An Expanded Two-Zn^{2+}-ions Motif Orchestrates pre-mRNA Maturation in the 3'-end Processing Endonuclease Machinery

Jure Borišek‡ and Alessandra Magistrato†*

‡National Institute of Chemistry, Hajdrihova 19, 1001 Ljubljana, Slovenia
†CNR-IOM-Democritos National Simulation Center c/o SISSA, Via Bonomea 265, 34136 Trieste, Italy

ABSTRACT: Eukaryotic precursor-messenger RNAs (pre-mRNAs) undergo extensive compositional pre-processing to become mature, protein-coding, mRNAs. Among the vital transformations underlying pre-mRNA maturation the 3'-end processing machinery (3EPM) orchestrates 3'-end pre-mRNA cleavage via a yet elusive catalytic mechanism. Here, all-atom simulations of a 350,000 atoms model of 3EPM disclose that its catalytic engine, the CPSF73 endonuclease, cleaves the 3'-end of pre-mRNA via an associative two-Zn^{2+}-ions aided mechanism, where the metals, beside activating the nucleophile and stabilizing the transition state, as in canonical two-Mg^{2+}-ions catalysis, assist the leaving group's protonation. In spite of the distinctive metal type content, an in depth structural and mechanistic inspection of two-Zn^{2+}-ions versus two-Mg^{2+}-ions dependent nucleases unlocks striking similarities between the expanded positive charge of their catalytic motifs, with the metals being assisted by second-shell basic residues/ metal ions. This catalytic architecture, hence, emerges as a critical prerequisite for a common and effective mechanism of phosphodiester bond hydrolysis in nuclease enzymes. Ostensibly, our outcomes unveil the salient molecular traits of 3EPM mechanism, providing tantalizing opportunities to harnessing this emerging drug target to fight the wide variety of human diseases associated with a deregulated pre-mRNA processing.

KEYWORDS: pre-mRNA, endonuclease, 3'-end processing, two Zn^{2+} ions, two Mg^{2+} ions, catalytic mechanism, QM/MM

INTRODUCTION

Most eukaryotic messenger RNA precursors (pre-mRNAs) require extensive compositional remodeling to be converted into mature, protein-coding mRNAs.[1] Among the essential manipulations underlying pre-mRNA maturation, the 3'-end cleavage and polyadenylation of pre-mRNA are critical processes for proper nuclear export, stability, translation of mRNAs, and for transcription termination of RNA polymerase II (Pol II).[2] The 3'-end processing of the three major classes of RNA Pol II transcripts (i.e. polyadenylated mRNA, histone mRNAs, and small nuclear RNA) is performed by three different machineries that share common structural and catalytic features.[2]

Recently, a cryogenic electron microscopy (cryo-EM) structure of the whole human histone pre-mRNA 3'-end processing machinery (3EPM) has provided ground-breaking information for dissecting the molecular terms of 3'-end cleavage of pre-mRNA.[3] This structure has elucidated an entangled protein/RNA network composed by nine proteins, among which the 73-kDa subunit of cleavage and polyadenylation specific factor (CPSF73), which is the catalytic engine for the 3'-end cleavage of pre-mRNA, and two RNA filaments: the histone pre-mRNA substrate and the U7 small nuclear (sn) RNA.[3] CPSF73 is a specific endonuclease belonging to the β-CASP family, composed by a β-CASP and a metallo-β-lacamase (MβL) domain, with only the latter being endowed with two-Zn^{2+}-ions dependent nuclease catalysis, at variance to most RNA and DNA processing enzymes/ribozyme, which, instead, rely on an extensively characterized two-Mg^{2+}-ions aided mechanism to perform this function.[4-8] The striking relevance of pre-mRNA 3'-end-processing is emphasized by the implications of its de-regulation into a wide variety of human diseases, including endocrine disruption, early-onset of myopia, acute myeloid leukemia and Ewing's sarcoma.[9-11] Accordingly, CPSF73 has been recognized as an emerging drug target for anti-cancer,[12] anti-inflammatory[13] and anti-protozoal therapies.[14] In this scenario, an atomic-level dissection of its catalytic mechanism may foster the development of novel cutting-edge therapeutic applications.

In this study, we resolved the mechanistic intricacies of 3'-end-processing of pre-mRNA in metazoans trough classical and quantum–classical (QM/MM) molecular dynamics (MD) simulations, in combination with enhanced sampling methods. Our findings unlock how CPSF73 promotes pre-mRNA 3'-end cleavage via the tripartite charge architecture of an expanded two-Zn^{2+}-ions motif that neatly emulates the mechanism of two-Mg^{2+}-ions aided nucleases.[15]

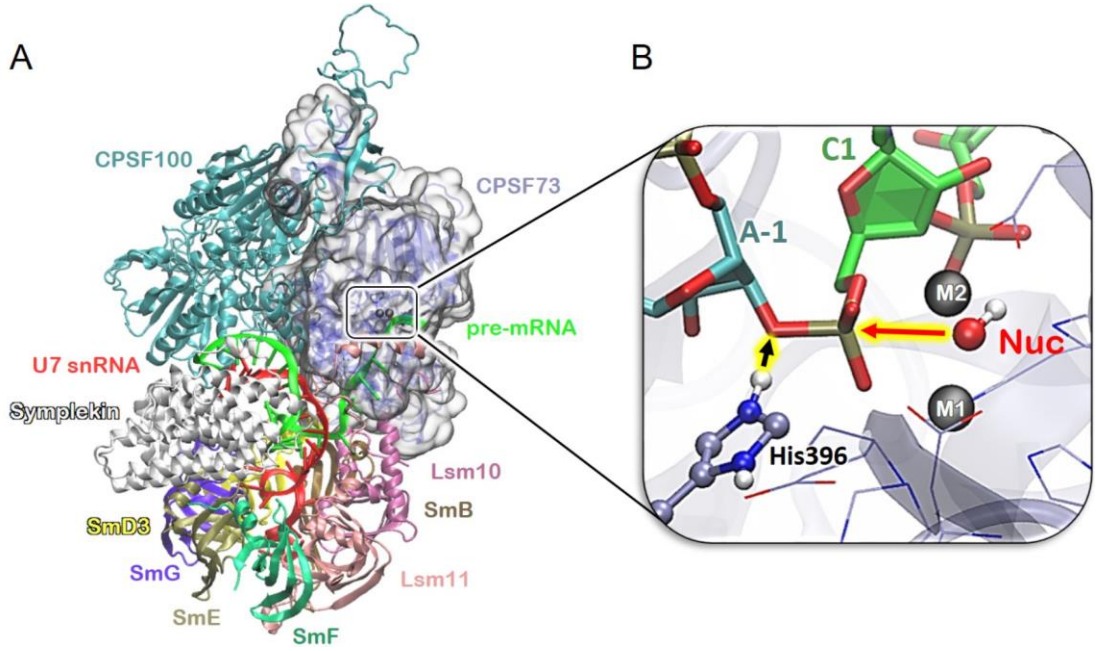


Figure 1. A) Model of human histone pre-mRNA 3'-end processing machinery built on the cryo-EM structure (PDB id: 6V4X). Proteins SmD3 (yellow), SmB (brown), Lsm10 (magenta), Lsm11 (pink), SmE (light brown), SmF (green), SmG (violet), CPSF100 (cyan), Symplekin (white), and RNAs U7 snRNA (red) and pre-mRNA (light green) are depicted as cartoons, whereas CPSF73 is shown as transparent gray surface with enclosed ice-blue cartoons. Zn^{2+} ions are represented as gray spheres. B) Inset of the catalytic site featuring the Zn^{2+} ions, labelled as M1 and M2, shown as gray van deer Waals spheres, the pre-mRNA strand, the hydroxide nucleophile, the residues coordinating the metal ions, and His396, shown in licorice and colored by atom name. Carbon atoms of the pre-mRNA are shown in green and cyan for the 5'- and 3'- end of the pre-mRNA, respectively.

RESULTS AND DISCUSSION

Classical and quantum classical (QM/MM) molecular dynamics (MD) simulations were done on an explicitly solvated model system (350,000 atoms) of human 3EPM extracted from its cryo-EM structure (PDB id: 6V4X).[3] Our model accounts for the entire resolved structure, encompassing the SmD3, SmB, Lsm10, Lsm11, SmE, SmF, SmG, CPSF73, CPSF100 and Symplekin proteins, the U7 snRNA and pre-mRNA strands, and two Zn^{2+} ions (Figure 1 and Table S1 of Supporting Information (SI)). Since the predicted pKa value of Glu204 for the starting structure of the apo (Zn^{2+} free) form of the protein was 7.85, which is only slightly higher than the physiological pH, we considered two models: one in which Glu204 is deprotonated ($Glu204^{-1}$), and a model in which Glu204 is neutral ($Glu204^{0}$).

After initial equilibration by the classical MD (~ 200 ns) simulations using the Amber ff14SB force field (FF) for proteins[16] and ff99+bsc0+χOL3 FF for the RNA strands[17] (see Computational Details), each model system was relaxed via QM/MM MD with the QM part treated at the DFT-BLYP[18] level of theory, similarly to previous studies.[19-21]

Our simulations reveal that the two catalytic Zn^{2+} ions coordinate the O^{Rp} and O^{sp} oxygens of the scissile phosphate, while histidine residues (His71, His73, His76, His158 and His418) and aspartate (Asp75) complete their plastic coordination spheres as in MβL enzymes (Figures 1 and S1).[3-4] Aiming to exhaustively assess the catalytic mechanism, we first placed a water molecule in the position of the oxygen atom, which in the cryo-EM structure bridges the two Zn^{2+} ions.[3] In the model with deprotonated Glu204 ($Glu204^{-1}$), the flanking hydrogen (H-) bonding residues Asp179 and His396 were both in ionized states ($Asp179^{-1}$ and $His396^{+1}$). After having equilibrated the system by classical MD and QM/MM MD simulations, we observe the spontaneous formation of a bridging hydroxide, with the water proton being transferred to His71 and ultimately to Asp179 (Figure S1 and Movie S1). This occurs within the first ps of the QM/MM MD simulation due to the strong Lewis acid character of Zn^{2+} ions.[22] Conversely, the H-Nδ@His396 is shared with Oε@Glu204, being mostly localized on the latter. (Figure S2A),

We equilibrated this newly formed reactant state with 200 ns of classical and subsequently with ~10 ps of QM/MM MD. As a result, the catalytic site is poised for catalysis with the bridging hydroxide ion (i.e. the putative nucleophile) placed at an optimal distance and orientation to attack the scissile phosphate P@C1 atom, whereas the leaving $O^{3'}$@A-1 atom of the substrate hydrogen (H-)bonds to a HεN-His396, which may therefore act as a general acid. Remarkably, after the classical and QM/MM MD equilibation the newly acquired H-OεAsp179 engages an hydrogen (H-)bond with Oε2@Glu204, with proton being shared, but mostly localized on Oε1@Asp179. As well, the H-Nδ@His396 is shared with Oε@Glu204, being mostly localized on the former (Figure S2B). This reactant (R) state is quite similar to the experimentally determined structure, differing from it only in the orientation of the sidechain

oxygens of Asp179 and Glu204 residues and the of M1 Zn^{2+} ion (Figure S3).

We next simulated the 3'-end pre-mRNA cleavage reaction by performing QM/MM MD free energy simulations with thermodynamic integration (see Computational Details). The nucleophilic attack of the O_{nuc} atom of the hydroxide ion to the P@C1 atom of scissile phosphate was accomplished by exploiting a reaction coordinate (RC) based on the difference between the distances of the forming (O_{nuc}@OH^- – P@C1) and the breaking bonds (P@C1 – $O3'$@A-1), distributed over 18 windows. As a result, 3EPM promotes pre-mRNA cleavage through the following steps: (i) at the reactant (**R**) state the hydroxide ion lies at a distance of ~ 3.30 Å from the scissile phosphate P@C1, at ~ 2.0 Å distance from both Zn^{2+} ions, and the O^{Rp} and O^{Sp}@C1 are bound by the Zn^{2+} ions M1 and M2, respectively; (ii) The proton transfer (**PT**) from an ionized His396 to $O3'$@A-1 oxygen of the leaving group starts occurring just before the transition state (**TS**) is reached (Figure 2); (iii) At the **TS**, the forming (O_{nuc}@OH^- – P@C1) and breaking bonds (P@C1 – $O3'$@A-1) display approximately the same lengths (~2.2 Å) and the nucleophile coordinates both Zn^{2+} ions; (iv) After the **TS**, the hydroxide ion loses its coordination to M2, and (v) the proton transfer from Nε@His396 to the O3'@A-1 ends when the product (**P**) starts forming (Figures 2-3, Figure S4 and Movie S2), in line with an associative mechanism (Figure S5). Interestingly, the H-Nε@His396 is shared with Glu204 along the reaction coordinate, being fully transferred to Nε@His396 after the reaction has occurred (Figure S6). We remark that this mechanism is different from the water-mediated pre-mRNA dissociative mechanism observed in group II intron ribozyme and in first transesterification step mediated by the spliceosome, where a dissociative pathway is operative.[20,23]

The Zn^{2+} ions coordination sphere is rather plastic, being endowed with four ligands for most of the simulation time, except in proximity of the **TS**, where the metals adopt five-coordinated configuration (Figure S7 and Table S2).

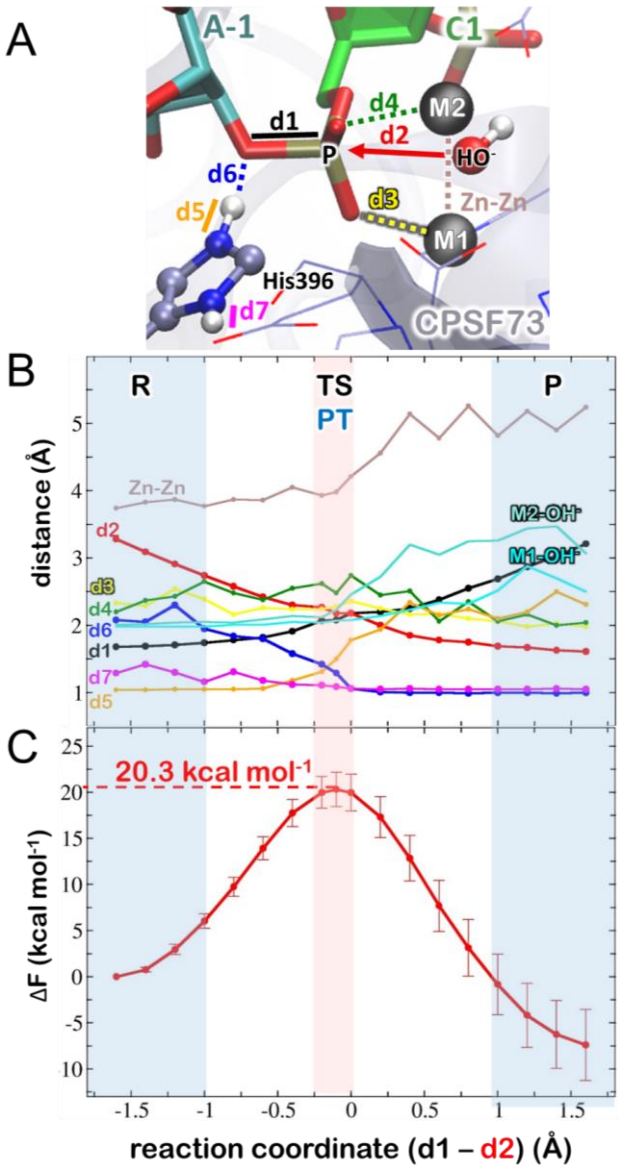


Figure 2. (A) Snapshot of the reactant state showing the definition of the critical bond distances. The Zn^{2+} ions are shown as gray spheres and labelled as M1 and M2. RNA is represented in licorice with green and cyan carbon atoms in the 5'- and 3'- end of the pre-mRNA, respectively. The protein is depicted as gray transparent cartoons, the general acid (His396) is shown in balls and sticks, while the surrounding residues are represented as lines. (B) Evolution of selected critical bond distances along the reaction coordinate. (C) Helmholtz free energy profile (ΔF [kcal/mol]) calculated from thermodynamic integration at the DFT-BLYP[24] level with a mixed double zeta/plane wave basis sets and Amber ff14SB/ff99+bsc0+χOL3[25] level for the QM and MM part, respectively. The areas corresponding to the reactant (**R**), transition state (**TS**), proton transfer (**PT**) and product (**P**) states are highlighted in blue, red and blue, respectively.

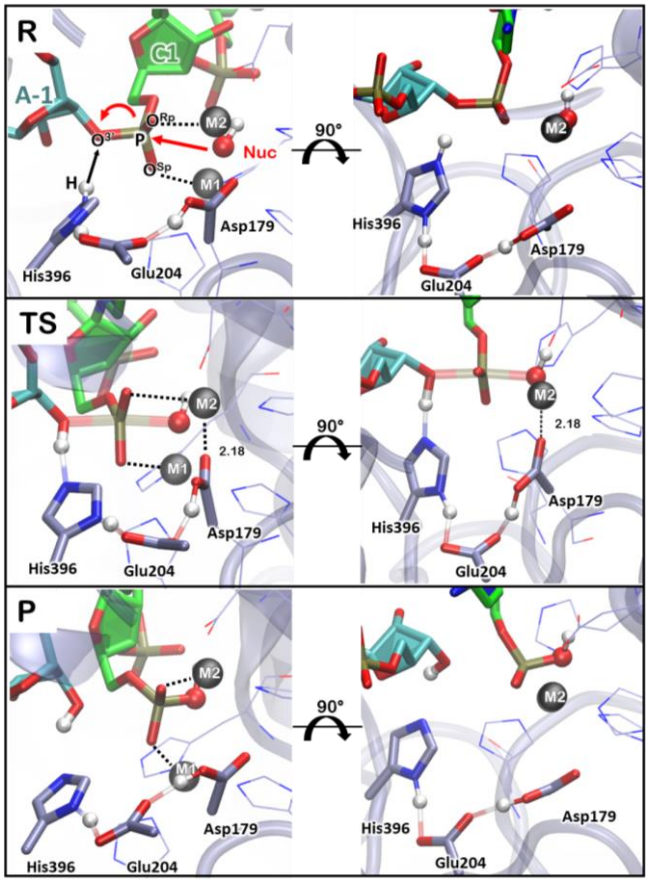


Figure 3. Representative snapshots of the reactant (**R**), transition state (**TS**) and product (**P**) structures of the 3'-end pre-mRNA processing reaction from a front and side point of view. The carbon atoms of the 5'- and 3'-end of pre-mRNA are depicted in green and cyan, respectively; His396, Glu204 and Asp179 residues are shown in licorice, key hydrogen atoms and Zn^{2+} ions are shown in white and in gray spheres, respectively. The other ligands coordinating the metals are shown as lines. Dashed lines depict the coordination of Zn^{2+} ions with the scissile phosphate oxygens O^{Sp} and O^{Rp}, and with Asp179.

The reaction occurs with a Helmholtz free energy barrier ($\Delta F^{\#}$) of 20.3 ± 1.9 kcal/mol (Figure 2C), as calculated via the blue moon ensemble method.[26] To our knowledge, the experimental catalytic rate of CPSF73 endonuclease of 3EPM has not yet been hitherto reported. Nevertheless, the experimental catalytic rate for ribonuclease Z, sharing with CPSF73 a nearly identical catalytic architecture,[4,27] is of 7.8 min^{-1}. This corresponds to a $\Delta F^{\#}$ of 18.7 kcal/mol, which is fully consistent with the $\Delta F^{\#}$ calculated here.

In order to inspect whether the thermodynamic integration calculations may have led to an over/underestimation of $\Delta F^{\#}$ owing to the choice of a monodimensional RC, we performed QM/MM metadynamics simulations[28] using two reaction coordinates (also referred as collective variables (CVs)). Here CV1, defined as the difference of the forming (O_{nuc}@OH^- – P@C1) and the breaking bond (P@C1 – O3'@A-1) distances, corresponds to the reaction coordinate used in the thermodynamic integration calculations, thus accounting for the nucleophilic attach of the hydroxide ion on the scissile phosphate. Conversely, CV2, defined as the distance from the His396 proton to O3'@A-1 (H-Nε@His396 - O3'@A-1), accounts for the proton transfer from the general acid to the leaving group. Remarkably, metadynamics simulations showed that the two CVs are partially interdependent, leading to a $\Delta F^{\#}$ of ~ 18.0 ± 2.2 kcal/mol (Figure 4), in excellent agreement with experimental catalytic rate of ribonuclease Z ($\Delta F^{\#}$ of 18.7 kcal/mol).[27]

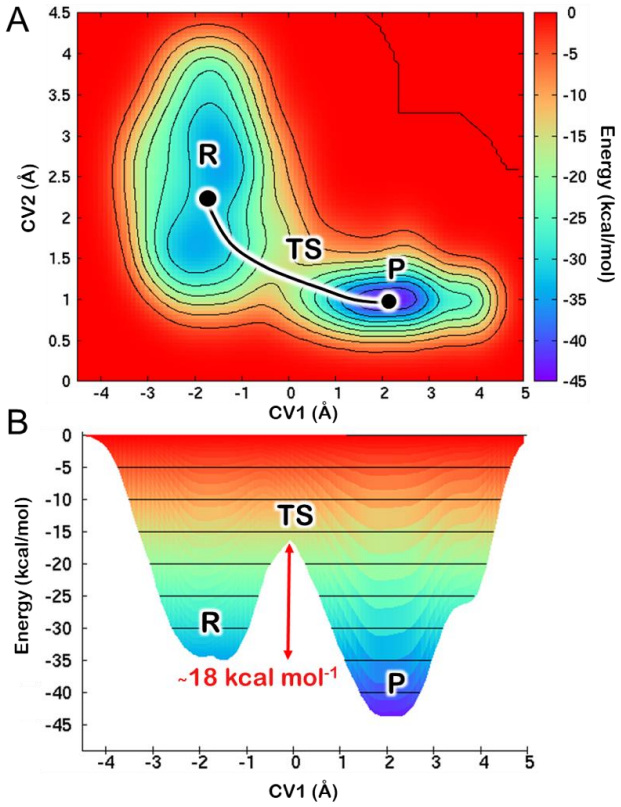


Figure 4. Free energy surface obtained from QM/MM metadynamics simulation showing the progress of reaction from reactants (**R**) via transition state (**TS**) to products (**P**). Collective variable 1 (CV1) denotes the nucleophilic attack on scissile phosphate, while CV2 accounts the proton transfer from H-Nε@His396 atom to the O3'@C-1 of the leaving group. Depicted are (A) CV1 versus CV2, where the free energy corresponds to the color bar, and (B) projection of the free energy (kcal/mol) vs CV1.

Although being transition metals, the zinc ions have been shown to form ionic bonds, thus subtracting charge density to coordinating ligands, suitably polarizing the reactants,[22] and stabilizing the TSs. Consistently, in CPSF73 the analysis of the dynamical RESP charges[29] along RC reveals a decrease of the O_{nuc}@OH^- charge as the nucleophile approaches the **TS**, owing to its persistent interaction with the metal ions and the incipient formation of its bond with the scissile phosphate. Complementarily, the Nε@His396/$O_{3'}$@C-1 charge decreases/increases near the **TS**, indicating the stabilization of the leaving group by His396 (Figure S8). At the **TS**, H-Nδ@His396 engages in an extended H-bond network, involving Glu204 and H-Oε1@Asp179 (Figure S6), which ultimately coordinates the M2 Zn^{2+} ion with its Oε2@Asp159 atom. Strikingly, this

metal appears to aptly act as an electron sink, contributing to activate of the general acid via this extended H-bond network.

In this catalytic mechanism Glu204 takes an active role in orienting the general acid His396, and, by H-bonding to the Zn^{2+} (M2)-coordinated Asp179 residue, it even contributes to polarize and increase the acidity of His396 (the general acid) (Figure 3). This is consistent with mutational studies pinpointing the vital role of the conserved His396 and Glu204 residues for CPSF73 catalysis.[30,31-32]

To further assess the importance of the Glu204 and His396 residues we also considered them in neutral form ($Glu204^0$ and $His396^0$) in the second model. Moreover, we assumed that a pre-activated hydroxyl nucleophile initially binds the Zn^{2+} ions, diffusing from the bulk water. As a result, Asp179 remains in its deprotonated form. We thus investigated the 3'-end cleavage reaction mechanism considering this second set of protonation states (i.e. $Glu204^0$, $Asp179^{-1}$ and $His396^0$). Even in this model, a proton is shared between Oε@Glu204 and Nδ@His396 (Figure S2C), being now prevalently localized on the first. Nevertheless, due to the presence of an ionized $Asp159^{-1}$, the extended H-bond network, heading from His396 towards Zn^{2+} M2, breaks. As a result, His396 can no longer effectively protonate the leaving group (Figure S9). This leads to a significant increase of the $\Delta F^{\#}$ reaction to 31.7 ± 2.3 kcal/mol (Figures S9A and B), further providing a rationale for the experimentally observed critical role of Glu204 in the CPSF73 catalysis.[4,33]

Stunningly, at variance to most DNA/RNA processing enzymes featuring a two-Mg^{2+}-ions motif, the β-CASP nucleases encompass a two-Zn^{2+}-ions catalytic scaffold, typical of MβL enzymes, and commonly exploited to cleave the β-lactam moiety of antibiotics. It is therefore unclear why CPSF73 rely on this rather atypical catalytic motif to perform RNA cleavage. It has been proposed that the CPSF73 scaffold, composed by the β-CASP and MβL domains, is instrumental to selectively recognize the 3'-end of pre-mRNA.[2] The use of a MβL domain may therefore be dictated by structural/selectivity requirements. Nevertheless, the MβL's active site is endowed with several histidine residues, which impose the recruitment of Zn^{2+} ions due to the extraordinary histidine's affinity towards this metal ion (see the analysis of metal binding sites in enzymes in Figure S10). In this scenario, it is tempting to suggest that CPSF73, after having selectively recognized pre-mRNA, adapts to efficiently perform its 3'-end cleavage trough a two-Zn^{2+}-ions aided catalysis.

In spite of the different metal type content, astonishing commonalities appear between two-Zn^{2+}/Mg^{2+}-ions dependent nucleases. Ostensibly, although being transition metals, the Zn^{2+} ions have been demonstrated to establish mostly ionic bonds, with the coordination bonds character being contributed by the 4s orbitals.[34] Hence, their geometry and coordination number are prevalently imposed by the protein environment.[35] This provides a rationale for the plastic Zn^{2+} ions coordination sphere observed in CPSF73, which partially resembles that of two-Mg^{2+}-ions aided enzymes. Nevertheless, the stronger Lewis acid character[22] of Zn^{2+} ions (Zn^{2+} have a lower positive charge compared to Mg^{2+} ions, i.e. Zn^{2+} ions have 0.2-0.7 e in CPSF73 (Figure S8) compared to 0.75-0.9 e of Mg^{2+} ions in spliceosome[21-22]) triggers the formation of a bridging hydroxide nucleophile, resulting into a distinctive catalytic architecture from most DNA/RNA processing enzymes.[5] In β-CASP nucleases, the metals bind the O^{Sp} and O^{Rp} oxygens of the scissile phosphate, rather than sharing a bridging O^{Sp} oxygen as in two-Mg^{2+}-ions aided nucleases, assuming an average inter metal distance of ~3.8 Å as compared to ~4.5 of the Mg^{2+} ions enzymes. In spite of these structural differences, the two-Zn^{2+}-ions in β-CASP nucleases activate the nucleophile, stabilize the reactant and the transition state, as in the general two-Mg^{2+}-ions catalysis.[2-3,15] Additionally, one metal also acts as an electron sink, assisting the protonation of the leaving group via a sophisticate H-bond network, which from a Zn^{2+} (M2)-coordinating residue (Asp179) heads to the general acid. Most importantly, the two-Zn^{2+}-ions architecture of CPSF73, expanded by the positively charged His396 residue, fits within an enlarged two-Mg^{2+}-ions catalytic scaffold, where the metals are complemented by nearby second-shell basic mono/divalent ions[20-21,36-37] and/or residues (Figure 5).[38] Indeed, emerging evidences pinpoint that a variety of nucleic acid processing enzymes, previously believed to rely on pure two-Mg^{2+}-ions catalysis,[39-43] harbor at least an additional functionally relevant positive charge, thus sharing with CPSF73, and other two-Zn^{2+}-dependent nucleases, an alike expanded positively charged catalytic architecture (Figure 5).

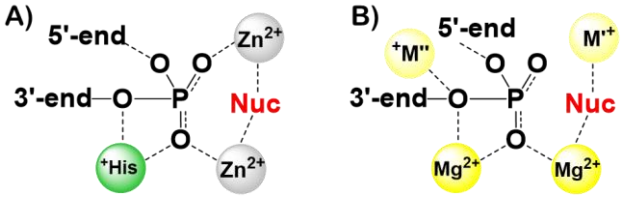


Figure 5. Schematic presentation of the positively charged hotspots in metallo-dependent nucleases relying on (A) a two-Zn^{2+}-ions motif and an ionized histidine and (B) on a two-Mg^{2+}-ions motif and additional positively charged metals (M' and/or M'') or protein residues. Nuc denotes the nucleophile.

CONCLUSIONS

In conclusion, our outcomes resolve the mechanistic intricacies underlying 3'-end processing of pre-mRNA in metazoans, disclosing that CPSF73 effectively promotes 3'-end pre-mRNA hydrolysis via an associative two-Zn^{2+}-ions aided mechanism, featuring a proton transfer from the His396 to the leaving group. The zinc ions, besides activating the hydroxide nucleophile and acting as a molecular slingshot to promote its attack, and stabilizing the **TS,** fulfill the unprecedented role of assisting the protonation of the leaving group via a sophisticate H-bond network heading from a Zn^{2+}-coordination ligand to the general acid. Stunningly, a detailed structural and mechanistic inspection of two-Zn^{2+}-ions and two-Mg^{2+}-ions dependent nucleases illuminate astonishing commonalities between the expanded positive charge architectures of the two enzymatic families,

evoking a common catalytic mechanism. Having CPSF73 recently emerged as an appealing drug-target,[2] our study sets the basis for harnessing the salient molecular traits of its mechanism to devise cutting edge therapeutic approaches against the variety of human diseases associated to a deregulated pre-mRNA 3'-end processing.

COMPUTATIONAL DETAILS

Structural models. All calculations were based on the cryogenic electron microscopy (cryo-EM) structure in which the histone pre-mRNA 3'-end processing machinery (3EPM) from human was trapped at its pre-catalytic state (PDB id: 6V4X).[3] Our models account for all the resolved components of cryo-EM structure, namely proteins SmD3, SmB, Lsm10, Lsm11, SmE, SmF, SmG, CPSF73, CPSF100, Symplekin, and RNA filaments U7 snRNA and pre-mRNA, and two Zn^{2+} ions (Figure 1).

The protonation states of amino acids were determined by Propka,[44] at physiological pH condition of 7.4. Histidine residues, coordinating the Zn^{2+} ions, were protonated at either Nε2 or Nδ1, and His396, which stabilizes the leaving group, was instead considered in its ionized form. However, Glu204 was predicted to have a pKa value of 7.85 in the cryo-EM structure of 3EPM (PDB id: 6V4X) and of 7.3 in the CPSF73 crystal structure (PDB id: 2I7V). Being these values an indication that Glu204 can exist in both neutral and deprotonated forms at physiological pH, we have built two models considering either protonation state of Glu204.

We initially considered a deprotonated Glu204 model ($Glu204^{-1}$, $Asp179^{-1}$ and $His396^{+1}$). In this model we placed a water molecule in the position of the oxygen atom, which in the cryo-EM structure bridges the two Zn^{2+} ions. In this first model the bridging water deprotonates during quantum-classical (QM/MM) MD simulation and Asp179 acquires its proton becoming neutral ($Asp179^{0}$). In the second model, we considered protonated Glu204 and a neutral His396 ($Glu204^{0}$, $His396^{0}$), and Asp179 in its ionized form ($Asp179^{-1}$). In this model, we assume that a pre-activated hydroxide diffuses to the active site from a bulk water, and is bridging the Zn^{2+} ions. Each model investigated was equilibrated by classical and subsequent QM/MM MD simulations.

Classical molecular dynamics simulations. Classical MD simulations were performed with Amber18 PMEMD software package[45] to prepare a reactive adduct suitable to undergo QM/MM MD[46-47] calculations. The AMBER-ff14SB force field (FF)[16] was employed for proteins, while the ff99+bsc0+χOL3 FF was used for RNAs,[17] since these are, overall, still the most validated FFs for protein/RNA systems,[48-49] and showed reliable results in our previous simulation studies of the spliceosome and group II Intron ribozyme.[20,36-37] We remark that this FF has documented shortcomings, but these may not become apparent in the limited time frame of the simulations performed in this study.[50-51] Zn^{2+} ions were described with the 12-6 LJ nonbonded model[52] to account for the plastic Zn^{2+} coordination sphere comprising 5/6 ligands in the Cryo-EM structure and the Na^{+} ion parameters were taken from Joung and Cheatham FF for monovalent ions.[53] The water molecules were included considering the superposition of the crystal structure of human CPSF73 protein (PDB id: 2I7V) to the same protein of the whole 3EPM considered here.[4] The system was solvated by a 12 Å layer of TIP3P[54] water molecules leading to a box of 135 x 198 x 149 $Å^3$ containing, besides the proteins and the nucleic acids, two Zn^{2+} ions and 72 Na^{+} counter ions counting up to 350 895 atoms (Table S1). The positions of the Na^{+} ions were determined on the basis of the system's electrostatic potential.

After the initial minimization, the system was heated up to 300K over 10 ns, while imposing positional restraints of 250 kcal/mol$Å^2$ on the heavy atoms. Subsequently, restraints were slowly removed and 200 ns of MD on the isothermal-isobaric ensemble *(*NPT) ensemble using periodic boundary condition*.* The temperature control (300K) was performed by Langevin thermostat[55] with a collision frequency of 1 ps^{-1}, and pressure control (1 atm) was accomplished by Berendsen barostat.[56] To preserve the experimentally determined geometry of the zinc coordination sphere due to the imperfect nature of FF for transition metal ions, we applied the distance restraints to ligand atoms of 32 kcal/mol$Å^2$ for M1 and M2 Zn^{2+} ions through the whole productive MD run, whereas all other atoms were unrestrained. The SHAKE algorithm[57] was used to constrain the bonds involving hydrogen atoms and the particle mesh Ewald method[58] to account for long-range electrostatic interactions with a cutoff of 10 Å. An integration time step of 2 fs was used.

QM/MM molecular dynamics simulations. After the initial model underwent classical MD simulations, the resulting structure was relaxed by means of unconstrained 10 ps of QM/MM Born - Oppenheimer MD simulations performed with the CP2K 6.1 program.[28] The QM region comprised the phosphate backbone of the substrate nucleobases A-1, C1 and sugar of A-1, Zn-coordinating endonuclease CPSF73 residues His71, His73, His76, His158, His418, Asp75, Asp179 and the leaving group stabilizing residue His396 together with Glu204, the nucleophile (water molecule or hydroxide ion) and two Zn^{2+} ions (109 atoms), while the rest of the system was treated at MM level with the same FF of the classical MD simulations. The QM region was simulated in a cubic box with sides of 24 Å and described at the DFT-BLYP[24] level by employing a dual Gaussian-type/plane waves basis set (GPW).[59] We employed a double zeta (MOLOPT) basis set,[60] along with an auxiliary PW basis set with a density cutoff of 320 Ry and Goedecker–Teter–Hutter (GTH) pseudopotentials.[61] This level of theory has been successfully employed in QM/MM MD simulations of biomolecules.[46,62] The valences of terminal QM atoms were saturated by using capping hydrogen atoms. All QM/MM MD simulations were performed using an integration time step of 0.5 fs in the NVT ensemble. Constant temperature was maintained by employing a Nosé–Hoover thermostat.[63] Dynamical RESP[29] charges were calculated using CP2K 6.1 program[28] for each reaction coordinate and averaged over the equilibrated part of the trajectories.

Free energy calculations

QM/MM Thermodynamic integration. The free energy profiles were obtained by integrating the constraint force acting along the selected distance(s) chosen as reaction coordinate (RC).[64] In details, Helmholtz free energy profile was calculated with the blue moon ensemble method.[26] The systems were simulated for 5-7 ps for each value of the RC. Average forces were calculated from data collected over the last 3 ps of each trajectory, similarly to previous studies.[7,19-21] The reaction mechanism was sampled through the 18 sequential windows. The RC was defined as the difference of forming (O_{nuc}@OH^- – P@C1) and breaking bond (P@C1 – $O^{3'}$@A-1) distances (Figure 2). Both, the first and second models were simulated with the same settings as described above. As a result, ~ 200 ps of cumulative biased and unbiased QM/MM MD simulations have been done in this study. The estimated standard error on each simulated window of free energy profile was calculated by error propagation analysis and overall error on the free energy barrier was assessed as the sum of the standard errors of all subjected window, as performed in previous studies. The resulting standard deviation is consistent with other QM/MM MD studies based on the same free energy methodology.[20-21,65]

QM/MM Metadynamics. The relaxed structure from 10 ps of QM/MM MD simulations of the first model was used as a starting point to additionally evaluate reaction mechanism and the free energy barrier by performing metadynamics (MTD) simulation, as implemented in the CP2K code.[28] MTD was performed considering two collective variables (CVs); CV1, similar to thermodynamic integration simulations, consisted of the difference of forming bond (O_{nuc}@OH^- – P@C1) and the breaking bond (P@C1 - $O^{3'}$@A-1) distances of the scissile phosphate, while CV2 accounted the proton transfer from the histidine to the leaving group (H-Nε@His396 - $O^{3'}$@A-1). Gaussian hills with a height of 0.6 kcal mol^{-1} and widths of 0.40 and 0.30 Å for CV1 and CV2, respectively, were added every 15 fs. Harmonic walls between -4 – 4 Å were used for CV1 and at 4 Å for CV2, respectively, to restrict the exploration of the free energy surface around the reactive region. The simulation was concluded after deposition of 823 hills, when the re-crossing event from reactants to products and back occurred 3 times. The standard deviation of two different time averages of the biased potential in the first and second part of the converged interval of the MTD simulation was used to estimate the standard error of the free energy profile.[66]

AUTHOR INFORMATION

Corresponding Author

*alessandra.magistrato@sissa.it

ASSOCIATED CONTENT

Supporting Information. Supporting Figures, Tables and Movies. This information is available free of charge on the ACS Publications website.

ACKNOWLEDGMENT

JB thanks the Slovenian Research Agency (funding no. P1-0017); AM thanks Italian Association for Cancer Research (IG 24514) for financial support.

REFERENCES

(1) Borišek, J.; Casalino, L.; Saltalamacchia, A.; Mays, S. G.; Malcovati, L.; Magistrato, A. Atomic-Level Mechanism of Pre-mRNA Splicing in Health and Disease. *Acc. Chem. Res.* **2021**, *54*, 144-154.

(2) Sun, Y. D.; Hamilton, K.; Tong, L. Recent molecular insights into canonical pre-mRNA 3 '-end processing. *Transcription* **2020**, *11*, 83-96.

(3) Sun, Y. D.; Zhang, Y. X.; Aik, W. S.; Yang, X. C.; Marzluff, W. F.; Walz, T.; Dominski, Z.; Tong, L. Structure of an active human histone pre-mRNA 3 '-end processing machinery. *Science* **2020**, *367*, 700-703.

(4) Mandel, C. R.; Kaneko, S.; Zhang, H. L.; Gebauer, D.; Vethantham, V.; Manley, J. L.; Tong, L. Polyadenylation factor CPSF-73 is the pre-mRNA 3 '-end-processing endonuclease. *Nature* **2006**, *444*, 953-956.

(5) Steitz, T. A.; Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6498-6502.

(6) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. *Acc. Chem. Res.* **2015**, *48*, 220-228.

(7) Sgrignani, J.; Magistrato, A. QM/MM MD Simulations on the Enzymatic Pathway of the Human Flap Endonuclease (hFEN1) Elucidating Common Cleavage Pathways to RNase H Enzymes. *ACS Catal.* **2015**, *5*, 3864-3875.

(8) Sgrignani, J.; Magistrato, A. The Structural Role of Mg2+ Ions in a Class I RNA Polymerase Ribozyme: A Molecular Simulation Study. *J. Phys. Chem. B* **2012**, *116*, 2259-2268.

(9) Rehfeld, A.; Plass, M.; Krogh, A.; Friis-Hansen, L. Alterations in polyadenylation and its implications for endocrine disease. *Front. Endocrinol.* **2013**, *4*, 53.

(10) Curinha, A.; Braz, S. O.; Pereira-Castro, I.; Cruz, A.; Moreira, A. Implications of polyadenylation in health and disease. *Nucleus* **2014**, *5*, 508-519.

(11) Ouyang, J. M.; Sun, W. M.; Xiao, X. S.; Li, S. Q.; Jia, X. Y.; Zhou, L.; Wang, P. F.; Zhang, Q. J. CPSF1 mutations are associated with early-onset high myopia and involved in retinal ganglion cell axon projection. *Hum. Mol. Genet.* **2019**, *28*, 1959-1970.

(12) Ross, N. T.; Lohmann, F.; Carbonneau, S.; Fazal, A.; Weihofen, W. A.; Gleim, S.; Salcius, M.; Sigoillot, F.; Henault, M.; Carl, S. H.; Rodriguez-Molina, J. B.; Miller, H. R.; Brittain, S. M.; Murphy, J.; Zambrowski, M.; Boynton, G.; Wang, Y.; Chen, A.; Molind, G. J.; Wilbertz, J. H.; Artus-Revel, C. G.; Jia, M.; Akinjiyan, F. A.; Turner, J.; Knehr, J.; Carbone, W.; Schuierer, S.; Reece-Hoyes, J. S.; Xie, K.; Saran, C.; Williams, E. T.; Roma, G.; Spencer, M.; Jenkins, J.; George, E. L.; Thomas, J. R.; Michaud, G.; Schirle, M.; Tallarico, J.; Passmore, L. A.; Chao, J. A.; Beckwith, R. E. J. CPSF3-dependent pre-mRNA processing as a druggable node in AML and Ewing's sarcoma. *Nat. Chem. Biol.* **2020**, *16*, 50-59.

(13) Kakegawa, J.; Sakane, N.; Suzuki, K.; Yoshida, T. JTE-607, a multiple cytokine production inhibitor, targets CPSF3 and inhibits pre-mRNA processing. *Biochem. Biophys. Res. Commun.* **2019**, *518*, 32-37.

(14) Sonoiki, E.; Ng, C. L.; Lee, M. C. S.; Guo, D. H.; Zhang, Y. K.; Zhou, Y.; Alley, M. R. K.; Ahyong, V.; Sanz, L. M.; Lafuente-Monasterio, M. J.; Dong, C.; Schupp, P. G.; Gut, J.; Legac, J.; Cooper, R. A.; Gamo, F. J.; DeRisi, J.; Freund, Y. R.; Fidock, D. A.; Rosenthal, P. J. A potent antimalarial benzoxaborole targets a Plasmodium falciparum cleavage and polyadenylation specificity factor homologue. *Nat Commun.* **2017**, *8*.

(15) Yang, W. Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011**, *44*, 1-93.

(16) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory. Comput.* **2015**, *11*, 3696-3713.

(17) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817-3829.

(18) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988**, *38*, 3098-3100.

(19) Simona, F.; Magistrato, A.; Dal Peraro, M.; Cavalli, A.; Vila, A. J.; Carloni, P. Common mechanistic features among metallo-beta-lactamases: a computational study of Aeromonas hydrophila CphA enzyme. *J. Biol. Chem.* **2009**, *284*, 28164-28171.

(20) 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*, 10374-10377.

(21) Borišek, J.; Magistrato, A. All-Atom Simulations Decrypt the Molecular Terms of RNA Catalysis in the Exon-Ligation Step of the Spliceosome. *ACS Catal.* **2020**, *10*, 5328-5334.

(22) Dudev, T.; Lim, C. Competition among Metal Ions for Protein Binding Sites: Determinants of Metal Ion Selectivity in Proteins. *Chem. Rev.* **2014**, *114*, 538-556.

(23) Huang, W. T.; Huang, Y.; Xu, J.; Liao, J. L. How Does the Spliceosome Catalyze Intron Lariat Formation? Insights from Quantum Mechanics/Molecular Mechanics Free-Energy Simulations. *J. Phys. Chem. B* **2019**, *123*, 6049-6055.

(24) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B.* **1988**, *37*, 785-789.

(25) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817-3829.

(26) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **1989**, *156*, 472-477.

(27) de la Sierra-Gallay, I. L.; Pellegrini, O.; Condon, C. Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z. *Nature* **2005**, *433*, 657-661.

(28) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: atomistic simulations of condensed matter systems. *WIRES Comput. Mol. Sci.* **2014**, *4*, 15-25.

(29) Laio, A.; VandeVondele, J.; Rothlisberger, U. D-RESP: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations. *J. Phys. Chem. B* **2002**, *106*, 7300-7307.

(30) Ryan, K.; Calvo, O.; Manley, J. L. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* **2004**, *10*, 565-573.

(31) Callebaut, I.; Moshous D Fau - Mornon, J.-P.; Mornon Jp Fau - de Villartay, J.-P.; de Villartay, J. P. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family.

(32) Dominski, Z.; Carpousis, A. J.; Clouet-d'Orval, B. Emergence of the β-CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. *Biochim. Biophys. Acta* **2013**, *1829*, 532-551.

(33) Ryan, K.; Calvo, O.; Manley, J. L. Evidence that polyadenylation factor CPSF-73 is the rnRNA 3 ' processing endonuclease. *RNA* **2004**, *10*, 565-573.

(34) Krezel, A.; Maret, W. The biological inorganic chemistry of zinc ions. *Arch. Biochem. Biophys.* **2016**, *611*, 3-19.

(35) Daniel, A. G.; Farrell, N. P. The dynamics of zinc sites in proteins: electronic basis for coordination sphere expansion at structural sites. *Metallomics* **2014**, *6*, 2230-2241.

(36) 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*, 6584-6589.

(37) Saltalamacchia, A.; Casalino, L.; Borišek, J.; Batista, V. S.; Rivalta, I.; Magistrato, A. Decrypting the Information Exchange Pathways across the Spliceosome Machinery. *J. Am. Chem. Soc.* **2020**, *142*, 8403–8411.

(38) Genna, V.; Colombo, M.; De Vivo, M.; Marcia, M. Second-Shell Basic Residues Expand the Two-Metal-Ion Architecture of DNA and RNA Processing Enzymes. *Structure* **2018**, *26*, 40-50.

(39) Ivanov, I.; Tainer, J. A.; McCammon, J. A. Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1465-1470.

(40) Prieto, J.; Redondo, P.; Merino, N.; Villate, M.; Montoya, G.; Blanco, F. J.; Molina, R. Structure of the I-SceI nuclease complexed with its dsDNA target and three catalytic metal ions. *Acta Crystallogr. F* **2016**, *72*, 473-479.

(41) AlMalki, F. A.; Flemming, C. S.; Zhang, J.; Feng, M.; Sedelnikova, S. E.; Ceska, T.; Rafferty, J. B.; Sayers, J. R.; Artymiuk, P. J. Direct observation of DNA threading in flap endonuclease complexes. *Nat. Struct. Mol. Biol.* **2016**, *23*, 640-646.

(42) Samara, N. L.; Yang, W. Cation trafficking propels RNA hydrolysis. *Nat. Struct. Mol. Biol.* **2018**, *25*, 715-721.

(43) Yang, W.; Weng, P. J.; Gao, Y. A new paradigm of DNA synthesis: three-metal-ion catalysis. *Cell. Biosci.* **2016**, *6*, 51.

(44) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *J. Chem. Theory. Comput.* **2011**, *7*, 525-537.

(45) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E. AMBER 2018 (University of California, San Francisco). **2018**.

(46) 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*, 6217-6263.

(47) Janoš, P.; Spinello, A.; Magistrato, A. All-atom simulations to studying metallodrugs/target interactions. *Curr. Opin. Chem. Biol.* **2020**, *61*, 1-8.

(48) Sponer, J.; Krepl, M.; Banas, P.; Kuhrova, P.; Zgarbova, M.; Jurecka, P.; Havrila, M.; Otyepka, M. How to understand atomistic molecular dynamics simulations of RNA and protein-RNA complexes? *WIREs RNA* **2017**, *8*.

(49) Cesari, A.; Bottaro, S.; Lindorff-Larsen, K.; Banáš, P.; Šponer, J.; Bussi, G. Fitting Corrections to an RNA Force Field Using Experimental Data. *J. Chem. Theory Comput.* **2019**, *15*, 3425-3431.

(50) Hanke, C. A.; Gohlke, H. Force Field Dependence of Riboswitch Dynamics. *Methods Enzymol.* **2015**, *553*, 163-191.

(51) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E1346-e1355.

(52) Li, P. F.; Song, L. F.; Merz, K. M. Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883-895.

(53) Joung, I. S.; Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020-9041.

(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(55) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **1992**, *32*, 523-535.

(56) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

(57) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

(58) Harvey, M. J.; De Fabritiis, G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. *J. Chem. Theory. Comput.* **2009**, *5*, 2371-2377.

(59) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103-128.

(60) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.

(61) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703-1710.

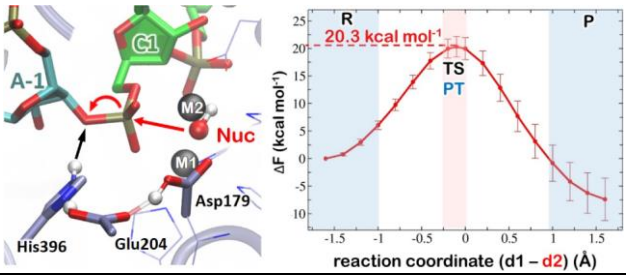
(62) Sgrignani, J.; Iannuzzi, M.; Magistrato, A. Role of Water in the Puzzling Mechanism of the Final Aromatization Step Promoted by the Human Aromatase Enzyme. Insights from QM/MM MD Simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2218-2226.

(63) Nose, S. A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511-519.

(64) Sprik, M.; Ciccotti, G. Free energy from constrained molecular dynamics. *J. Chem. Phys.* **1998**, *109*, 7737-7744.

(65) 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**, 13596-13605.

(66) Bisha, I.; Rodriguez, A.; Laio, A.; Magistrato, A. Metadynamics Simulations Reveal a Na+ Independent Exiting Path of Galactose for the Inward-Facing Conformation of vSGLT. *PLoS Comp. Biol.* **2014**, *10*, e1004017.


C1
A-1
M2
Nuc
M1
Asp179
His396
Glu204
R
P
20.3 kcal mol-1
TS
PT
ΔF (kcal mol-1)
reaction coordinate (d1 – d2) (Å)