Structure-based domain assignment in *Leishmania infantum* EndoG: characterization of a pH-dependent regulatory switch and a C-terminal extension that largely dictates DNA substrate preferences

Cristina Oliva[1], Pedro A. Sánchez-Murcia[2], Eva Rico[1], Ana Bravo[2], Margarita Menéndez[3], Federico Gago[2,*] and Antonio Jiménez-Ruiz[1,*]

[1]Departamento de Biología de Sistemas, Universidad de Alcalá, E-28805 Alcalá de Henares, Madrid, Spain, [2]Departamento de Ciencias Biomédicas y "Unidad Asociada IQM-CSIC", Universidad de Alcalá, E-28805 Alcalá de Henares, Madrid, Spain and [3]Instituto de Química Física Rocasolano, Consejo Superior de Investigaciones Científicas (CSIC), E-28006 Madrid, Spain

ABSTRACT

Mitochondrial endonuclease G from *Leishmania infantum* (*Li*EndoG) participates in the degradation of double-stranded DNA (dsDNA) during parasite cell death and is catalytically inactive at a pH of 8.0 or above. The presence, in the primary sequence, of an acidic amino acid-rich insertion exclusive to trypanosomatids and its spatial position in a homology-built model of *Li*EndoG led us to postulate that this peptide stretch might act as a pH sensor for self-inhibition. We found that a *Li*EndoG variant lacking residues 145–180 is indeed far more active than its wild-type counterpart at pH values >7.0. In addition, we discovered that (i) *Li*EndoG exists as a homodimer, (ii) replacement of Ser211 in the active-site SRGH motif with the canonical aspartate from the DRGH motif of other nucleases leads to a catalytically deficient enzyme, (iii) the activity of the S211D variant can be restored upon the concomitant replacement of Ala247 with Arg and (iv) a C-terminal extension is responsible for the observed preferential cleavage of single-stranded DNA (ssDNA) and ssDNA–dsDNA junctions. Taken together, our results support the view that *Li*EndoG is a multidomain molecular machine whose nuclease activity can be subtly modulated or even abrogated through architectural changes brought about by environmental conditions and interaction with other binding partners.

INTRODUCTION

Sugar nonspecific endonucleases belong to a family of enzymes that cleave single- and double-stranded DNA (ssDNA and dsDNA, respectively) and/or RNA molecules with little or no sequence specificity. Among them, endonuclease G (EndoG) is highly conserved in eukaryotes [1] and is targeted to the mitochondrial intermembrane space from where it is released in response to a death stimulus. Upon subsequent translocation to the nucleus, EndoG is thought to participate in the process of DNA degradation that is associated with programmed cell death [2].

Organisms have sophisticated mechanisms for genome protection against self-destructive nuclease activity. Although EndoG is usually silenced by compartmentalization into mitochondria, additional protective strategies have evolved in nature, such as activation via cross-linking of cysteine residues, as in the nucleases secreted by *Serratia marcescens* [3] and *Vibrio vulnificus* [4], or active site blockade, as exemplified by the specific protein inhibitors co-crystallized with nucleases from *Anabaena spiroides* [5] and *Drosophila melanogaster* [6]. These mechanisms avoid irreversible damage caused by stray EndoG.

In the last decade, the 3D structures of several eukaryotic endo/exo-nucleases from different organisms have been solved and deposited in the Protein Data Bank (PDB) [7]. From the structural point of view, the common core of these metallonucleases is made up of two antiparallel β-sheets and one α-helix that contains an essential and strictly conserved asparagine residue whose carboxamide oxygen is used to coordinate to a divalent metal ion. This cation is usually magnesium although manganese or, in the case of human EXOG [8], calcium is also catalytically competent.

*To whom correspondence should be addressed. Tel: +34 918 855 109; Fax: +34 918 854 585; Email: antonio.jimenez@uah.es
Correspondence may also be addressed to Federico Gago. Tel: +34 918 854 514; Fax: +34 918 854 591; Email: federico.gago@uah.es

We previously showed that the protozoan parasite *Leishmania infantum* expresses a protein, *Li*EndoG, with structural and functional characteristics that strongly resemble those of endonucleases G from higher eukaryotes, including mitochondrial localization and migration to the nucleus in response to a death stimulus (9). More recently, we demonstrated that *Li*EndoG also has a pro-life role in the parasite, as indicated by the reduced growth rates observed in *Li*EndoG hemi-knockouts and the cytotoxicity of the *Li*EndoG inhibitor Lei49 on promastigotes and amastigotes, which can be counteracted by *Li*EndoG overexpression (10).

Herein, guided and supported by an atomic three-dimensional model of the homodimeric enzyme, we seek a rationale for the occurrence of the SRGH motif in EndoGs from trypanosomatids in place of the canonical DRGH and pursue the assignment of specific functions to several discrete domains that appear to be unique to these enzymes. Two of these assigned functions are validated in *Li*EndoG through the experimental generation and biochemical characterization of several protein variants whose enzymatic activities on a variety of substrates are compared to that of the wild-type enzyme.

MATERIALS AND METHODS

Molecular modeling of *Li*EndoG

The dimers observed in the crystal structures of *Drosophila melanogaster* EndoG (*Dm*EndoG, PDB id. 3ISM) (6), the inactive H148A (11), H148A/F122A (12) and H148A/Q130K (13) variants of *Caenorhabditis elegans* EndoG (*Ce*EndoG, a.k.a. CED-3 protease suppressor-6 [CPS-6], PDB ids. 3S5B, 5GKC and 4QN0, respectively), and human mitochondrial endo-exonuclease G (EXOG, PDB id. 4A1N) (11) are considered to be valid models for eukaryotic EndoGs because of the high conservation, from yeast to human, of many residues that are crucial for both activity and dimerization (6,14). Comparison and inspection of the incomplete homology models provided by the threading methods implemented in the Phyre2 (15,16) and Swiss-Model (17) web servers allowed us to construct a consensus 'gapped' model comprising residues 99–144, 181–264, 301–315 and 364–439 of *Li*EndoG (UniProtKB entry B0ZTD3) that was elaborated further with the aid of additional iterative and incremental modeling, as described below. Although human EXOG has been biochemically characterized in detail (18), the three-dimensional coordinates deposited in the PDB for this protein remain unpublished and suffer from some uncertainties. For this reason, we refined them against the accompanying electron density map using Coot (19) and Phenix (20,21). The resulting model (Supplementary Figure S1) was then used as a template by the one-to-one threading methods implemented in the Phyre (15) and SwissModel (17) servers to generate a consensus 3-helix bundle for the C-terminal domain (CTD) of *Li*EndoG. Structure visualization, molecular editing, and figure preparation were performed with PyMOL (22).

No suitable templates were found for building a homology model of inserts 1−3 (Figure 1) but the algorithms (23,24) implemented in Robetta (25) were able to produce a reasonable structure for ***insert 1*** (residues 145–180), which comprises the (Asp/Glu)-rich region exclusive to nucleases from trypanosomatids, and we tentatively modeled the polyglutamine tract in **insert 2** as an α-helix (26,27). The flap in ***insert 3***, one β-strand of which is proposed to form an intermolecular β-sheet in the dimer, was constructed by following an incremental approach that took into account the putatively equivalent positioning at the tip of the β-hairpin of [321]RPDG[324] in *Li*EndoG and [214]EGDG[217], [221]EDDG[224] and [205]RGDG[208] in *Ce*EndoG, *Dm*EndoG and EXOG, respectively. The Random Coordinate Descent (RCD) loop closure algorithm (28) was used profusely for the *ab initio* modeling of the remaining gaps.

The two putative disulfide bonds between residues 100–116 and 434–439, earlier predicted by the DISULFIND server (29) on the basis of the primary sequence (9), and now found to be feasible in 3D given the spatial closeness and suitable side-chain orientation of the two cysteines making up each pair (Figure 1), were incorporated into the structural model of each monomer. The catalytic Mg^{2+} ion and the water molecules completing its octahedral solvation shell were introduced in agreement with the geometries observed in both *Dm*EndoG (6) and EXOG (11), which are coincident with that described previously for our simplified model of the active site of *Li*EndoG (9) based on *Vibrio vulnificus* nuclease (*Vvn*) (30). Finally, all the domains were integrated into a *Li*EndoG dimer that was visually and computationally inspected. Minimal manual adjustments alleviated some steric clashes involving some side-chain rotamers and protein backbone atoms. The resulting protein structure was then refined in the AMBER force field (31) by carrying out, essentially as described (30), a series of Cα-restrained energy minimizations followed by molecular dynamics (MD) simulations at 300 K in a box of TIP3P water molecules and counterions using the *pmemd* module of the AMBER 14.0 suite of programs (31).

The complexes of wild-type *Vvn* and its inactive H80A variant with a dsDNA octamer (32) and a dsDNA 16-mer (33), respectively, as well as the complex between the H148A/F122A variant of *Ce*EndoG and a 5′-TTTTTGT-3′ ssDNA (12), provided the structural information necessary for building the initial *Li*EndoG:DNA complexes prior to refinement using the same molecular mechanics methods that were employed on the protein alone. In this respect, further crystallographic refinement of the *Ce*EndoG(H148A/F122A):ssDNA complex (PDB code: 5GKP) against the accompanying electron density map using Coot (19) and Phenix (20,21) allowed us to unambiguously trace most of the intact octanucleotide chains bound to both subunits and place all the bases in the right register and correct positions (Supplementary Figure S2 and Table S1). This step was instrumental in deriving a reliable molecular model of *Li*EndoG:ssDNA in which the residues that interact with the phosphate backbone can be unequivocally identified (Table 1). As a template for our ssDNA–dsDNA junction we used the single-crystal structure of a DNA Holliday junction model (34) (PDB code: 4GS2) and changed base composition by means of the *ad hoc* utility implemented in Coot (19).
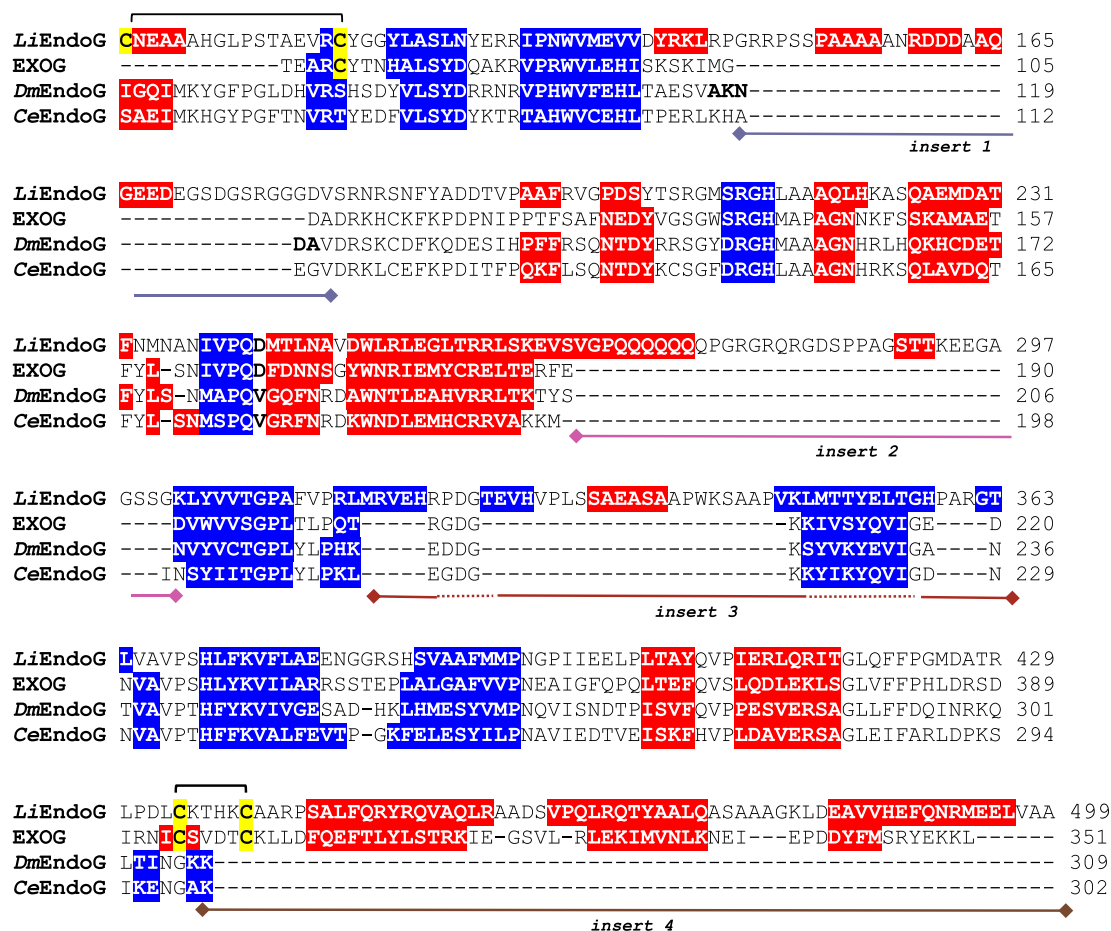
```
LiEndoG   CNEAAAHGLPSTAEVRCYGGYLASLNYERRIPNWVMEVVDYRKLRPGRRPSSPAAAAANRDDDAAQ   165
EXOG      -----------TEARCYTNHALSYDQAKRVPRWVLEHISKSKIMG------------------   105
DmEndoG   IGQIMKYGFPGLDHVRSHSDYVLSYDRRNRVPHWVFEHLTAESVAKN------------------   119
CeEndoG   SAEIMKHGYPGFTNVRTYEDFVLSYDYKTRTAHWVCEHLTPERLKHA------------------   112
                                                                          insert 1

LiEndoG   GEEDEGSDGSRGGGDVSRNRSNFYADDTVPAAFRVGPDSYTSRGMSRGHLAAAQLHKASQAEMDAT   231
EXOG      -------------DADRKHCKFKPDPNIPPTFSAFNEDYVGSGWSRGHMAPAGNNKFSSKAMAET   157
DmEndoG   -------------DAVDRSKCDFKQDESIHPFFRSQNTDYRRSGYDRGHMAAAGNHRLHQKHCDET   172
CeEndoG   -------------EGVDRKLCEFKPDITFPQKFLSQNTDYKCSGFDRGHLAAAGNHRKSQLAVDQT   165

LiEndoG   FNMNANIVPQDMTLNAVDWLRLEGLTRRLSKEVSVGPQQQQQQQPGRGRQRGDSPPAGSTTKEEGA   297
EXOG      FYL-SNIVPQDFDNNSGYWNRIEMYCRELTERFE-----------------------------   190
DmEndoG   FYLS-NMAPQVGQFNRDAWNTLEAHVRRLTKTYS-----------------------------   206
CeEndoG   FYL-SNMSPQVGRFNRDKWNDLEMHCRRVAKKM-----------------------------   198
                                                                          insert 2

LiEndoG   GSSGKLYVVTGPAFVPRLMRVEHRPDGTEVHVPLSSAEASAAPWKSAAPVKLMTTYELTGHPARGT   363
EXOG      ----DVWVVSGPLTLPQT-----RGDG-----------------------KKIVSYQVIGE----D   220
DmEndoG   ----NVYVCTGPLYLPHK-----EDDG-----------------------KSYVKYEVIGA----N   236
CeEndoG   ---INSYIITGPLYLPKL-----EGDG-----------------------KKYIKYQVIGD----N   229
                                                                          insert 3

LiEndoG   LVAVPSHLFKVFLAEENGGRSHSVAAFMMPNGPIIEELPLTAYQVPIERLQRITGLQFFPGMDATR   429
EXOG      NVAVPSHLYKVILARSSTEPLALGAFVVPNEAIGFQPQLTEFQVSLQDLEKLSGLVFFPHLDRSD   389
DmEndoG   TVAVPTHFYKVIVGESAD-HKLHMESYVMPNQVISNDTPISVFQVPPESVERSAGLLFFDQINRKQ   301
CeEndoG   NVAVPTHFFKVALFEVTP-GKFELESYILPNAVIEDTVEISKFHVPLDAVERSAGLEIFARLDPKS   294

LiEndoG   LPDLCKTHKCAARPSALFQRYRQVAQLRAADSVPQLRQTYAALQASAAAGKLDEAVVHEFQNRMEELVAA   499
EXOG      IRNICSVDTCKLLDFQEFTLYLSTRKIE-GSVL-RLEKIMVNLKNEI---EPDDYFMSRYEKKL------   351
DmEndoG   LTINGKK-----------------------------------------------------------   309
CeEndoG   IKENGAK-----------------------------------------------------------   302
                                                                          insert 4
```

Figure 1. Structure-based multiple sequence alignment, as provided by the updated Dali server (70), of our *Li*EndoG model against the three most suitable templates found in the Protein Data Bank. α-helices and β-strands are boxed in a red or blue background, respectively. Note the presence of the four insertions (◆◆) in *Li*EndoG for which the only counterpart in the other three enzymes is the EXOG CTD, which shows some homology to *insert 4*. The two rotated square brackets connecting the cysteines boxed in a yellow background represent the two putative disulfide bonds. The AKNDA sequence from *Dm*EndoG highlighted in bold was used to replace the putative self-inhibitory domain (*insert 1*) in the *Li*EndoGΔi variant.

Table 1. Protein atoms involved in phosphate recognition of the oligonucleotides to be cleaved by *Ce*EndoG, EXOG and *Li*EndoG[a]

Nuclease	DNA phosphodiester bond																	
	-5		-4		-3		-2		-1		0		+1		+2		+3	
	OP1	OP2	OP1	OP2	OP1	OP2	OP1	OP2	OP1	OP2	OP1[b]	OP2	OP1	OP2	OP1	OP2	OP1	OP2
*Ce*EndoG	—	—	—	—	—	—	(NH1 R192)[c]	NH2 R156	NH G153 NH2 R192	NH1 R117	Mg^{2+}	NH2 R117	NE R146	NH2 R146	NH C141	NZ K140	—	NZ K140
EXOG	[L302][d]	[E184]	[L309]	[L309]	[E317]	[L302]	NH1 R183	NE R314 NK148 OH Y310	NH G145 NH2 R183	NH1 R109	Mg^{2+}	NH2 R109	NE R138	NH2 R138	NH G133	[V132]	—	[V132]
*Li*EndoG	(NE R442)	(NH1 R259)	(NH1 R449)	(NH2 R442) (NH2 R449)	(NH2 R457)	(NH2 R442)	NH2 R258	OH Y450	NH Q219 NH2 R258	NH1 R183	Mg^{2+}	NH2 R183	NE R212	NH2 R212	NH S207	(OG1 T206)	—	(OG S207)

[a]The reference frame for numbering the phosphodiester bonds is upstream (–) and downstream (+) of site 0 (i.e. the cleavage site, which is defined as the bond linking the Mg^{2+}-bound O3' of one nucleotide to the P atom of the next nucleotide). OP1 and OP2 are the phosphate oxygens involved in recognition.
[b]The Mg^{2+} ion is included for completeness and as an aid for establishing the frame of reference.
[c]Residue names in brackets are thought to be involved in phosphate recognition but these particular contacts are not observed in the reference crystal structure.
[d]Residue names in square brackets are included for completeness and comparative purposes but they cannot establish a favorable interaction with a phosphate group.

Expression of wild-type *Li*EndoG and engineered variants

Escherichia coli Rosetta (DE3) pLysS bacterial cells were transformed with pET28a-endog. Expression of *Li*EndoG was then induced in LB medium at 0.6 OD_{600nm} with 1 mM IPTG at 37°C during 3 h. Because most of the expressed protein was found in the non-soluble fraction, *E. coli* cells were lysed in a pH 8.0 buffer containing 8 M urea, 100 mM NaH_2PO_4, 10 mM imidazole, and 10 mM Tris. The cleared lysate was loaded onto a HisTrap Ni–NTA resin column (GE Healthcare) in which the protein was refolded by progressive removal of urea prior to elution with 500 mM imidazole.

The DNA coding for the *Li*EndoG variant lacking amino acids 440–510 (*Li*EndoG△^{CTD}) was obtained by PCR amplification of the *Li*EndoG gene sequence from its start position up to the nucleotides encoding amino acid 439. This PCR-amplified gene was cloned into the BamHI/XhoI site of the expression vector pET-28a.

For *Li*EndoG△i construction, the sequence corresponding to the putative self-inhibitory domain encoding amino acids 145–180 was replaced by nucleotides encoding amino acids AKNDA to make this region resemble the positionally equivalent region of *Dm*EndoG (see Figure 1). This was achieved by two separate PCR amplifications followed by a fusion PCR via overlap sequences. The fusion construct was also cloned into the BamHI/XhoI site of the expression vector pET-28a.

The S211A, S211D, H214A, R212A, S211D/A247R and D(160−162)A variants were generated using the QuikChange XL-Site-Directed mutagenesis kit (Agilent) and oligos that were designed following the manufacturer's instructions. The presence of the desired mutations was checked by DNA sequencing.

Fluorescence-based nuclease activity assay

The enzymatic activity of *Li*EndoG was monitored by measuring the increase in fluorescence obtained upon digestion of a dsDNA probe constructed by hybridization of the oligonucleotides FAM-5′-CTGTCGCTACCTGTGG-3′ and 5′-CCACAGGTAGCGACAG-3′-TAMRA. The intrinsic fluorescence of 6-carboxyfluorescein (FAM) is quenched by the proximity of tetramethylrhodamine (TAMRA) in the undigested probe. Digestion of any of the strands causes separation of the fluorophore from the quencher which is apparent by emission of a fluorescent signal. The dsDNA probe (30 pmol) was digested with the different nucleases (1 ng/μl) in 10 mM NaAc, 10 mM MES, 20 mM Tris–HCl and 10 mM $MgCl_2$. The pH value was adjusted to 6.5 unless otherwise indicated. The reactions were monitored either in a Victor 1420 Multilabel Counter (Wallac) or in an EnSpire Microplate Reader, at excitation and emission wavelengths of 492 nm and 517 or 495 and 520 nm, respectively.

Nuclease activity assay on supercoiled circular DNA

pRSETa plasmid DNA (0.3 μg) was digested in a final volume of 20 μl pH 6.5 buffer containing 22.5 mM Tris, 50 mM NaCl, MES 10 mM, NaAc 10 mM, 2.5 mM $MgCl_2$ and 5% glycerol during 1 h at 16°C using the amounts of enzyme

indicated in the figure. The digested DNA was resolved on 1.0% (w/v) agarose gels, stained with ethidium bromide and visualized under UV light.

Activity assays using radioactively labeled oligonucleotide substrates

The ssDNA oligonucleotides shown in Table 2 were purchased from Invitrogen and radioactively labeled at the 5′-end with [γ^{32}P]-ATP using polynucleotide kinase (Thermo Scientific). Unincorporated radionucleotides were removed using the QIAquick Nucleotide Removal kit (Qiagen). To generate the labeled dsDNA substrates, each corresponding radioactively labeled ssDNA was incubated in a 1:1.25 molar ratio with its complementary strand in 20 mM Tris–HCl (pH 7.5) and 50 mM NaCl. Annealing was achieved by heating the solution at 95°C for 5 min and then cooling down slowly (–1°C per minute) to 25°C.

Enzymatic cleavage assays were done at 37°C in a pH 6.5 buffer containing 22.5 mM Tris, 50 mM NaCl, MES 10 mM, NaAc 10 mM, 2.5 mM $MgCl_2$ and 5% glycerol in a final volume of 10 μl. Cleavage products were resolved by 7 M urea 20% polyacrylamide gel electrophoresis (PAGE). DNase I and exonuclease T7 (Thermo Scientific) were used as positive controls of endonuclease and exonuclease activities, respectively.

Analytical ultracentrifugation

Sedimentation velocity experiments were run at 45 000 rpm using cells with double sector Epon-charcoal centrepieces. Differential sedimentation coefficients were calculated by least-squares boundary modeling of the experimental data with the program SEDFIT. The experimental values were corrected to the viscosity and density of water at 20°C ($s_{20,w}$) (35). Sedimentation equilibrium experiments were carried out at different rotor speeds, from 6500 to 11 000 rpm, as described previously (36), and weight-average molecular weights ($M_{w,app}$) were calculated using the HETERO-ANALYSIS program, version 1.1.2 (http://www.biotech.uconn.edu/auf/). All measurements were performed in 50 mM MES, 500 mM NaCl, 0.4 mM KCl buffer (pH 6.5) at 20°C using protein concentrations ranging from 0.07 to 0.77 mg/ml in an Optima XL-A analytical ultracentrifuge (Beckman Coulter) with an AN50-Ti rotor. The partial specific volume of *Li*EndoG△^{CTD} and the buffer's density and viscosity were estimated from the amino acid sequence and the buffer composition, respectively, using the SEDNTERP program (37).

RESULTS

An updated *Li*EndoG molecular model

Our initial homology model for *Li*EndoG (9) afforded enough detail about the active site environment but provided a reliable structure only for the catalytic core of a single monomer. To follow up on this work, we elaborated on our model further and then focused on four conspicuous regions for which no functional assignment has been reported yet: 145–180 (***insert 1***), 265–300 (***insert 2***), 316–363 (***insert 3***) and 440–510 (***insert 4***). *Insert 1*, which appears to be unique

Table 2. Oligonucleotides used for radioactive labeling on the 5′-end

dsDNA probe[a]
forward 5′–**GCCGCTAACAGAGACGCGGAGGCCGCGCAGGGTGAAGAAGAC**–3′
reverse 5′–**GTCTTCTTCACCCTGCGCGGCCTCCGCGTCTCTGTTAGCGGC**–3′
ssDNA probe
5′–CGCCAGATCTTCCC–3′
dsDNA with ssDNA overhangs
forward 5′–
TTTTTTTTTTTTTTTTTTTTTGG**TTAATACGACTCACTATAGGAGAACCTTAAGGTTTAACTTTAAGACCC**CCCCCCCCCCCCCCCCCCCCC–3′
reverse 5′-TTTTTTTTTT**GGGTCTTAAAGTTAAACCTTAAGGTTCTCCTATAGTGAGTCGTATTAA**–3′

[a]Nucleotides involved in base-pair formation are highlighted in bold.

to trypanosomatids (Figure 1), contains numerous aspartic and glutamic acid residues that are clustered into two distinct regions. *Insert 2* comprises a Q,G-rich region that is identified as part of a compositionally biased (38) and intrinsically disordered domain (39). *Insert 3* represents an expansion of a β-hairpin region at the dimerization interface that is much shorter in the reference templates. *Insert 4* is a C-terminal extension that, of all available templates, is only present, albeit with low sequence identity, in the CTD of human EXOG. Our updated model is now a homodimer that incorporates all insertions and provides a structural rationale to the functions experimentally assigned to inserts 1 and 4.

Active site signature. The active site of *Li*EndoG consists of a ββα-metal finger that contains an SRGH signature (9) in place of the canonical DRGH ProSite motif (PDOC00821) characteristic of most other members of this family of nucleases (4). In the apo state, the catalytic Mg^{2+} ion is coordinated directly by the carboxamide oxygen of the Asn246 side chain and five water molecules, four of which are hydrogen bonded to the carboxylate oxygens of Glu254, the imidazole Nδ of His214 and the side-chain carboxamide of Gln241 (9).

Dimerization domain. The 3D structural superposition of *Dm*EndoG, *Ce*EndoG and EXOG demonstrates that the interface residues involved in dimer stabilization are highly conserved in these three enzymes. Since a multiple sequence alignment shows that many of these residues are also present in *Li*EndoG (Figure 1) we considered *Li*EndoG to be, in all likelihood, a homodimer as well. This view is strongly supported by results from analytical ultracentrifugation proving that the biologically active form of a *Li*EndoG lacking only the CTD (*Li*EndoGΔ^CTD) is indeed a homodimer (Figure 2). Given this premise, it is noteworthy that the two active sites are displayed, as is true for the template dimeric enzymes as well, on opposite sides of the dimerization interface (Figure 3; LiEndoG_apo.pdb in Supplementary material).

Disulfide bridges. Our updated *Li*EndoG molecular model also gave us additional confidence on the existence—for which there is no precedent in any of the other nucleases used as templates—of the two disulfide bridges postulated earlier (9), one between Cys100 and Cys116 and another linking Cys434 and Cys439 (Figure 1). With respect to the latter, which is just upstream of the CTD, the positionally equivalent thiol groups of Cys294
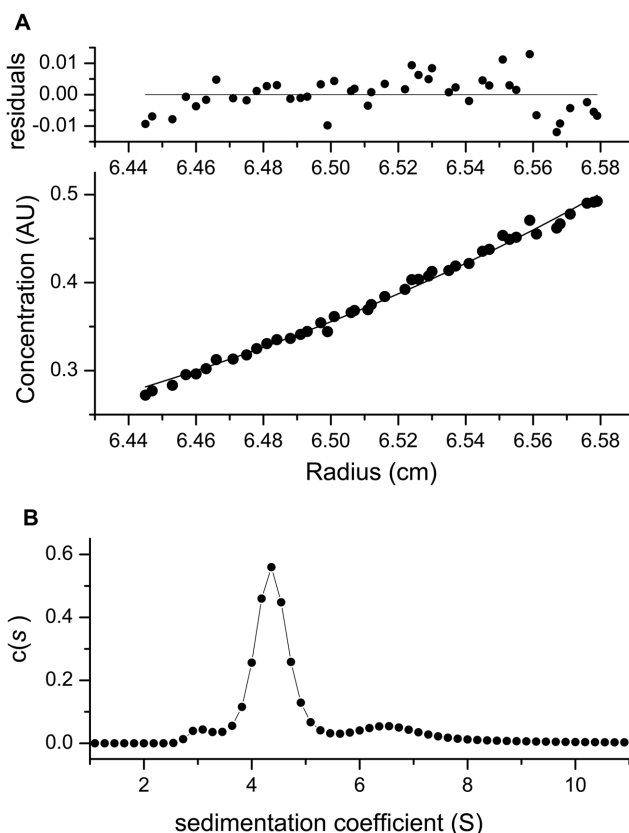

Figure 2. *Li*EndoG quaternary structure. (**A**) Sedimentation equilibrium profile of *Li*EndoG (0.77 mg/ml) in 50 mM MES, 500 mM NaCl, 0.4 mM KCl buffer (pH 6.5) at 8,000 rpm and 20°C. The lower part shows the experimental data (circles) and the best fit (solid line) to a single dimeric species of 91 ± 2 kDa. The upper part shows residuals of the theoretical fit. (**B**) Distribution of experimental sedimentation coefficients of the same sample in sedimentation velocity experiments (45 000 rpm, 20°C). Data shown are representative of at least four independent experiments.

and Cys299 in PDB entry 4A1N appear close enough for a disulfide bond to form under oxidative conditions but this EXOG protein was expressed in the reducing environment of the *E. coli* cytoplasm (40,41) and no details about the experimental procedures are publicly available (11). Remarkably, the amino acid sequence of yeast endonuclease G Nuc1p (*Sc*Nuc1p) (42,43) also shows two cysteines in equivalent positions (Supplementary Figure S3), which suggests that the existence of a disulfide bond between these two residues is highly plausible and may play a functional role. In any case, it must be noted that all of
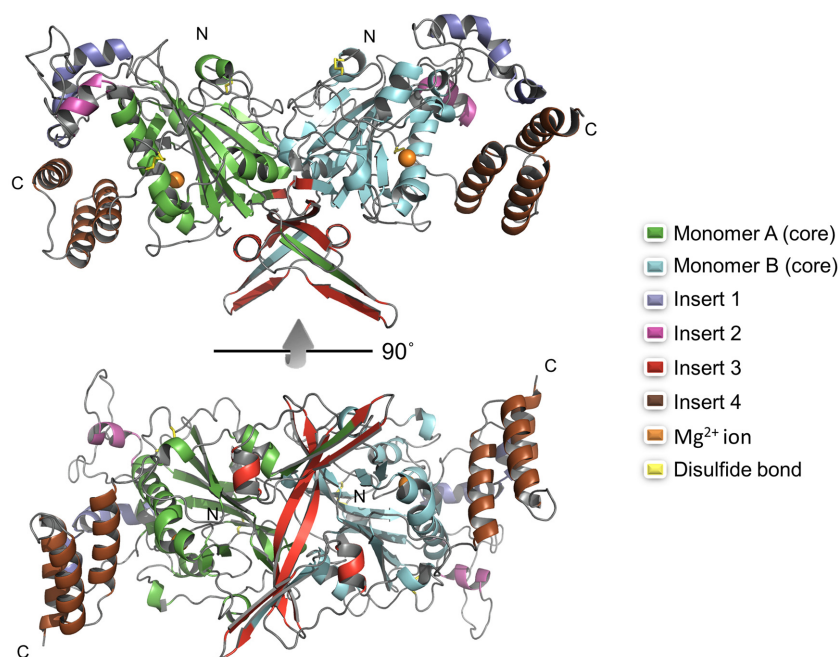
Figure 3. Molecular model of homodimeric *Li*EndoG. Two side views differing in a 90° rotation, as indicated, are shown. The central cores of the two monomers are colored in green and cyan. The Mg^{2+} ions bound in each active site, close to the slightly distorted α-helix, are displayed as orange spheres. Inserts 1, 2, 3 and 4 in both monomers are colored in blue, purple, red and brown, respectively. The putative disulfide bonds are displayed as yellow sticks. Note the extended interfacial β-strands giving rise to a long antiparallel β-sheet in the dyad axis.

these postulated cystines are intrasubunit and we do not find any structural clue pointing to the possible existence of intermonomeric crosslinks in *Li*EndoG.

Insert 1 (residues 145–180). *Dm*EndoG's activity is known to be negatively regulated by an inhibitory protein (*Dm*EndoGI) that binds to it with subpicomolar affinity and consists of two highly homologous domains linked by a polypeptide chain. Similarly, the activity of the cyanobacterial, nonspecific and extracellular nuclease NucA is kept in check by the inhibitor NuiA. The structural resolution of *Dm*EndoG:*Dm*EndoGI (6) and NucA:NuiA (5) complexes by X-ray crystallography (PDB entries: 3ISM and 2O3B, respectively) has provided valuable insights into the binding determinants and mechanisms of inhibition, which are totally unrelated. Thus, *Dm*EndoGI docks a loop region rich in acidic residues into the oligonucleotide-binding groove of *Dm*EndoG to form a plethora of salt bridges with Arg side-chains (6) whereas NuiA makes use of its C-terminal threonine to directly coordinate to the catalytic Mg^{2+} cation (5,44). To the best of our knowledge, no such inhibitors have been described in *Leishmania* but the presence of this insertion in the *Li*EndoG sequence (Figure 1) made us suspect that this region could adopt a definite secondary structure (Figure 3) which, under certain conditions, might block the active site and prevent binding of the negatively charged polynucleotides. The modeling of this ∼30-aa segment as a candidate domain for self-regulation of the nuclease activity was far from trivial because no suitable templates could be found in the whole PDB. Nonetheless, the algorithm implemented in the Robetta web server (25) filled in the ^{145}Pro−Asp^{180} gap and provided us with a suitable geometry for this partially disordered region, which likely in-

cludes a short α-helix encompassing the polyalanine expansion (45).

Insert 2 (residues 265–300) represents a sequence stretch which extends from the C-terminus of the α-helix that contains the catalytic Asn246 up to the N-terminus of a β-strand that makes up the long central intermonomeric β-sheet. We tentatively modeled the polyglutamine tract as an α-helix (26,27), filled in the gap (28), and relaxed the conformation of the resulting loop by means of a restrained MD simulation. This region is spatially close to *insert 1* (Figure 3) and contains a positively charged ^{277}GRGRQR^{282} segment that we think provides affinity for polyphosphate binding.

Insert 3 (residues 316–363) is found within a region that is structurally conserved in all homologous enzymes used as templates (Figure 1). Thus, EXOG, *Dm*EndoG and *Ce*EndoG show a common 'flap' or 'wing' that protrudes from the globular domain of each monomer and is located at one end of an antiparallel β-sheet, the longer β-strand of which (residues 347–358) contributes to the dimerization interface by forming an intermolecular β-sheet (Figure 3). Significantly, in *Li*EndoG the Thr352 residues from each subunit come close together midway on one side of this β-sheet whereas the ends appear to be stabilized by a salt bridge between Lys348 in one monomer and Glu354 in the other. This electrostatic interaction seems to be highly relevant for dimer stabilization because it is also conserved in homodimeric *Dm*EndoG (6) whereas the positionally equivalent residues in human EXOG (18) and *Ce*EndoG are Lys209:Gln215 and Lys218:Gln224, respectively, also allowing for the formation of hydrogen bonds. On the other side of this β-sheet, the Tyr353 residues from both subunits

also meet, and their phenol rings are involved in strong hydrogen bonding interactions with ion-pairing side-chains protruding from the [127]ERR[129] loop, which also appears to be key for dimerization.

On the basis of the multiple sequence alignment shown in Figure 1, we predict that residues 314–328 in *Li*EndoG make up an extended β-hairpin. A sequence similarity search using FASTA (46) on the UniProt Knowledgebase (47) revealed that the amino acid stretch immediately downstream is unique to *Leishmania*. To model this region, we reasoned that the [321]RPDG[324] stretch must be located at the tip of the hairpin because both Asp323 and Gly324 are structurally conserved in *Dm*EndoG, *Ce*EndoG and human EXOG as the central residues in the distinct flaps (or 'wings') that, together in the dimer, resemble a saddle. We believe that this turn is potentially involved in crucial protein-protein interactions because a similar loop with an identical [18]RPDGT[22] sequence at the tip is also present in the cytosolic protein PYM (PDB entry: 1RK8) (48), in which it defines the effector-binding domain to the core of the exon junction protein Mago-Y14 that participates in postsplicing mRNA metabolism (49).

Interestingly, placing [321]RPDG[324] in *Li*EndoG at the turn of the β-hairpin suggested to us that [316]MRVEH[320] and [325]TEVHV[329] could also be part of the antiparallel β-strand. By carrying out this modeling work, we managed to reduce the length of sequence without an assigned structure. Next, we realized that [333]SAEASAAPWKSA[344] in *insert 3* is highly similar to the [1184]SKTASASPWKSA[1195] sequence present in an amphiphilic α-helical region of the cytoskeletal protein αII spectrin that has been shown by X-ray crystallography (PDB entry: 2FOT) to fit snugly within the hydrophobic binding groove of Ca^{2+}-bound calmodulin (50). We could then hypothesize that amino acids 314−344 in each *Li*EndoG monomer adopt a ββα structure that allows extension of the common flap in the eukaryotic family of EndoG nucleases while maintaining a similar fold in this region (Figure 3). We found indirect support to this assumption in the crystal structure of TM1367 protein from *Thermotoga maritima* (51), which includes a sequence identical to [316]MRVE[319] of *Li*EndoG at an N-terminal β$_1$β$_2$α$_1$ motif. The fact that this stretch is flanked by two prolines is suggestive of the need for a forced turn in the secondary structure so as to fit in this insertion while keeping the overall shape of the rest of the protein. Although these findings are surely not enough to guarantee the goodness of our model, we are confident about the feasibility of these structural features, which pave the way for further research in the absence of an X-ray crystal structure.

Insert 4. Multiple sequence alignments and homology modeling of residues 440–510 of *Li*EndoG consistently found a similar CTD in human EXOG (18) as the only structural template available in the PDB, on the basis of its predicted secondary structure, despite its relatively low sequence identity (11). Unfortunately, due to poor electron density, a number of side chains were missing from the model deposited in this database and reasonable doubts arose about the plausibility of the interhelical loop regions, particularly involving helix 2. Likewise, we could not be sure about the exact composition and relative orientation of the second and third helices. Refinement of PDB entry 4A1N using Coot (19) and Phenix (20,21) substantially improved the figures of merit ($R_{work} = 0.191$ and $R_{free} = 0.250$) and provided a Ramachandran map free of outliers (Supplementary Figure S1) and a MolProbity (52) overall score of 2.49. The optimized sequence alignment obtained when this refined structure—with the right register of amino acids—was used as input for the one-to-one threading method implemented in Phyre (16) gave us extra confidence that the CTD of *Li*EndoG consists of a three-helix bundle that places helix 1 spatially close to the DNA-binding groove.

Substrate recognition and binding. Re-refinement of PDB entry 5GKP allowed us to trace the position of nucleotides 3−8 of 5'-d(TTTTTTGT)-3' in chain C and 1−7 in chain D bound to monomers A and B of the H148A/F122A variant of *Ce*EndoG, respectively (Supplementary Figure S2). The Mg^{2+} located within the ββα-metal motif is coordinated by the phosphate oxygens O3' and OP1 of T5 and T6, respectively, in subunit A, and O3' and OP1 of T2 and T3 in subunit B. Thus, we considered the phosphodiester bond between these O3' atoms and the phosphorus of the following nucleotide as site 0 for cleavage (Table 1). Protein residues involved in fixating the phosphates upstream and downstream of this site were also identified. A best-fit root-mean-square Cα superposition of EXOG and *Li*EndoG onto *Ce*EndoG then allowed us to locate the equivalent residues in these two enzymes and find a rationale for the differences in substrate selectivity, as will be discussed below.

Mutations leading to replacement of Ser211 with Ala or Asp in the active site strongly affect *Li*EndoG DNA cleavage efficiency

The SRGH active site motif in *Li*EndoG is rather unique among nonspecific ββα-metal finger nucleases (53), which usually contain the 'canonical' DRGH signature. To assess the importance of Ser211 in the DNA cleavage ability of *Li*EndoG, we expressed the Ser211Ala and Ser211Asp variants and assayed their nuclease activities. Whereas the Ser211Ala variant showed only a moderate reduction in activity, replacement of this Ser with the 'canonical' Asp resulted in a dramatic loss of catalytic efficiency (Figure 4). This observation is in consonance with results reported for the activity on ssDNA of human EXOG upon replacement of the equivalent Ser137 with Asp (18).

Inspection of our template structures revealed that the active site Asp in the ββα-domains of *Dm*EndoG and *Ce*EndoG establishes a strong electrostatic interaction with an arginine located at the N-terminus of an α-helix. Because in *Li*EndoG an Ala occupies this position, we thought that a second Ala→Arg mutation could restore the lost activity in the S211D variant. Indeed, when we expressed and tested the S211D/A247R double mutant, we found that 75% of the activity of the wild-type enzyme had been recovered (Figure 4). This finding supports the view that in *Li*EndoG$^{S211D/A247R}$ (as happens with 'true' DRGH-containing ββα-metal finger nucleases) the guanidinium of Arg247 is used to anchor one of the side-chain carboxylate oxygens of Asp211 in place so that the other oxygen
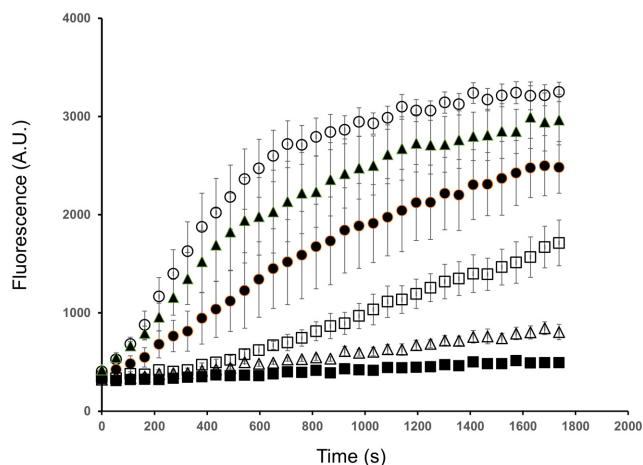
Figure 4. Time course of the digestion of the 5′-FAM/3′-TAMRA-labeled 16-bp dsDNA probe by wild-type *Li*EndoG (○) and by the protein variants generated by site-directed mutagenesis of the coding sequence: *Li*EndoG^{S211A} (●), *Li*EndoG^{S211D} (□), *Li*EndoG^{H214A} (■), *Li*EndoG^{R212A} (△), and *Li*EndoG$^{S211D/A247R}$ (▲) over 30 min of incubation. Error bars represent the standard error obtained from four independent experiments ($n = 4$).

hydrogen bonds to the carboxamide nitrogen of the active site Asn246. In the wild-type *Li*EndoG this carboxamide nitrogen is hydrogen bonded to the hydroxyl oxygen of Ser211, which points away from the hydrophobic side chain of Met243.

For comparative purposes and completeness, we also replaced His214 with Ala. As expected, this H214A variant was unable to degrade DNA, given the crucial role of this residue as the general base that activates a water molecule for the nucleophilic in-line attack on the phosphorous atom during the nuclease reaction (30). Similarly, replacement of Arg212 with Ala generates a variant that is virtually inactive (Figure 4).

*Li*EndoG is endowed with a pH-sensitive self-inhibitory domain

Insert 1 is located in close proximity to the active site and approximately in the same spatial region that is occupied by *Dm*EndoGI in the *Dm*EndoG:*Dm*EndoGI complex (6) (Figure 3). The presence of numerous aspartic and glutamic acid residues in the sequence of this EndoG inhibitor, as well as previous results from our laboratory demonstrating that *Li*EndoG is inhibited at pH 8.0, led us to think that this stretch could be acting as a self-inhibitory domain responsible for the pH-dependence of this enzyme. To test this hypothesis, the Asp/Glu-rich stretch comprising residues 145–180, which is exclusive to trypanosomatids (Supplementary Figure S4), was replaced by the sequence AKNDA that occupies the equivalent position in *Dm*EndoG (Figure 1). The activity of this *Li*EndoGΔi variant was compared to that of the wild-type enzyme over a range of pH values. The results obtained (Figure 5A) demonstrate that deletion of this region abolishes inhibition at pH values of 7.5−8. Indeed, whereas the wild-type enzyme is extremely sensitive to 0.5−1.5 pH increments, *Li*EndoGΔi preserves its catalytic proficiency over the whole range of pH values studied.

Once the pH-dependent modulation of this self-inhibitory domain was demonstrated, we focused on analysing the relevance of some of the negatively charged residues present in this sequence stretch. We replaced the three contiguous aspartic residues located in the central region of the domain (^{160}DDD^{162}) with three alanines to generate *Li*EndoG$^{D(160-162)A}$ and the nuclease activity of this variant was studied at different pH values. The results shown in Figure 5B demonstrate that these three aspartates are essential for *Li*EndoG's self-inhibition as the *Li*EndoG$^{D(160−162)A}$ variant essentially recapitulates the behavior of *Li*EndoGΔi over the range of pHs tested.

Deletion of the C-terminal domain affects the DNA cleavage pattern

Alignment of EndoG sequences from trypanosomatids with those from endo/exonucleases of other organisms such as *D. melanogaster*, *C. elegans*, *S. cerevisiae* and *H. sapiens* (Supplementary Figure S5) reveals the presence of an extended CTD with low sequence identity but predicted high structural similarity to the CTD of human EXOG. To define the role of this CTD, we generated a truncated *Li*EndoG protein lacking the last 70 residues (*Li*EndoGΔ^{CTD}), which was expressed, refolded, and purified following the same procedures that were used to obtain wild-type *Li*EndoG.

The equilibrium sedimentation profiles of *Li*EndoGΔ^{CTD} could be fitted to a single ideal species model (Figure 2A) with an apparent molecular mass ($M_{w,app}$) of 91 ± 2 kDa (mean of six independent experiments), which is 1.98-fold the value measured by mass spectrometry. Sedimentation velocity experiments carried out in parallel showed the presence of three different species (Figure 2B), the major one being compatible with a dimer ($s_{20,w} = 5.0 ± 0.07$ S; $M_w \sim 81$ kDa) and the other two with a monomer ($s_{20,w} = 3.2 ± 0.1$ S; $M_w \sim 44$ kDa) and a tetramer ($s_{20,w} = 7.6 ± 0.2$ S; $M_w \sim 159$ kDa; mean of four independent experiments). Their percentages were consistent with the $M_{w,app}$ calculated from equilibrium experiments and an upper limit of 14 nM was estimated for the dimer dissociation constant from the percentages of species (around 81%, 4% and 15%, respectively) measured at the lowest *Li*EndoGΔ^{CTD} concentration used.

When the activities of wild-type *Li*EndoG and *Li*EndoGΔ^{CTD} on plasmid DNA were compared (Figure 6A), similar nicking of the supercoiled substrate was observed although the concentration of *Li*EndoGΔ^{CTD} had to be increased fourfold to reach an equivalent efficiency. Accordingly, in all subsequent comparative experiments, protein concentrations were adjusted so as to maintain similar enzymatic activities.

The nuclease activity of both enzymes was then assayed on a 5′-^{32}P-labeled 42-bp-long dsDNA helix and the generated fragments were subjected to denaturing PAGE. Digestion of the same probe with DNase I was included as a control. Both *Li*EndoG and *Li*EndoGΔ^{CTD} produced DNA fragments of many different sizes (Figure 6B). As expected from digestion of a 5′-labeled probe by an endonuclease with no sequence specificity, fragment size decreased on increasing the amount of enzyme. Low enzyme concentrations gave rise to an almost random distribution of DNA
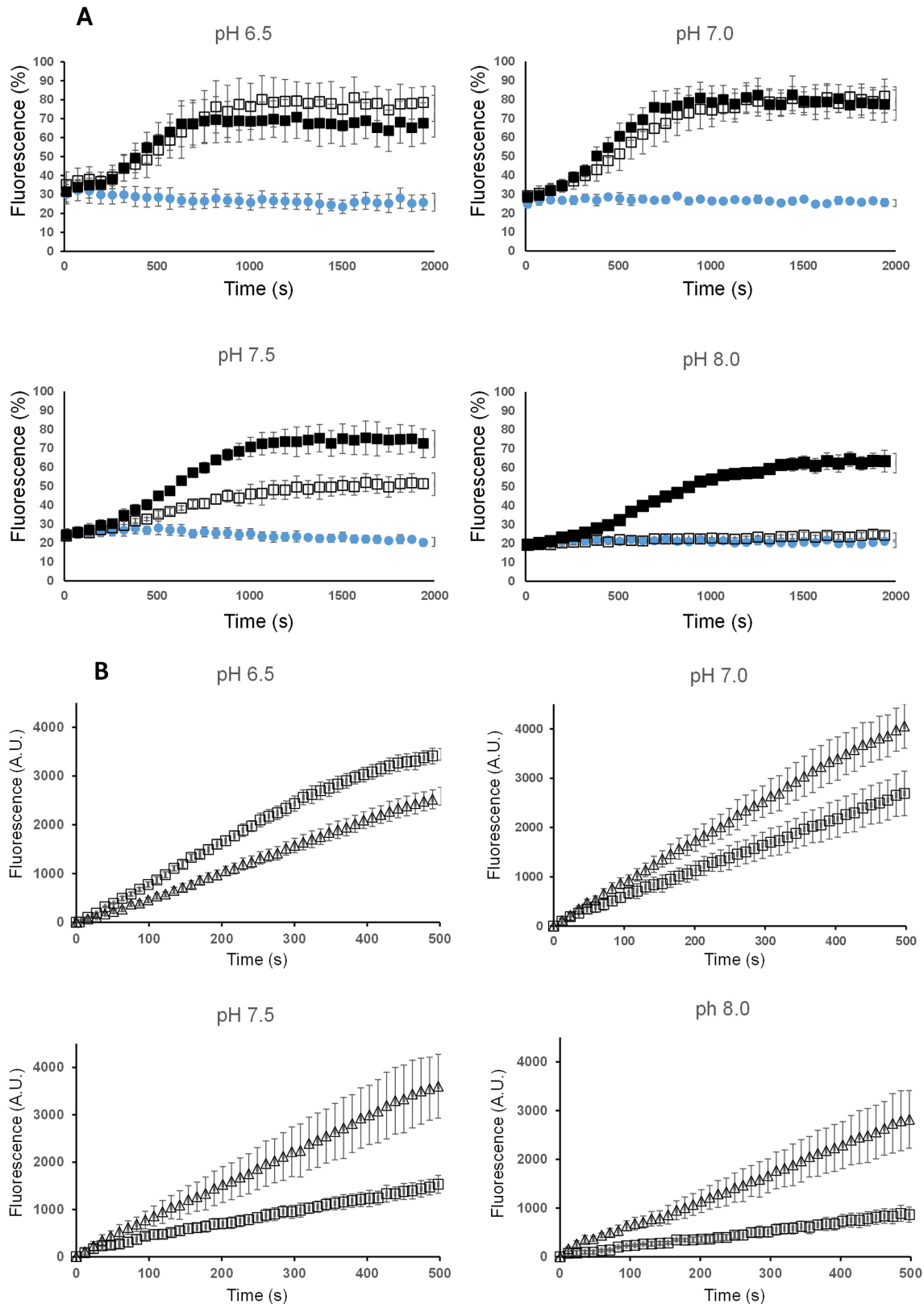
Figure 5. (**A**) Comparison of the endonucleolytic activities on a dually 5′-FAM/3′-TAMRA-labeled dsDNA probe of wild-type *Li*EndoG (□) and *Li*EndoGΔi (■) at different pH values over >30 min of incubation. (●) Fluorescence of the probe in the absence of enzyme. All data are normalized against the fluorescence values of the 5′FAM-labeled ssDNA probe. (**B**) Activity of wild-type *Li*EndoG (□), and *Li*EndoG^D(160-162)A (△) on the same probe at different pH values over the first 10 min of incubation. Error bars represent the standard error obtained from three independent experiments (*n* = 3).
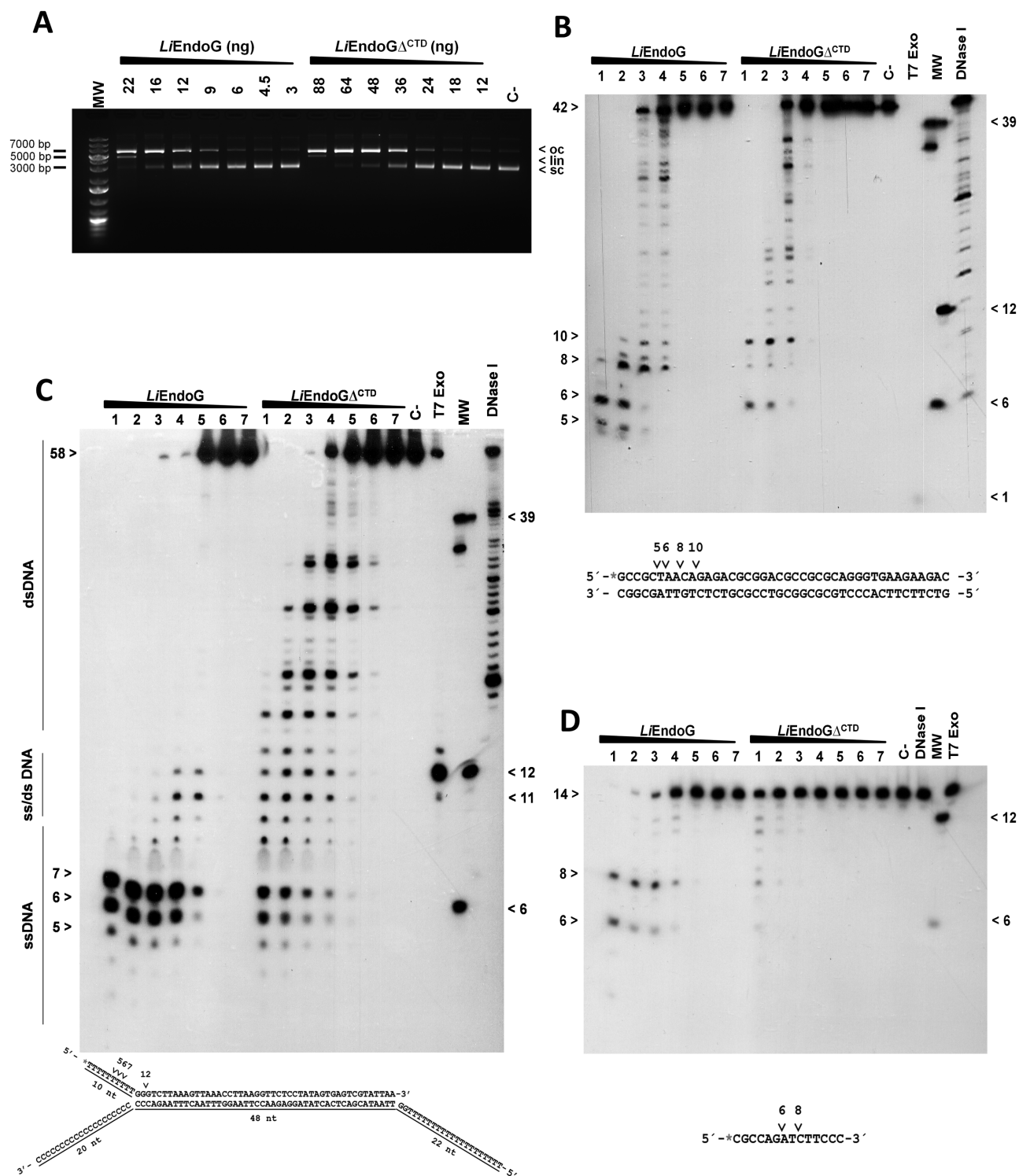
Figure 6. DNA digestion products of wild-type *Li*EndoG and *Li*EndoGΔ^CTD. (**A**) Agarose gel electrophoresis of supercoiled plasmid DNA in the absence of nucleases (C-) and after digestion with increasing amounts of each enzyme. Due to removal of torsional stress by the endonucleolytic cleavage of one of the two DNA strands, supercoiled molecules (sc) transform into open circular forms (oc). Digestion at the highest enzyme concentrations generates a faint band corresponding to linear molecules (lin). (**B–D**) Denaturing PAGE bands showing the different ^{32}P-labeled DNA probes (sequences indicated in each figure) and DNA fragments after incubation with either the digestion buffer in the absence of protein (C-) or increasing (from right to left) concentrations of enzyme. The red asterisk denotes the radioactively labeled position. Fragments generated after digestion by either T7 exonuclease or DNase I are shown on the right-hand side, as well as the 12-nt and 6-nt-long fragments loaded in the gel as molecular weight (Mw) markers.

fragments that progressively disappeared as the concentration of enzyme increased. The smallest DNA fragments detected were 5-, 6-, 8- and 10-mers for *Li*EndoG and 6- and 10-mers for *Li*EndoGΔ^{CTD}. A remarkable difference was that *Li*EndoG quickly generated small-sized fragments. Note that approximately half of the radioactive signal is associated to an octanucleotide, *lane 3*, even when there is still a large proportion of undigested probe. In contrast, upon incubation with *Li*EndoGΔ^{CTD} high concentrations of short oligonucleotides are visible only after all the original DNA probe has been completely digested.

To test the activity of both enzymes on ssDNA and dsDNA, we selected a probe with a 48-nt-long central dsDNA stem flanked on each side by a differently sized ssDNA overhang (Figure 6C). Strikingly different behaviors were found: the wild type showed an extraordinary affinity for the ssDNA region, as revealed by the complete absence of DNA fragments longer than 12 nucleotides (Figure 6C), whereas this particular preference for ssDNA was not observed for *Li*EndoGΔ^{CTD}, which gave rise to a degradation pattern quite similar to that observed upon cleavage of a dsDNA substrate possessing blunt-end termini (Figure 6B). In addition, only *Li*EndoG showed a definite preference for cleaving the probe at positions 7 and 11, as fragments of these two sizes were the first to be observed upon increasing concentrations of the enzyme (*lane 5*). Since position 11 marks precisely the first nucleotide of the dsDNA stem, it seems feasible that *Li*EndoG can eliminate 5′ overhangs appended to a dsDNA, an activity typically displayed by flap endonucleases. Once generated, the 11-nt-long fragment is quickly cleaved and shorter ones appear but 5−6-mers are not degraded further.

Finally, to reinforce the idea that the CTD is responsible for the preferential activity of *Li*EndoG on ssDNA, we selected as substrate a short ssDNA fragment unable to form any dsDNA structure. The results shown in Figure 6D reveal that *Li*EndoGΔ^{CTD} displays such a low affinity for ssDNA that only a negligible degradation pattern is observed at the highest enzyme concentration assayed (*lane 1*, *Li*EndoGΔ^{CTD}). In stark contrast, this probe is readily cleaved by wild-type *Li*EndoG (starting in *lane 4*).

Analysis of the final products of DNA digestion by *Li*EndoG consistently shows that the smallest ssDNA substrate that is cleaved needs to be longer than a pentamer or a hexamer. We take this as an indication, in agreement with our molecular models (Figure 7, Supplementary Figure S6, LiEndoG_ssDNA.pdb and LiEndoG_ssDNA+dsDNA.pdb in Supplementary material), that shorter oligonucleotides do not bind with sufficient affinity because they lose some relevant interactions with the enzyme (Table 1).

DISCUSSION

We previously reported (9) that the core of the catalytic domain of *Li*EndoG is likely to display an overall similarity to other structurally well-characterized extracellular nucleases such as those from *Serratia marcescens* (54) and *Anabaena* sp. (55). These two proteins belong to the ββα-metal superfamily of nucleases which share a structurally conserved active site scaffold and utilize a divalent metal

ion to catalyze the hydrolysis of the phosphodiester bond (56). We proposed that, in *Li*EndoG, the hydrated Mg^{2+} ion is coordinated directly through the side-chain carboxamide oxygen of Asn246 and also that one of the Mg^{2+}-bound water molecules is hydrogen-bonded to the imidazole Nδ of His214, which behaves as a general base to activate a water molecule for the nucleophilic in-line attack on the phosphorous atom of DNA or RNA substrates (9). This catalytic His214 is part of the 'SRGH' motif characteristic of a second family of sugar non-specific nucleases in higher eukaryotes, represented by EXOG, which also have an additional CTD consisting of ∼70 amino acids (18). Human EXOG and EndoG were shown to have complementary and partially overlapping enzymatic activities. When Ser137 in the catalytic site of EXOG was replaced by the aspartic acid residue present in the canonical 'DRGH' motif the enzymatic activity of the resulting S137D variant was lower than that of the wild type (18). In our hands, the DNA cleavage ability of the S211D variant of *Li*EndoG was strongly reduced. This was intriguing, especially in view of the additional finding that this nuclease activity was affected only slightly in the S211A variant (Figure 4). Therefore, we thought that the deleterious effect of Asp211 was mostly due to the negative charge in the side chain of this amino acid. Re-examination of the macromolecular structures revealed that in *Dm*EndoG and *Ce*EndoG the carboxylate of the equivalent Asp is involved in a salt bridge with the guanidinium of Arg181, the spatial counterpart of which is an alanine (Ala247) in *Li*EndoG. When an arginine was introduced at this position, the nuclease activity of the S211D/A247R variant arising from the double mutation was ∼75% of that of the wild-type enzyme. Therefore, we propose that the carboxylate of Asp211 disorganizes the active site and faces the nucleotide phosphates unless it is firmly anchored in place by the guanidinium of Arg247.

As we demonstrated earlier, *Li*EndoG plays a dual role in both cell death and survival (9,10). Due to the importance of controlling these sugar nonspecific nucleases, different molecular mechanisms have been found in nature to finely modulate their activity. Among them, cellular compartmentalization, the existence of specific inhibiting proteins, and regulation by pH are good examples (5,6,44). For *Li*EndoG we have demonstrated abolition of nuclease activity at high pH values (pH $\geq$ 8.0). In this respect, different approaches indicate pH values close to 8 in the mitochondrial matrix and a reduction of about 1 pH unit in the cytosol (57,58). Accordingly, *Li*EndoG would be almost completely inhibited inside the mitochondrion but could become activated upon its release to the cytosol during cell death. This activation could be further potentiated by the decrease in cytosolic pH values observed during apoptosis-like processes (2,59). However, since both cytosolic and mitochondrial pHs can be strongly reduced during activation of cells with Ca^{2+}-mobilizing agonists (58), the activity of *Li*EndoG inside the mitochondrion might be stimulated in different cellular conditions. In fact, any decrease in mitochondrial activity would decrease proton transport rates and cause a reduction in the pH within the matrix that, eventually, would indirectly activate *Li*EndoG.

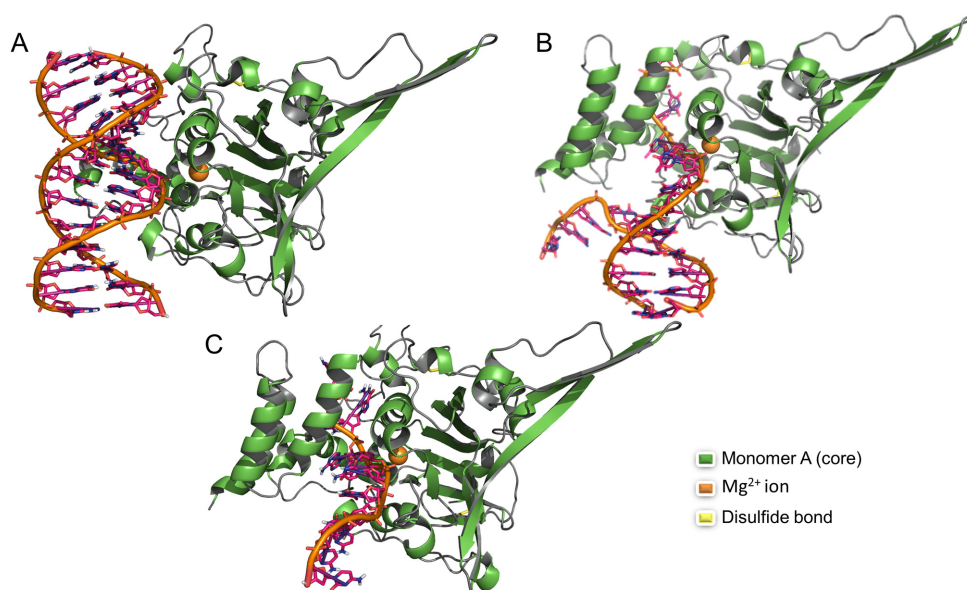Our updated molecular model of *Li*EndoG indicates that, in all probability, the domain corresponding to

Figure 7. Molecular models of (**A**) *Li*EndoGΔ^CTD in complex with a dsDNA molecule, (**B**) wild-type *Li*EndoG in complex with a ssDNA–dsDNA junction and (**C**) wild-type *Li*EndoG in complex with a ssDNA oligonucleotide. Only one protein monomer is shown for clarity and a rainbow coloring scheme has been chosen to facilitate tracing of the backbone from the N-terminus (blue) to the C-terminus (red). The corresponding DNA sequences are displayed in Table 2. The dotted arrows in (B) and (C) point to residues F447, Y450 and R451 in helix 1 of the CTD (shown as sticks).

residues 145−180 is located in the same spatial region that is occupied by *Dm*EndoGI in its complex with *Dm*EndoG (6). To demonstrate the self-inhibitory function of this unique amino acid stretch, we first expressed a variant enzyme in which this domain was deleted (*Li*EndoGΔi) and then measured the nuclease activity of this engineered protein in the presence of a 5′-FAM/3′-TAMRA-labeled dsDNA oligonucleotide at different pH values ranging from 6.5 to 8.0. As shown in Figure 5A, the nuclease activity of *Li*EndoGΔi is not reduced at basic pH values (7.0−8.0). It must be pointed out that the activity of both the wild-type enzyme and this variant is maximum at pH 6.5. Accordingly, deletion of this domain does not shift the optimum pH value but broadens the range of pH values at which this enzyme is active. Furthermore, the putative role of the three central Asp residues (^{160}DDD^{162}) in the mechanism of self-inhibition was assessed by site-directed mutagenesis. The finding that replacement of these three aspartates with three alanines renders a protein (*Li*EndoG$^{D(160-162)A}$) that is still active at pH 8.0 reinforces the relevance of these acidic residues and points to an effective pH-dependent self-regulatory mechanism in *Li*EndoG.

Our experimental results indicate that the C-terminal extension present in *Li*EndoG, which –to the best of our knowledge—has a known structural counterpart only in human mitochondrial EXOG, is responsible for its high affinity toward ssDNA. Remarkably, deletion of this domain in EXOG or replacement of Gly277, closely upstream of the CTD, with valine results in variants that are much less active or as inactive on ssDNA as the H140A enzyme, respectively (18). This indicates the relevance of both the presence and the positioning of this domain in EXOG for its enzymatic activity. Furthermore, when working on ssDNA or dsDNA substrates, EXOG produces di- and mononucleotides or mostly dinucleotides, respectively, but

this 5′–3′ exonuclease activity is abolished in the G277V and △^CTD variants (18). Since no mono- or dinucleotides are detected upon digestion with *Li*EndoG, despite the presence of a CTD and significant sequence similarity with EXOG, *Li*EndoG cannot be considered a *bona fide* exonuclease. In any case, it seems clear that the substrate preferences are in both cases dictated by the presence of this C-terminal extension.

The modeling work presented here leads us to think that the CTD of *Li*EndoG folds into a three-helix bundle similar to that found in human EXOG and different from that reported for DnaT, a primosomal protein that is required for the stalled replication fork restart in *E. coli* (60). The unpublished crystallographic work on EXOG (11), improved upon by our own refinement (Supplementary Figure S1 and Table S1), sheds light on the preferential cleavage by these enzymes because it shows in atomic detail the spatial relation between the CTD and the nucleotide-binding groove in the vicinity of the active site. When the EXOG structure is best-fit superimposed onto that of the *Ce*EndoG(H148A/Q130K) in its complex with ssDNA, insight is gained about the contacts established between these nucleases and this type of substrate. This knowledge can then be safely extrapolated to *Li*EndoG to complement what is already known about its interaction with dsDNA (9). Thus, we learnt that the three helices in the CTD bundle are arranged in such a way that, when positioned close to the active site, they prevent binding of a dsDNA to the enzyme's active site (Figure 7). Nonetheless, passage of a ssDNA is still possible, hence the affinity of *Li*EndoG for ssDNA–dsDNA junctions. At the junction, the displaced unpaired DNA strand is bent at a single phosphodiester bond by the loop connecting helices 1 and 2 of the CTD in a manner reminiscent of that described for helices α2 and α3 of human Exo1 (61). The hydroxyl group of the strictly con-

served Ser461, at the tip of this loop, and the guanidinium of Arg492 in helix 3, can theoretically hydrogen bond to a phosphate in the displaced DNA strand, so as to stabilize the unpaired state (Figure 7B), with the aid of the side chains of Asn184 and Asn187.

In addition, helix 1 of the CTD projects the side chains of the strictly conserved Phe447, Tyr450 and Arg451 toward the ssDNA substrate. We propose that the side chain of Tyr450, with the structural support of Phe447, plays the roles of (i) preventing the stacking of the base in the second nucleotide 5′-upstream of that undergoing the cleavage −in a manner reminiscent of the *hydrophobic wedge* observed in FEN-1 family members (61), (ii) establishing a hydrogen bond with a phosphate oxygen in the nucleotidic substrate and (iii) introducing a turn in the ssDNA to be cleaved such that upstream nucleotides are directed toward the polyglutamine domain in *insert 2*, thus accounting for the much higher affinity for this type of substrate of the wild-type enzyme relative to that of its $\triangle^{CTD}$ counterpart (Figure 6). Although our model strongly suggests that *insert 2* could behave as a ssDNA-binding domain, we cannot assign to it a definite structure with confidence because it is predicted as an intrinsically disordered region and lacks homology with the limited set of proteins with known structure and similar function (62).

Since this CTD is present neither in *Dm*EndoG nor in *Ce*EndoG, both lacking exonuclease activity, and also because the 5′-3′ exonuclease activity is thought to be essential in human EXOG for DNA recombination and repair (18), we suspected that *Li*EndoG might be competent as an exonucleolytic enzyme or have the capacity to trim bifurcated or branched DNA molecules containing free 5′-single-stranded ends (i.e. 5′-flaps) such as those emerging from replication, recombination and repair processes (63). The CTD of *Li*EndoG directs substrates containing ssDNA toward the active site of the enzyme. Upon its removal, *Li*EndoG$\triangle^{CTD}$ essentially behaves as do human EndoG and *Ce*EndoG (no preference for ssDNA) (8), which naturally lack this C-terminal extension and are regarded as endonucleases (Figure 1). This behavior can be rationalized because it is the C-terminal three-helix bundle, presumably when held in place with the aid of the disulfide linkage between Cys434 and Cys439, that confers affinity for ssDNA, be it alone or as part of a dsDNA-ssDNA junction. In the absence of the CTD in front of the active site, either because it is displaced by a hinge motion from the main body of the enzyme or because it has been deleted, *Li*EndoG can bind dsDNA in the active site and behaves as a typical endonuclease. In connection with this reasoning, the reported high amounts of catalytically incompetent molecules and the opposed differences in activity on a hairpin substrate of recombinant *Ce*EndoG (∼3-fold) and EXOG (∼8-fold) enzymes expressed in *E. coli* when they are purified either from the soluble fraction or from inclusion bodies followed by refolding (8) could be related to the correct formation of these disulfide bonds.

Interestingly, the preferential ability of *Li*EndoG to produce 5−7-mers (Figure 6B−D) rather than the mono- and dinucleotides characteristically produced by EXOG, despite the presence of a structurally similar CTD in both enzymes, can be mostly ascribed to a correlated mutation (64) resulting in the reciprocal positional exchange of positively and negatively charged amino acids in the Arg457:Glu488 and Glu317:Arg346 pairs that stabilize the interaction between helices 1 and 3 through a salt bridge in the respective enzymes (Supplementary Figure S5A). Arg457 is involved in substrate phosphate recognition three nucleotides upstream of the phosphodiester bond that is cleaved in the active site of *Li*EndoG (Table 1) but this favourable electrostatic interaction cannot take place in EXOG through the carboxylate of the positionally equivalent Glu317. Instead, a reciprocal Ala↔Arg exchange at positions 454 and 314 makes it possible for Arg314 in EXOG (PDB numbering), but not for Ala454 in *Li*EndoG, to presumably recognize a terminal 5′-phosphate. In contrast, the positions occupied by Arg442, Arg449 and Arg259 in *Li*EndoG, which can recognize a terminal 5′-phosphate 4−5 nucleotides upstream of the cleavage site (Table 1), are occupied in EXOG by Leu302, Leu309 and Glu184, which are incapable of such recognition. This structural rationale accounts well for the observed differences in nuclease activities exerted by these enzymes (8).

Because the nuclease activity in the S211D variant of *Li*EndoG was restituted up to 75% of that of the wild-type enzyme upon the concomitant introduction of a A247R replacement, and because EXOG also has (i) an ^{137}SRGH^{140} signature, (ii) a hydrophobic residue (Phe168) next to Ser137, and (iii) a similar CTD, we decided to check the amino acid composition of EXOG at the position equivalent to that of Ala247, which turns out to be a serine (Ser172). Interestingly, the distance (∼10.0−14.0 Å) between the C_β atoms of Ser172 and Arg314 in the X-ray crystal structure of EXOG (or Ala247 and Arg454 in our *Li*EndoG model) is less than or almost equal to twice the length of an Arg side chain. This fact virtually rules out, on electrostatic and steric grounds, the possibility of having two Arg residues simultaneously in these positions and strongly suggests that these sequence differences, which spatially cluster around the 'wedge' Tyr350, go hand in hand (Supplementary Figure S5). Thus, our observations provide a rationale for the reported substrate differences among members of this superfamily of nucleases, which appear to be dictated by the intimate relation in these enzymes between the nuclease domain and the CTD facing the active site cleft.

Remarkably, *Sc*Nuc1p (42,43) can produce not only endonucleolytic products but also dinucleotides due to preferential cleavage close to the 5′-end (8). We note that the CTD of this enzyme, although similarly linked through a (presumably) redox-dependent disulfide bridge to the globular protein, is much shorter than the CTD of both EXOG and *Li*EndoG and unlikely to adopt an α-helical structure. The three-dimensional structure of *Sc*Nuc1p has not been reported yet but we believe that its CTD can fold similarly to the synaptic CTD (residues 220–276) of Maedi-Visna virus integrase, which has been recently solved both in isolation as a dimer by X-ray crystallography at 1.78 Å resolution (UniProtKB entry P35956, PDB code 5LLJ) and, at much lower resolution by cryo-electron microscopy (PDB code 5M0R), as part of the supramolecular assembly that mediates lentiviral DNA integration (65). In this complex, the β$_1$-β$_2$ loops of two CTDs get inserted into expanded DNA major grooves and contribute to target DNA bending. The

Arg-rich cluster present in the CTD of *Sc*Nuc1p could conceivably play the role, if properly positioned, of recognizing the terminal 5′-phosphate located two nucleotides upstream of the cleavage site.

The finding that EndoG from trypanosomatids removes 5′ overhangs from dsDNA in a highly specific and efficient manner suggests some similarities with CRN-1, a *C. elegans* ortholog of flap endonuclease FEN-1 that localizes to nuclei and cooperates with *Ce*EndoG to promote stepwise DNA fragmentation utilizing the endonuclease activity of *Ce*EndoG and both the 5′-3′ exonuclease and gap-dependent endonuclease activities of CRN-1 (66). Cooperation between both enzymes has been proposed to play a critical role in switching the state of cells from DNA replication/repair to DNA degradation during apoptosis. The existence and properties of a DNA flap endonuclease from *Trypanosoma cruzi* (*Tc*FEN1) have just been reported (67) and orthologs have been detected in *L. major* (67) and *L. donovani* (68). *Tc*FEN1 localizes to the nucleus in transfected epimastigotes and its overexpression increases both parasite proliferation and survival against exogenous oxygen peroxide. *Ld*EndoG and *Ld*FEN1 have been shown to co-immunoprecipitate in nuclear extracts following treatment of the parasites with a topoisomerase IB inhibitor (68). Whether or not *Li*EndoG and *Li*FEN1 also cooperate in processing flap fragments during DNA replication and repair is not known yet but *Li*EndoG on its own might well have an activity equivalent to that of *Ce*EndoG and CRN-1 acting together.

In conclusion, the much larger size of *Li*EndoG compared to that of other EndoGs (9) is due to sequence insertions that are not found outside the trypanosomatid family and provide this enzyme with rather unique properties. In the present work, we have identified and characterized the role of two of these insertions, namely a pH-regulated self-inhibitory domain and an extended CTD that confers substrate selectivity for ssDNA, both as a free molecule and as part of a ssDNA–dsDNA junction. Consequently, *Li*EndoG emerges as an endonuclease with the added structure-specific ability to cut preferentially at the base of a ssDNA overhang, in a manner reminiscent of that carried out by two-metal-ion flap endonucleases (69). In addition, the elongated flaps (or 'wings') that make up a saddle-shaped part of the dimerization domain are suggestive of the participation of this enzyme in multiprotein assemblies. All in all, our study paves the way for the characterization of other domains and prompts the investigation of *Li*EndoG's binding partners.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Low,R.L. (2003) Mitochondrial Endonuclease G function in apoptosis and mtDNA metabolism: a historical perspective. *Mitochondrion*, **2**, 225–236.
2. Ramsdale,M. (2012) Programmed cell death in the cellular differentiation of microbial eukaryotes. *Curr. Opin. Microbiol.*, **15**, 646–652.
3. Ball,T.K., Suh,Y. and Benedik,M.J. (1992) Disulfide bonds are required for *Serratia marcescens* nuclease activity. *Nucleic Acids Res.*, **20**, 4971–4974.
4. Wu,S.-I., Lo,S.-K., Shao,C.-P., Tsai,H.-W. and Hor,L.-I. (2001) Cloning and characterization of a periplasmic nuclease of *Vibrio vulnificus* and its role in preventing uptake of foreign DNA. *Appl. Environ. Microbiol.*, **67**, 82–88.
5. Ghosh,M., Meiss,G., Pingoud,A.M., London,R.E. and Pedersen,L.C. (2007) The nuclease A-inhibitor complex is characterized by a novel metal ion bridge. *J. Biol. Chem.*, **282**, 5682–5690.
6. Loll,B., Gebhardt,M., Wahle,E. and Meinhart,A. (2009) Crystal structure of the EndoG/EndoGI complex: mechanism of EndoG inhibition. *Nucleic Acids Res.*, **37**, 7312–7320.
7. Rose,P.W., Prlic,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J. *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
8. Kieper,J., Lauber,C., Gimadutdinow,O., Urbanska,A., Cymerman,I., Ghosh,M., Szczesny,B. and Meiss,G. (2010) Production and characterization of recombinant protein preparations of endonuclease G-homologs from yeast, *C. elegans* and humans. *Protein Expression Purif.*, **73**, 99–106.
9. Rico,E., Alzate,J.F., Arias,A.A., Moreno,D., Clos,J., Gago,F., Moreno,I., Domínguez,M. and Jiménez-Ruiz,A. (2009) *Leishmania infantum* expresses a mitochondrial nuclease homologous to EndoG that migrates to the nucleus in response to an apoptotic stimulus. *Mol. Biochem. Parasitol.*, **163**, 28–38.
10. Rico,E., Oliva,C., Gutierrez,K.J., Alzate,J.F., Genes,C.M., Moreno,D., Casanova,E., Gigante,A., Pérez-Pérez,M.-J., Camarasa,M.-J. *et al.* (2014) *Leishmania infantum* EndoG is an endo/exo-nuclease essential for parasite survival. *PLoS ONE*, **9**, e89526.
11. Welin,M., Moche,M., Arrowsmith,C.H., Berglund,H., Bountra,C., Collins,R., Edwards,A.M., Flodin,S., Graslund,S., Hammarstrom,M. *et al.* (2012) Human mitochondrial endo-exonuclease. *Protein Data Bank*, doi:10.2210/pdb4a1n/pdb.
12. Lin,J.L., Wu,C.C., Yang,W.Z. and Yuan,H.S. (2016) Crystal structure of endonuclease G in complex with DNA reveals how it nonspecifically degrades DNA as a homodimer. *Nucleic Acids Res.*, **44**, 10480–10490.
13. Lin,J.L., Nakagawa,A., Skeen-Gaar,R., Yang,W.Z., Zhao,P., Zhang,Z., Ge,X., Mitani,S., Xue,D. and Yuan,H.S. (2016) Oxidative stress impairs cell death by repressing the nuclease activity of mitochondrial endonuclease G. *Cell Rep*, **16**, 279–287.
14. Lin,J.L.J., Nakagawa,A., Lin,C.L., Hsiao,Y.-Y., Yang,W.-Z., Wang,Y.-T., Doudeva,L.G., Skeen-Gaar,R.R., Xue,D. and Yuan,H.S. (2012) Structural insights into apoptotic DNA degradation by CED-3 Protease Suppressor-6 (CPS-6) from *Caenorhabditis elegans*. *J. Biol. Chem.*, **287**, 7110–7120.
15. Kelley,L.A. and Sternberg,M.J.E. (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.

16. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.

17. Biasini,M., Bienert,S., Waterhouse,A., Arnold,K., Studer,G., Schmidt,T., Kiefer,F., Gallo Cassarino,T., Bertoni,M., Bordoli,L. *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252–W258.

18. Cymerman,I.A., Chung,I., Beckmann,B.M., Bujnicki,J.M. and Meiss,G. (2008) EXOG, a novel paralog of Endonuclease G in higher eukaryotes. *Nucleic Acids Res.*, **36**, 1369–1379.

19. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.

20. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Echols,N., Headd,J.J., Hung,L.W., Jain,S., Kapral,G.J., Grosse Kunstleve,R.W. *et al.* (2011) The Phenix software for automated determination of macromolecular structures. *Methods*, **55**, 94–106.

21. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

22. DeLano,W.L. (2015) *The PyMOL Molecular Graphics System, version 1.8.2.0*. Schrödinger, LLC.

23. Stein,A. and Kortemme,T. (2013) Improvements to robotics-inspired conformational sampling in Rosetta. *PLoS One*, **8**, e63090.

24. Mandell,D.J., Coutsias,E.A. and Kortemme,T. (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods*, **6**, 551–552.

25. Kim,D.E., Chivian,D. and Baker,D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.*, **32**, W526–W531.

26. Miller,J., Rutenber,E. and Muchowski,P.J. (2009) Polyglutamine dances the conformational cha-cha-cha. *Structure*, **17**, 1151–1153.

27. Robertson,A.L., Horne,J., Ellisdon,A.M., Thomas,B., Scanlon,M.J. and Bottomley,S.P. (2008) The structural impact of a polyglutamine tract is location-dependent. *Biophys. J.*, **95**, 5922–5930.

28. López-Blanco,J.R., Canosa-Valls,A.J., Li,Y. and Chacón,P. (2016) RCD+: Fast loop modeling server. *Nucleic Acids Res.*, **44**, W395–W400.

29. Ceroni,A., Passerini,A., Vullo,A. and Frasconi,P. (2006) DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res.*, **34**, W177–W181.

30. Bueren-Calabuig,J.A., Coderch,C., Rico,E., Jiménez-Ruiz,A. and Gago,F. (2011) Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: *Vibrio vulnificus* periplasmic nuclease as a test case. *ChemBioChem*, **12**, 2615–2622.

31. Case,D.A.T.A.D., Cheatham,T.E. III, Simmerling,C.L., Wang,J., Duke,R.E., Luo,R., Walker,R.C., Zhang,W., Merz,K.M., Roberts,B. *et al.* (2012) University of California.

32. Li,C.L., Hor,L.I., Chang,Z.F., Tsai,L.C., Yang,W.Z. and Yuan,H.S. (2003) DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site. *EMBO J.*, **22**, 4014–4025.

33. Wang,Y.T., Yang,W.J., Li,C.L., Doudeva,L.G. and Yuan,H.S. (2007) Structural basis for sequence-dependent DNA cleavage by nonspecific endonucleases. *Nucleic Acids Res.*, **35**, 584–594.

34. Carter,M., Voth,A.R., Scholfield,M.R., Rummel,B., Sowers,L.C. and Ho,P.S. (2013) Enthalpy-entropy compensation in biomolecular halogen bonds measured in DNA junctions. *Biochemistry*, **52**, 4891–4903.

35. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.

36. Monterroso,B., Saiz,J.L., Garcia,P., Garcia,J.L. and Menendez,M. (2008) Insights into the structure-function relationships of pneumococcal cell wall lysozymes, LytC and Cpl-1. *J. Biol. Chem.*, **283**, 28618–28628.

37. Laue,T.M., Shah,B.D., Ridgeway,T.M. and Pelletier,S.L. (1992) In: Harding,SERA and Horton,JC (eds). *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Royal Soc. Chem., Cambridge, pp. 90–125.

38. Wootton,J.C. and Federhen,S. (1996) Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.*, **266**, 554–571.

39. Ward,J.J., McGuffin,L.J., Bryson,K., Buxton,B.F. and Jones,D.T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**, 2138–2139.

40. de Marco,A. (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb. Cell Fact.*, **8**, 26.

41. Stewart,E.J., Aslund,F. and Beckwith,J. (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J.*, **17**, 5543–5550.

42. Buttner,S., Eisenberg,T., Carmona-Gutierrez,D., Ruli,D., Knauer,H., Ruckenstuhl,C., Sigrist,C., Wissing,S., Kollroser,M., Frohlich,K.U. *et al.* (2007) Endonuclease G regulates budding yeast life and death. *Mol. Cell*, **25**, 233–246.

43. Burhans,W.C. and Weinberger,M. (2007) Yeast endonuclease G: complex matters of death, and of life. *Mol. Cell*, **25**, 323–325.

44. Meiss,G., Gimadutdinow,O., Haberland,B. and Pingoud,A. (2000) Mechanism of DNA cleavage by the DNA/RNA-non-specific *Anabaena* sp. PCC 7120 endonuclease NucA and its inhibition by NuiA. *J. Mol. Biol.*, **297**, 521–534.

45. Polling,S., Ormsby,A.R., Wood,R.J., Lee,K., Shoubridge,C., Hughes,J.N., Thomas,P.Q., Griffin,M.D., Hill,A.F., Bowden,Q. *et al.* (2015) Polyalanine expansions drive a shift into α-helical clusters without amyloid-fibril formation. *Nat. Struct. Mol. Biol.*, **22**, 1008–1015.

46. McWilliam,H., Li,W., Uludag,M., Squizzato,S., Park,Y.M., Buso,N., Cowley,A.P. and Lopez,R. (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.*, **41**, W597–W600.

47. UniProt,C. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.

48. Bono,F., Ebert,J., Unterholzner,L., Güttler,T., Izaurralde,E. and Conti,E. (2004) Molecular insights into the interaction of PYM with the Mago–Y14 core of the exon junction complex. *EMBO Rep.*, **5**, 304–310.

49. Le Hir,H., Sauliere,J. and Wang,Z. (2016) The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.*, **17**, 41–54.

50. Simonovic,M., Zhang,Z., Cianci,C.D., Steitz,T.A. and Morrow,J.S. (2006) Structure of the calmodulin αII-spectrin complex provides insight into the regulation of cell plasticity. *J. Biol. Chem.*, **281**, 34333–34340.

51. Jin,K.K., Krishna,S.S., Schwarzenbacher,R., McMullan,D., Abdubek,P., Agarwalla,S., Ambing,E., Axelrod,H., Canaves,J.M., Chiu,H.J. *et al.* (2006) Crystal structure of TM1367 from *Thermotoga maritima* at 1.90 Å resolution reveals an atypical member of the cyclophilin (peptidylprolyl isomerase) fold. *Proteins*, **63**, 1112–1118.

52. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

53. Hsia,K.-C., Li,C.-L. and Yuan,H.S. (2005) Structural and functional insight into sugar-nonspecific nucleases in host defense. *Curr. Opin. Struct. Biol.*, **15**, 126–134.

54. Lunin,V.Y., Levdikov,V.M., Shlyapnikov,S.V., Blagova,E.V., Lunin,V.V., Wilson,K.S. and Mikhailov,A.M. (1997) Three-dimensional structure of *Serratia marcescens* nuclease at 1.7 Å resolution and mechanism of its action. *FEBS Lett.*, **412**, 217–222.

55. Ghosh,M., Meiss,G., Pingoud,A., London,R.E. and Pedersen,L.C. (2005) Structural insights into the mechanism of nuclease A, a ββα metal nuclease from *Anabaena*. *J. Biol. Chem.*, **280**, 27990–27997.

56. Dupureur,C.M. (2008) Roles of metal ions in nucleases. *Curr. Opin. Chem. Biol.*, **12**, 250–255.

57. Llopis,J., McCaffery,J.M., Miyawaki,A., Farquhar,M.G. and Tsien,R.Y. (1998) Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 6803–6808.

58. Santo-Domingo,J. and Demaurex,N. (2012) The renaissance of mitochondrial pH. *J. Gen. Physiol.*, **139**, 415–423.

59. Gannavaram,S. and Debrabant,A. (2012) Programmed cell death in *Leishmania*: biochemical evidence and role in parasite infectivity. *Front. Cell Infect. Microbiol.*, **2**, 95.

60. Liu,Z., Chen,P., Wang,X., Cai,G., Niu,L., Teng,M. and Li,X. (2014) Crystal structure of DnaT84-153-dT10 ssDNA complex reveals a novel single-stranded DNA binding mode. *Nucleic Acids Res.*, **42**, 9470–9483.

61. Orans,J., McSweeney,E.A., Iyer,R.R., Hast,M.A., Hellinga,H.W., Modrich,P. and Beese,L.S. (2011) Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell*, **145**, 212–223.

62. Dickey,T.H., Altschuler,S.E. and Wuttke,D.S. (2013) Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure*, **21**, 1074–1084.

63. Ceska,T.A. and Sayers,J.R. (1998) Structure-specific DNA cleavage by 5′ nucleases. *Trends Biochem. Sci.*, **23**, 331–336.

64. Pazos,F., Helmer-Citterich,M., Ausiello,G. and Valencia,A. (1997) Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.*, **271**, 511–523.

65. Ballandras-Colas,A., Maskell,D.P., Serrao,E., Locke,J., Swuec,P., Jonsson,S.R., Kotecha,A., Cook,N.J., Pye,V.E., Taylor,I.A. *et al.* (2017) A supramolecular assembly mediates lentiviral DNA integration. *Science*, **355**, 93–95.

66. Parrish,J.Z., Yang,C., Shen,B. and Xue,D. (2003) CRN-1, a *Caenorhabditis elegans* FEN-1 homologue, cooperates with CPS-6/EndoG to promote apoptotic DNA degradation. *EMBO J.*, **22**, 3451–3460.

67. Ponce,I., Aldunate,C., Valenzuela,L., Sepulveda,S., Garrido,G., Kemmerling,U., Cabrera,G. and Galanti,N. (2016) A flap endonuclease (TcFEN1) is involved in *Trypanosoma cruzi* cell proliferation, DNA repair, and parasite survival. *J. Cell. Biochem.*, **118**, 1722–1732.

68. BoseDasgupta,S., Das,B.B., Sengupta,S., Ganguly,A., Roy,A., Dey,S., Tripathi,G., Dinda,B. and Majumder,H.K. (2008) The caspase-independent algorithm of programmed cell death in *Leishmania* induced by baicalein: the role of LdEndoG, LdFEN-1 and LdTatD as a DNA 'degradesome'. *Cell Death Differ.*, **15**, 1629–1640.

69. Balakrishnan,L. and Bambara,R.A. (2013) Flap endonuclease 1. *Annu. Rev. Biochem.*, **82**, 119–138.

70. Holm,L. and Laakso,L.M. (2016) Dali server update. *Nucleic Acids Res.*, **44**, W351–W355.