: 35617357]

(56). Mondal S; Mukherjee S; Acharya S; Bagchi B Unfolding of Dynamical Events in the Early Stage of Insulin Dimer Dissociation. J. Phys. Chem. B 2021, 125 (29), 7958–7966. [PubMed: 34260242]

(57). Raiteri P; Laio A; Gervasio FL; Micheletti C; Parrinello M Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. J. Phys. Chem. B 2006, 110 (8), 3533–3539. [PubMed: 16494409]

(58). Laio A; Parrinello M Escaping Free-Energy Minima. Proc. Natl. Acad. Sci. U. S. A 2002, 99 (20), 12562–12566. [PubMed: 12271136]

(59). Cisneros GA; Perera L; García-Díaz M; Bebenek K; Kunkel TA; Pedersen LG Catalytic Mechanism of Human DNA Polymerase λ with Mg2+ and Mn2+ from Ab Initio Quantum Mechanical/Molecular Mechanical Studies. DNA Repair (Amst) 2008, 7 (11), 1824–1834. [PubMed: 18692600]

(60). Li Y; Freudenthal BD; Beard WA; Wilson SH; Schlick T Optimal and Variant Metal-Ion Routes in DNA Polymerase β's Conformational Pathways. J. Am. Chem. Soc 2014, 136 (9), 3630–3639. [PubMed: 24511902]

(61). Yang L; Beard WA; Wilson SH; Broyde S; Schlick T Polymerase β Simulations Suggest That Arg258 Rotation Is a Slow Step Rather than Large Subdomain Motions per Se. J. Mol. Biol 2002, 317 (5), 651–671. [PubMed: 11955015]

(62). Yang L; Beard WA; Wilson SH; Roux B; Broyde S; Schlick T Local Deformations Revealed by Dynamics Simulations of DNA Polymerase β with DNA Mismatches at the Primer Terminus. J. Mol. Biol 2002, 321 (3), 459–478. [PubMed: 12162959]

(63). Yang L; Beard WA; Wilson SH; Broyde S; Schlick T Highly Organized but Pliant Active Site of DNA Polymerase β: Compensatory Mechanisms in Mutant Enzymes Revealed by Dynamics Simulations and Energy Analyses. Biophys. J 2004, 86 (6), 3392–3408. [PubMed: 15189842]

(64). Yang L; Arora K; Beard WA; Wilson SH; Schlick T Critical Role of Magnesium Ions in DNA Polymerase β's Closing and Active Site Assembly. J. Am. Chem. Soc 2004, 126 (27), 8441–8453. [PubMed: 15238001]

(65). Radhakrishnan R; Schlick T Fidelity Discrimination in DNA Polymerase β: Differing Closing Profiles for a Mismatched (G:A) versus Matched (G:C) Base Pair. J. Am. Chem. Soc 2005, 127 (38), 13245–13252. [PubMed: 16173754]

(66). Andrés Cisneros G; Perera L; Sehaaper RM; Pedersen LC; London RE; Pedersen LG; Darden TA Reaction Mechanism of the ε Subunit of E. Coli DNA Polymerase III: Insights into Active Site Metal Coordination and Catalytically Significant Residues. J. Am. Chem. Soc 2009, 131 (4), 1550–1556. [PubMed: 19119875]

(67). Genna V; Vidossich P; Ippoliti E; Carloni P; De Vivo M . A Self-Activated Mechanism for Nucleic Acid Polymerization Catalyzed by DNA/RNA Polymerases. J. Am. Chem. Soc 2016, 138 (44), 14592–14598. [PubMed: 27530537]

(68). Kim T; Freudenthal BD; Beard WA; Wilson SH; Schlick T Insertion of Oxidized Nucleotide Triggers Rapid DNA Polymerase Opening. Nucleic Acids Res 2016, 44 (9), 4409–4424. [PubMed: 27034465]

(69). Kirmizialtin S; Nguyen V; Johnson KA; Elber R How Conformational Dynamics of DNA Polymerase Select Correct Substrates: Experiments and Simulations. Structure 2012, 20 (4), 618–627. [PubMed: 22483109]

(70). Meli M; Sustarsic M; Craggs TD; Kapanidis AN; Colombo G DNA Polymerase Conformational Dynamics and the Role of Fidelity-Conferring Residues: Insights from Computational Simulations. Front. Mol. Biosci 2016, 3, 20. [PubMed: 27303671]

(71). Bergonzo C; Campbell AJ; Walker RC; Simmerling C A Partial Nudged Elastic Band Implementation for Use with Large or Explicitly Solvated Systems. Int. J. Quantum Chem 2009, 109 (15), 3781–3790. [PubMed: 20148191]

(72). Ghoreishi D; Cerutti DS; Fallon Z; Simmerling C; Roitberg AE Fast Implementation of the Nudged Elastic Band Method in AMBER. J. Chem. Theory Comput 2019, 15 (8), 4699–4707. [PubMed: 31314523]

(73). Geronimo I; Vidossich P; De Vivo M Local Structural Dynamics at the Metal-Centered Catalytic Site of Polymerases Is Critical for Fidelity. ACS Catal 2021, 11 (22), 14110–14121.

(74). Papasaikas P; Valcárcel J The Spliceosome: The Ultimate RNA Chaperone and Sculptor. Trends Biochem. Sci 2016, 41 (1), 33–45. [PubMed: 26682498]

(75). Chow LT; Gelinas RE; Broker TR; Roberts RJ An Amazing Sequence Arrangement at the 5′ Ends of Adenovirus 2 Messenger RNA. Cell 1977, 12 (1), 1–8. [PubMed: 902310]

(76). Wan R; Bai R; Shi Y Molecular Choreography of Pre-mRNA Splicing by the Spliceosome. Curr. Opin. Struct. Biol 2019, 59, 124–133. [PubMed: 31476650]

(77). Sponer J; Bussi G; Krepl M; Banas P; Bottaro S; Cunha RA; Gil-Ley A; Pinamonti G; Poblete S; Jurečka P; Walter NG; Otyepka M RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview. Chem. Rev 2018, 118 (8), 4177–4338. [PubMed: 29297679]

(78). Fica SM; Montoya G; Carlomagno T Cryo-EM Snapshots of the Human Spliceosome Reveal Structural Adaptions for Splicing Regulation Curr. Opin. Struct. Biol 2020, 65, 139–148. [PubMed: 32717639]

(79). Galej WP; Toor N; Newman AJ; Nagai K Molecular Mechanism and Evolution of Nuclear Pre-mRNA and Group II Intron Splicing: Insights from Cryo-Electron Microscopy Structures. Chem. Rev 2018, 118 (8), 4156–4176. [PubMed: 29377672]

(80). Palermo G; Casalino L; Magistrato A; Andrew McCammon J Understanding the Mechanistic Basis of Non-Coding RNA through Molecular Dynamics Simulations. J. Struct. Biol 2019, 206 (3), 267–279. [PubMed: 30880083]

(81). Casalino L; Palermo G; Spinello A; Rothlisberger U; Magistrato A All-Atom Simulations Disentangle the Functional Dynamics Underlying Gene Maturation in the Intron Lariat Spliceosome. Proc. Natl. Acad. Sci. U. S. A 2018, 115 (26), 6584–6589. [PubMed: 29891649]

(82). Zhang X; Zhan X; Yan C; Zhang W; Liu D; Lei J; Shi Y Structures of the Human Spliceosomes before and after Release of the Ligated Exon. Cell Res 2019, 29 (4), 274–285. [PubMed: 30728453]

(83). Yan C; Hang J; Wan R; Huang M; Wong CCL; Shi Y Structure of a Yeast Spliceosome at 3.6-Angstrom Resolution. Science 2015, 349 (6253), 1182–1191. [PubMed: 26292707]

(84). Marcia M; Pyle AM Visualizing Group II Intron Catalysis through the Stages of Splicing. Cell 2012, 151 (3), 497–507. [PubMed: 23101623]

(85). Saltalamacchia A; Casalino L; Borišek J; Batista VS; Rivalta I; Magistrato A Decrypting the Information Exchange Pathways across the Spliceosome Machinery. J. Am. Chem. Soc 2020, 142 (18), 8403–8411. [PubMed: 32275149]

(86). Brunk E; Rothlisberger U Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. Chem. Rev 2015, 115 (12), 6217–6263. [PubMed: 25880693]

(87). Casalino L; Palermo G; Rothlisberger U; Magistrato A Who Activates the Nucleophile in Ribozyme Catalysis? An Answer from the Splicing Mechanism of Group II Introns. J. Am. Chem. Soc 2016, 138 (33), 10374–10377. [PubMed: 27309711]

(88). Kumar A; Satpati P Structure-Based Thermodynamics of Ion Selectivity (Mg2+ versus Ca2+ and K+ versus Na+) in the Active Site of the Eukaryotic Lariat Group II Intron from Algae Pylaiella Littoralis. Phys. Chem. Chem. Phys 2022, 24 (39), 24192–24202. [PubMed: 36168989]

(89). Casalino L; Palermo G; Abdurakhmonova N; Rothlisberger U; Magistrato A Development of Site-Specific Mg2+-RNA Force Field Parameters: A Dream or Reality? Guidelines from Combined Molecular Dynamics and Quantum Mechanics Simulations. J. Chem. Theory Comput 2017, 13 (1), 340–352. [PubMed: 28001405]

(90). Huang W; Huang Y; Xu J; Liao J Lou. How Does the Spliceosome Catalyze Intron Lariat Formation? Insights from Quantum Mechanics/Molecular Mechanics Free-Energy Simulations. J. Phys. Chem. B 2019, 123 (28), 6049–6055. [PubMed: 31267745]

(91). Bai R; Yan C; Wan R; Lei J; Shi Y Structure of the Post-Catalytic Spliceosome from Saccharomyces Cerevisiae. Cell 2017, 171 (7), 1589–1598.e8. [PubMed: 29153833]

(92). Steitz TA; Steitz JA A General Two-Metal-Ion Mechanism for Catalytic RNA. Proc. Natl. Acad. Sci. U. S. A 1993, 90 (14), 6498–6502. [PubMed: 8341661]

(93). Palermo G; Cavalli A; Klein ML; Alfonso-Prieto M; Dal Peraro M; De Vivo M Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. Acc. Chem. Res 2015, 48 (2), 220–228. [PubMed: 25590654]

(94). Berta D; Buigues PJ; Badaoui M; Rosta E Cations in Motion: QM/MM Studies of the Dynamic and Electrostatic Roles of H+ and Mg2+ Ions in Enzyme Reactions. Curr. Opin. Struct. Biol 2020, 61, 198–206. [PubMed: 32065923]

(95). Bonnal SC; López-Oreja I; Valcárcel J Roles and Mechanisms of Alternative Splicing in Cancer — Implications for Care. Nat. Rev. Clin. Oncol 2020, 17 (8), 457–474. [PubMed: 32303702]

(96). Cretu C; Schmitzová J; Ponce-Salvatierra A; Dybkov O; De Laurentiis EI; Sharma K; Will CL; Urlaub H; Lührmann R; Pena V Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. Mol. Cell 2016, 64 (2), 307–319. [PubMed: 27720643]

(97). Patnaik MM; Lasho TL; Finke CM; Hanson CA; Hodnefield JM; Knudson RA; Ketterling RP; Pardanani A; Tefferi A Spliceosome Mutations Involving SRSF2, SF3B1, and U2AF35 in Chronic Myelomonocytic Leukemia: Prevalence, Clinical Correlates, and Prognostic Relevance. Am. J. Hematol 2013, 88 (3), 201–206. [PubMed: 23335386]

(98). Landau DA; Carter SL; Stojanov P; McKenna A; Stevenson K; Lawrence MS; Sougnez C; Stewart C; Sivachenko A; Wang L; Wan Y; Zhang W; Shukla SA; Vartanov A; Fernandes SM; Saksena G; Cibulskis K; Tesar B; Gabriel S; Hacohen N; Meyerson M; Lander ES; Neuberg D; Brown JR; Getz G; Wu CJ Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. Cell 2013, 152 (4), 714–726. [PubMed: 23415222]

(99). Cretu C; Agrawal AA; Cook A; Will CL; Fekkes P; Smith PG; Lührmann R; Larsen N; Buonamici S; Pena V Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds. Mol. Cell 2018, 70 (2), 265–273.e8. [PubMed: 29656923]

(100). Finci LI; Zhang X; Huang X; Zhou Q; Tsai J; Teng T; Agrawal A; Chan B; Irwin S; Karr C; Cook A; Zhu P; Reynolds D; Smith PG; Fekkes P; Buonamici S; Larsen NA The Cryo-EM Structure of the SF3b Spliceosome Complex Bound to a Splicing Modulator Reveals a Pre-mRNA Substrate Competitive Mechanism of Action. Genes Dev 2018, 32 (3–4), 309–320. [PubMed: 29491137]

(101). Borišek J; Saltalamacchia A; Spinello A; Magistrato A Exploiting Cryo-EM Structural Information and All-Atom Simulations to Decrypt the Molecular Mechanism of Splicing Modulators. J. Chem. Inf. Model 2020, 60 (5), 2510–2521. [PubMed: 31539251]

(102). Yan C; Dodd T; He Y; Tainer JA; Tsutakawa SE; Ivanov I Transcription Preinitiation Complex Structure and Dynamics Provide Insight into Genetic Diseases. Nat. Struct. Mol. Biol 2019, 26 (6), 397–406. [PubMed: 31110295]

(103). Singharoy A; Teo I; McGreevy R; Stone JE; Zhao J; Schulten K Molecular Dynamics-Based Refinement and Validation for Sub-5 Å Cryo-Electron Microscopy Maps. Elife 2016, 5, e16105. [PubMed: 27383269]

(104). Dienemann C; Schwalb B; Schilbach S; Cramer P Promoter Distortion and Opening in the RNA Polymerase II Cleft. Mol. Cell 2019, 73 (1), 97–106.e4. [PubMed: 30472190]

(105). Noé F; Fischer S Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules. Curr. Opin. Struct. Biol 2008, 18 (2), 154–162. [PubMed: 18378442]

(106). Silva D-A; Weiss DR; Pardo Avila F; Da L-T; Levitt M; Wang D; Huang X Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution. Proc. Natl. Acad. Sci 2014, 111 (21), 7665–7670. [PubMed: 24753580]

(107). Da LT; Pardo-Avila F; Xu L; Silva DA; Zhang L; Gao X; Wang D; Huang X Bridge Helix Bending Promotes RNA Polymerase II Backtracking through a Critical and Conserved Threonine Residue. Nat. Commun 2016, 7 (1), 1–10.

(108). Weiss DR; Levitt M Can Morphing Methods Predict Intermediate Structures? J. Mol. Biol 2009, 385 (2), 665–674. [PubMed: 18996395]

(109). Dangkulwanich M; Ishibashi T; Liu S; Kireeva ML; Lubkowska L; Kashlev M; Bustamante CJ Complete Dissection of Transcription Elongation Reveals Slow Translocation of RNA Polymerase II in a Linear Ratchet Mechanism. Elife 2013, 2, e00971. [PubMed: 24066225]

(110). Marrink SJ; Risselada HJ; Yefimov S; Tieleman DP; De Vries AH The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. J. Phys. Chem. B 2007, 111 (27), 7812–7824. [PubMed: 17569554]

(111). Monticelli L; Kandasamy SK; Periole X; Larson RG; Tieleman DP; Marrink SJ The MARTINI Coarse-Grained Force Field: Extension to Proteins. J. Chem. Theory Comput 2008, 4 (5), 819–834. [PubMed: 26621095]

(112). Torrie GM; Valleau JP Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. J. Comput. Phys 1977, 23 (2), 187–199.

(113). Eun C; Ortiz-Sánchez JM; Da L; Wang D; McCammon JA Molecular Dynamics Simulation Study of Conformational Changes of Transcription Factor TFIIS during RNA Polymerase II Transcriptional Arrest and Reactivation. PLoS One 2014, 9 (5), e97975. [PubMed: 24842057]

(114). Schmeing TM; Ramakrishnan V What Recent Ribosome Structures Have Revealed about the Mechanism of Translation. Nature 2009, 461 (7268), 1234–1242. [PubMed: 19838167]

(115). Frank J The Translation Elongation Cycle—Capturing Multiple States by Cryo-Electron Microscopy. Philos. Trans. R. Soc. B Biol. Sci 2017, 372 (1716), 20160180

(116). Sanbonmatsu KY; Joseph S; Tung CS Simulating Movement of tRNA into the Ribosome during Decoding. Proc. Natl. Acad. Sci. U. S. A 2005, 102 (44), 15854–15859. [PubMed: 16249344]

(117). Lind C; Åqvist J Principles of Start Codon Recognition in Eukaryotic Translation Initiation. Nucleic Acids Res 2016, 44 (17), 8425–8432. [PubMed: 27280974]

(118). Satpati P; Sund J; Åqvist J Structure-Based Energetics of MRNA Decoding on the Ribosome. Biochemistry 2014, 53 (10), 1714–1722. [PubMed: 24564511]

(119). Zeng X; Chugh J; Casiano-Negroni A; Al-Hashimi HM; Brooks CL Flipping of the Ribosomal A-Site Adenines Provides a Basis for TRNA Selection. J. Mol. Biol 2014, 426 (19), 3201–3213. [PubMed: 24813122]

(120). Panecka J; Mura C; Trylska J Interplay of the Bacterial Ribosomal A-Site, S12 Protein Mutations and Paromomycin Binding: A Molecular Dynamics Study. PLoS One 2014, 9 (11), e111811. [PubMed: 25379961]

(121). Sothiselvam S; Liu B; Han W; Ramu H; Klepacki D; Atkinson GC; Brauer A; Remm M; Tenson T; Schulten K; Vaźquez-Laslop N; Mankin AS Macrolide Antibiotics Allosterically Predispose the Ribosome for Translation Arrest. Proc. Natl. Acad. Sci. U. S. A 2014, 111 (27), 9804–9809. [PubMed: 24961372]

(122). Small MC; Lopes P; Andrade RB; MacKerell AD Impact of Ribosomal Modification on the Binding of the Antibiotic Telithromycin Using a Combined Grand Canonical Monte Carlo/Molecular Dynamics Simulation Approach. PLOS Comput. Biol 2013, 9 (6), e1003113. [PubMed: 23785274]

(123). Aleksandrov A; Simonson T Molecular Dynamics Simulations of the 30S Ribosomal Subunit Reveal a Preferred Tetracycline Binding Site. J. Am. Chem. Soc 2008, 130 (4), 1114–1115. [PubMed: 18219717]

(124). Romanowska J; McCammon JA; Trylska J Understanding the Origins of Bacterial Resistance to Aminoglycosides through Molecular Dynamics Mutational Study of the Ribosomal A-Site. PLOS Comput. Biol 2011, 7 (7), e1002099. [PubMed: 21814503]

(125). Whitford PC; Blanchard SC; Cate JHD; Sanbonmatsu KY Connecting the Kinetics and Energy Landscape of tRNA Translocation on the Ribosome. PLOS Comput. Biol 2013, 9 (3), e1003003. [PubMed: 23555233]

(126). Munro JB; Altman RB; Tung CS; Sanbonmatsu KY; Blanchard SC A Fast Dynamic Mode of the EF-G-Bound Ribosome. EMBO J 2010, 29 (4), 770. [PubMed: 20033061]

(127). Khade PK; Joseph S Messenger RNA Interactions in the Decoding Center Control the Rate of Translocation. Nat. Struct. Mol. Biol 2011, 18 (11), 1300–1302. [PubMed: 22020300]

(128). Bock LV; Blau C; Schröder GF; Davydov II; Fischer N; Stark H; Rodnina MV; Vaiana AC; Grubmüller H Energy Barriers and Driving Forces in tRNA Translocation through the Ribosome. Nat. Struct. Mol. Biol 2013, 20 (12), 1390–1396. [PubMed: 24186064]

(129). Whitford PC; Noel JK; Gosavi S; Schug A; Sanbonmatsu KY; Onuchic JN An All-Atom Structure-Based Potential for Proteins: Bridging Minimal Models with All-Atom Empirical Forcefields. Proteins Struct. Funct. Bioinforma 2009, 75 (2), 430–441.

(130). Nguyen K; Whitford PC Steric Interactions Lead to Collective Tilting Motion in the Ribosome during mRNA–tRNA Translocation. Nat. Commun 2016, 7 (1), 1–9.

(131). Lind C; Sund J; Åqvist J Codon-Reading Specificities of Mitochondrial Release Factors and Translation Termination at Non-Standard Stop Codons. Nat. Commun 2013, 4, 2940. [PubMed: 24352605]

(132). Guzel P; Yildirim HZ; Yuce M; Kurkcuoglu O Exploring Allosteric Signaling in the Exit Tunnel of the Bacterial Ribosome by Molecular Dynamics Simulations and Residue Network Model. Front. Mol. Biosci 2020, 7, 586075. [PubMed: 33102529]

(133). Doudna JA; Charpentier E The New Frontier of Genome Engineering with CRISPR-Cas9. Science 2014, 346 (6213), 1258096. [PubMed: 25430774]

(134). Hsu PD; Lander ES; Zhang F Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 2014, 157 (6), 1262–1278. [PubMed: 24906146]

(135). Chen JS; Doudna JA The Chemistry of Cas9 and Its CRISPR Colleagues. Nat. Rev. Chem 2017, 1 (10), 1–15.

(136). Saha A; Arantes PR; Palermo G Dynamics and Mechanisms of CRISPR-Cas9 through the Lens of Computational Methods. Curr. Opin. Struct. Biol 2022, 75, 102400. [PubMed: 35689914]

(137). East KW; Skeens E; Cui JY; Belato HB; Mitchell B; Hsu R; Batista VS; Palermo G; Lisi GP NMR and Computational Methods for Molecular Resolution of Allosteric Pathways in Enzyme Complexes. Biophys. Rev 2019, 12 (1), 155–174. [PubMed: 31838649]

(138). Palermo G; Miao Y; Walker RC; Jinek M; McCammon JA Striking Plasticity of CRISPR-Cas9 and Key Role of Non-Target DNA, as Revealed by Molecular Simulations. ACS Cent. Sci 2016, 2 (10), 756–763. [PubMed: 27800559]

(139). Chen JS; Dagdas YS; Kleinstiver BP; Welch MM; Sousa AA; Harrington LB; Sternberg SH; Joung JK; Yildiz A; Doudna JA Enhanced Proofreading Governs CRISPR–Cas9 Targeting Accuracy. Nature 2017, 550 (7676), 407–410. [PubMed: 28931002]

(140). Dagdas YS; Chen JS; Sternberg SH; Doudna JA; Yildiz A A Conformational Checkpoint between DNA Binding and Cleavage by CRISPR-Cas9. Sci. Adv 2017, 3 (8), eaao0027. [PubMed: 28808686]

(141). Wang J; Arantes PR; Bhattarai A; Hsu RV; Pawnikar S; Huang Y ming M; Palermo G; Miao Y Gaussian Accelerated Molecular Dynamics: Principles and Applications. Wiley Interdiscip. Rev. Comput. Mol. Sci 2021, 11 (5), e1521. [PubMed: 34899998]

(142). Palermo G; Miao Y; Walker RC; Jinek M; McCammon JA CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. Proc. Natl. Acad. Sci. U. S. A 2017, 114 (28), 7260–7265. [PubMed: 28652374]

(143). Nierzwicki Ł; Arantes PR; Saha A; Palermo G Establishing the Allosteric Mechanism in CRISPR-Cas9. Wiley Interdiscip. Rev. Comput. Mol. Sci 2021, 11 (3), e1503. [PubMed: 34322166]

(144). East KW; Newton JC; Morzan UN; Narkhede YB; Acharya A; Skeens E; Jogl G; Batista VS; Palermo G; Lisi GP Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. J. Am. Chem. Soc 2020, 142 (3), 1348–1358. [PubMed: 31885264]

(145). Nierzwicki Ł; East KW; Morzan UN; Arantes PR; Batista VS; Lisi GP; Palermo G Enhanced Specificity Mutations Perturb Allosteric Signaling in CRISPR-Cas9. Elife 2021, 10, e73601. [PubMed: 34908530]

(146). Nierzwicki Ł; East KW; Binz JM; Hsu RV; Ahsan M; Arantes PR; Skeens E; Pacesa M; Jinek M; Lisi GP; Palermo G Principles of Target DNA Cleavage and the Role of Mg2+ in the Catalysis of CRISPR–Cas9. Nat. Catal 2022, 5 (10), 912–922. [PubMed: 36778082]

(147). Casalino L; Nierzwicki Ł; Jinek M; Palermo G Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. ACS Catal 2020, 10 (22), 13596–13605. [PubMed: 33520346]

(148). Zhu X; Clarke R; Puppala AK; Chittori S; Merk A; Merrill BJ; Simonović M; Subramaniam S Cryo-EM Structures Reveal Coordinated Domain Motions That Govern DNA Cleavage by Cas9. Nat. Struct. Mol. Biol 2019, 26 (8), 679–685. [PubMed: 31285607]

(149). Anders C; Niewoehner O; Duerst A; Jinek M Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease. Nature 2014, 513 (7519), 569–573. [PubMed: 25079318]

(150). Ricci CG; Chen JS; Miao Y; Jinek M; Doudna JA; McCammon JA; Palermo G Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. ACS Cent. Sci 2019, 5 (4), 651–662. [PubMed: 31041385]

(151). Jinek M; Chylinski K; Fonfara I; Hauer M; Doudna JA; Charpentier E A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 2012, 337 (6096), 816–821. [PubMed: 22745249]

(152). Jiang F; Taylor DW; Chen JS; Kornfeld JE; Zhou K; Thompson AJ; Nogales E; Doudna JA Structures of a CRISPR-Cas9 R-Loop Complex Primed for DNA Cleavage. Science 2016, 351 (6275), 867–871. [PubMed: 26841432]

(153). Carter EA; Ciccotti G; Hynes JT; Kapral R Constrained Reaction Coordinate Dynamics for the Simulation of Rare Events. Chem. Phys. Lett 1989, 156 (5), 472–477.

(154). Bravo JPK; Liu M. Sen; Hibshman GN; Dangerfield TL; Jung K; McCool RS; Johnson KA; Taylor DW Structural Basis for Mismatch Surveillance by CRISPR–Cas9. Nature 2022, 603 (7900), 343–347. [PubMed: 35236982]

(155). Pacesa M; Loeff L; Querques I; Muckenfuss LM; Sawicka M; Jinek M R-Loop Formation and Conformational Activation Mechanisms of Cas9. Nature 2022, 609 (7925), 191–196. [PubMed: 36002571]

(156). Yoon H; Zhao LN; Warshel A Exploring the Catalytic Mechanism of Cas9 Using Information Inferred from Endonuclease VII. ACS Catal 2019, 9 (2), 1329–1336. [PubMed: 34046245]

(157). Pacesa M; Lin CH; Cléry A; Saha A; Arantes PR; Bargsten K; Irby MJ; Allain FHT; Palermo G; Cameron P; Donohoue PD; Jinek M Structural Basis for Cas9 Off-Target Activity. Cell 2022, 185 (22), 4067–4081.e21. [PubMed: 36306733]

(158). Saha A; Arantes PR; Hsu RV; Narkhede YB; Jinek M; Palermo G Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a 2020, 60 (12), 6427–6437

(159). Saha A; Ahsan M; Arantes PR; Schmitz Michael; Chanez C; Jinek M; Palermo G An Alpha-Helical Lid Guides the Target DNA toward Catalysis in CRISPR-Cas12a. bioRxiv 2022, 2022.09.05.506663

(160). Jorgensen WL; Maxwell DS; Tirado-Rives J Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J. Am. Chem. Soc 1996, 118 (45), 11225–11236.

(161). 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 P A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. J. Comput. Chem 2003, 24 (16), 1999–2012. [PubMed: 14531054]

(162). Christen M; Hünenberger PH; Bakowies D; Baron R; Bürgi R; Geerke DP; Heinz TN; Kastenholz MA; Kräutler V; Oostenbrink C; Peter C; Trzesniak D; Van Gunsteren WF The GROMOS Software for Biomolecular Simulation: GROMOS05. J. Comput. Chem 2005, 26 (16), 1719–1751. [PubMed: 16211540]

(163). MacKerell AD; Bashford D; Bellott M; Dunbrack RL; Evanseck JD; Field MJ; Fischer S; Gao J; Guo H; Ha S; Joseph-McCarthy D; Kuchnir L; Kuczera K; Lau FTK; Mattos C; Michnick S; Ngo T; Nguyen DT; Prodhom B; Reiher WE; Roux B; Schlenkrich M; Smith JC; Stote R; Straub J; Watanabe M; Wiórkiewicz-Kuczera J; Yin D; Karplus M All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J. Phys. Chem. B 1998, 102 (18), 3586–3616. [PubMed: 24889800]

(164). Cornell WD; Cieplak P; Bayly CI; Gould IR; Merz KM; Ferguson DM; Spellmeyer DC; Fox T; Caldwell JW; Kollman PA A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc 1995, 117 (19), 5179–5197.

(165). Cheatham TE; Cieplak P; Kollman PA A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. J. Biomol. Struct. Dyn 1999, 16 (4), 845–862. [PubMed: 10217454]

(166). Wang J; Cieplak P; Kollman PA How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? Keywords: Additive Force Field; Nonadditive Force Field; Restrained Electrostatic Potential (RESP); Torsional Angle Parameterization. J. Comput. Chem 2000, 21 (12), 1049–1074.

(167). Zgarbová M; Otyepka M; Šponer J; Mládek A; Banáš P; Cheatham TE; Jurečka P Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. J. Chem. Theory Comput 2011, 7 (9), 2886–2902. [PubMed: 21921995]

(168). Xu Y; Vanommeslaeghe K; Aleksandrov A; MacKerell AD; Nilsson L Additive CHARMM Force Field for Naturally Occurring Modified Ribonucleotides. J. Comput. Chem 2016, 37 (10), 896–912. [PubMed: 26841080]

(169). Denning EJ; Priyakumar UD; Nilsson L; MacKerell AD Impact of 2’-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. J. Comput. Chem 2011, 32 (9), 1929–1943. [PubMed: 21469161]

(170). Ivani I; Dans PD; Noy A; Pérez A; Faustino I; Hospital A; Walther J; Andrio P; Goñi R; Balaceanu A; Portella G; Battistini F; Gelpí JL; González C; Vendruscolo M; Laughton CA; Harris SA; Case DA; Orozco M Parmbsc1: A Refined Force Field for DNA Simulations. Nat. Methods 2016, 13 (1), 55–58. [PubMed: 26569599]

(171). Zgarbová M; Šponer J; Otyepka M; Cheatham TE; Galindo-Murillo R; Jurečka P Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. J. Chem. Theory Comput 2015, 11 (12), 5723–5736. [PubMed: 26588601]

(172). Galindo-Murillo R; Robertson JC; Zgarbová M; Šponer J; Otyepka M; Jurečka P; Cheatham TE Assessing the Current State of Amber Force Field Modifications for DNA. J. Chem. Theory Comput 2016, 12 (8), 4114–4127. [PubMed: 27300587]

(173). Hart K; Foloppe N; Baker CM; Denning EJ; Nilsson L; MacKerell AD Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. J. Chem. Theory Comput 2012, 8 (1), 348–362. [PubMed: 22368531]

(174). Lemkul JA; MacKerell AD Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA. J. Chem. Theory Comput 2017, 13 (5), 2072–2085. [PubMed: 28398748]

(175). Yoo J; Aksimentiev A Improved Parametrization of Li +, Na +, K +, and Mg 2+ Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. J. Phys. Chem. Lett 2012, 3 (1), 45–50.

(176). Li P; Merz KM Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. J. Chem. Theory Comput 2014, 10 (1), 289–297. [PubMed: 24659926]

(177). Li P; Roberts BP; Chakravorty DK; Merz KM Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. J. Chem. Theory Comput 2013, 9 (6), 2733–2748. [PubMed: 23914143]

(178). Panteva MT; Giambaşu GM; York DM Force Field for Mg2+, Mn2+, Zn2+, and Cd2+ Ions That Have Balanced Interactions with Nucleic Acids. J. Phys. Chem. B 2015, 119 (50), 15460–15470. [PubMed: 26583536]

(179). Li Z; Song LF; Li P; Merz KM Systematic Parametrization of Divalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. J. Chem. Theory Comput 2020, 16 (7), 4429–4442. [PubMed: 32510956]

(180). Sengupta A; Li Z; Song LF; Li P; Merz KM Parameterization of Monovalent Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. J. Chem. Inf. Model 2021, 61 (2), 869–880. [PubMed: 33538599]

(181). Grotz KK; Cruz-León S; Schwierz N Optimized Magnesium Force Field Parameters for Biomolecular Simulations with Accurate Solvation, Ion-Binding, and Water-Exchange Properties. J. Chem. Theory Comput 2021, 17 (4), 2530–2540. [PubMed: 33720710]

(182). Grotz KK; Schwierz N Optimized Magnesium Force Field Parameters for Biomolecular Simulations with Accurate Solvation, Ion-Binding, and Water-Exchange Properties in SPC/E, TIP3P-Fb, TIP4P/2005, TIP4P-Ew, and TIP4P-D. J. Chem. Theory Comput 2022, 18 (1), 526–537. [PubMed: 34881568]

(183). Zhang Y; Jiang Y; Peng J; Zhang H Rational Design of Nonbonded Point Charge Models for Divalent Metal Cations with Lennard-Jones 12–6 Potential. J. Chem. Inf. Model 2021, 61 (8), 4031–4044. [PubMed: 34313132]

(184). Shi K; Li Z; Anstine DM; Tang D; Colina CM; Sholl DS; Siepmann JI; Snurr RQ Two-Dimensional Energy Histograms as Features for Machine Learning to Predict Adsorption in Diverse Nanoporous Materials. J. Chem. Theory Comput 2023.

(185). Li P; Merz KM Metal Ion Modeling Using Classical Mechanics. Chem. Rev 2017, 117 (3), 1564–1686. [PubMed: 28045509]

(186). Gao J Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. Acc. Chem. Res 1996, 29 (6), 298–305

(187). Brunk E; Ashari N; Athri P; Campomanes P; De Carvalho FF; Curchod BFE; Diamantis P; Doemer M; Garrec J; Laktionov A; Micciarelli M; Neri M; Palermo G; Penfold TJ; Vanni S; Tavernelli I; Rothlisberger U Pushing the Frontiers of First-Principles Based Computer

Simulations of Chemical and Biological Systems. Chimia (Aarau) 2011, 65 (9), 667. [PubMed: 22026176]

(188). Nierzwicki L; Ahsan M; Palermo G The Electronic Structure of Genome Editors from the First Principles. Electron. Struct 2023, 5, 014003. [PubMed: 36926635]

(189). Laio A; Parrinello M Escaping Free-Energy Minima. Proc. Natl. Acad. Sci. U. S. A 2002, 99 (20), 12562–12566. [PubMed: 12271136]

(190). Marx D; Hutter J Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods; Cambridge University Press, 2009.

(191). Car R; Parrinello M Unified Approach for Molecular Dynamics and Density-Functional Theory. Phys. Rev. Lett 1985, 55 (22), 2471. [PubMed: 10032153]

(192). Nerenberg PS; Head-Gordon T New Developments in Force Fields for Biomolecular Simulations. Curr. Opin. Struct. Biol 2018, 49, 129–138. [PubMed: 29477047]

(193). Lemkul JA; Huang J; Roux B; Mackerell AD An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. Chem. Rev 2016, 116 (9), 4983–5013. [PubMed: 26815602]

(194). Ponder JW; Wu C; Ren P; Pande VS; Chodera JD; Schnieders MJ; Haque I; Mobley DL; Lambrecht DS; Distasio RA; Head-Gordon M; Clark GNI; Johnson ME; Head-Gordon T Current Status of the AMOEBA Polarizable Force Field. J. Phys. Chem. B 2010, 114 (8), 2549–2564. [PubMed: 20136072]

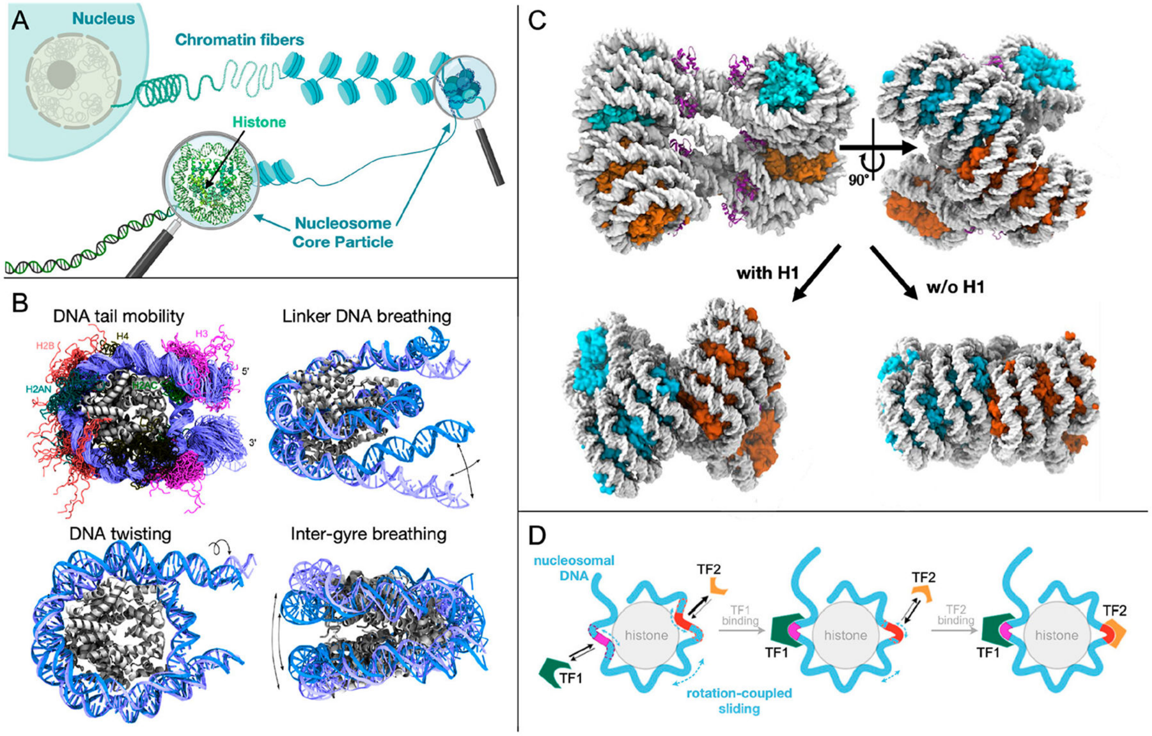


Figure 1. Nucleosome structural assemblies and dynamical components.

(A) Schema showing the basic units of chromatin and nucleosome core particle organization. The nucleosome core particle consists of H2A, H2B, H3, and H4 core histone proteins wrapped by DNA. Created with BioRender.com. **(B)** Large amplitude dynamical motions of the NCP were observed from ~1 μs long unbiased MD simulations[24,36]. Arrows indicate the type of motions between two representative snapshots (shown in blue and violet). Adapted with permission from Journal of Molecular Biology, 433, Huertas & Cojocaru, "Breaths, Twists, and Turns of Atomistic Nucleosomes", 166744.[[24]] Copyright 2021 Elsevier Ltd. **(C)** Structure of a 30-nm chromatin fiber subjected to all-atom molecular simulations by Woods et al.[17] The H1 linker histones (purple) were shown to inhibit DNA flexibility, stabilize tetra-nucleosomal units, and increased chromatin compaction (with H1). Upon the removal of H1 (w/o H1), a significant destabilization of the structure is observed, as the fiber adopts less strained and untwisted states. Adapted with permission from Journal of Molecular Biology, 433, Woods et al., "The Dynamic Influence of Linker Histone Saturation within the Poly-Nucleosome Array", 166902.[[17]] Copyright 2021 Elsevier Ltd. **(D)** Schematic representation of the allosteric regulation in the NCP induced by Transcription Factor (TF) binding. The binding of the first TF (TF1) restricts the sliding motion of the nucleosomal DNA and changes the exposure extent of a distant binding site for a second TF (TF2). Adapted with permission from Tan & Takada.[[20]] Copyright 2020 National Academy of Sciences.

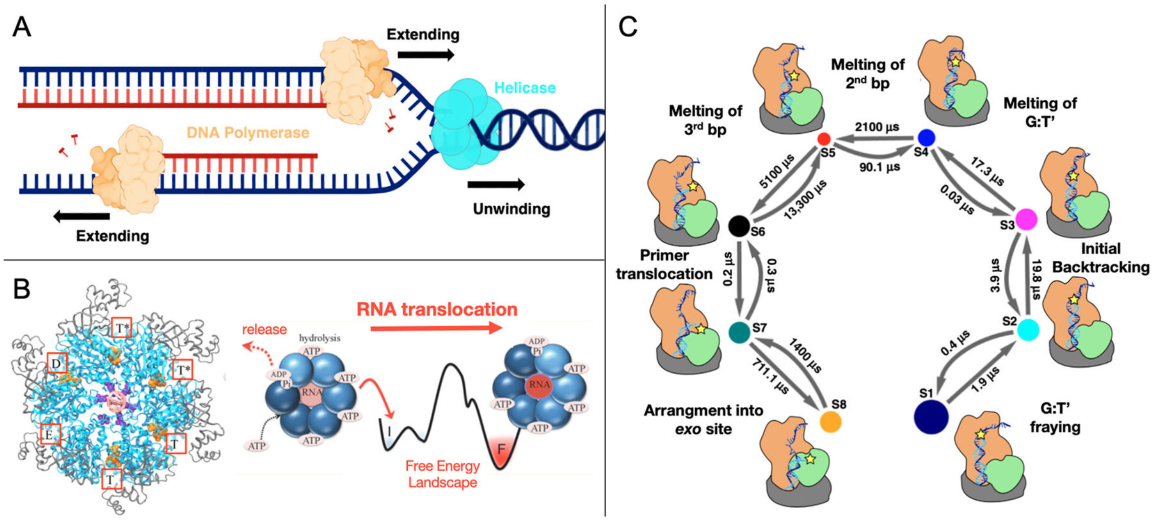


Figure 2.

Replication machinery. **(A)** Schematic representation of the replication process. The helicase enzymes bind and unwind the DNA strands, while the DNA polymerases synthesize the complementary sequence in each DNA strand. Created with BioRender.com **(B)** Structure of Rho hexameric helicase and the proposed rotary reaction mechanism. Each Rho subunit consists of a C-terminal and N-terminal domain shown in blue and grey respectively. The ATP binding sites are located at the inter−subunit interfaces, labeled according to the state of ATP at the respective sites viz. hydrolysis-competent (T*), ATP-bound (T), "old" product (D), and empty (E) state. The release of the hydrolysis product (ADP + Pi) triggers the force-generating process of Rho. Adapted with permission from Ma & Schulten[50]. Copyright 2015 American Chemical Society. **(C)** Markov State Modelling (MSM) derived complete kinetic model for the *pol*–to–*exo* conformational switching. The model shows the macrostates (S1–S8) denoted by circles where arrows indicate the transition rates between the states. Reprinted with permission from Dodd et al., Macmillan Publishers Ltd: *Nat. Commun.* [52], Copyright 2020.

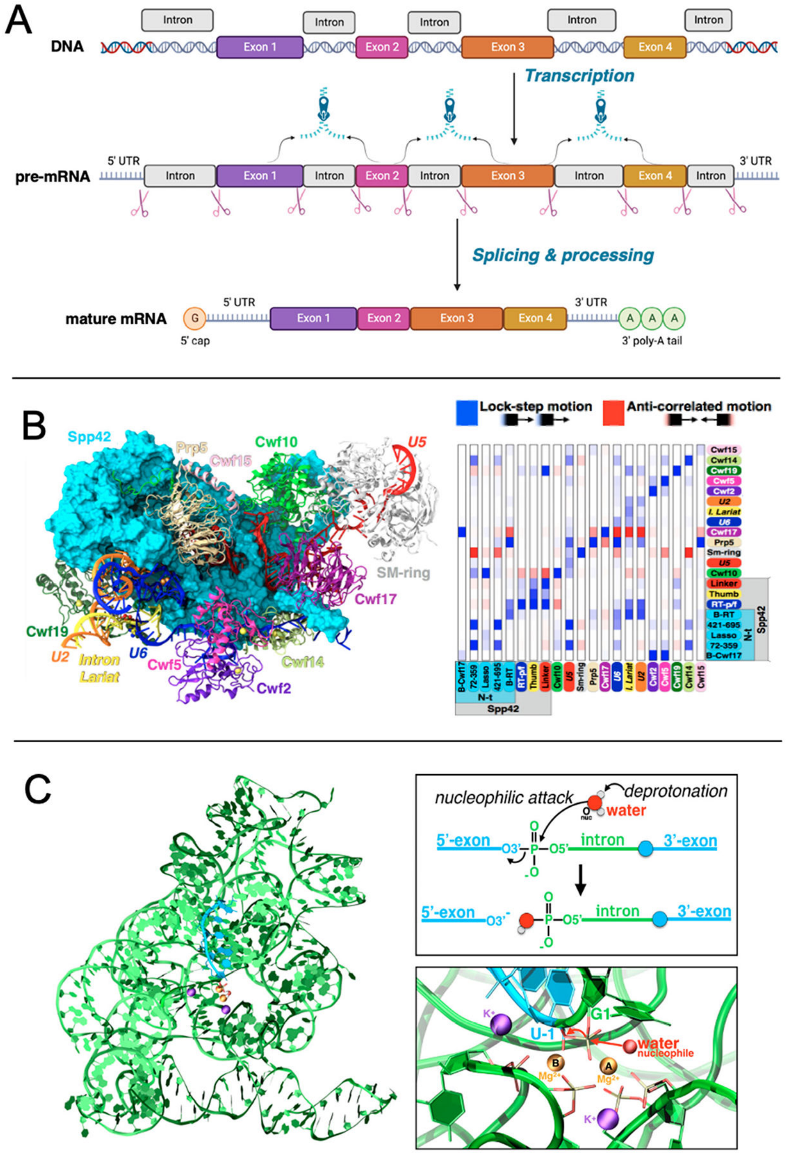


Figure 3. Splicing mechanism and functional dynamics of spliceosomal complexes.

(A) Illustration of the RNA splicing process. Created with BioRender.com **(B)** Left: model of the intron lariat spliceosome from the yeast *S. pombe* cryo-EM structure (PDB: 3JB9)[83]. The spliceosomal components are shown as surfaces and cartoons using different colors. Right: histogram reporting the (per-column) normalized correlation scores (CSs) calculated for each pair of spliceosomal components. Adapted with permission from Casalino et al. [81] Copyright 2018 National Academy of Sciences. **(C)** Left: structure of the *Oceanobacillus iheyensis* group II intron before splicing (PDB: 4FAQ)[84]. The intron (green) and exon (cyan) are shown as ribbons. Right: reaction schema of the first hydrolysis step of splicing (top) and the respective arrangement at the catalytic site (bottom). Adapted with permission from Casalino et al.[81] Copyright 2016 American Chemical Society.

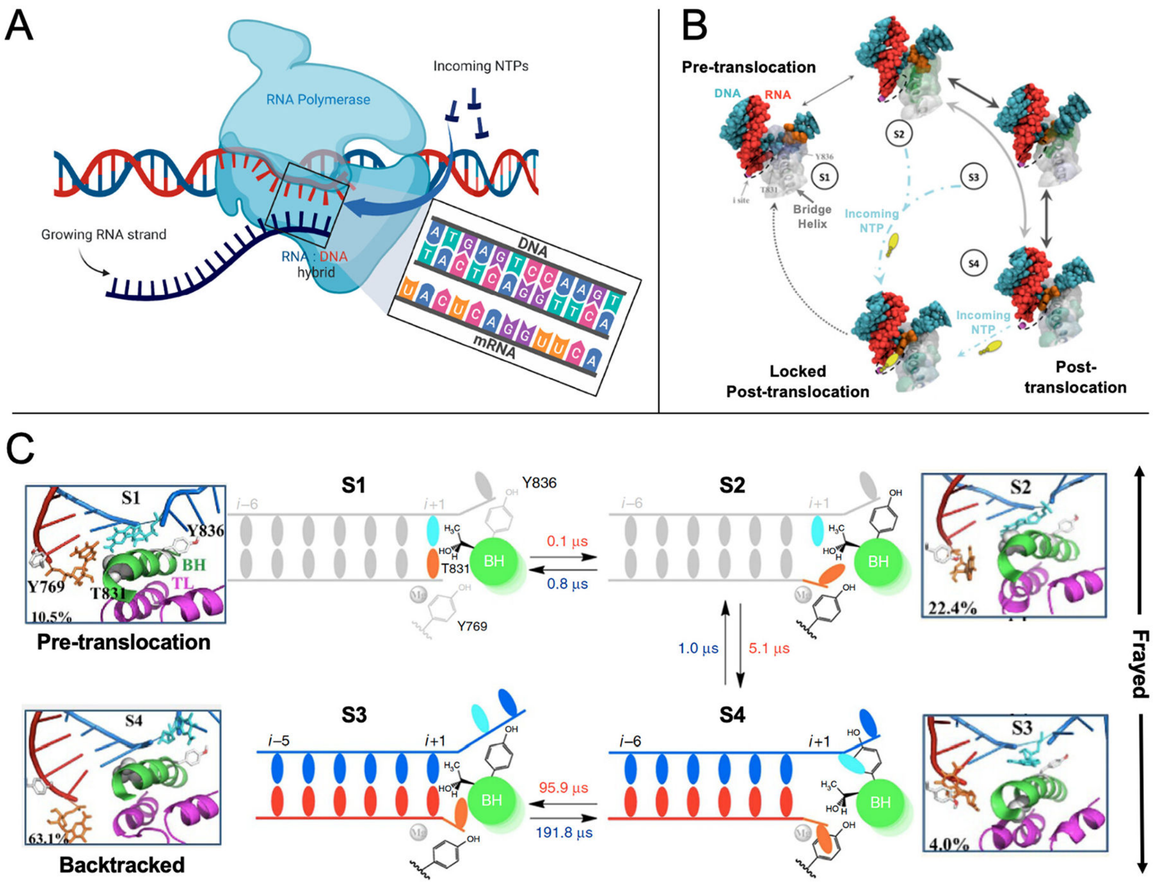


Figure 4. Transcription machineries.

(A) Schematic representation of the transcription process. Created with BioRender.com **(B)** The model of eukaryotic RNA polymerase (Pol II) translocation cycle from MD simulations of Silva et al.[106] At the pre-translocation state (S1), the bridge helix (BH) interacts with the ‘i+1’ DNA nucleotide, facilitating the motion of the RNA:DNA hybrid toward the post-translocation state. At S2, upstream RNA:DNA hybrid backbone is translocated, while the transition nucleotide (TN) lags. The empty active site here allows the entry of incoming NTP to the ‘i’ site. At S3, the oscillation of the BH facilitates the TN cross-over, while maintaining strong interactions mainly with residue T831. Finally, at S4, the TN moves to its final i+1 post-translocation position. Adapted with permission from Silva et al.[[106]] Copyright 2014 National Academy of Sciences. **(C)** Stepwise model of backtracking mechanism in Pol II elongation complex (EC)[107]. BH when bent towards the i+1 site, acts as a checkpoint to examine the stability of the base pairs in the active site using residue T831 as a sensing probe (state S1). Upon mismatches, this bending can trigger the motion of the RNA 3’-end nucleotide to the frayed state where the nucleotide stacks directly to residue Y769 reaching state S2. In the S3 state, the BH residue Y836 stacks with DNA TN, while residue Y769 stacks with the RNA 3’-end nucleotide through their aromatic rings. Finally, the backtracking of the RNA:DNA hybrid takes place (state S4). Reprinted with permission from Da et al., Macmillan Publishers Ltd: *Nat. Commun*. [[107]], Copyright 2016.

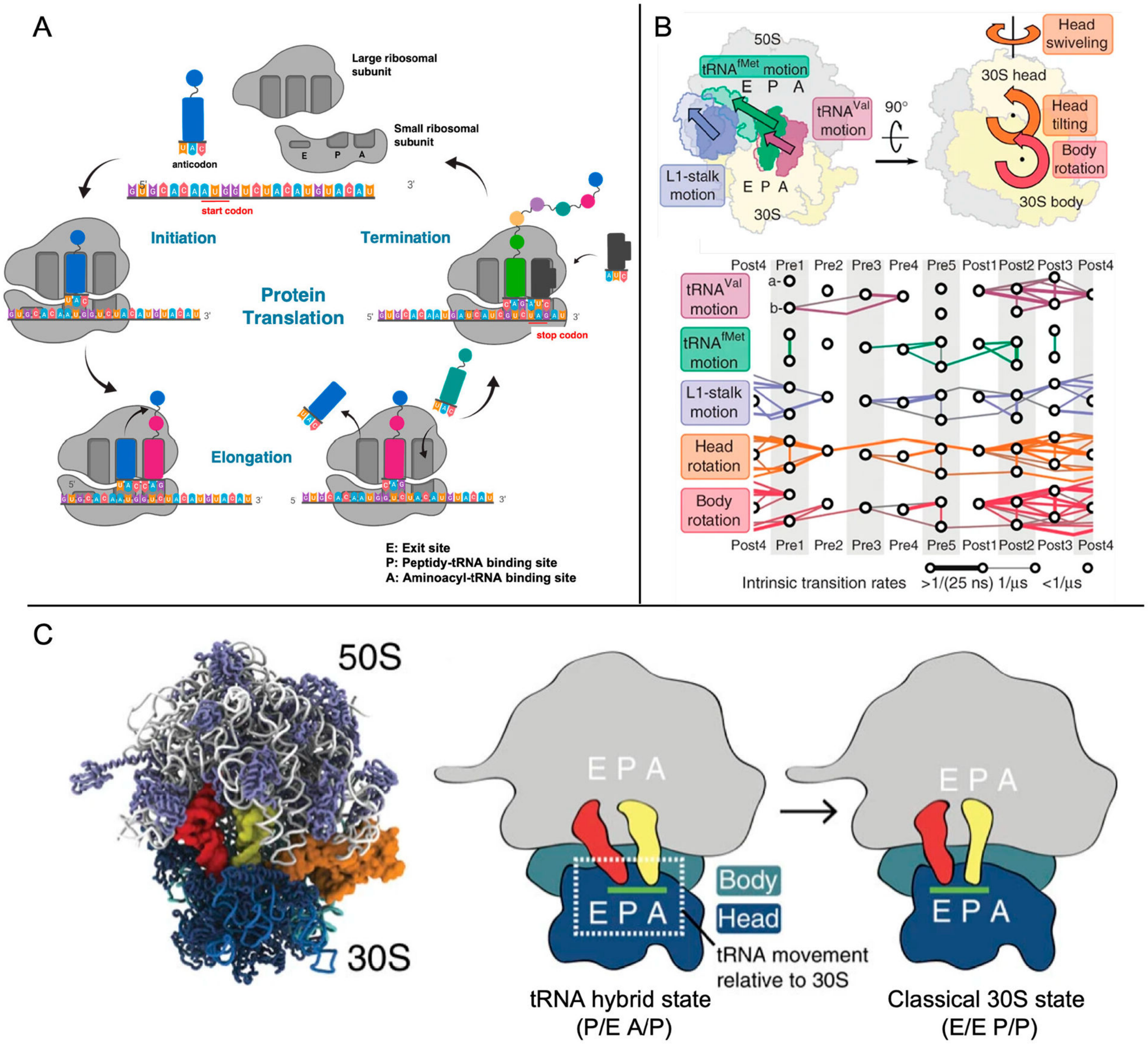


Figure 5. Ribosomal complexes

(A) General view of the eukaryotic protein translation. Firstly, an initiation complex is formed with the small 30S subunit, the initiator tRNA, and the initiation factors followed by the addition of the large 50S subunit to form an intact ribosome. This complex consists of three tRNA holding sites from 3' to 5', designated as the A-site (holds aminoacyl-tRNA or termination release factors), P-site (holds a peptidyl tRNA), and E-site (holds a free tRNA). The elongation process involves a series of steps comprising mRNA movement through the ribosome with the tRNAs translocating across the holding sites with the formation of new peptide bonds at each step. The process continues until a stop codon is encountered. Adapted from "Protein Translation", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates. **(B)** The structural transitions and rotations accompanying the tRNA translocation are indicated on a model of the ribosomal complex using arrows (top). The estimated transition rates, derived from MD simulations, between different states are schematically reported (bottom). Reprinted with permission from Bock et al., Macmillan Publishers Ltd: *Nat. Struct. Mol. Biol.*[128] Copyright 2013. **(C)** Structural description of tRNA translocation studied by Nguyen *et al.* The structure of the 70S ribosome with two tRNAs (red, yellow) and the elongation factor, EF-G (orange) is shown on the left, followed by the schematic representation of tRNA conformations before and after translocation on the 30S subunit. Reprinted with permission from Bock et al., Macmillan Publishers Ltd: *Nat. Commun.*[130] Copyright 2016.

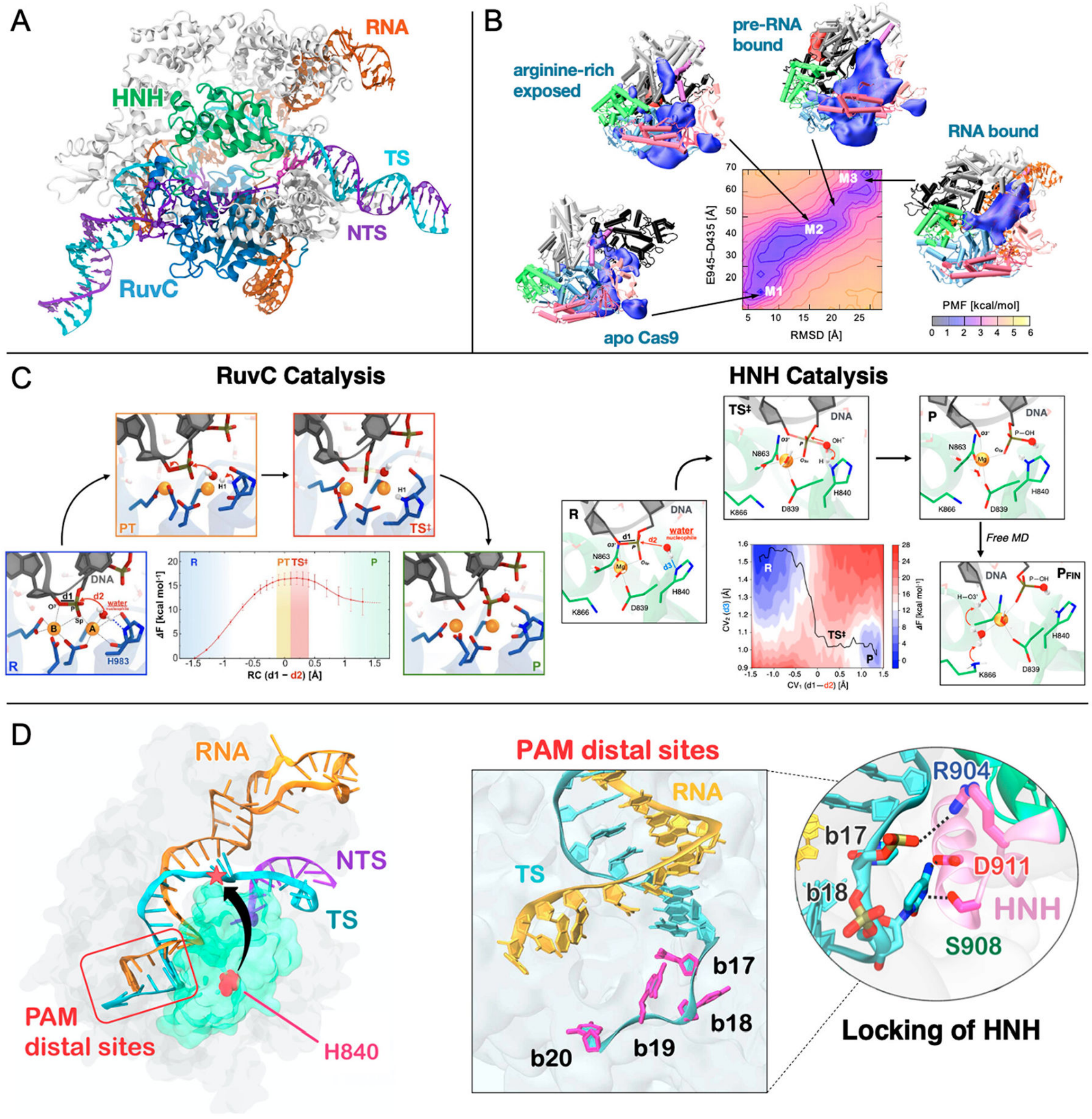


Figure 6. Gene-editing complexes.

(A) Overview of the *Streptococcus Pyogenes* CRISPR-Cas9 system. The Cas9 protein is shown as ribbons, highlighting the catalytic domains HNH (green) and RuvC (blue). The guide RNA (orange), target DNA (cyan), and non-target DNA (violet) are shown as ribbons. The system is shown in its catalytically active form, prone to perform DNA cleavages (PDB: 6O0Y)[148]. **(B)** Free energy profile of the conformational transition underlying RNA binding. The Potential of Mean Force (PMF) is plotted along the E945-D435 distance (used in FRET studies) and the RMSD between the apo Cas9 and the RNA-bound system. The PMF reveals three local minima, M1: crystallographic state of the apo Cas9, M2: pre-RNA bound intermediate states characterized by the solvent-exposed R-rich helix, and, M3: the RNA bound state. Adapted with permission from Palermo et al.[[142]] Copyright 2017 National Academy of Sciences. **(C)** Catalytic mechanism of DNA cleavage in CRISPR-Cas9. Structural and energetic features showing reactant (R), proton transfer (PT), transition state (TS), and product (P) of non-target DNA cleavage by RuvC (left) and target DNA strand cleavage by HNH reactant (right). Adapted with permission from Casalino et al.[147] Copyright 2020 American Chemical Society. Reprinted with permission from Nierzwicki et al., Macmillan Publishers Ltd: *Nat. Catal.*[146] Copyright 2022. **(D)** X-ray structure of CRISPR-Cas9, highlighting the location of PAM-distal sites with respect to the inactive HNH domain (i.e., the catalytic H840 is far from the cleavage site, PDB: 4UN3)[149]. The right panel shows the opening of the RNA:DNA hybrid in the presence of four base pair mismatches at PAM distal sites (magenta), and the locking interactions between the DNA TS and the HNH domain. Adapted with permission from Ricci et al.[150]. Copyright 2019 American Chemical Society.