Fine tuning of the catalytic activity of colicin E7 nuclease domain by systematic N-terminal mutations

Eszter Németh,[1,2] Tamás Körtvélyesi,[2] Peter W. Thulstrup,[3] Hans E.M. Christensen,[4] Milan Kožíšek,[5] Kyosuke Nagata,[6] Anikó Czene,[7] and Béla Gyurcsik[1,7]*

[1]Department of Inorganic and Analytical Chemistry, University of Szeged, 6720 Szeged, Hungary

[2]Department of Physical Chemistry and Material Sciences, University of Szeged, 6720 Szeged, Hungary

[3]Department of Chemistry, University of Copenhagen, 2100 Copenhagen, Denmark

[4]Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[5]Institute of Organic Chemistry and Biochemistry, Gilead Sciences and IOCB Research Center Prague, Academy of Sciences of the Czech Republic, 166 10 Prague 6, Czech Republic

[6]Nagata Special Laboratory, Faculty of Medicine, University of Tsukuba, Tsukuba 305-8575, Japan

[7]MTA-SzTE Bioinorganic Chemistry Research Group of Hungarian Academy of Sciences, 6720 Szeged, Hungary

*Correspondence to: Béla Gyurcsik, Department of Inorganic and Analytical Chemistry, University of Szeged, Dóm tér 7, H-6720 Szeged, Hungary. Tel: +36-62544335. Fax: +36-625443430. E-mail: gyurcsik@chem.u-szeged.hu

Running title: **Fine tuning colicin E7 nuclease activity**

Total number of manuscript pages: 20, supplementary material pages: 8, tables: 2, and figures: 7

Electronic supplementary material for this manuscript is available as ESM.pdf.

Abstract: The nuclease domain of colicin E7 (NColE7) promotes the nonspecific cleavage of nucleic acids at its C-terminal HNH motif. Interestingly, the deletion of four N-terminal residues (446-449 NColE7 = KRNK) resulted in complete loss of the enzyme activity. R447A mutation was reported to decrease the nuclease activity, but a detailed analysis of the role of the highly positive and flexible N-terminus is still missing. Here we present the study of four mutants, with a decreased activity in the following order: NColE7 >> KGNK > KGNG ~ GGNK > GGNG. At the same time the folding, the metal-ion, and the DNA-binding affinity were unaffected by the mutations as revealed by linear and circular dichroism spectroscopy, isothermal calorimetric titrations and gel mobility shift experiments. Semiempirical quantum chemical calculations and molecular dynamics simulations revealed that K446, K449 and/or the N-terminal amino group are able to approach the active centre in the absence of the other positively charged residues. The results suggested a complex role of the N-terminus in the catalytic process that could be exploited in the design of a controlled nuclease.

Keywords: DNA cleavage; flow linear dichroism; ITC; positively charged N-terminal residues; Zn^{2+} binding

50-75-word statement, written for a broader audience:

NColE7, a nonspecific metallonuclease binds a Zn^{2+}-ion in its C-terminal catalytic centre. The quality, number and position of positively charged residues at the flexible N-terminus (446-KRNK-449) influence the catalytic activity. This feature could be exploited for allosteric control and modulation of the activity in an artificial nuclease. We provide a systematic study on the effect of N-terminal mutations on protein function along with the folding, metal ion and DNA binding.

Introduction

Colicin E7 (ColE7) is a nuclease toxin produced by *Escherichia coli* for protection from related bacteria.[1, 2] It has three functional domains: the receptor binding, the membrane translocation and the nuclease domain.[3, 4] The nuclease domain (NColE7) enters the cytoplasm of the target cell after the cleavage of ColE7 by a specific periplasmic protease recognizing the R447 residue.[5] NColE7 kills the attacked cell through the nonspecific digestion of the chromosomal DNA.[2] The host cell is protected against nuclease activity by coexpressing an immunity protein (Im7) that inhibits the substrate binding of NColE7.[6, 7]

The mechanism of DNA hydrolysis catalyzed by nuclease colicins was discussed extensively. The proposals largely depend on the quality of the metal ion.[6, 8-14] The catalytic centre is a ββα-metal binding HNH motif[15-17] at the C-terminus of these enzymes (Fig. 1). The metal cofactor in the active site of NColE7 is a Zn^{2+}-ion.[9] In contrast, Zn^{2+} was inhibitory for NColE9 - a related nuclease with high sequence identity.[11, 12] The presence of the metal ion is not necessary for substrate binding, but is essential for the hydrolytic reaction.[18] It binds and electrostatically activates the scissile phosphodiester bond. In NColE7 the Zn^{2+}-ion is coordinated by the side-chains of three histidine residues (H544, H569 and H573). A well conserved histidine (H545 in NColE7) is responsible for generating the nucleophilic hydroxide by deprotonating a water molecule (Fig. S1.).[9, 19, 20] The general acid residue that protonates the leaving group was not yet identified. In most nucleases it is a water molecule coordinated to the metal ion, which is usually Mg^{2+}.[21] However, such a coordinated water molecule wasn't detected in the distorted tetrahedral arrangement around the Zn^{2+}-ion in NColE7/DNA complex. R538, Q542 and H569 were speculated to provide the proton for the leaving group in NColE7[22] in analogy to NColE9.[23] Recently a shuttle mechanism was suggested in which the leaving group is protonated by the hydrogen ion originating from the same water molecule that initiated the nucleophilic attack.[24]

Fig. 1 near here

The N-terminal R447 and the C-terminal HNH motif are close in space in NColE7 [Fig. 1(A)]. This arginine was proposed to increase the DNA binding affinity.[5] It is intriguing why this single residue – lying outside the DNA-binding helices – influences the catalytic reaction. In NColE9 the arginine corresponding to R447 in NColE7 was supposed to bind the substrate and stabilize the pentavalent transition state.[25, 26] Based on the crystal structures of Vvn – an another HNH nuclease – it was also hypothesized that the arginine side-chain binds and stabilizes the cleaved DNA to decelerate the reverse reaction.[8, 27] R447A mutation in NColE7 significantly reduced its *in vitro* DNase activity,[5] while previously we showed that the cytotoxicity of NColE7 is completely lost upon deletion of the KRNK sequence (residues 446-449) at the N-terminus.[28] This suggests that the presence of a positively charged residue at the N-terminus is essential for the catalytic activity and the K446 and/or K449 residues may partly take over the role of the missing arginine. Positively charged residues are often found in a similar orientation close to the active site in the published crystal structures of HNH-nucleases.[28] This property could be generally exploited in the design and fine tuning of the catalytic activity of an artificial nuclease with intramolecular allosteric control.[29] Therefore, the purpose of this study was to investigate the effect of the mutations of the N-terminal positive residues on the catalytic activity. NColE7 and its four mutants (with KGNK, KGNG, GGNK and GGNG amino acids within the 446-449 segment of the sequence - see [Fig. 1(B)]) were expressed, purified and studied by experimental (mass spectrometry, isothermal calorimetric titration, agarose gel mobility shift assays, circular and linear dichroism spectroscopy) and computational (semiempirical quantum chemical calculations and molecular dynamics) techniques.

Results

Preparation and DNA cleavage of NColE7 and its mutants

The GST fusion forms of NColE7 and its mutants were co-expressed with the Im7 immunity protein to avoid cytotoxicity. The complexes were purified with GST affinity chromatography and the GST-tag was cleaved on column (Fig. S2.) The nucleases were separated from the immunity protein by adjusting the pH to 3.0.[7] At this pH the His side-chains are protonated and consequently, the purified proteins did not contain metal ion, as confirmed by their mass spectra. Upon addition of one equivalent zinc(II)-acetate, however, all nucleases were detected in their holo forms (Table I). The GPLGSPEF sequence encoded by pGEX-6P-1 vector remained at the N-terminus of the NColE7 variants after the Human rhinovirus C3 protease cleavage. This sequence was considered not to interfere with the properties investigated below.[28]

Table I. near here

The cleavage of the pUC19 plasmid by the NColE7 variant proteins in their Zn^{2+}-loaded form was monitored by agarose gel electrophoresis. The supercoiled form of the plasmid already disappeared at the first measurement point (approximately 2 minutes after mixing the solutions) in the presence of NColE7, while it was still detectable after 140 minutes in the DNA solutions containing the NColE7 mutants (Fig. 2.)

Fig. 2 near here

Flow linear dichroism (FLD) as a non-invasive technique was simultaneously applied for monitoring the digestion of long linear *calf thymus* DNA chains. The signal intensity of the FLD spectra depends on the ability of the long filiform DNA molecules to align, i.e. on the length, conformation and rigidity of the double helix.[30-32] Due to the cleavage by nucleases the DNA chains shortened and aquired a lower degree of orientation in the Couette flow cell. Thus, the absolute value of FLD intensity decreased during the measurement (Fig.

S3), while the absorbance was also constant during the reaction, indicating that no DNA was lost by precipitation (Fig. S4). The absolute value of FLD intensities at 260 nm plotted as a function of time in Fig. 3 allows for the comparison of the nuclease activity of the mutants under the same conditions provided.

Fig. 3 near here

Folding and Zn^{2+} binding of the proteins

The synchrotron radiation circular dichroism (SRCD) spectra of the various mutants are shown in Fig. 4(A). The shape of the spectra for all studied proteins carrying mutations in the region of KRNK amino acids is similar to the spectrum of NColE7. The interaction with metal ion [Fig. 4(B)] does not yield significant protein refolding. It is likely that the structure of the NColE7 variants is preformed for the metal ion binding, *i.e.* the purified proteins were properly folded and capable of strong Zn^{2+} binding. The change upon Zn^{2+} binding was similar for all mutants and to that of NColE9 published earlier.[33] Addition of excess Zn^{2+} did not affect the structure.

Fig. 4 near here

Quantitative data on the Zn^{2+} binding of mutants were obtained by isothermal microcalorimetric titrations (ITC). An exotherm heat effect was observed at the beginning of the titrations. This effect could not be avoided and no suitable model for analysis was found that could well describe the whole curve. Therefore, in each case the first part of the curve was ignored during the analysis (Fig. 5). Similar curves with even more intense changes at the beginning of the titration of the zinc transporter YiiP protein with $ZnCl_2$ [34] were observed, presumably due to the formation of oligomeric complexes. Although dimerization of the NColE7 variants may also occur,[7] further investigations are needed to understand this phenomenon.

Fig. 5 near here

DNA binding

Based on the earlier results, we expected that the N-terminal mutations will also significantly influence the DNA binding of the proteins. The presence of the metal ion in the active site of NColE7 is necessary for the catalytic activity, but it is not required for the DNA binding.[18] Accordingly, EDTA was applied to prevent the cleavage of the DNA in the experiments below. CT-DNA solution was titrated with the proteins (Fig. 6). Interaction with proteins lead to a decreased flow linear dichroism signal of DNA due to the change in the overall shape and thus, to the decrease of the orientation ability of the molecules. The cleavage of DNA was excluded based on the comparison with gel electrophoresis experiments. The UV absorption spectra confirmed that precipitation of the DNA-protein complexes did not significantly affect the data (Fig. S5). The order of the DNA binding based on the protein concentration required to decrease the FLD signal of DNA to its half intensity is as follows: NColE7 (1.0 μM) > KGNG (1.8 μM) ~ KGNK (2.0 μM) ~ GGNK (2.4 μM) > GGNG (3.7 μM). The FLD data, however, reflect the combined effect of the formation of the protein/DNA complexes and the subsequent conformation change of the DNA. Therefore, the apparent K_d values related to DNA binding were calculated from the agarose gel electrophoresis experiments carried out with a 13bp DNA fragment (Fig. S6). K_d values between 0.15 and 0.32 μM were obtained for the DNA binding affinities of NColE7 and its mutants against a 13 bp fragment.

Fig. 6 near here

Molecular dynamics

Molecular dynamics (MD) simulations were performed with the NColE7, KGNK, KGNG, GGNK and GGNG protein sequences starting from the residue 446, ignoring the 8

amino acids part from the GST cleavage in the experiments [Fig. 1(B)]. The models contained Zn^{2+}-bound proteins, with the metal ion in the HNH motif also coordinated by a phosphate ion. The root mean square deviation (RMSD) from the reference structure at 0 ps is higher for the mutants than for NColE7 (Fig. S7). The lysine containing mutants are similar to each other, but the increased RMSD for the GGNG mutant indicated that this structure changed to the largest extent. It was shown that in the KGNK and KGNG mutants the side-chain of K446, while in GGNK and GGNG the N-terminal amino group approached the phosphate ion (Figs. 7 (A) and S7). This supports the experimental observation that these positively charged groups might be able to partially take over the role of the arginine and can promote the catalytic activity. However, for GGNG only a few structures with such conformation were found, and due to the high flexibility of this chain, the N-terminus often turned out from the active site.

Fig. 7 near here

Semi-empirical calculations

In order to get more information about the fine structure of the active site, we optimized the geometry of the proteins by semiempirical quantum chemical calculations, handling the whole molecule on the PM6 level in MOPAC2009 (MOZYME method) with implicit water surroundings. The alignment to the optimized structure of NColE7 yielded an RMSD of 0.086 nm for KGNG, 0.087 nm for GGNK and 0.113 nm for GGNG. Small deviations from NColE7 occurred mainly at the mutated N-termini, but could also be detected in the loop between the two β-strands of the HNH motif. The phosphate ion mimicking the scissile phosphodiester bond was also displaced [Fig. 7 (B)]. This effect can be related to the decreased activity and supports the experimental data.

Discussion

The nuclease domain of ColE7 (NColE7) and its four N-terminal mutants [Fig. 1 (B)] were expressed and purified. Gel electrophoresis and FLD spectroscopic experiments demonstrated that the catalytic activity of the mutants significantly decreased compared to NColE7. The results revealed a substantial decrease of the enzymatic activity (Figs. 2 and 3) due to the R447G mutation, in agreement with the previous observations with R447E and R447A mutants.[5] According to LD results, the observed order of nuclease activity was NColE7 >> KGNK > GGNK ~ KGNG > GGNG (Fig. 3). This order implies that K446 and K449 can promote the reaction in the absence of R447 to a low extent. The catalytic activity was not completely abolished in any mutant. The activity of the GGNG mutant in contrast to the ΔN4-NColE7 protein[28] can be explained either by the positive N-terminal amino group at position G438 and/or the interactions of the backbone. The N-terminal part interacts directly with the DNA-binding helix, as well as with the loop of HNH motif [Fig. 1(A)]. The structure of this loop is essential for cleavage [35-37].

The reason for the dramatic change in enzymatic activity was systematically studied. As shown by CD-spectroscopy (Fig. 4), the mutations did not significantly affect the distribution of the secondary elements in the structure of the protein, suggesting that all four mutants maintained the native NColE7 folding. Similar result was obtained by the comparison of CD spectra of the WT ColE7 with those of the K446E, R447E and K446E/R447E mutants in the presence of Im7.[5]

According to the ITC results (Fig. 5) the K_d values are in nM range, similarly as it was estimated for the Zn^{2+} binding of NColE9.[38] The data collected in Table II demonstrate that all the mutants bound Zn^{2+}-ion as strongly as NColE7 with 1:1 stoichiometry. The binding was enthalpy-favored resulting in large ΔH values ranging from –40 to –30 kcal/mol. The

mutations had no obvious effect on Zn^{2+}-binding: K_d ~ 13 nM was obtained for all of the proteins within the experimental error suggesting that the difference in metal binding can not be the reason for the decreased activity of the mutants compared to NColE7.

Table II. near here

In an earlier publication the significantly increased K_M value of the R447A mutant compared to NColE7 was related to its decreased DNA binding affinity.[5] The K_d values range between ~ 0.15 μM and 0.32 μM for NColE7 and its variants (Table II), revealing that the changes at the N-terminus did not significantly affect the DNA binding affinity. This is in accordance with the fact that the strong DNA binding part of NColE7 is the central helical section.[9, 35, 36, 39] Only small differences were detected by FLD spectroscopy and agarose gel mobility shift assays in the DNA binding of the mutants and NColE7 (Fig. 6). These results imply the functional role of the N-terminal positively charged groups either by taking part in the molecular mechanism of the reaction, or influencing the conformation and electronic properties of the bound DNA molecule.

The computational results showed that there is a cooperation between the N- and C-termini of the protein. MD simulations pointed out that the N-terminal amino group and the lysine side-chains are capable to bind the scissile phosphate in the absence of R447 [Figs. 7(A) and S7]. According to semiempirical quantum chemical calculations the active site of the mutants is slightly distorted, and thus the scissile phosphate is displaced [Fig. 7(B)].

Both experimental and computational data revealed that R447 is the most efficient, but the K446 and K449 side-chains and/or the N-terminal amino group can also promote the catalytic reaction. The phenomenon resembles to an intramolecular allosteric process in that the N-terminal flexible chain has to approach the C-terminal active site during reaction. This knowledge may be applied to modulate the catalytic activity of an NColE7 based artificial nuclease.

Materials and methods

Cloning, protein expression and purification

The pQE70 plasmid containing the NColE7 and Im7 immunity protein genes was a generous gift from prof. K.-F. Chak, Institute of Biochemistry and Molecular Biology, National Yang Ming University, Taipei, Taiwan.[1, 2] From this template the primers applied in PCR (collected in Table S1.) amplified DNA segments including the genes of the native or mutated NColE7 as well as the Im7 protein. The obtained fragments were cloned into a pGEX-6P-1 vector (GE Healthcare BioSci.) providing an N-terminal glutathione-S-transferase (GST) affinity fusion tag. The plasmids were transformed into *E. coli* DH10B cells for DNA cloning and into *E. coli* BL21 (DE3) cells for protein production in 3×650 ml LB/Amp culture. At OD_{600} 0.6-0.7 the protein expression was induced with isopropyl β-*D*-1-thiogalactoside (IPTG, final concentration 0.1 mM), and the incubation was continued for further two hours at 25 °C. Cells were then sedimented by centrifugation and the pellets were resuspended in 50 ml PBS buffer (0.14 M NaCl, 2.7 mM KCl, 10.0 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) by sonication. The soluble fractions were loaded on a GST affinity chromatography column (GSTPrepFF16/10, GE Healthcare BioSci.). The fusion proteins were cleaved on column with Human rhinovirus C3 protease[40] - sold as PreScission protease by GE Healthcare - to remove the GST tag. 20 ml of 10 μM protease in PBS was loaded on the column and the reaction was continued overnight at 4 °C or 2 hours at room temperature followed by the elution with PBS. The fractions containing the nuclease mutants in complex with the immunity protein as well as the protease were collected. To disrupt the interaction between the NColE7 mutants and the Im7 protein the pH was adjusted to 3.0 after a 3× dilution with a 20 mM Gly/HCl buffer. The components were then separated on a Sepharose

SP FF 16/10 cation exchange column with a binding buffer 20 mM Gly/HCl pH = 3.0 and a gradient of 0-2 M NaCl in 30×column volume (CV). The immunity protein was eluted at pH = 8.0 with PBS conatining 2M NaCl. The fractions of the nucleases or Im7 were concentrated by Amicon ultrafilter with 5 kDa cutoff, and the buffer was exchanged to 20 mM HEPES, pH = 7.7. The sequences of the purified mutant proteins in comparison with NColE7 are shown in Fig. 1(B).

Nano-electrospray ionization mass spectrometry (nano-ESI-MS)

Mass spectra were obtained on a LCT Premier (Waters) instrument equipped with a Nanoflow Electrospray Ionization (nano-ESI) source and a time-of-flight (TOF) analyzer. The instrument was operated in positive ion mode and it was calibrated using 100 mg/ml CsI in 50 % 2-propanol in the *m/z* range from 600 to 12000. Samples were sprayed from middle size Au/Pd-coated borosilicate glass capillary needles (Proxeon) loaded with 3 µl protein solution. The protein concentration was between 10 – 20 µM in 500 mM ammonium acetate (Sigma) buffer. For the study of metal binding, 1 equivalent Zn^{2+}-acetate was added to the protein samples before measurement. The de-salting of the protein solution and buffer exchange to the volatile buffer was done using Micro BioSpin chromatography column (BioRad). The needle voltage was typically around 1200 V and 50 V cone voltage was applied, with a cone gas maintained at 20 L/h and the source temperature was maintained at 50 °C. The recorded *m/z* data were deconvoluted using the MassLynx^{TM} v4.1 (Waters) software equipped with the MaxEnt1 algorithm. The high charge states of the multiply charged spectrum, ranging from +10 to +17, were used to calculate the apparent mass.

Isothermal calorimetry (ITC)

Isothermal calorimetric titrations were performed on a MicroCal Auto ITC-200 (GE) instrument. The protein samples (~ 50 μM) were prepared by 12 hour dialysis in 7000 MWCO Thermo Scientific Slide-A-Lyzer casettes, against 20 mM cacodylate buffer, pH = 7.0. $ZnCl_2$ was dissolved in the same buffer. The ionization enthalpy of cacodylate – applied as buffer – is close to zero (–0.47 kcal/mol[41]), so the contribution of protonation/deprotonation processes, if there were any present, was negligible in the observed ΔH. The dilution heat of $ZnCl_2$ with the buffer was determined for each experiment and the integrated data of dilution heats were subtracted from the corresponding data of protein titrations. A control titration of the KGNK mutant with plain buffer was done showing no significant effects (data not shown). The enthalpy change during the titrations of 200 μl protein solutions with 2 μl aliquots of 400 μM $ZnCl_2$ up to 40 μl (spacing of 240 s) can be consequently attributed to the metal binding or competition processes. Instead of degassing the samples before titration the plates were shortly centrifuged.

Gel mobility shift assays

In the protein-DNA binding studies the concentration of the 13 base pair oligonucleotide was 0.2 μM, and the protein final concentration ranged between 0 and 7 μM. The ^{31}P-radiolabelled DNA was hybridized from one single oligonucleotide with complementary sequence at each end forming a loop. The samples were run on 6% native PAGE at 4 °C . The reaction mixture contained 4 mM NaCl, 4 mM HEPES buffer (pH=7.9) and 100 μM EDTA to inhibit DNA digestion. For DNA cleavage studies 450 ng/well 2686 base pairs pUC19 plasmid DNA was applied, and the protein concentration was 2.8 μM. Zn^{2+}-ions were added to proteins prior the reaction in 1:1 molar ratio. 10 μl of the reaction was loaded onto an ethidium bromide containing 1% agarose gel. The electrophoresis was performed in a Bio-Rad Wide Mini Sub Cell® GT system at 6.7 V/cm in TAE buffer (40 mM

Tris, 20 mM acetic acid, and 1 mM EDTA, pH = 8.0). For comparison 6 µl of the Bio-Rad 1 kbp marker DNA was also loaded to the gel.

Circular dichroism spectroscopy

The Synchrotron Radiation Circular Dichroism (SRCD) spectra were recorded at the SRCD facility at the CD1 beamline at the Institute for Storage Ring Facilities (ISA), University of Aarhus, Denmark.[42, 43] All spectra were recorded with 1 nm steps and a dwell time of 2 s per step, in 100.4 µm quartz cells (SUPRASIL, Hellma GmbH, Germany). The concentration of the protein solutions was 3.2×10^{-5} M in 10 mM HEPES, pH = 7.7.

Linear dichroism spectroscopy

Flow linear dichroism spectra were measured on a Jasco-815 CD spectrophotometer equipped for linear dichroism spectroscopy (LD), using a microvolume Couette flow device as described in.[44] An additional quartz lens was mounted to allow for focusing onto the sample in the Couette cell, which was positioned as close to the photomultiplier as possible in the J-815 sample compartment. In the Couette cell, an outer quartz cylinder rotates, and an inner quartz rod is static, the annular gap is 0.25 mm giving a combined light path of 0.5 mm. Water circulation through the metal block of the flow device thermostated the cell to 298.0 K. The sample volume was 70 µl and 3000 rpm rotation was applied. The optical bandwidth was 1 nm and the spectra were recorded in continuous mode between 190 and 400 nm with 50 nm/min scanning speed, 1 s integration time, 0.5 nm data pitch, and with 15 L/min nitrogen flow. LD of double stranded DNA samples yields a characteristic negative signal at the absorbance maximum of the DNA bases (*ca.* 260 nm) since the transition moments of the base π- π^* transitions are located in the plane of the bases, and as these are oriented orthogonally with respect to the axis of DNA helix.[45]

Protein-DNA binding was studied with a 130 μM (final concentration calculated for base pairs) *calf thymus* CT-DNA sample. The mixture contained 60 μM EDTA, 17 mM KH_2PO_4, 2.4 mM HEPES (the pH of both buffers was adjusted to 7.7) and 0-5 μM protein. The solutions were incubated for 10 min before recording the spectra. The incubation time did not influence the results.

The cleavage of 130 μM (bp) CT-DNA by different mutants was followed for 8 hours (1 spectrum/h). The proteins (40 μM) were incubated with one equivalent of $Zn(Ac)_2$ (40 μM) for 30 min at room temperature before the reaction started. The final concentration of both the protein and Zn^{2+}-ions was 0.5 μM. 720 μl of reaction mixture was incubated at 37 °C and in each hour 70 μl aliquot was taken for LD test. All solutions were in mixed buffers of 17 mM KH_2PO_4 and 2.4 mM HEPES, the pH of both buffers was adjusted to 7.7. The spectra were smoothed with the means-movement method, convolution width of 11.

Computational methods

Initial conformation of NColE7 and the shortened mutants was obtained from a crystal structure 1M08.[7] The original N-terminus of NColE7 was restored by an M446K mutation. In all models the proteins had uncapped termini (i.e. NH_3^+ and COO^- groups).

Molecular dynamics (MD) calculations were performed with GROMACS 4.05,[46] with the force field Gromos 53a6.[47] The ionizable residues were charged according to the default pK_a values at pH = 7.2 detected by PropKa 3.0.[48] Each protein was placed in a cubic box with edge size of *ca.* 8 nm, and solvated by explicit SPC/E water model containing about 16000 equilibrated water molecules. The system was neutralized with Cl^- ions replacing water molecules according to the electrostatic potential of the system points. Energy minimization was carried out with the steepest descent method. 200 ps position restrained dynamics was performed in NVT ensemble to equilibrate the system (solvate and generate initial velocities

with Maxwell distribution) including explicit water molecules. 25 ns MD simulations were performed in the NPT ensemble with periodic box conditions. The integration step was 2 fs. The temperature was set to 300 K and isotropic Berendsen p-coupling and T-coupling was applied. For Coulomb interactions PME was applied with 0.9 nm cut-off for electrostatic and 1.6 nm for van der Waals interactions. The dielectric constant was set to 1.0. The LINCS constraint algorithm was used. Trajectories were analyzed starting at 500 ps.

Semi-empirical quantum chemical computations were performed on the proteins including a Zn^{2+} and a phosphate ion with the PM6 method implemented in MOPAC2009.[49-52] Localized molecular orbitals were applied by the MOZYME[52] model. The solvation was considered by COSMO method[53] with the dielectric constant of 78.4. The geometry optimization was carried out by the L-BFGS method after the initial minimization of the hydrogen positions. The gradient norm was set to 1.0 kcal/mol/Å. The thermodynamic parameters were computed with the PM6 method, then recalculated in one SCF cycle with the PM6-DH2 correction.

Electronic Supplementary Information is available for this manuscript: Table S1. Sequences of the primers applied for the amplification of the genes of NColE7 variants. **Figure S1.** Proposed mechanism of DNA-cleavage by NColE7. **Figure S2.** SDS-PAGE monitoring of protein purification. CT-DNA cleavage monitored by FLD **Figure S3.** and UV-Vis spectroscopy **Figure S4.** The cleavage of 130 μM (base pairs) CT-DNA by different nucleases (0.5 μM) as followed by flow linear dichroism spectroscopy. **Figure S5. A)** Flow linear dichroism calibration curve of CT-DNA in the 55-130 μM concentration range. **B)** Absorption changes during the titration of CT-DNA with NColE7 and its mutants. **Figure S6.** Titration of a radiolabelled 13 bp DNA with NColE7 and GGNG. **Figure S7.** Results of MD simulations: **A)** RMSD (root mean square deviation) of the studied proteins, as compared to the starting structure of the simulation. **B)** Distance of the positively charged residues from

the metal ion in the active centre in the MD simulation of the zinc-bound proteins containing a phosphate ion in the active center.

Acknowledgement

This work has received support through the Hungarian Science Foundation (OTKA-NKTH CK80850), European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.2/B-10/1-2010-0012, TÁMOP 4.2.4.A/2-11-1-2012-0001 'National Excellence Program', TÁMOP-4.2.2.C-11/1/KONV 2012-010, TÁMOP-4.2.2.A-11/1/KONV-2012-0047. The computational resources at HPC of University Szeged and financial support from the European Community's Seventh Framework Programme (FP7/2007-2013) CALIPSO under grant agreement nº 312284 is also gratefully acknowledged. B.G. thanks JSPS, while E.N. to Danish Ministry of Science, Innovation and Higher Education for the fellowship provided. MK was partially funded by EU OPPC program CZ.2.16/3.1.00/24016. All authors disclose any potential sources of conflict of interest.

References

1. Chak K, Kuo W, Lu F, James R (1991) Cloning and Characterization of the Cole7 Plasmid. J Gen Microbiol 137:91-100.

2. Lin Y, Liao C, Liang P, Yuan H, Chak K (2004) Involvement of colicin in the limited protection of the colicin producing cells against bacteriophage. Biochem Biophys Res Commun 318:81-87.

3. Liao C, Hsiao K, Liu Y, Leng P, Yuen HS, Chak K (2001) Processing of DNase Domain during Translocation of Colicin E7 across the Membrane of Escherichia coli. Biochem Biophys Res Commun 284:556-562.

4. Cheng YS, Shi Z, Doudeva LG, Yang WZ, Chak KF, Yuan HS (2006) High-resolution crystal structure of a truncated ColE7 translocation domain: implications for colicin transport across membranes. J Mol Biol 356:22-31.

5. Shi Z, Chak K, Yuan H (2005) Identification of an essential cleavage site in ColE7 required for import and killing of cells. J Biol Chem 280:24663-24668.

6. Sui M, Tsai L, Hsia K, Doudeva L, Ku W, Han G, Yuan H (2002) Metal ions and phosphate binding in the H-N-H motif: Crystal structures of the nuclease domain of ColE7/Im7 in complex with a phosphate ion and different divalent metal ions. Protein Sci 11:2947-2957.

7. Cheng Y, Hsia K, Doudeva LG, Chak K, Yuan HS (2002) The Crystal Structure of the Nuclease Domain of Colicin E7 Suggests a Mechanism for Binding to Double-stranded DNA by the H–N–H Endonucleases. J Mol Biol 324:227-236.

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

9. Doudeva L, Huang D, Hsia K, Shi Z, Li C, Shen Y, Cheng Y, Yuan H (2006) Crystal structural analysis and metal-dependent stability and activity studies of the ColE7 endonuclease domain in complex with DNA/Zn2+ or inhibitor/Ni2+. Protein Sci 15:269-280.

10. Pommer AJ, Cal S, Keeble AH, Walker D, Evans SJ, Kühlmann UC, Cooper A, Connolly BA, Hemmings AM, Moore GR, et al (2001) Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. J Mol Biol 314:735-749.

11. Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K, Riley M, Slatin S, Cavard D (2007) Colicin biology. Microbiol Mol Biol Rev 71:158-229.

12. Papadakos G, Wojdyla JA, Kleanthous C (2012) Nuclease colicins and their immunity proteins. Q Rev Biophys 45:57-103.

13. Galburt EA, Stoddard BL (2002) Catalytic mechanisms of restriction and homing endonucleases. Biochemistry 41:13851-13860.

14. Stoddard BL (2005) Homing endonuclease structure and function. Q Rev Biophys 38:49-95.

15. Mehta P, Katta K, Krishnaswamy S (2004) HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily. Protein Sci 13:295-300.

16. Orlowski J, Bujnicki JM (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. Nucleic Acids Res 36:3552-3569.

17. Veluchamy A, Mary S, Acharya V, Mehta P, Deva T, Krishnaswamy S (2009) HNHDb: a database on pattern based classification of HNH domains reveals functional relevance of sequence patterns and domain associations. Bioinformation 4:80-83.

18. Ku W, Liu Y, Hsu Y, Liao C, Liang P, Yuan H, Chak K (2002) The zinc ion in the HNH motif of the endonuclease domain of colicin E7 is not required for DNA binding but is essential for DNA hydrolysis. Nucleic Acids Res 30:1670-1678.

19. Eastberg JH, Eklund J, Monnat R, Jr., Stoddard BL (2007) Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. Biochemistry 46:7215-7225.

20. Midon M, Gimadutdinow O, Meiss G, Friedhoff P, Pingoud A (2012) Chemical rescue of active site mutants of S. pneumoniae surface endonuclease EndA and other nucleases of the HNH family by imidazole. Chembiochem 13:713-721.

21. Dupureur CM (2008) Roles of metal ions in nucleases. Curr Opin Chem Biol 12:250-255.

22. Ko T, Liao C, Ku W, Chak K, Yuan HS (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. Structure 7:91-102.

23. Garinot-Schneider C, Pommer AJ, Moore GR, Kleanthous C, James R (1996) Identification of Putative Active-site Residues in the DNase Domain of Colicin E9 by Random Mutagenesis. J Mol Biol 260:731-742.

24. Bueren-Calabuig JA, Coderch C, Rico E, Jimenez-Ruiz A, Gago F (2011) Mechanistic insight into the catalytic activity of betabetaalpha-metallonucleases from computer simulations: Vibrio vulnificus periplasmic nuclease as a test case. Chembiochem 12:2615-2622.

25. Walker DC, Georgiou T, Pommer AJ, Walker D, Moore GR, Kleanthous C, James R (2002) Mutagenic scan of the H-N-H motif of colicin E9: implications for the mechanistic enzymology of colicins, homing enzymes and apoptotic endonucleases. Nucleic Acids Res 30:3225-3234.

26. Maté MJ, Kleanthous C (2004) Structure-based Analysis of the Metal-dependent Mechanism of H-N-H Endonucleases. J Biol Chem 279:34763-34769.

27. Li CL, Hor LI, Chang ZF, Tsai LC, Yang WZ, Yuan HS (2003) DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site. EMBO J 22:4014-4025.

28. Czene A, Németh E, Zóka IG, Jakab-Simon NI, Körtvélyesi T, Nagata K, Christensen HEM, Gyurcsik B (2013) The role of the N-terminal loop in the function of the colicin E7 nuclease domain. J Biol Inorg Chem 18:309-321.

29. Gyurcsik B, Czene A (2011) Towards artificial metallonucleases for gene therapy: recent advances and new perspectives. Future Med Chem 3:1935-1966.

30. Matsuoka Y, Nielsen PE, Norden BJ (1984) On the structure of active chromatin. A flow linear dichroism study of chromatin fractionated by nuclease digestion. FEBS Lett 169:309-312.

31. Gololobov GV, Chernova EA, Schourov DV, Smirnov IV, Kudelina IA, Gabibov AG (1995) Cleavage of supercoiled plasmid DNA by autoantibody Fab fragment: application of the flow linear dichroism technique. Proc Natl Acad Sci U S A 92:254-257.

32. Hicks MR, Rodger A, Thomas CM, Batt SM, Dafforn TR (2006) Restriction enzyme kinetics monitored by UV linear dichroism. Biochemistry 45:8912-8917.

33. van den Bremer ET, Jiskoot W, James R, Moore GR, Kleanthous C, Heck AJ, Maier CS (2002) Probing metal ion binding and conformational properties of the colicin E9 endonuclease by electrospray ionization time-of-flight mass spectrometry. Protein Sci 11:1738-1752.

34. Chao Y, Fu D (2004) Thermodynamic studies of the mechanism of metal binding to the Escherichia coli zinc transporter YiiP. J Biol Chem 279:17173-17180.

35. Hsia K, Chak K, Liang P, Cheng Y, Ku W, Yuan HS (2004) DNA Binding and Degradation by the HNH Protein ColE7. Structure 12:205-214.

36. Wang Y, Yang W, Li C, Doudeva LG, Yuan HS (2007) Structural basis for sequence-dependent DNA cleavage by nonspecific endonucleases. Nucleic Acids Res 35:584-594.

37. Huang H, Yuan HS (2007) The Conserved Asparagine in the HNH Motif Serves an Important Structural Role in Metal Finger Endonucleases. J Mol Biol 368:812-821.

38. Pommer AJ, Kuhlmann UC, Cooper A, Hemmings AM, Moore GR, James R, Kleanthous C (1999) Homing in on the role of transition metals in the HNH motif of colicin endonucleases. J Biol Chem 274:27153-27160.

39. Levin KB, Dym O, Albeck S, Magdassi S, Keeble AH, Kleanthous C, Tawfik DS (2009) Following evolutionary paths to protein-protein interactions with high affinity and selectivity. Nat Struct Mol Biol 16:1049-U67.

40. Walker P, Leong L, NG P, Tan S, Waller S, Murphy D, Porter A (1994) Efficient and Rapid Affinity Purification of Proteins using Recombinant Fusion Proteases. Bio-Technology 12:601-605.

41. Fukada H, Takahashi K (1998) Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. Proteins 33:159-166.

42. Miles AJ, Hoffmann SV, Tao Y, Janes RW, Wallace BA (2007) Synchrotron Radiation Circular Dichroism (SRCD) spectroscopy: New beamlines and new applications in biology. J Spectroscopy 21:245-255.

43. Miles AJ, Janes RW, Brown A, Clarke DT, Sutherland JC, Tao Y, Wallace BA, Hoffmann SV (2008) Light flux density threshold at which protein denaturation is induced by synchrotron radiation circular dichroism beamlines. J Synchrotron Radiat 15:420-422.

44. Marrington R, Dafforn T, Halsall D, Rodger A (2004) Micro-volume Couette flow sample orientation for absorbance and fluorescence linear dichroism. Biophys J 87:2002-2012.

45. Norden B, Kubista M, Kurucsev T (1992) Linear Dichroism Spectroscopy of Nucleic-Acids. Q Rev Biophys 25:51-170.

46. Hess B, Kutzner C, van dS, Lindahl E (2008) GROMACS 4:â€‰ Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4:435-447.

47. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656-1676.

48. Olsson MHM, SÃ¸ndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J Chem Theory Comput 7:525-537.

49. Stewart JJP (2009) Application of the PM6 method to modeling proteins. J Mol Model 15:765-805.

50. Stewart JJP (2008) MOPAC2009, Colorado Springs, Stewart Computational Chemistry, http://openmopac.net.

51. Stewart JJP (2007) Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. J Mol Model 13:1173-1213.

52. Stewart JJP (1996) Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. Int J Quantum Chem 58:133-146.

53. Klamt A, Schüürmann G (1993) Cosmo - a New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. J Chem Soc Perkin Trans 2:799-805.

Figure legends

Figure 1. (A) Crystal structure of NColE7 (PDB: 1MZ8[6]) in complex with a Zn^{2+} and a phosphate ion. The HNH motif is in orange and the N-terminal loop is in blue. Among the N-terminal amino acids, R447 is the closest residue to the phosphate ion that forms a bridge between the Zn^{2+}-ion and R447. Hydrogen bonds of the backbone of the N-terminus are also shown. (B) Sequences of the purified NColE7 variants. The proteins containing the black part of the sequence denoted by NColE7 will be referred to as wild type (WT) NColE7 in the following text. The four mutants are named KGNK, KGNG, GGNK and GGNG based on the 446-449 segments of the sequence, respectively. The changes compared to the NColE7 sequence are highlighted in red. The remaining sequence after the Human rhinovirus C3 protease cleavage encoded by pGEX-6P-1 plasmid is written in blue.

Figure 2. Digestion of 28 nM pUC19 (that is 74 μM calculated for base pairs) by 2.8 μM NColE7 mutants, incubated with one equivalent zinc(II)-acetate before mixing with DNA. The samples were kept at 37 °C and run subsequently on 1% agarose gel. In control experiments (data not shown) cleavage of the plasmid DNA incubated with only zinc(II)-acetate was not observed. 1 kb Molecular Ruler (BioRad) served as the reference.

Figure 3. Cleavage of 130 μM (base pairs) CT-DNA by different nucleases (0.5 μM) followed by flow linear dichroism spectroscopy. The proteins were pre-incubated for 30 min in the presence of one equivalent $ZnCl_2$. The LD signal intensity at 260 nm is taken from baseline corrected spectra of aliquots of the stock solutions, incubated at 37 °C. Control

experiments included CT-DNA incubated at 37 °C without or in the presence of 0.5 μM $ZnCl_2$, 0.5 μM NColE7 and 60 μM EDTA or 0.5 μM NColE7 and 1 μM Im7.

Figure 4. (A) SRCD spectra of the NColE7 variant proteins. (B) SRCD spectra of WT NColE7 in the absence (two independent measurements) and upon the addition of Zn^{2+}-ions. All spectra were normalized to the same 31.6 μM protein concentration.

Figure 5. Microcalorimetric titration of 50 μM KGNG in 20 mM cacodylate buffer, pH = 7.0. The red points at the beginning were ignored during the curve evaluation.

Figure 6. The effect of the CT-DNA binding of NColE7 and the four mutants on the flow linear dichroism signal intensity at 260 nm. 130 μM CT-DNA (calculated for base pairs) was incubated with 0-5 μM protein and 60 μM EDTA in each case.

Figure 7. (A) Snapshots at 20 ns of the simulations of the mutant proteins: KGNK in red, KGNG in blue, GGNK in cyan and GGNG in green. The Zn^{2+}-ions are shown by spheres and the phosphate ion by sticks. (B) Optimized structure of the mutants and the WT NColE7. NColE7 is in grey, KGNG in blue, GGNK in cyan, and GGNG in green. Molecules were aligned by PyMOL.

Table I. Comparison of the expected and measured molecular masses of the purified proteins. The numbers refer to the holo form of the WT NColE7 and its mutants containing one Zn^{2+}-ion. The spectra were recorded after short incubation of the proteins with one equivalent of zinc(II)-acetate.

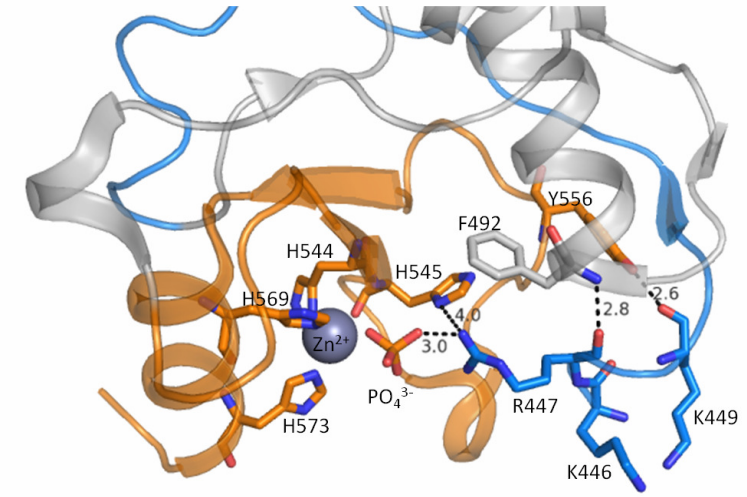
Protein	Expected mass Da	Measured mass Da
NColE7	15875.2	15874.9
KGNK	15776.1	15776.2
KGNG	15705.0	15704.5
GGNK	15705.0	15704.8
GGNG	15633.9	15632.8
Im7	10961.1	10961.5 / 10412.8[a]

[a] The immunity protein was found in two forms corresponding to the full length sequence and with four histidines missing from the His-tag at the C-terminus of the protein.

Table II. Zn^{2+} and DNA binding affinities of NColE7 variants. Micro ITC titrations of 50 μM proteins with $ZnCl_2$ stock solution were performed in 20 mM cacodylate buffer at pH = 7.0. DNA binding of the 13 bp DNA was monitored by agarose gel electrophoresis. For the analysis a model with one Zn^{2+} and one DNA binding site was applied. K_d is the apparent dissociation constant under the given condition. The fitting errors are also provided.

Protein	Zn^{2+} binding		DNA binding
	Stoichiometry	K_d nM	K_d μM
NColE7	1.2	16.6 ± 3.5	0.16 ± 0.05
KGNK	0.9	13.2 ± 1.5	0.15 ± 0.05
KGNG	1.0	13.4 ± 1.4	0.32 ± 0.07
GGNK	1.2	11.9 ± 2.6	0.30 ± 0.06
GGNG	0.9	12.8 ± 1.9	0.30 ± 0.06

A



B

```
NColE7    438-GPLGSPEFKRNKPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
KGNK      438-GPLGSPEFKGNKPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
KGNG      438-GPLGSPEFKGNGPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
GGNK      438-GPLGSPEFGGNKPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
GGNG      438-GPLGSPEFGGNGPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL

NColE7    SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGK-576
KGNK      SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGK-576
KGNG      SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGK-576
GGNK      SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGK-576
GGNG      SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGK-576
```

Figure 1.

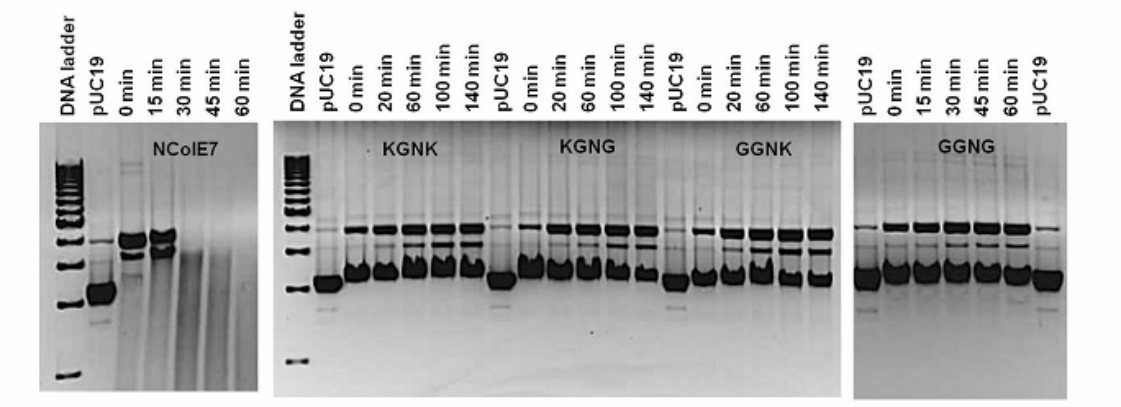

DNA ladder
pUC19
0 min
15 min
30 min
45 min
60 min
NColE7
DNA ladder
pUC19
0 min
20 min
60 min
100 min
140 min
pUC19
0 min
20 min
60 min
100 min
140 min
pUC19
0 min
20 min
60 min
100 min
140 min
KGNK
KGNG
GGNK
pUC19
0 min
15 min
30 min
45 min
60 min
pUC19
GGNG

Figure 2.

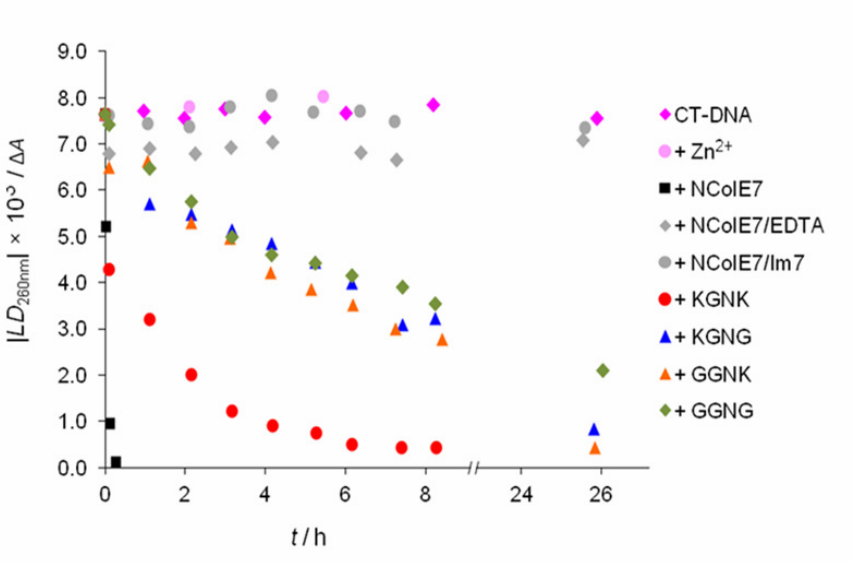
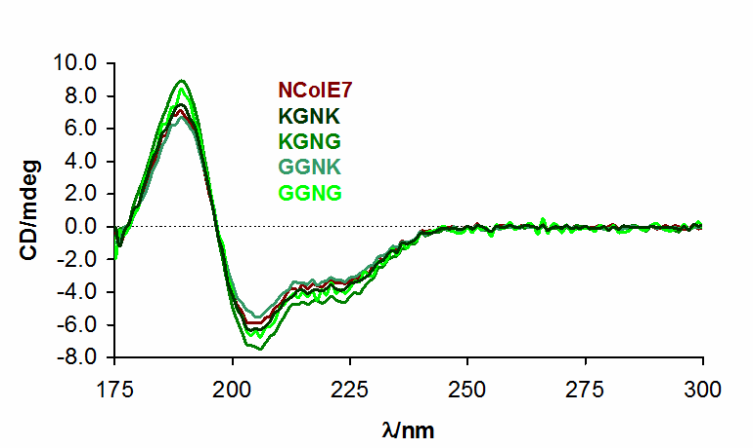

9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
|LD260nm| × 10³ / ΔA
0
2
4
6
8
24
26
t / h
CT-DNA
+ Zn2+
+ NCoIE7
+ NCoIE7/EDTA
+ NCoIE7/Im7
+ KGNK
+ KGNG
+ GGNK
+ GGNG

Figure 3.

A


NColE7
KGNK
KGNG
GGNK
GGNG
CD/mdeg
λ/nm
10.0
8.0
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0
-8.0
175
200
225
250
275
300

B

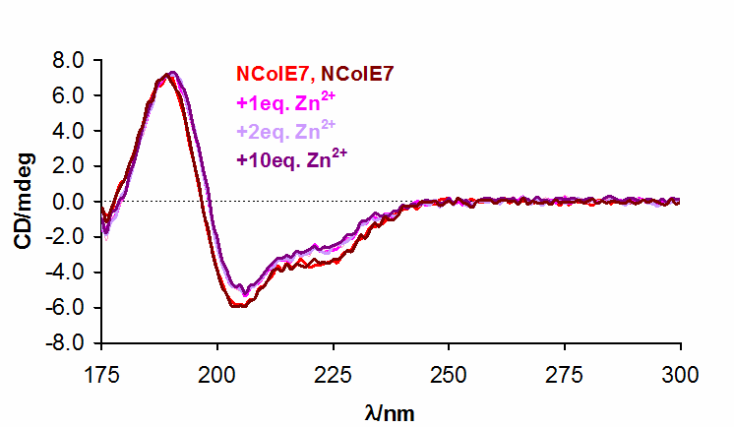

NColE7, NColE7
+1eq. Zn2+
+2eq. Zn2+
+10eq. Zn2+
CD/mdeg
λ/nm
8.0
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0
-8.0
175
200
225
250
275
300

Figure 4.

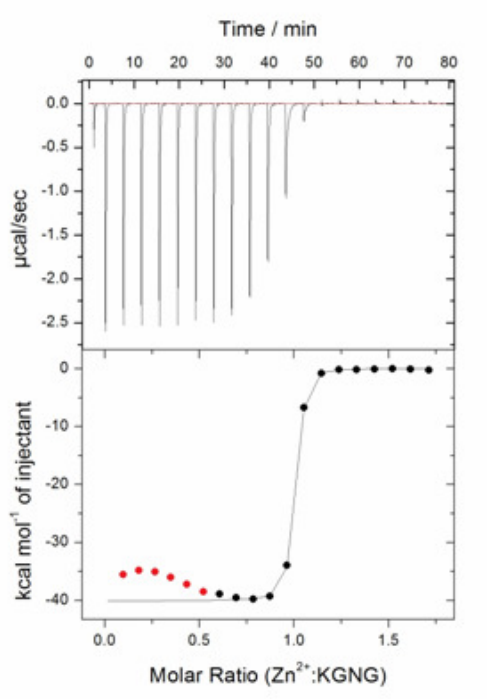

Time / min
0
10
20
30
40
50
60
70
80
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
µcal/sec
0
-10
-20
-30
-40
kcal mol-1 of injectant
0.0
0.5
1.0
1.5
Molar Ratio (Zn2+:KGNG)

Figure 5.

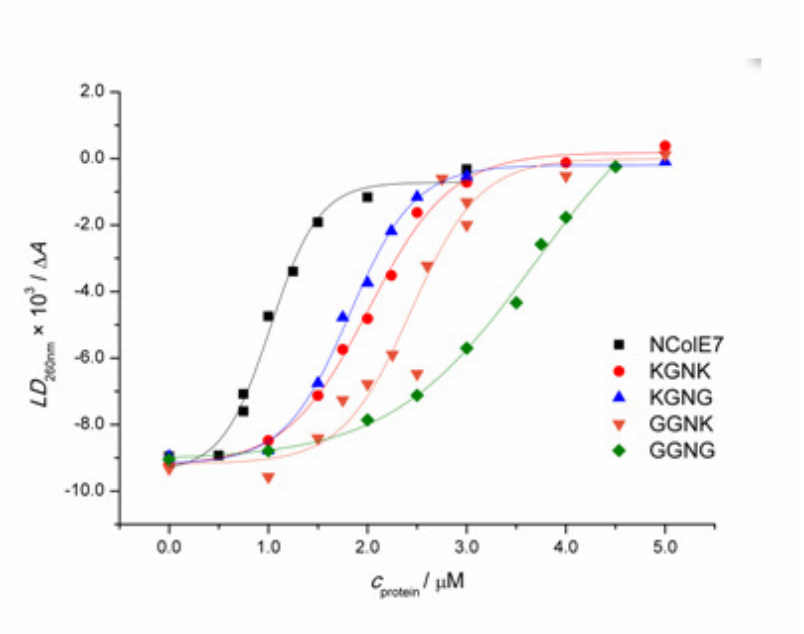
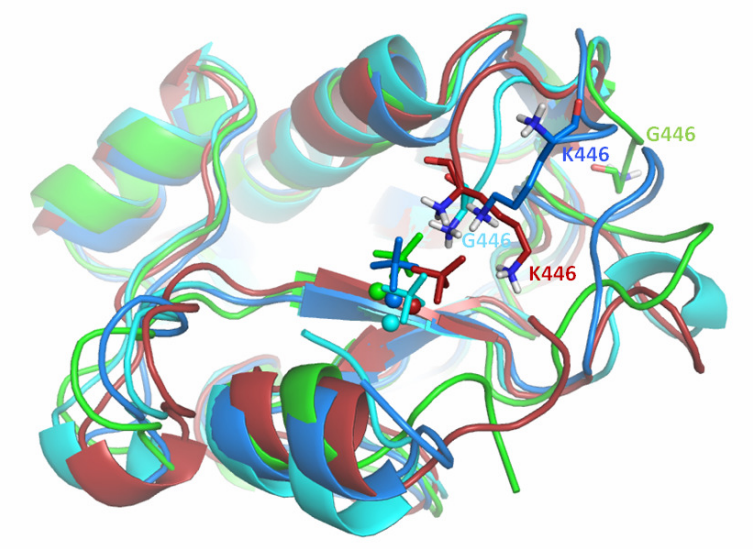

2.0
0.0
-2.0
-4.0
-6.0
-8.0
-10.0
0.0
1.0
2.0
3.0
4.0
5.0
$LD_{260nm} \times 10^3 / \Delta A$
$c_{protein}$ / μM
NColE7
KGNK
KGNG
GGNK
GGNG

Figure 6.

A



B

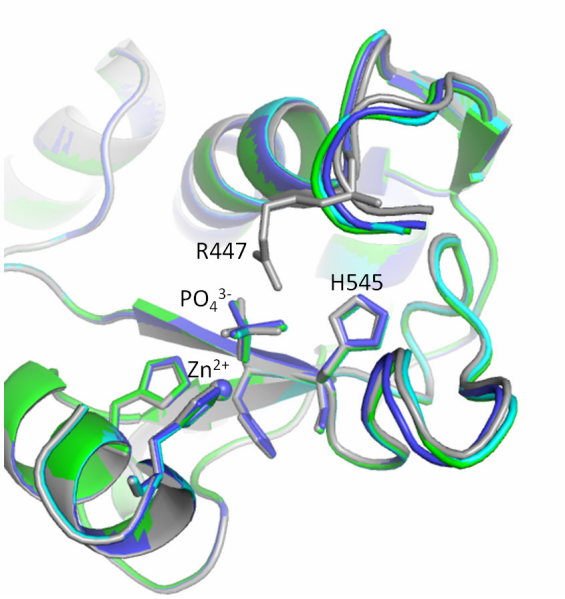


Figure 7.