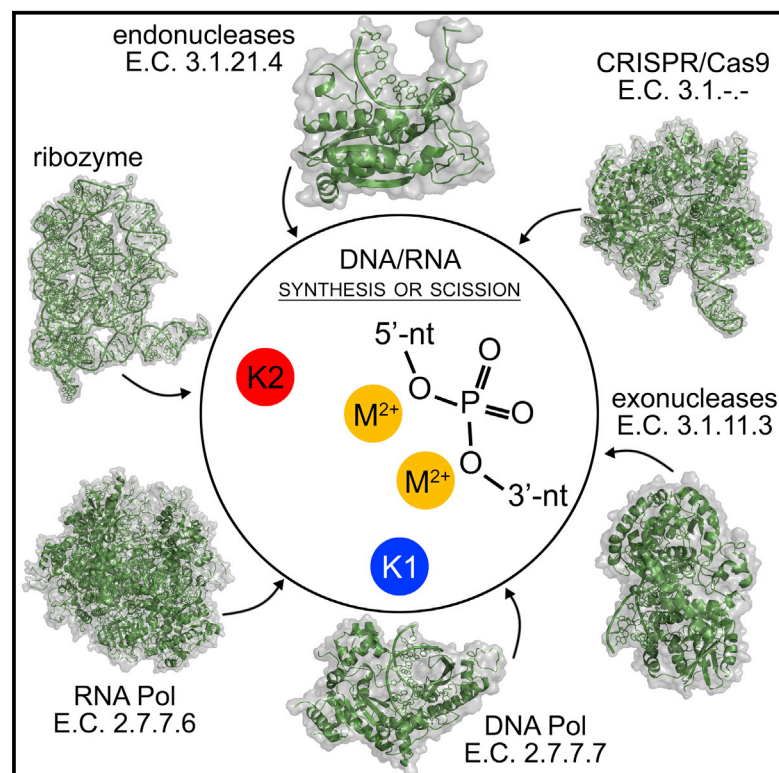


Structure

Second-Shell Basic Residues Expand the Two-Metal-Ion Architecture of DNA and RNA Processing Enzymes

Graphical Abstract



Authors

Vito Genna, Matteo Colombo,
Marco De Vivo, Marco Marcia

Correspondence

marco.devivo@iit.it (M.D.V.),
mmarcia@embl.fr (M.M.)

In Brief

Genna et al. reveal structural conservation of basic amino acids and monovalent cations in the active site of two-metal-ion enzymes and ribozymes, which catalyze the vital biological reactions of DNA and RNA synthesis and scission. The work offers insights into fundamental two-metal-ion machineries such as CRISPR/Cas9, the spliceosome, and DNA/RNA polymerases.

Highlights

- Basic residues in the active site of two-metal-ion enzymes are structurally conserved
- These residues are also conserved in evolution
- Mutagenesis suggests these residues may exert an effect on DNA- and RNA processing
- Our work offers insights into CRISPR/Cas9, spliceosome, and DNA/RNA polymerases



Second-Shell Basic Residues Expand the Two-Metal-Ion Architecture of DNA and RNA Processing Enzymes

Vito Genna,^{1,4} Matteo Colombo,^{2,4} Marco De Vivo,^{1,3,*} and Marco Marcia^{2,5,*}

¹Laboratory of Molecular Modelling & Drug Discovery, Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genoa, Italy

²European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, Grenoble 38042, France

³IAS-5/INM-9 Computational Biomedicine Forschungszentrum Jülich, Wilhelm-Johnen-Straße, 52428 Jülich, Germany

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: marco.devivo@iit.it (M.D.V.), mmarcia@embl.fr (M.M.)

<https://doi.org/10.1016/j.str.2017.11.008>

SUMMARY

Synthesis and scission of phosphodiester bonds in DNA and RNA regulate vital processes within the cell. Enzymes that catalyze these reactions operate mostly via the recognized two-metal-ion mechanism. Our analysis reveals that basic amino acids and monovalent cations occupy structurally conserved positions nearby the active site of many two-metal-ion enzymes for which high-resolution (<3 Å) structures are known, including DNA and RNA polymerases, nucleases such as Cas9, and splicing ribozymes. Integrating multiple-sequence and structural alignments with molecular dynamics simulations, electrostatic potential maps, and mutational data, we found that these elements always interact with the substrates, suggesting that they may play an active role for catalysis, in addition to their electrostatic contribution. We discuss possible mechanistic implications of this expanded two-metal-ion architecture, including inferences on medium-resolution cryoelectron microscopy structures. Ultimately, our analysis may inspire future experiments and strategies for enzyme engineering or drug design to modulate nucleic acid processing.

INTRODUCTION

Enzymatic cleavage and formation of phosphodiester bonds in DNA and RNA is central to life and health. These reactions allow processing nucleic acids during DNA replication, DNA recombination, DNA repair, transcription, splicing, and defense from pathogens (Yang et al., 2006). All these key chemical processes are controlled by vital cellular machineries including both protein and RNA enzymes (Strater et al., 1996), such as endo- and exonucleases, DNA and RNA polymerases, and ribozymes (i.e., group II intron and spliceosome) (Table 1). Independent of their biopolymer scaffold, these cellular machineries display a surprising degree of structural similarity (Livesay et al., 2003;

Yang et al., 2006), suggesting convergence of their enzymatic reaction mechanism. Indeed, Steitz and Steitz (1993) first described the so-called two-metal-ion mechanism, which is shared among most nucleic acid-processing enzymes. According to this general mechanism, divalent metal ions (typically Mg, Mn, and Zn) are chelated by acidic groups of strictly conserved catalytic residues and the phosphodiester substrate is properly oriented via three specific binding sites (sites 1, 2, and 3) (Steitz and Steitz, 1993). Then, one metal (usually referred to as metal A, M_A) stabilizes the activated nucleophile, while the second metal (metal B, M_B) assists the release of the leaving group. Together, M_A and M_B stabilize the trigonal bipyramidal intermediate formed at the transition state (Steitz and Steitz, 1993). Notably, some enzymes such as the $\beta\beta\alpha$ -Me and HNH nucleases may operate with only one catalytic metal ion (M_A , which is structurally conserved), while a basic amino acid replaces the missing M_B for catalysis (Yang, 2008). In all cases, a first-shell structural architecture centered on two conserved and positively charged elements located in the catalytic site is crucial for efficient DNA and RNA processing (Palermo et al., 2015).

Over the years, 3D structures have revealed additional important catalytic elements in specific classes of two-metal-ion enzymes. For example, in DNA polymerases a third transiently bound divalent cation and a positively charged, highly flexible residue (Arg61 in the human DNA polymerase- η [Pol- η]) facilitate product release (Gao and Yang, 2016; Genna et al., 2016a; Nakamura et al., 2012). However, so far no catalytic element other than M_A - M_B and their coordinating acidic residues appeared conserved across different classes of two-metal-ion enzymes.

In this respect, recent crystal structures of self-splicing group II intron ribozymes revealed that, besides M_A - M_B and surrounding conserved nucleotides, two potassium ions (K1 and K2) located in the vicinity of the active site are essential for catalysis (Figures 1A and 2; Table 1) (Marcia et al., 2013a, 2013b; Marcia and Pyle, 2012, 2014). Interestingly, the active site of this RNA enzyme is strikingly similar to that of the protein enzyme BamHI, a type II restriction endonuclease, despite their fundamentally different biopolymer scaffold (Marcia and Pyle, 2012). This surprising structural similarity at the active site suggested the possible existence of enzymatic amino acid counterparts of K1 and K2 in BamHI (Marcia and Pyle, 2012). The presence of K1- and K2-like amino acids in protein enzymes that process DNA and RNA would infer a more



Table 1. Representative Two-Metal-Ion Enzymes that Possess Positively Charged Elements in the Second Coordination Shell of the Active Site

Enzyme	PDB ID	Resolution (Å)	Enzymatic Classification (E.C.)	K1	d1 (Å) ^a	d-ac (Å) ^b	Mutants	Functional Defect	K2	d2 (Å) ^a	d-Sub (Å) ^c	Mutants	Functional Defect
Group II intron (<i>O. iheyensis</i>)	4FAR	2.86	ribozyme	K1	3.83	2.60	K to Li, Na, Cs (Marcia and Pyle, 2012)	distortion of active site, splicing defect	K2	7.93	2.78	K to Li, Na, Cs (Marcia and Pyle, 2012)	distortion of active site, splicing defect
BamHI (<i>B. amyloliquefaciens</i>)	2BAM	2.00	3.1.21.4	Lys61	5.87	via Tyr65 (3.08 Å)	computational electrostatic model (Sun et al., 2003)	NA	Lys126	6.23	2.90	computational electrostatic model (Sun et al., 2003)	NA
Exo-λ (<i>Escherichia virus lambda</i>)	4WUZ/3SM4	2.38/1.88	3.1.11.3	Lys131	4.44	2.97	K131A (Zhang et al., 2011)	no exonuclease activity	Arg28	9.55	2.96	R28A (Zhang et al., 2011)	no exonuclease activity
Cas9 (<i>S. pyogenes</i>)	4CMQ/5F9R	3.09/3.40	3.1.--	Lys974	6.05	3.08	K974A (Zhang et al., 2016)	altered substrate binding and kinetics	Lys968	8.82	3.79	K968A (Zhang et al., 2016)	altered substrate binding and kinetics
DNA Pol-η (<i>H. sapiens</i>)	4ECS	1.95	2.7.7.7	Lys231	3.98	2.68	K279A in yeast (Johnson et al., 2003) ^d	decreased efficiency	Lys224	4.71	3.43	NA	NA
FluA, endonuclease PA subunit (influenza A virus)	2W69	2.05	2.7.7.48	His41	2.91	3.47	NA		Arg84	6.74	x ^e	NA	
				Lys134	4.17	2.77	K134A (Crepin et al., 2010)	reduced thermal stability and no activity	Lys137	13.33	x ^e	K137A (Crepin et al., 2010)	decreased activity

The complete list of enzymes analyzed in this work is reported in [Tables S1](#) and [S2](#).

^aClosest distance between the ion/amino acid corresponding to K1 or K2 and the M_A-M_B center.

^bClosest distance between the ion/amino acid corresponding to K1 and the acidic residues that chelate M_A-M_B.

^cClosest distance between the ion/amino acid corresponding to K2 and the substrate.

^dSee [Table S1](#).

^eNo substrate in the active site; NA, not available.

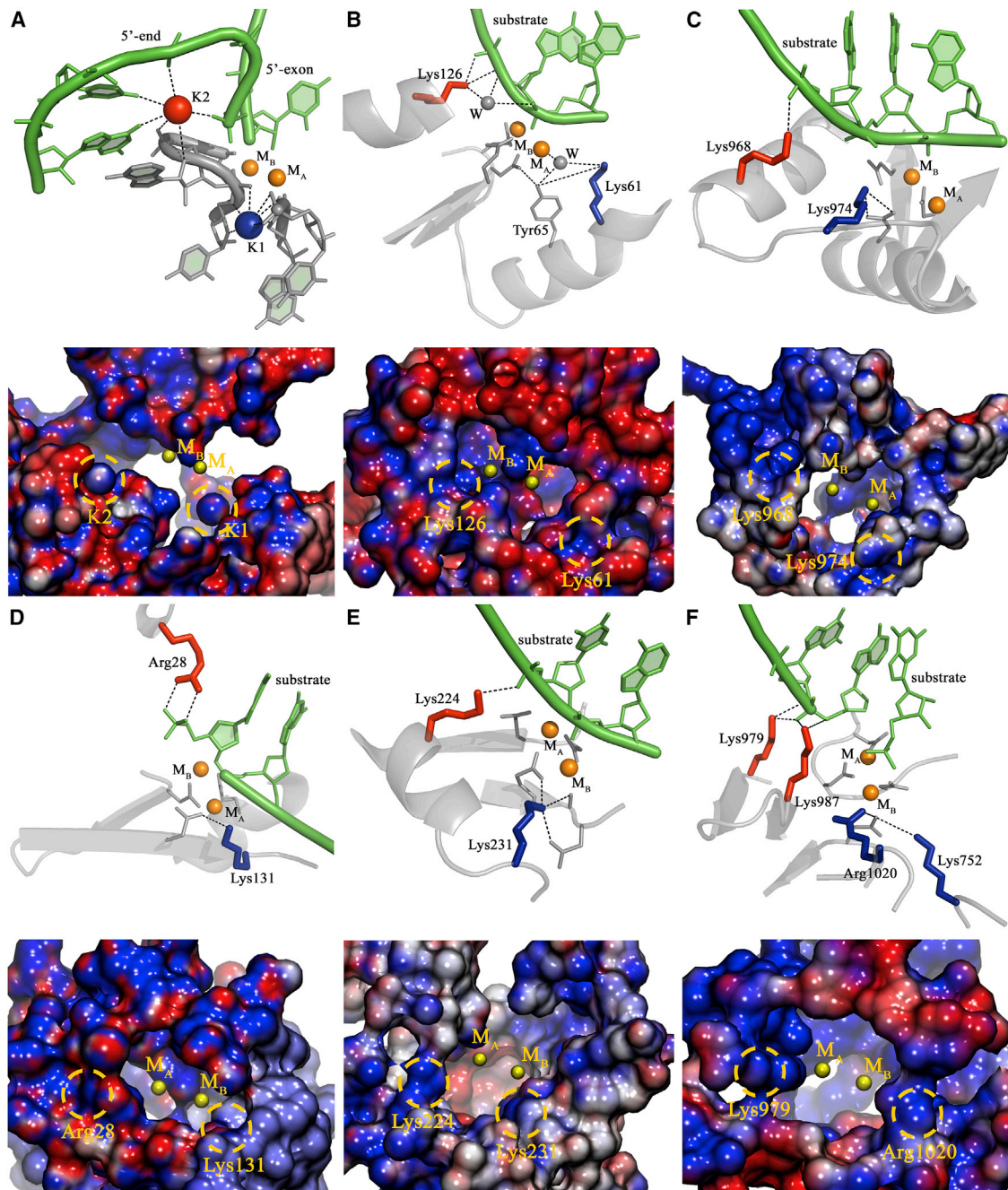


Figure 1. Structural Comparison of the Catalytic Sites of Representative Two-Metal-Ion Enzymes

(A) Group II intron (PDB: 4FAR).

(B) Restriction endonuclease BamHI (PDB: 2BAM).

(C) SpyCas9 (RuvC active site modeled from PDB: 4CMQ and 5F9R).

(D) Exonuclease Exo- λ (modeled from PDB: 4WUZ and 3SM4).

(E) DNA polymerase Pol- η (PDB: 4ECS).

(F) RNA polymerase Pol-II (PDB: 2E2H, predicted from medium-resolution structures, see Table S2).

Each top panel presents the structure of the active site of the respective molecules in ribbon-stick representation. End each bottom panel presents the molecular surface surrounding the two-metal-ion active sites (radii 15 Å) colored by intensity of electrostatic potential (-10 kT/e [darkest red] to $+10$ kT/e [darkest blue]). Catalytic divalent ions (M_A - M_B) are depicted as orange spheres, substrates are represented in green, and the backbone of the corresponding macromolecule is in gray, with acidic residues coordinating M_A - M_B represented as sticks. K1- and K2-like elements are depicted as spheres (for ions) or sticks (for amino acids) in blue and in red, respectively. "W" indicates water molecules. Black dashed lines indicate ionic or hydrogen-bond interactions established by K1- and K2-like elements with active site residues and reactants. K1-like elements rigidify the residues that coordinate M_A - M_B , while K2-like elements orient the substrates into the active site. All structures have been drawn using PyMOL.

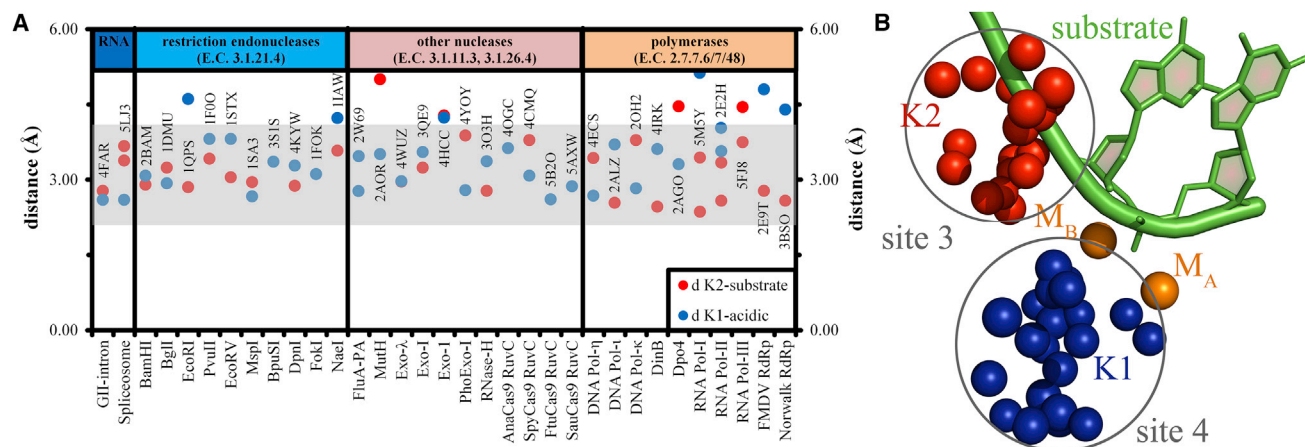


Figure 2. Positioning of K1- and K2-like Elements with Respect to the Two-Metal-Centered Active Site

(A) The plot reports distances in angstroms of K1-like elements from the acidic residues that coordinate M_A - M_B (d K1-acidic, blue dots) and of K2-like elements from the substrate (d K2-substrate, red dots). The gray shade covers the optimal range of distances for hydrogen bonds and ionic interactions (Ippolito et al., 1990). Several outliers correspond to structures solved at low resolution (i.e., PDB: 5FJ8) or with no metals in the active site (i.e., PDB: 1IAW). For enzymes in which K1 contacts acidic residues indirectly (PDB: 2BAM, 1DMU, 1QPS, and 2ALZ), we reported the closest distance to the linking residue (see also Tables 1 and S1). For structures where K2 is present, but the substrate is not resolved in the PDB file (PDB: 3S1S, 1FOK, 4OGC, 5B2O, and 5AXW), we did not plot any data point (see Tables 1 and S1). In exo- λ structure PDB: 4WUZ, the “d K1-acidic” and “d K2-substrate” data points overlap and only the blue dot is actually visible. The names of the enzymes are on the x axis, corresponding PDB codes are indicated on top of every data point. Enzymes are arranged by classes and their respective Enzymatic Classification (E.C.) numbers are indicated on the top of the graph.

(B) Overlap of the structures reported in (A), aligned manually in Coot using the substrate and the two-metal-ion center as a guide. For clarity, substrate (green cartoon) and M_A - M_B center (orange spheres) are represented only for endonuclease BamHI. Blue and red spheres represent the position of the K1- and K2-like elements from all structures, respectively. For clarity, we only represent the position of the potassium ion (for ribozymes), of the epsilon nitrogen atom (for lysine residues), or of the guanidinium carbon atom (for arginine residues), as appropriate. Gray circles represent spheres of a radius of 4.0 Å (K1) and 3.5 Å (K2) and identify sites 4 (K1) and 3 (K2), respectively, as described in Figure 4 and in the main text.

extended set of functional components for the enzymatic processing of nucleic acids through the two-metal-ion mechanism.

Intrigued by this hypothesis, here we examined a broad set of DNA/RNA processing enzymes, comprising splicing machineries such as the group II intron and the spliceosome, nucleases such as the contemporary Cas9 genome-editing tool, and DNA and RNA polymerases. Remarkably, in all these enzymes we identified two positively charged residues with a conserved spatial localization in the second coordination shell of the two-metal-aided catalytic site similar to K1 and K2 in the group II intron ribozyme. These structural elements define a larger two-metal-ion-centered enzymatic structure, which potentially reflects a common strategy of multiple two-metal-ion enzyme classes to ensure fidelity, substrate specificity, and catalytic efficiency for DNA and RNA processing.

RESULTS

Second-Shell Positively Charged Residues Define Geometry and Electrostatic Potential of the Active Site of Two-Metal-Ion Enzymes

In the group II intron, potassium ion K1 anchors the residues that coordinate M_A - M_B , while potassium ion K2 stabilizes the 5'-splice junction before catalysis and the scissile phosphate after catalysis (Marcia and Pyle, 2012, 2014; Marcia et al., 2013b). K1 and K2 produce two discrete regions of positive electrostatic potential (+7.4 kT/e) at 4–8 Å from the active site, which interrupts the negative potential created by first-shell coordinators of M_A - M_B (Figure 1A). Replacement of K1 with smaller

(i.e., lithium and sodium) or bigger (i.e., cesium) ions prevents M_A - M_B binding, thus affecting catalysis (Marcia and Pyle, 2012). In summary, the K1 and K2 ions rigidify the architecture of the M_A - M_B binding pocket, mediate functional conformational changes during the splicing cycle, and stabilize reactants in the pre- and post-catalytic states (Marcia and Pyle, 2012).

Stiffening of the active site, modulation of its electrostatic environment, and orientation of the reactants relative to the M_A - M_B center are needed by two-metal-ion enzymes to ensure specificity in substrate recognition and to augment the fidelity of the reaction (Hanoian et al., 2015; Jeltsch et al., 1993; Kurpiewski et al., 2004; Ramachandrakurup et al., 2016; Warshel et al., 2006). Therefore, we reasoned that counterparts of K1 that rigidify the first coordination shell of M_A - M_B and counterparts of K2 that help orienting the substrates into the active site may exist in other families of two-metal-ion enzymes.

We thus selected two-metal-ion enzymes belonging to six different enzymatic classes, chosen so that there are high-resolution (≤ 3 Å) structures in the PDB having an integer metal ion center for at least one enzyme of every considered class (Table 1). This selection resulted in a collection of 49 3D structures, which we systematically compared to identify putative K1 and K2-like residues in their active sites (Figure 2; Table S1).

Functional Monovalent Cations of the Group II Intron Match Functional Basic Amino Acids of Endo- and Exonucleases, Including Cas9

We first superposed the coordinates of *Oceanobacillus iheyensis* group II intron with those of substrate-bound BamHI. Two

BamHI lysines match the group II intron K1 and K2 ions (Figure 1B). Specifically, Lys61 engages in a hydrogen-bond network with catalytic Asp94 via Tyr65 (Marcia and Pyle, 2012), resembling K1, and Lys126 forms a direct contact with the substrate DNA, analogously to K2 (Table 1). Lys61 and Lys126 generate an electrostatic potential distribution around the active site, analogous to that observed for K1 and K2 ions in the group II intron (+8.5 kT/e; Figure 1B). Notably, Lys61 and Lys126 are strictly conserved in different bacterial species (Data S1) and aid cognate DNA binding via large amplitude motions (Uyar et al., 2011) and favorable Coulombic interactions (Sun et al., 2003) (Table 1).

Interestingly, a similar structural architecture with positively charged residues arranged in the second coordination shell of M_A - M_B can also be observed: (1) in other orthodox PD-(D/E)XK restriction endonucleases (Pingoud and Jeltsch, 2001), such as EcoRI, which produces 5' overhangs, BglI, which produces 3' overhangs, and PvuII, which produces blunt ends (Data S1; Figure S1; Table S1); and (2) in other subfamilies of bacterial restriction endonucleases, such as class IIP endonuclease MspI, class IIG endonuclease BpuSI, class IIM endonuclease DpnI, class IIS endonuclease FokI, and class IIE endonuclease NaeI (Table S1). Such second-shell positively charged residues are highly conserved for all these enzymes (Data S1) and participate in DNA binding and catalysis (Jeltsch et al., 1993; Kurpiewski et al., 2004; Ramachandrakurup et al., 2016) (Table S1). Other bacterial endonucleases, such as MthH, which is involved in DNA repair (Pingoud and Jeltsch, 2001), or viral endonucleases, such as the PA subunit of influenza virus RNA polymerases (FluA, FluB, FluC), which is involved in cap-snatching (Dias et al., 2009), also present evolutionarily conserved residues that impair catalysis if mutated at K1- and K2-like positions nearby their two-metal-ion active sites (Data S1; Figure S1; Table S1).

Second, we identified K1- and K2-like residues in the RuvC domain of Cas9, an RNA-guided DNA bacterial endonuclease associated with the CRISPR type II adaptive immunity system (Jiang and Doudna, 2015) and nowadays largely exploited for genome editing (Wright et al., 2016). A complete picture of the RuvC active site of Cas9 can be obtained by superposition of a *Streptococcus pyogenes* Cas9 (SpyCas9) structure that displays the M_A - M_B center, but no substrates (PDB: 4CMQ), with a SpyCas9 structure that displays the substrate DNA but no divalent metals (PDB: 5F9R) (Jinek et al., 2014). Such superposition suggests that Lys974 and Lys968 occupy analogous structural locations as the K1 and the K2 ions of the group II intron, respectively (Figure 1C). Lys974 and Lys968 also produce two regions of positive electrostatic potential on the surface of SpyCas9 (+9.8 kT/e and +13.0 kT/e, respectively) discontinuing the negative potential (−9.5 kT/e) that surrounds the catalytic metals (Figure 1C). Interestingly, Lys974 and Lys968 have precise counterparts in the Cas9 active site of *Actinomyces naeslundii* (AnaCas9), *Francisella tularensis* (FtuCas9), and *Staphylococcus aureus* (SauCas9), and in RNase-H, which is evolutionarily related to the RuvC domain of Cas9 (De Vivo et al., 2008; Ho et al., 2010; Jiang and Doudna, 2015; Jinek et al., 2014) (Figure S1; Data S1).

Finally, K1- and K2-like residues are also persistently conserved among exonucleases. *Escherichia* virus lambda exonuclease (Exo-λ), which catalyzes 5' to 3' exonucleolytic

cleavage in dsDNA (Pingoud and Jeltsch, 2001), presents Lys131 in a position analogous to the K1 ion and Arg28 in a position analogous to K2 (Figure 1D). Both residues are strictly conserved (Data S1), abolish catalysis if mutated (Zhang et al., 2011) (Table 1), and have parallel counterparts in LHK-exonuclease and RecE (Yang et al., 2011). Counterparts of Lys131 and Arg28 in human exonuclease-I (Exo-I), which catalyzes 3' to 5' exonucleolytic cleavage in dsDNA, are Lys85 and Arg92, respectively. Intriguingly, Arg92 forms a bifurcated H-bond interaction with both scissile and 5' end phosphate groups, as observed for K2 ions in group II introns. Finally, similar structural determinants are also present in prokaryotic exonucleases, such as Exo-1 of the mesophilic host *Escherichia coli* (conserved His181 and Arg165 in K1- and K2-like positions, respectively) and Exo-I of the extremophilic host *Pyrococcus horikoshii* (PhoExo-I, Arg142, and Lys136 in K1- and K2-like positions, respectively).

K1/K2-like Residues Are Also Similarly Located Nearby the Catalytic Site of DNA and RNA Polymerases

Having assessed that K1- and K2-like residues are recurrent in enzymes that catalyze scission of phosphodiester bonds, we then analyzed enzymes that catalyze the synthesis of phosphodiester bonds, such as DNA and RNA polymerases. Prominent classes of DNA and RNA polymerases, which satisfy our selection criteria (3D structures available at a resolution higher than 3 Å and displaying an integer metal ion center, see above), are viral RNA-directed RNA polymerases (RdRps) (Baltimore, 1971) and Y-family DNA polymerases, which are involved in DNA repair (Genna et al., 2016a, 2016b; McCulloch et al., 2004; Nakamura et al., 2012; Patra et al., 2015).

We first considered viral RdRps. Norwalk virus RNA polymerase presents Lys374 in a position equivalent to K1 and Arg392 in a position analogous to K2 (Figure S1). Structure-based sequence alignments suggest that Lys374 and Arg392 are highly conserved in RdRps from *Caliciviridae* and from other single-stranded positive RNA viruses (viral group IV, according to the Baltimore classification [Baltimore, 1971]), i.e., *Picornaviridae* (Data S1) and *Flaviviridae* (Table S1). In other viral groups, the presence of such residues is less evident from sequence alignments, but structural superposition helps identifying putative K1- and K2-like counterparts. For instance, in the RdRps from *Orthomixoviridae* and *Bunyaviridae*, the role of K1 and K2 may be played by residues consistently located in ultraconserved polymerase motifs D, F, and G, which are common to all single-stranded negative RNA viruses (viral group V) and to the reverse transcriptase of retroviruses (viral group VI; Table S1) (Gerlach et al., 2015; Muller et al., 1994).

Second, we analyzed the human Y-family DNA Pol-η. Here, Lys231 anchors M_A - M_B in a position analogous to the K1 ion, while Lys224 orients the phosphate backbone in the proximity of the cleavage site in a position analogous to K2 (Figure 1E). Both Lys231 and Lys224 are strictly conserved (Data S1) and are essential for substrate binding and stabilization (Ummat et al., 2012) (Table 1). For instance, Lys231 in the pre-reactive complex contributes to properly position the incoming substrate aiding catalysis, while in the post-reactive state it counterbalances the negative charge of departing pyrophosphate (Genna et al., 2016a). The surface electrostatic map of Pol-η confirms

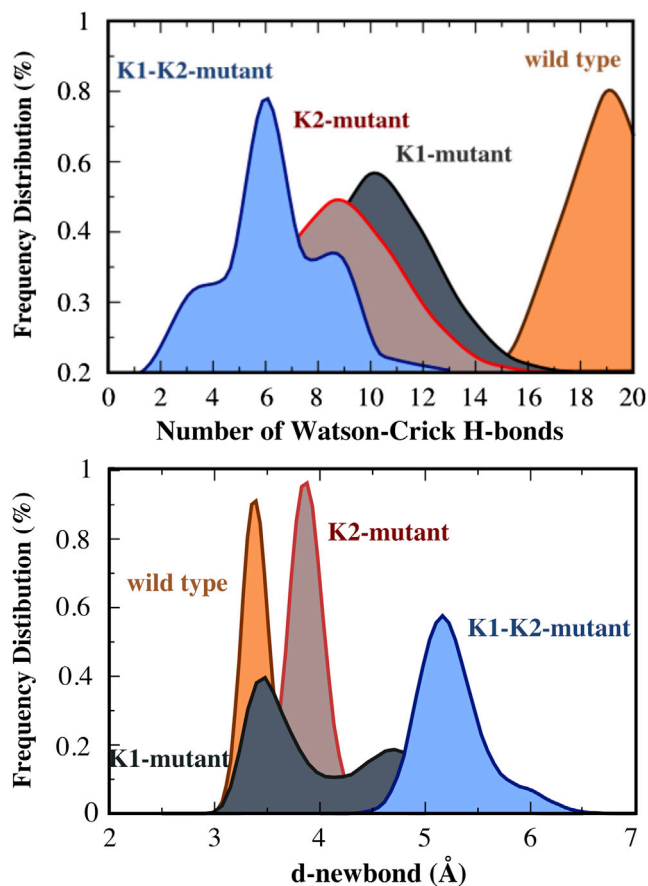


Figure 3. Molecular Dynamics Simulations of DNA Polymerase Pol- η Mutated at the K1- and K2-Sites

Top: frequency distribution of the Watson-Crick inter-strand H-bonds in wild-type Pol- η , and in Pol- η mutated at the K1 site (K231A, K1-mutant), at the K2 site (K224A, K2-mutant), and at both K1 and K2 sites (K224A/K231A, K1-K2-mutant). Bottom: frequency distribution of the length *d-newbond*, defined in the main text (Genna et al., 2016a).

that Lys231 and Lys224 produce discrete regions of positive electrostatic potential at 8–9 Å from the catalytic metals, mimicking the electrostatic effect of K1- and K2-like residues in endo- and exonucleases. Moreover, Lys231 and Lys224 have precise counterparts in other human Y-family DNA polymerases, such as DNA polymerase- ι (Pol- ι) and DNA polymerase- κ (Pol- κ), in yeast DNA polymerases, such as Pol- η from *S. cerevisiae*, and in mesophilic and thermophilic bacterial DNA polymerases, such as *E. coli* DinB and *S. solfataricus* Dbh and DPO4 (Table S1; Figure S2) (Uljon et al., 2004). Human and yeast Pol- η , human Pol- ι , human Pol- κ , *S. solfataricus* Dpo4, and *E. coli* DinB represent a particularly interesting subset of enzymes within our analysis, because all these enzymes share a common ancestor but have diverged over one billion years in evolution. Moreover, all their 3D structures are available at high resolution and in a catalytically active conformation, displaying an integer active site and substrate. Thus, this subset of two-metal-ion enzymes represents a curated dataset to analyze the conservation in structural distribution of K1- and K2-like residues around the two-metal ion active site. A superposition of these

structures remarkably shows that both K1- and K2-like residues localize within 1.5-Å-radius spheres at about 4.5 Å from the two-metal-ion center (Figure S2), thus suggesting that their structural position has been preserved over evolution.

We further assessed the structural role of K1- and K2-like elements in the active site through extensive (~1.5 μ s) classical molecular dynamics (MD) simulations of human wild-type DNA Pol- η in complex with the substrate dsDNA and the incoming dATP, and of three Pol- η mutants derived from such a wild-type system (K231A, K224A, and K224A/K231A mutants, see the STAR Methods). We found that, along our MD simulations, both K231A and K224A mutations lead to displacement of the substrate dsDNA, to a lower number of inter-strand Watson-Crick H-bonds in the dsDNA, and to amplified fluctuations of *d-newbond*, i.e., the distance between the two reactive groups, the 3'-O⁻ of the terminal base of the DNA and the α -phosphorus of dATP (Genna et al., 2016a) (Figures 3 and S3). In the Pol- η K231A mutant, the dsDNA reached a root-mean-square deviation (RMSD) value of 2.5 ± 0.4 Å, which is ~1 Å higher with respect to the wild-type form. Concomitantly, the number of the Watson-Crick inter-strand H-bonds dropped to 11 from the optimal value of 20 observed in the wild-type system (Figure 3). This distortion, which indicates partial dsDNA unfolding, causes the disruption of the Michaelis-Menten complex, as indicated also by fluctuations of *d-newbond*, which reached an average value of 4.0 ± 0.6 (i.e., ~0.7 Å longer if compared with the wild-type system, where *d-newbond* is 3.3 Å [Genna et al., 2016a]). Also, in the Pol- η K224A mutant, we observed similar conformational changes. The dsDNA reached an RMSD value of 3.8 ± 0.3 Å and a reduced number of nine Watson-Crick inter-strand H-bonds. The *d-newbond* length was 3.8 ± 0.1 Å, which suggests a conformation likely less prone to react (Figures 3 and S3). Finally, the most severe distortions were observed in the Pol- η K224A/K231A double mutant, where the dsDNA substrate reached an RMSD value of 3.8 ± 0.8 Å, the inter-strand H-bonds dropped to 6, and *d-newbond* reached an average value of 5.5 ± 0.4 Å after only ~300 ns (Figures 3 and S3). Taken together, these data indicate significant distortions of the dsDNA upon mutation of K1- and/or K2-like elements.

Common Characteristics of K1- and K2-Counterparts in Nucleic Acid-Processing Enzymes

Our work suggests that K1- and K2-like residues are a common feature of nucleases and polymerases. Some K1- and K2-like residues play idiosyncratic roles in specific enzyme classes. For instance, K1-like Lys144 of BglI and Lys131 of Exo- λ activate the reaction nucleophile (Yang et al., 2011), while K2-like Arg61 of Pol- η facilitates pyrophosphate departure after catalysis (Genna et al., 2016a). However, we noted that all K1- and K2-like residues found in the classes of two-metal-ion enzymes considered here share the following characteristics.

First, we show that they occupy analogous structural positions in the second coordination shell of the active site. A superposition of all structures analyzed in our work provides a quantitative assessment of the level of structural conservation of K1- and K2-like elements. All K1-like residues localize within a 4.0-Å-radius sphere at about 6.0 Å from the M_A - M_B center, and all K2-like residues localize within a 3.5-Å-radius sphere at about 8.0 Å from the M_A - M_B center (Figure 2). In particular, the

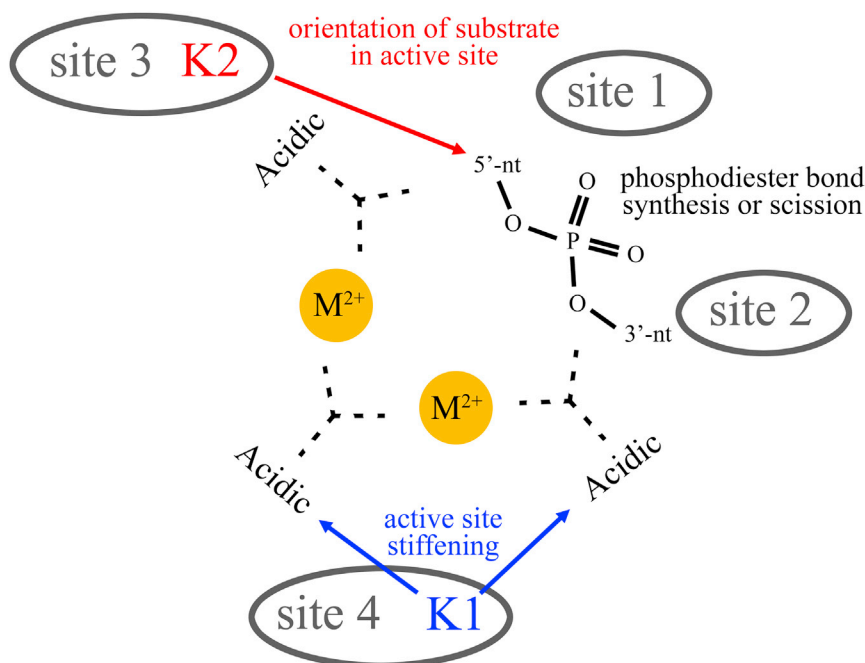


Figure 4. Structural Model Recapitulating Positions and Possible Roles of K1- and K2-like Positive Charges in the Active Site of Two-Metal-Ion Enzymes

In the active site, strictly conserved acidic residues (“acidic”) coordinate two divalent ions (M^{2+}) that catalyze synthesis or scission of phosphodiester bonds in the DNA/RNA substrate. “5'-nt” and “3'-nt” indicate nucleotides on the 5' and 3' side of the reaction site, respectively. In all enzymes discussed in this work, the active site is completed by the presence of two positively charged elements that correspond to the group II intron K1 and K2 ions. These structural elements are either basic amino acids or monovalent cations. These elements likely contribute to rigidify the active site for optimal M^{2+} binding and to orient the substrate for catalysis. Sites 1, 2, and 3 are structural positions previously identified by Steitz and Steitz (1993), and site 4 is a structural site identified as described in the main text (see the Results section). Sites 3 and 4 encompass K2- and K1-like residues, respectively.

K2-like residue forms ionic or hydrogen-bonding interactions with the reaction substrate and/or product (Figure 2; Tables 1 and S1). Within the typical active site geometry of two-metal-ion enzymes (Steitz and Steitz, 1993), K2-like thus occupies “site 3” (Figure 4). Instead, the K1-like residue forms ionic or hydrogen-bonding interactions with residues of the first coordination shell of M_A - M_B (Figure 2; Tables 1 and S1). As such, the K1-like residue occupies a site hereafter named “site 4.” In this expanded two-metal-ion-centered architecture, site 4 is structurally juxtaposed to sites 1 and 2, which flank the scissile phosphate on the 5'- and the 3'-sides, respectively, and it is typically located closer to the 3' end side of the substrate and thus closer to site 2 than to site 1. Relative to the substrate backbone, site 4 is consistently located opposite the nitrogenous base moieties (Figure 4).

Second, as expected for such basic residues, we found that K1- and K2-like analogs produce discrete peaks of second-shell-positive electrostatic potential, which discontinues in a similar manner the negative potential surrounding M_A - M_B (Figure 1). Finally, we demonstrate that K1- and K2-like residues are highly conserved for each enzymatic class (Data S1), and we report evidence from available literature, suggesting that their mutations are associated to functional defects across the entire dataset (Tables 1 and S1).

DISCUSSION

Here, we have analyzed six classes of RNA and protein enzymes that follow a two-metal-ion mechanism of catalysis, including self-splicing group II introns, nucleases, and DNA/RNA polymerases. In our work, we have specifically considered high-resolution structures (<3 Å) that display all reactants and metal ions in the enzymatic active site. By multiple-sequence and structural alignments, molecular modeling, MD simulations, and electrostatic potential maps, we have identified two second-shell and

positively charged elements (basic amino acids or monovalent cations) that are located in the vicinity of the two-metal-aided active site. In all the analyzed structures, despite the different size, shape, composition, and biological function of the respective enzymes, we have found that these elements share four common characteristics: they occupy analogous structural positions, they help modulating both the electrostatics in the surroundings of the active site and substrate stability, they are evolutionarily conserved, and they are associated to functional defects if mutated (see Tables 1, S1, and S2). These observations and results allow distinguishing K1- and K2-like residues from any other basic amino acid present in the proximity of two-metal-centered active sites.

Also, these common properties of the K1- and K2-like residues can serve as a guide to identify analogous residues in other classes of two-metal-ion enzymes, particularly those for which structures can be determined only at medium-low resolution (>3 Å). To exemplify the predictive potential of our analysis of K1- and K2-analogs, we have analyzed the recent 3.9–4.0 Å resolution X-ray and cryoelectron microscopy structures of DNA-directed RNA polymerases (DdRp), which synthesize RNA in eukaryotic cells (Cramer et al., 2001; Tafur et al., 2016), and the 3.6–5.8 Å cryoelectron microscopy structures of the spliceosome, a mega-Dalton machinery which catalyzes nuclear pre-mRNA splicing in eukaryotes and which is evolutionarily linked to the group II intron (Fica et al., 2014; Pyle and Lambowitz, 2006; Sharp, 1991). In these enzymes, we can detect specific basic residues at the putative K1 and K2 sites. In those sites where two basic residues could potentially act as K1 (or K2) analogs, we are presenting both alternatives below (Table S2). Such ambiguity likely derives from the fact that the DdRps and spliceosome structures are still solved at only medium resolution, that ions in the M_A - M_B center are missing in some structures (i.e., RNA polymerase I [Pol-I] and RNA polymerase III [Pol-III]), and that enzymatic assays on mutants are not available yet.

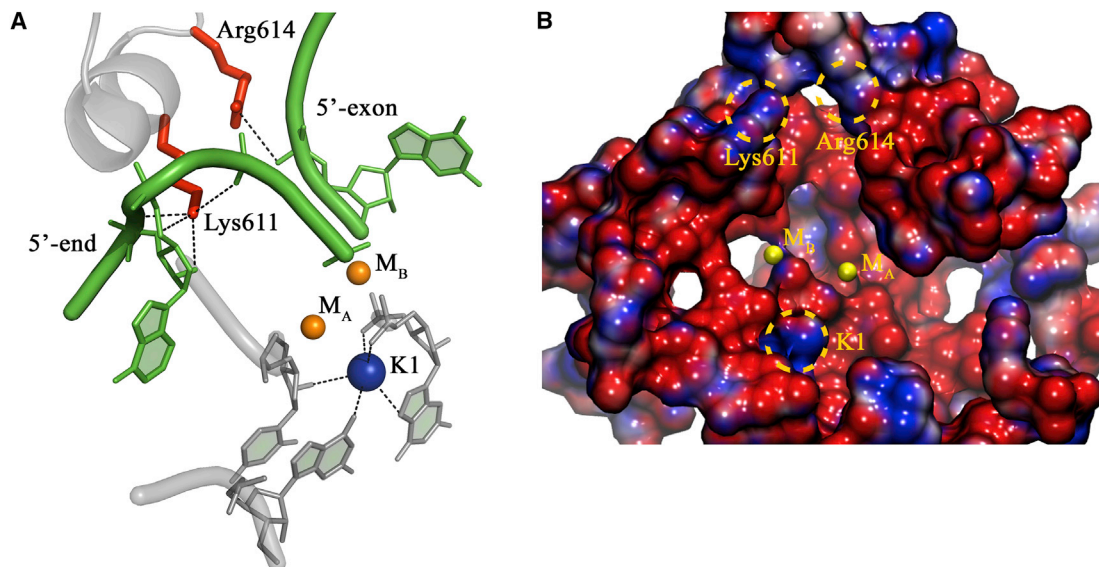


Figure 5. Prediction of K1- and K2-like Residues in Medium-Resolution Cryoelectron microscopy Structures of the Spliceosome

(A) Representation of the active site of the spliceosomal C complex from *S. cerevisiae* immediately after branching (PDB: 5LJ3).

(B) Electrostatic distribution around the active site. Structural elements are oriented and color coded as in Figure 1. A putative potassium ion at the K1-site is indicated (K1), although not modeled in the actual structure (see main text for details).

We suggest that in *S. cerevisiae* RNA polymerase II (Pol-II) (Cramer et al., 2001), Lys752 (Rbp1 subunit), and/or Arg1020 (Rbp2) occupy the K1 position, and Lys979 and/or Lys987 (Rbp2) occupy the K2 position (Figure 1F). Mutation of these amino acids is lethal in yeast (Table S2). Moreover, these amino acids are strictly conserved across Pol-II from yeast to humans, and they are also conserved in RNA Pol-I and Pol-III (Tafur et al., 2016) (Data S1). Instead, in the spliceosome, we noticed that a putative K1 cavity essentially identical to the one of group II intron exists between U6-snRNA residues G52, A59, G60, and U80 (Figure 5; Table S2). Although not modeled in the current structures, a potassium ion could optimally fit this putative K1 site at least at certain steps of the catalytic cycle, which would explain the spliceosomal dependence on potassium for splicing and debranching (Hardy et al., 1984; Tseng and Cheng, 2013). In addition, Lys611 (PROCN domain of the Prp8 subunit [Staub et al., 2004], Data S1), which stabilizes the curvature of the intron backbone (U2 and U4) proximal to the splicing junction in the C complex, and/or conserved Arg614 (Prp8), which makes contacts with the exon in the C^{*} complex (Figures 2 and 5; Table S2), seem to constitute optimal spliceosomal counterparts of K2. These residues produce regions of positive electrostatic potential at 7–9 Å from the active site (+6.7 kT/e), similar to the K2 ion in group II intron (Figure 5).

Notably, our analysis also spotted enzymatic classes that do not possess K1- and/or K2-like residues. For instance, class IIF endonucleases present K2 counterparts (i.e., Arg102 in NgoMIV, Arg218 in Sfil, and Arg152 in SgrAI, respectively), but no K1 counterparts, the role of the latter possibly being replaced by conserved lysines that directly coordinate M_A-M_B (i.e., Lys187 in NgoMIV, Lys102 in Sfil, and Lys242 in SgrAI) or by other non-basic residues (Table S1). DNA polymerase β (Pol-β) also represents an exception, because in this enzyme the M_A-M_B-binding acidic residues are not coordinated by residues near

the putative K1-site (Arg183 or Arg149), but rather by the K2-like residue (Arg254; Table S1). A third notable case is represented by LAGLIDADG homing endonuclease I-Dmol, an enzyme that performs two DNA hydrolysis reactions using two asymmetric active sites and three divalent metal ions (M_A, M_B, and M_C). In the first and faster reaction, the non-coding DNA strand is cleaved with support of M_A and M_C, while in the second and slower reaction, the coding DNA strand is cleaved with support of M_B and M_C. Our analysis shows that functionally crucial Lys120 (Molina et al., 2015) occupies the K1-like position and Lys130 the K2-like position of the first active site (Table S1). Instead, in the second active site, only a K2-like residue (Lys 43), but no K1-like residue could be identified (Table S1). The asymmetry of the two active sites of I-Dmol, and their consequently different catalytic efficiencies (Molina et al., 2015), is thus perfectly mirrored by a different architecture of the K1 and K2 positions, with only the most efficient active site presenting both K1 and K2 positions occupied by a basic amino acid. These considerations suggest that K1- and K2-like residues have been specifically selected by diverse, yet not all, two-metal-ion enzyme classes for DNA/RNA processing. Further analysis is thus needed now to clarify why certain two-metal-ion enzymes preserved K1- and K2-like residues, and elucidate their exact catalytic role for DNA and RNA processing.

Preserved spatial localization of K1- and K2-like residues and catalytic defects caused by their mutation further suggest that the somewhat obvious presence of such positively charged residues proximal to the anionic nucleic acid backbone may actually be coupled to a functional role (Tables 1 and S1). Possibly, K1- and K2-like residues affect DNA and RNA processing reactions directly, by participating in the chemical or physical steps of catalysis. For instance, in group II introns K1 and K2 mediate conformational changes, such as the transition from the first to the second step of splicing (Marcia and Pyle, 2012),

and in polymerases these residues help translocation and nucleotide addition (Genna et al., 2016b; Wu et al., 2014; Zhao et al., 2013). Alternatively, K1 and K2 may play indirect functional roles, by contributing to stabilize the overall architecture of the active site. For example, these residues may help rigidifying the coordination center of metals M_A - M_B , while also correctly orienting the substrates into the active site, as suggested by our MD simulations (Figures 3 and S3). Future validation is clearly required to confirm or disprove such mechanistic hypotheses. In this respect, future mutagenesis and engineering studies on K1- and K2-like residues may have prospective applicability in designing enzymes with improved substrate selectivity or higher substrate specificity, which would have immediate biotechnological impact, e.g., in the optimization of CRISPR-Cas9 genome-editing machineries. Furthermore, design of small molecules that bind to K1- and K2-like sites may lead to identify antimicrobial or antitumoral drugs with improved specificity against two-metal-ion enzymes (Berdis, 2008). Finally, reproducing the chemical architecture of the K1-K2- M_A - M_B phosphorus-oxide cluster formed by the group II intron may lead to develop nucleic acid-processing biocatalysts, such as DNAzymes or RNAzymes with important applications in nanoengineering (Li et al., 2011). We therefore hope that our structural observations will now stimulate further studies to clarify possible functional implications of K1- and K2-like residues in two-metal-ion enzymes for DNA and RNA processing.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
 - Sequence and Structural Alignments
 - Calculations of Electrostatic Potential Maps
 - Molecular Dynamics Simulations

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, four tables, and one data file and can be found with this article online at <https://doi.org/10.1016/j.str.2017.11.008>.

AUTHOR CONTRIBUTIONS

M.M. and M.D.V. conceived and supervised the project. M.M. performed sequence and structural alignments. V.G. performed the MD analyses and calculated the electrostatic potential maps. M.C. performed the analyses of K1 and K2 mutants. M.M. and M.D.V. wrote the manuscript with contributions from all authors.

ACKNOWLEDGMENTS

We thank all members of the Marcia and De Vivo labs for helpful discussion. We thank Dr. Stephen Cusack (EMBL Grenoble), Prof. Giovanni Capranico (University of Bologna), and Prof. Andrea Cavalli (University of Bologna and Italian Institute of Technology) for critical reading of the manuscript. We also thank Dr. Wojciech Galej (EMBL Grenoble) for sharing the coordinates of the spliceosomal C* complex from *S. cerevisiae* (PDB: 5MPS) before release. Work in the Marcia lab is partly funded by the Agence Nationale de la Recherche

(ANR-15-CE11-0003-01) and by the Agence Nationale de Recherche sur le Sida et les hépatites virales (ANRS) (ECTZ18552), and uses the platforms of the Grenoble Instruct Center (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). M.C. is funded by the EI3POD postdoctoral program (EMBL/EU Marie Curie Actions Cofund). M.D.V. thanks the Italian Association for Cancer Research (AIRC) for financial support ("IG 18883").

Received: June 30, 2017

Revised: September 12, 2017

Accepted: November 12, 2017

Published: December 7, 2017

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Strathern et al. (2013); Treich et al. (1992).

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* *215*, 403–410.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* *98*, 10037–10041.
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* *35*, 235–241.
- Berdis, A.J. (2008). DNA polymerases as therapeutic targets. *Biochemistry* *47*, 8253–8260.
- Berendsen, H.J.C., Vanderspoel, D., and Vandrunen, R. (1995). Gromacs - a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Comm.* *91*, 43–56.
- Besler, B.H., Merz, K.M., and Kollman, P.A. (1990). Atomic charges derived from semiempirical methods. *J. Comput. Chem.* *11*, 431–439.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863–1876.
- Crepin, T., Dias, A., Palencia, A., Swale, C., Cusack, S., and Ruigrok, R.W. (2010). Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J. Virol.* *84*, 9096–9104.
- Dal Peraro, M., Spiegel, K., Lamoureux, G., De Vivo, M., DeGrado, W.F., and Klein, M.L. (2007). Modeling the charge distribution at metal sites in proteins for molecular dynamics simulations. *J. Struct. Biol.* *157*, 444–453.
- De Vivo, M., Dal Peraro, M., and Klein, M.L. (2008). Phosphodiester cleavage in ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. *J. Am. Chem. Soc.* *130*, 10955–10962.
- Dias, A., Bouvier, D., Crepin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S., and Ruigrok, R.W. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* *458*, 914–918.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* *60*, 2126–2132.
- Fica, S.M., Mefford, M.A., Piccirilli, J.A., and Staley, J.P. (2014). Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* *21*, 464–471.
- Gao, Y., and Yang, W. (2016). Capture of a third Mg(2+) is essential for catalyzing DNA synthesis. *Science* *352*, 1334–1337.
- Genna, V., Gaspari, R., Dal Peraro, M., and De Vivo, M. (2016a). Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA polymerase- ϵ . *Nucleic Acids Res.* *44*, 2827–2836.
- Genna, V., Vidossich, P., Ippoliti, E., Carloni, P., and De Vivo, M. (2016b). A self-activated mechanism for nucleic acid polymerization Catalyzed by DNA/RNA polymerases. *J. Am. Chem. Soc.* *138*, 14592–14598.

- Gerlach, P., Malet, H., Cusack, S., and Reguera, J. (2015). Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell* **161**, 1267–1279.
- Grest, G.S., and Kremer, K. (1986). Molecular-dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A* **33**, 3628–3631.
- Hanoian, P., Liu, C.T., Hammes-Schiffer, S., and Benkovic, S. (2015). Perspectives on electrostatics and conformational motions in enzyme catalysis. *Acc. Chem. Res.* **48**, 482–489.
- Hardy, S.F., Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1984). Cofactor requirements of splicing of purified messenger RNA precursors. *Nature* **308**, 375–377.
- Ho, M.H., De Vivo, M., Dal Peraro, M., and Klein, M.L. (2010). Understanding the effect of magnesium ion concentration on the catalytic activity of ribonuclease H through computation: does a third metal binding site modulate endonuclease catalysis? *J. Am. Chem. Soc.* **132**, 13702–13712.
- Ippolito, J.A., Alexander, R.S., and Christianson, D.W. (1990). Hydrogen bond stereochemistry in protein structure and function. *J. Mol. Biol.* **215**, 457–471.
- Jeltsch, A., Alves, J., Wolfes, H., Maass, G., and Pingoud, A. (1993). Substrate-assisted catalysis in the cleavage of DNA by the EcoRI and EcoRV restriction enzymes. *Proc. Natl. Acad. Sci. USA* **90**, 8499–8503.
- Jiang, F., and Doudna, J.A. (2015). The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **30**, 100–111.
- Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997.
- Johnson, R.E., Trincão, J., Aggarwal, A.K., Prakash, S., and Prakash, L. (2003). Deoxynucleotide triphosphate binding mode conserved in Y family DNA polymerases. *Mol. Cell. Biol.* **23**, 3008–3012.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
- Kazantsev, A.V., Krivenko, A.A., and Pace, N.R. (2009). Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA* **15**, 266–276.
- Kurpiewski, M.R., Engler, L.E., Wozniak, L.A., Kobylanska, A., Koziolkiewicz, M., Stec, W.J., and Jen-Jacobson, L. (2004). Mechanisms of coupling between DNA recognition specificity and catalysis in EcoRI endonuclease. *Structure* **12**, 1775–1788.
- Li, H., Labean, T.H., and Leong, K.W. (2011). Nucleic acid-based nanoengineering: novel structures for biomedical applications. *Interface Focus* **1**, 702–724.
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958.
- Livesay, D.R., Jambeck, P., Rojnuckarin, A., and Subramaniam, S. (2003). Conservation of electrostatic properties within enzyme families and superfamilies. *Biochemistry* **42**, 3464–3473.
- Marcia, M., Ermiler, U., Peng, G., and Michel, H. (2010). A new structure-based classification of sulfide:quinone oxidoreductases. *Proteins* **78**, 1073–1083.
- Marcia, M., Humphris-Narayanan, E., Keating, K.S., Somarowthu, S., Rajashankar, K., and Pyle, A.M. (2013a). Solving nucleic acid structures by molecular replacement: examples from group II intron studies. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 2174–2185.
- Marcia, M., and Pyle, A.M. (2012). Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**, 497–507.
- Marcia, M., and Pyle, A.M. (2014). Principles of ion recognition in RNA: insights from the group II intron structures. *RNA* **20**, 516–527.
- Marcia, M., Somarowthu, S., and Pyle, A.M. (2013b). Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob. DNA* **4**, 14.
- McCulloch, S.D., Kokoska, R.J., Masutani, C., Iwai, S., Hanaoka, F., and Kunkel, T.A. (2004). Preferential cis-syn thymine dimer bypass by DNA polymerase ϵ occurs with biased fidelity. *Nature* **428**, 97–100.
- Misra, V.K., and Draper, D.E. (2001). A thermodynamic framework for Mg²⁺ binding to RNA. *Proc. Natl. Acad. Sci. USA* **98**, 12456–12461.
- Molina, R., Stella, S., Redondo, P., Gomez, H., Marcaida, M.J., Orozco, M., Prieto, J., and Montoya, G. (2015). Visualizing phosphodiester-bond hydrolysis by an endonuclease. *Nat. Struct. Mol. Biol.* **22**, 65–72.
- Muller, R., Poch, O., Delarue, M., Bishop, D.H., and Bouloy, M. (1994). Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J. Gen. Virol.* **75**, 1345–1352.
- Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y.J., and Yang, W. (2012). Watching DNA polymerase ϵ make a phosphodiester bond. *Nature* **487**, 196–201.
- Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
- Palermo, G., Cavalli, A., Klein, M.L., Alfonso-Prieto, M., Dal Peraro, M., and De Vivo, M. (2015). Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.* **48**, 220–228.
- Patra, A., Banerjee, S., Johnson Salyard, T.L., Malik, C.K., Christov, P.P., Rizzo, C.J., Stone, M.P., and Egli, M. (2015). Structural basis for error-free bypass of the 5-N-methylformamidopyrimidine-dG lesion by human DNA polymerase ϵ and *Sulfolobus solfataricus* P2 polymerase IV. *J. Am. Chem. Soc.* **137**, 7011–7014.
- Pingoud, A., and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727.
- Pyle, A.M., and Lambowitz, A.M. (2006). Group II introns: ribozymes that splice RNA and invade DNA. In *The RNA World*, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor Press), pp. 469–505.
- Ramachandrakurup, S., Ammapalli, S., and Ramakrishnan, V. (2016). Investigating the effect of charged amino acids on DNA conformation in EcoRI-DNA complex: a molecular dynamics simulation study. *J. Biomol. Struct. Dyn.* 1–15.
- Schrödinger. (2010). The PyMOL Molecular Graphics System, Version 1.3r1 (Schrödinger LLC).
- Sharp, P.A. (1991). Five easy pieces. *Science* **254**, 663.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.
- Staub, E., Fizev, P., Rosenthal, A., and Hinemann, B. (2004). Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *Bioessays* **26**, 567–581.
- Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **90**, 6498–6502.
- Strater, N., Lipscomb, W.N., Klabunde, T., and Krebs, B. (1996). Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew. Chem. Int. Ed.* **35**, 2024–2055.
- Strathern, J., Malagon, F., Irvin, J., Gotte, D., Shafer, B., Kireeva, M., Lubkowska, L., Jin, D.J., and Kashlev, M. (2013). The fidelity of transcription: RPB1 (RPO21) mutations that increase transcriptional slippage in *S. cerevisiae*. *J. Biol. Chem.* **288**, 2689–2699.
- Sun, J., Viadiu, H., Aggarwal, A.K., and Weinstein, H. (2003). Energetic and structural considerations for the mechanism of protein sliding along DNA in the nonspecific BamHI-DNA complex. *Biophys. J.* **84**, 3317–3325.
- Tafur, L., Sadian, Y., Hoffmann, N.A., Jakobi, A.J., Wetzels, R., Hagen, W.J., Sachse, C., and Muller, C.W. (2016). Molecular structures of transcribing RNA polymerase I. *Mol. Cell* **64**, 1135–1143.
- Treich, I., Carles, C., Sentenac, A., and Riva, M. (1992). Determination of lysine residues affinity labeled in the active site of yeast RNA polymerase II(B) by mutagenesis. *Nucleic Acids Res.* **20**, 4721–4725.
- Tseng, C.K., and Cheng, S.C. (2013). The spliceosome catalyzes debranching in competition with reverse of the first chemical reaction. *RNA* **19**, 971–981.
- Uljon, S.N., Johnson, R.E., Edwards, T.A., Prakash, S., Prakash, L., and Aggarwal, A.K. (2004). Crystal structure of the catalytic core of human DNA polymerase κ . *Structure* **12**, 1395–1404.

- Ummat, A., Silverstein, T.D., Jain, R., Buku, A., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K. (2012). Human DNA polymerase eta is prealigned for dNTP binding and catalysis. *J. Mol. Biol.* *415*, 627–634.
- Uyar, A., Kurkcuoglu, O., Nilsson, L., and Doruker, P. (2011). The elastic network model reveals a consistent picture on intrinsic functional dynamics of type II restriction endonucleases. *Phys. Biol.* *8*, 056001.
- Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* *25*, 1157–1174.
- Warshel, A., Sharma, P.K., Kato, M., Xiang, Y., Liu, H., and Olsson, M.H. (2006). Electrostatic basis for enzyme catalysis. *Chem. Rev.* *106*, 3210–3235.
- Wright, A.V., Nunez, J.K., and Doudna, J.A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* *164*, 29–44.
- Wu, S., Beard, W.A., Pedersen, L.G., and Wilson, S.H. (2014). Structural comparison of DNA polymerase architecture suggests a nucleotide gateway to the polymerase active site. *Chem. Rev.* *114*, 2759–2774.
- Yang, W. (2008). An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* *15*, 1228–1231.
- Yang, W., Chen, W.Y., Wang, H., Ho, J.W.S., Huang, J.D., Woo, P.C.Y., Lau, S.K.P., Yuen, K.Y., Zhang, Q.L., Zhou, W.H., et al. (2011). Structural and functional insight into the mechanism of an alkaline exonuclease from *Laribacter hongkongensis*. *Nucleic Acids Res.* *39*, 9803–9819.
- Yang, W., Lee, J.Y., and Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol. Cell* *22*, 5–13.
- Zhang, F., Gao, L., Zetsche, B., and Slaymaker, I. (2016). CRISPR enzyme mutations reducing off-target effects. Patent EP3129393 A1, filed June 17, 2016, and published February 15, 2017.
- Zhang, J., McCabe, K.A., and Bell, C.E. (2011). Crystal structures of lambda exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. *Proc. Natl. Acad. Sci. USA* *108*, 11872–11877.
- Zhao, Y., Gregory, M.T., Biertumpfel, C., Hua, Y.J., Hanaoka, F., and Yang, W. (2013). Mechanism of somatic hypermutation at the WA motif by human DNA polymerase eta. *Proc. Natl. Acad. Sci. USA* *110*, 8146–8151.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
BLAST	Altschul et al., 1990	https://blast.ncbi.nlm.nih.gov/Blast.cgi
ClustalOmega	Sievers et al., 2011	http://www.ebi.ac.uk/Tools/msa/clustalo/
T-COFFEE	Notredame et al., 2000	http://tcoffee.org.cat/
COOT	Emsley and Cowtan, 2004	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
PyMOL	Schrödinger, 2010	https://pymol.org/
APBS	Baker et al., 2001	http://www.poissonboltzmann.org/
AMBER	Wang et al., 2004	http://ambermd.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Marco Marcia (mmarcia@embl.fr).

METHOD DETAILS

Sequence and Structural Alignments

Representative sequences of each enzyme have been selected in BLAST (Altschul et al., 1990) and multiple-sequence alignments have been performed in ClustalOmega (Sievers et al., 2011). Structure-based sequence alignments have been performed using T-COFFEE (Notredame et al., 2000), as described previously (Marcia et al., 2010).

Structural alignments have been performed manually in Coot (Emsley and Cowtan, 2004) using the di-nuclear metal center M_A - M_B , the substrates, and/or the coordinating acidic residues in the first resolution shell of M_A - M_B as a guide. The figures depicting the structures were drawn using PyMOL (Schrödinger, 2010).

Calculations of Electrostatic Potential Maps

Calculations of the electrostatic molecular surfaces were performed by solving a nonlinear Poisson-Boltzman equation with APBS (Baker et al., 2001) as described previously (Kazantsev et al., 2009). For all the systems, we used a multilevel grid approach with a fixed grid length of 129 points in each spatial direction obtaining a grid spacing of ≤ 0.365 Å. To generate the electrostatic maps, all biomolecules were treated with a low dielectric medium ($\epsilon = 2$) (Misra and Draper, 2001) and the surrounding solvent as a high dielectric continuum ($\epsilon = 78.54$). The ionic strength was set to 150 mM corresponding to a Debye length of 8 Å, while temperature was set to 300 K. AMBER package (Wang et al., 2004) was used to add hydrogen atoms to crystallographic models and the AMBER ff99SB-ILDN force field (Lindorff-Larsen et al., 2010) was used to derive atomic radii and charges. In order to diminish the sensitivity of computations to the grid setup, we have used cubic B-spline discretization method to map the charges of our systems into the grid points and the nine-point harmonic averaging approach to the surface-based dielectric and ion-accessibility coefficients. Finally, we applied Dirichlet boundary conditions by using multiple Debye-Hückel functionality.

Molecular Dynamics Simulations

Structural Models

We considered four different systems: human wild type DNA Pol- η [PDB id.: 4ECS, (Genna et al., 2016a)], and human DNA Pol- η mutated at the K1 site (K231A), at the K2 site (K224A), and at both K1 and K2 sites (K224A/K231A). All the systems are a ternary complex where human DNA Pol- η is bound to the dsDNA substrate and the incoming dATP.

Molecular Dynamics Simulation Set Up

The all-atom AMBER/parm99SB-ILDN (Lindorff-Larsen et al., 2010) force field was adopted for the Pol- η in complex with dsDNA, whereas dATP was treated with the general amber force field (Wang et al., 2004). The atomic charges were derived by fitting the electrostatic potential according to the Merz-Singh-Kollman scheme (Besler et al., 1990) also known as the RESP fitting procedure (Table S3). The length of all covalent bonds, including hydrogen atoms, was set using the LINCS algorithm, allowing a time-integration step of 2 fs. All simulations were performed using Gromacs 4.6.1 code (Berendsen et al., 1995). Long-range electrostatic interactions were calculated with the particle mesh Ewald method with a real space cutoff of 12 Å. Periodic boundary conditions in the three directions of Cartesian space were applied. Constant temperature (310 K) was imposed using Langevin dynamics (Grest and

Kremer, 1986) with a damping coefficient of 1 ps. A constant pressure of 1 atm was maintained with Langevin-Piston dynamics with a 200 fs decay period and a 50 fs time constant. The metal active site was treated with a flexible non-bonded approach based on the 'atoms in molecules' partitioning scheme of the DFT-BLYP electronic density of the active site [Table S3, (Dal Peraro et al., 2007)]. We could thus consider the charge-transfer interactions between Mg^{2+} ions and their ligands, permitting possible structural rearrangements at the active site during the MD simulations. All the simulated systems were hydrated using TIP3P (Jorgensen et al., 1983) water molecules. A total of 7 Mg^{2+} ions were added to each system to reach a final concentration of ~ 1 mM, while Na^+ and Cl^- ions were added to neutralize the total charge. The size of the final systems was approximately $115 \times 95 \times 93 \text{ \AA}^3$, with $\sim 35,000$ water molecules, resulting in a total number of $\sim 102,000$ atoms each.

We adopted the following simulation protocol: the systems were minimized using a steepest-descent algorithm and then slowly heated up to 310 K in 10 ns for a total of 2000 steps. The first 50 ns of production run are considered as the equilibration phase. Approximately ~ 500 ns of MD simulations were collected in the NPT ensemble for each of the three systems, resulting in a total of $\sim 1.5 \mu s$ of dynamics. Coordinates of the systems were collected every 5 ps for each run. Statistics were collected considering the equilibrated trajectories only, thus discarding the first ~ 50 ns of simulation for all the systems.

Confidence Interval (C.I.) Analysis

To determine whether the *d-newbond* distribution are significantly different with respect to the different mutant systems (human DNA Pol- η K231A, K224A, and K224A/K231A) we performed a confidence interval (C.I.) analysis. The null hypothesis is expressed by $\mu_X = \mu_Y$, where μ represents the mean of the distribution, while X and Y are two different systems among the mutated. In order to test the null hypothesis, we computed the distributions of the *d-newbond* differences (d_{X-Y}) for all possible cases and evaluated the mean μ and the associated standard error σ_E . Values are reported in the Table S4, together with the C.I. at which the null hypothesis can be rejected. The populations of *d-newbond* in DNA Pol- η K231A, K224A, and K224A/K231A were 90001, 90001, 83540, respectively, and were obtained by sampling the MD trajectory every 10 ps, for a total of 1.5 μs .

Structure, Volume 26

Supplemental Information

**Second-Shell Basic Residues Expand
the Two-Metal-Ion Architecture of DNA
and RNA Processing Enzymes**

Vito Genna, Matteo Colombo, Marco De Vivo, and Marco Marcia

1 SUPPLEMENTAL INFORMATION FOR:

2
3 **Second-shell basic residues expand the two-metal-ion**
4 **architecture of DNA and RNA processing enzymes.**

5 Vito Genna^{1,#}, Matteo Colombo^{2,#}, Marco De Vivo^{1,3,*}, Marco Marcia^{2,*}

6
7 *To whom correspondence should be addressed. E-mail: marco.devivo@iit.it; mmarcia@embl.fr
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

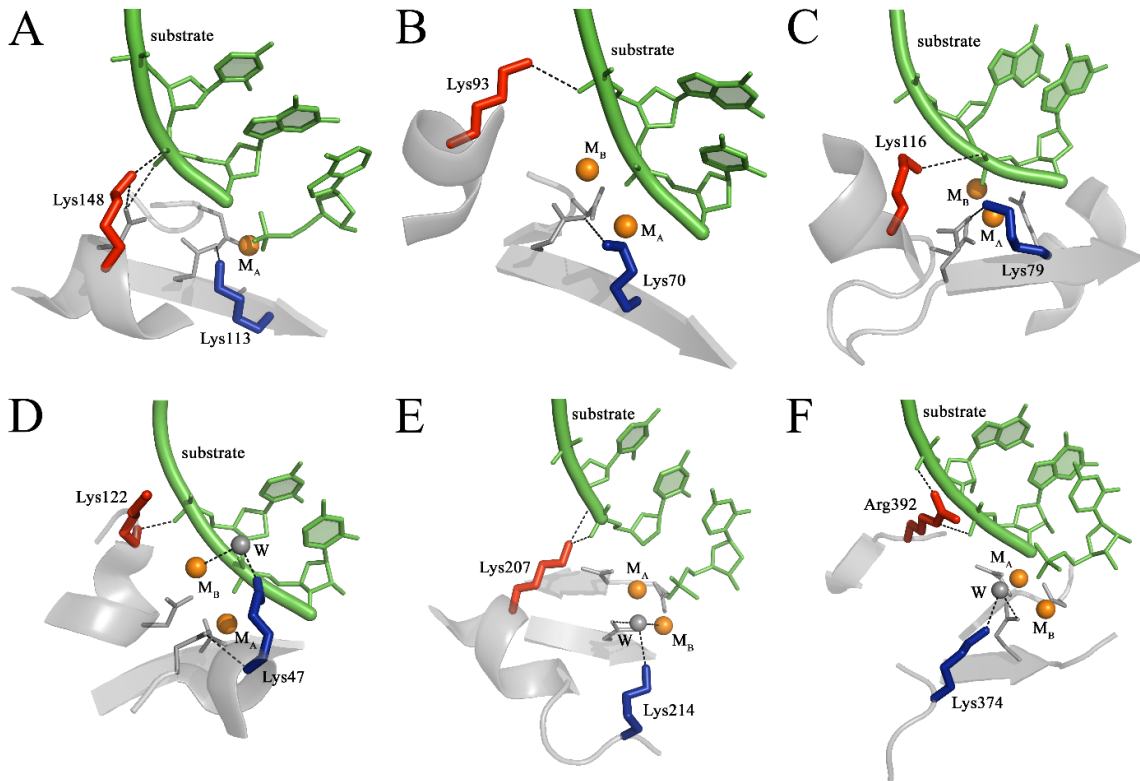
29 The supplemental information file contains:

- 30 - Fig. S1 (Related to Fig. 1)
- 31 - Fig. S2 (Related to Fig. 2)
- 32 - Fig. S3 (Related to Fig. 3)
- 33 - Table S2 (Related to Table 1)
- 34 - Table S3 (Related to Fig. 3)
- 35 - Table S4 (Related to Fig. 3)

1 **Supplemental Figures and Tables**

2 **Fig. S1: Additional examples of nucleases and polymerases possessing K1- and K2-like residues**
3 (Related to Fig. 1). A) Restriction endonuclease EcoRI (PDB id.: 1QPS); B) Restriction endonuclease
4 PvuII (1F00); C) Endonuclease MutH (2AOR); D) RNase-H (3O3H); E) DNA polymerase Pol- α
5 (2ALZ); and F) Norwalk virus RdRp (PDB id.: 3BSO). Structural elements are oriented and color coded
6 as in Fig. 1.

7

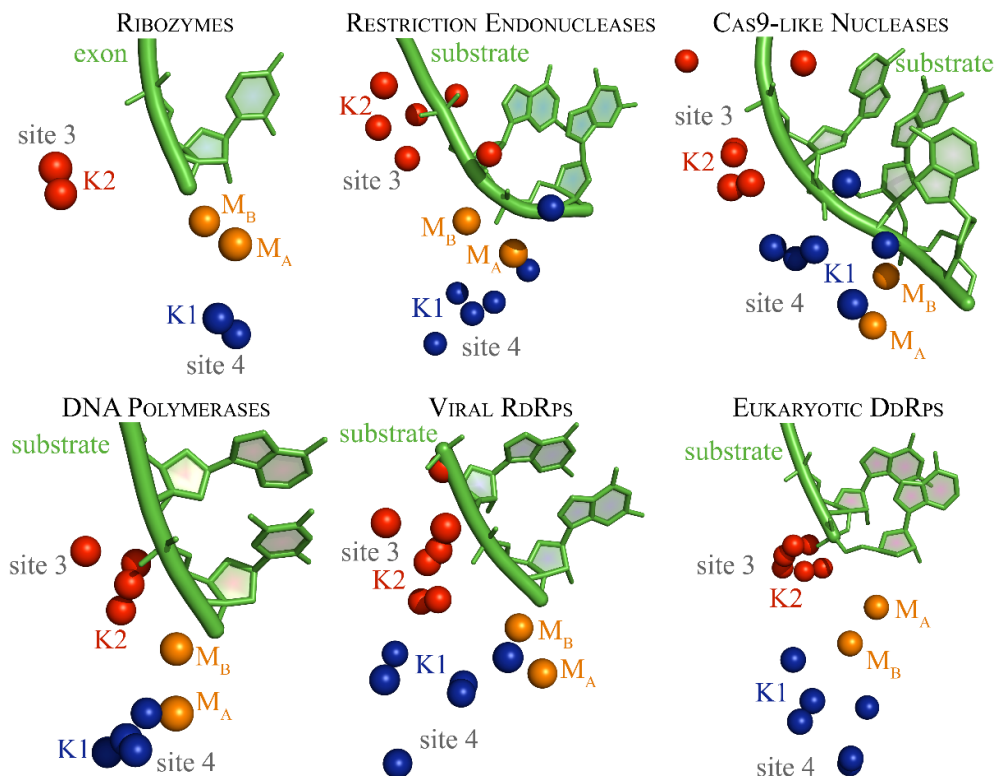


8

9

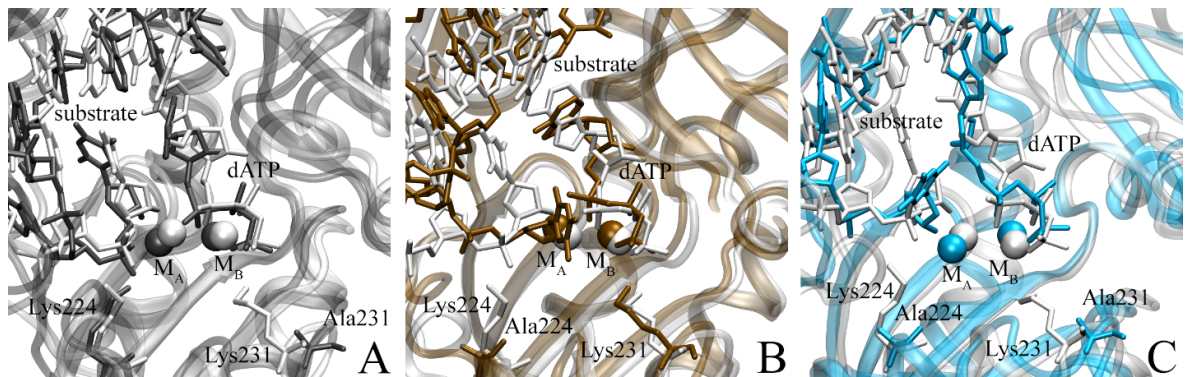
10

1 **Fig. S2: Structural overlap of K1- and K2-like residues in selected subsets of two-metal-ion**
2 **enzymes** (Related to Fig. 2). **Subset of ribozymes:** overlap of the structures of the group II intron
3 (4FAR) and of the spliceosome (5LJ3, putative K1 ion and K2-like Lys611, see also Fig. 5 for details).
4 Only the exon (substrate) and M_A - M_B center of 4FAR are represented for clarity. **Subset of restriction**
5 **endonucleases:** overlap of the structures of BamHI (2BAM), PvuII (1F0O), BpuSI (3S1S), MspI
6 (1SA3), FokI (1FOK), and NaeI (1IAW). Only the substrate and M_A - M_B center of 2BAM are
7 represented for clarity. **Subset of Cas9-like nucleases:** overlap of the structures of the RuvC domain
8 of Cas9 nucleases from *S. pyogenes* (4CMQ/5F9R, see also Figure 1C), *A. naeslundii* (4OGC), *F.*
9 *tularensis* (5B2O), and *S. aureus* (5AXW), and of RNase-H from *T. maritima* (3O3H) and *H. sapiens*
10 (3PUF). Only the substrate and M_A - M_B center of *S. pyogenes* Cas9 are represented for clarity. **Curated**
11 **subset of evolutionarily-related DNA polymerases:** overlap of the structures of human (PDB id 4ECS)
12 and yeast (3MFI) DNA polymerase η , human DNA polymerase ι (2ALZ), human DNA polymerase κ
13 (2OH2), *S. solfataricus* Dpo4 (2AGO), and *E. coli* DinB (4IRK). Only the substrate and M_A - M_B center
14 of 4ECS are represented for clarity. **Subset of viral RdRps:** overlap of the structures of the RdRps
15 from Norwalk virus (3BSO), Sapporo virus (2UUW), Rabbit haemorrhagic disease virus (1KHV),
16 Poliovirus (1RDR), Foot-and-mouth disease virus (2E9T), Hepatitis C virus (4E78), and Dengue virus
17 (2J7U). Only the substrate and M_A - M_B center of 3BSO are represented for clarity. **Subset of eukaryotic**
18 **DdRps:** overlap of the structures of yeast Pol-I (5M5Y), Pol-II (2E2H), and Pol-III (5FJ8). Only the
19 substrate and M_A - M_B center of 2E2H are represented for clarity. The structures were aligned and
20 represented as described in Fig. 2. Sites 3 and 4 are defined in Fig. 4 and in the main text.



21

1 **Fig. S3: Structural superimposition of the wild type and the mutated Pol- η in complex with**
2 **dsDNA and dATP from molecular dynamics simulations** (Related to Fig. 3). A) Superimposition
3 between human wild-type DNA Pol- η (white) and K231A mutant (grey). B) Superimposition between
4 human wild-type DNA Pol- η (white) and K224A mutant (ochre). C) Superimposition between human
5 wild-type DNA Pol- η (white) and K224A/K231A mutant (cyan). Spheres indicate magnesium ions
6 while the nucleic acid substrate, dATP, and the K1- and K2-like residues are reported in licorice. These
7 superpositions show the significant distortion of the reactant state upon mutations of K1- and/or K2-
8 like residues.
9



1 **Table S2. Prediction of K1- and K2-like residues in other classes of two-metal-ion enzymes (Related to Table 1).**

2

Enzyme	PDB id	Resolution (Å)	Enzymatic Classification (E.C.)	K1	d1 (Å) ^a	d-ac (Å) ^b	mutants	functional defect	K2	d2 (Å) ^a	d-sub (Å) ^c	mutants	functional defect
Spliceosome (<i>S. cerevisiae</i>)	5LJ3	3.80	ribozyme	K17 ^d	~3.8	~2.6	n.a.	n.a.	Lys611 (Prp8)	9.08	3.67	n.a.	n.a.
									Arg614 (Prp8)	9.88	3.38	n.a.	n.a.
RNA Pol-I (<i>S. cerevisiae</i>)	5M5Y	4.00	2.7.7.6	Lys934 (RPA190)	x ^e	8.49	n.a.	n.a.	Lys916 (RPA135)	x ^e	3.44	n.a.	n.a.
				Arg957 (RPA135)	x ^e	5.13	n.a.	n.a.	Lys924 (RPA135)	x ^e	2.36	n.a.	n.a.
RNA Pol-II (<i>S. cerevisiae</i>)	2E2H	3.95	2.7.7.6	Lys752 (Rbp1)	5.40	4.03	K752L (Strathern et al., 2013)	Lethal	Lys979 (Rbp2)	8.86	2.58	K979R (Treich et al., 1992)	Lethal
				Arg1020 (Rbp2)	3.12	3.57	n.a.	n.a.	Lys987 (Rbp2)	7.13	3.34	K987R (Treich et al., 1992)	Lethal
RNA Pol-III (<i>S. cerevisiae</i>)	5FJ8	3.90	2.7.7.6	Lys800 (Rpc1)	x ^e	9.71	n.a.	n.a.	Lys911 (Rpc2)	x ^e	3.75	n.a.	n.a.
				Arg952(Rpc2)	x ^e	7.55	n.a.	n.a.	Lys919 (Rpc2)	x ^e	4.45	n.a.	n.a.

3

4 ^a closest distance between the ion / amino acid corresponding to K1 or K2 and the M_A-M_B center; ^b closest distance between the ion / amino acid corresponding to K1 and the acidic residues that chelate M_A-M_B; ^c closest
5 distance between the ion / amino acid corresponding to K2 and the substrate; ^d putative, see main text for details; ^e no substrate or ions in the active site; n.a. = not available;

1 **Table S3: Calculated RESP charges** (Related to Fig. 3). Left: The QM electrostatic potential (ESP),
 2 required during the RESP fitting procedure to derive dATP atom charges, was calculated at the HF/6-
 3 31G* level of theory by using Gaussian 09. Right: Corrected charges for metal ions and ligands with
 4 respect to AMBER RESP charges. Inset: ball and stick representation of dATP with corresponding
 5 atomic nomenclature.

Label	Charge	Atom	δq	q^{eff}	q^{*}_{AMBER}
PA	1.630320	Mg _A -Mg _B	-0.25	1.75	2.00
PB	1.752583	O δ 1-2 (Asp13)	0.05	-0.74	-0.80
PG	1.631219	O δ 1-2 (Asp115)	0.07	-0.73	-0.80
C5'	0.145097	O ϵ 1-2 (Glu116)	0.05	-0.78	-0.82
O5	-0.653778	O (Met14)	0.06	-0.50	-0.56
C4'	0.080100				
O4	-0.436726				
C3'	0.079202				
O3	-0.651155				
C2'	0.114262				
C1'	0.217376				
N1	-0.864491				
O1A	-0.885583				
O1B	-0.880119				
O1G	-0.897715				
C2	0.622012				
O2A	-0.895529				
O2B	-0.855092				
O2G	-0.846348				
N3	-0.797823				
O3A	-0.899244				
O3B	-0.903616				
O3G	-0.883397				
C4	0.430437				
C5	-0.267981				
C6	0.685211				
N6	-1.042088				
N7	-0.584816				
C8	0.346741				
N9	-0.312571				
H01	0.060952				
H02	0.068952				
H03	0.083337				
H04	0.058165				
H05	0.387645				
H06	0.063558				
H07	0.063558				
H08	0.135477				
H22	0.400770				
H13	0.051333				
H25	0.043781				
H27	0.406164				

Deoxyadenosine triphosphate (dATP)

6
 7
 8
 9
 10 **Table S4: Calculated d-newbond differences** (Related to Fig. 3). Distributions of d-newbond
 11 differences ($d_{X,Y}$) for all MD systems with mean μ , standard error σ_E , and confidence interval (C.I.).

	μ (Å)	σ_E (Å)	C.I.
d-newbond (K2-K1)	0.02	0.0000447	> 99%
d-newbond (K1K2-K2)	0.14	0.0000545	> 99%
d-newbond (K2-K1K2)	0.14	0.0000545	> 99%