

On the role of the N-terminal loop in the function of Colicin E7 nuclease domain

Anikó Czene,^{[a],[b]} Eszter Németh,^{[a],[c]} István G. Zóka,^[a] Noémi I. Jakab-Simon,^{[a],[d]} Tamás Körtvélyesi,^[c] Kyosuke Nagata,^[e] Hans E.M. Christensen,^[d] and Béla Gyurcsik*^{[a],[b]}

^a*Department of Inorganic and Analytical Chemistry, University of Szeged, Dóm tér 7, H-6720 Szeged, Hungary*

^b*Bioinorganic Chemistry Research Group of Hungarian Academy of Sciences, Dóm tér 7, H-6720 Szeged, Hungary*

^c*Department of Physical Chemistry and Material Sciences, University of Szeged, Aradi Vértanúk tere 1, H-6720 Szeged, Hungary*

^d*Department of Chemistry, Technical University of Denmark, Kemitorvet, Building 207, 2800 Kgs. Lyngby, Denmark*

^e*Department of Infection Biology, Graduate School of Comprehensive Human Sciences and Faculty of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Japan*

*Corresponding author:

Béla Gyurcsik: Tel: +36-62-544-335, Fax: +36-62-544-340,

E-mail: gyurcsik@chem.u-szeged.hu

Abstract Colicin E7 (ColE7) is a metallonuclease toxin of *E. coli* belonging to the HNH superfamily of nucleases. It contains highly conserved amino acids in its HHX₁₄NX₈HX₃H ββα-type metal ion-binding C-terminal active centre. However, the proximity of the arginine at the N-terminus of the nuclease domain (NCole7, 446-576) is necessary for the hydrolytic activity. This poses a possibility of an allosteric activation control in this protein. To get more information on this phenomenon two protein mutants were expressed, i.e. the N-terminal 4 and 25 amino acids from the nuclease domain of ColE7 were removed. The effect of the N-terminal truncation on the Zn²⁺-ion and DNA binding, as well as, on the activity was investigated in this study by mass spectrometry, SRCD and fluorescence spectroscopy, and agarose gel mobility shift assays. The dynamics of protein backbone movement was simulated by molecular dynamics. Semiempirical quantum chemical calculations were performed to get better insight into the structure of the active centre. The longer protein interacted both with Zn²⁺-ion and DNA stronger than its shorter counterpart. The results were explained by the structural stabilization effect of the N-terminal amino acids on the catalytic centre. In agreement with this, the absence of the N-terminal sequences resulted in significantly increased movement of the backbone atoms: in ΔN25-NCole7 the amino acid strings between residues 485-487, 511-515 and 570-571, while in ΔN4-NCole7 those between 467-468, 530-535 and 570-571 compared to that of the native NCole7.

Keywords Metallonuclease, Colicin E7, N-terminally truncated mutants, zinc(II)-binding

Introduction

Colicin E7 (ColE7) is a metallonuclease toxin of *E. coli* [1]. Its role is to protect the host cell from other related bacteria and bacteriophages [2] by degradation of their chromosomal DNA during environmental stress. In order to exert cell killing activity, ColE7 has to get across both the outer and the inner cell membrane, facilitated by the receptor-binding and translocation domains [3, 4]. The host cell itself is protected by the simultaneously expressed immunity protein Im7 blocking the DNA-binding site [5, 6] of the NColE7 domain due to tight interactions based on charge-complementarity [7–12].

ColE7 belongs to the HNH superfamily of nucleases [13–15] possessing a 30-40 amino acids long $\beta\beta\alpha$ -type metal ion-binding motif in their active centre. The His and Asn amino acids are highly conserved within the sequence $\text{HHX}_{14}\text{NX}_8\text{HX}_3\text{H}$ corresponding to this motif at the C-terminal region of bacterial colicins and pyocins [16, 17]. At the same time, the HNH motif is found in various regions of a wide range of enzymes including group I homing endonucleases (e.g. I-Hmu-I [18]), procaryotic extracellular nucleases (Nuclease A [19]) and also in increasing number of restriction endonucleases (e.g. MnlI [20], KpnI [21, 22], HphI [23], Eco31I [24], Hpy99I [25]). Sequences are collected in the HNH [26] and Pfam databases [27]. The first His (H) out of the name-giving HNH amino acids acts as the general base in DNA hydrolysis. The Asn (N) residue plays structural role constraining the HNH loop by extensive hydrogen-bonding interactions [14, 28]. The third conserved residue out of the HNH string is a metal-binding His (H). The HNH-motif of ColE7 binds to the 3' site of the scissile phosphate in the minor groove of the DNA while the other parts of the nuclease domain provide strong, nonspecific binding within the major groove [16, 29] similarly to Colicin E9 [30, 31]. As such, NColE7 catalyses the nonspecific hydrolysis of nucleic acids.

There is still a debate about the role of the different divalent metal ions in colicin nucleases [30–38]. In NColE7 three His amino acid side-chains bind a metal cofactor, which is most probably Zn^{2+} -ion under physiological conditions [32], but the apo protein could be reactivated to a different extent by other divalent metal ions like Mn^{2+} , Ni^{2+} , Co^{2+} , Cu^{2+} , Mg^{2+} , Ca^{2+} and Sr^{2+} [5, 39]. The metal ion having a free coordination site has an essential multiple roles in DNA-cleavage: it binds to the scissile phosphodiester, polarizes the P-O bond for nucleophilic attack, stabilizes the phosphoanion transition state and the leaving group. As mentioned above, the attacking nucleophilic OH^- is supposed to be generated by the most conserved His residue of the HNH motif - not coordinating to the Zn^{2+} -ion. The hydrolytic reaction is also facilitated by the 19 degree bending of the DNA due to the protein-

binding [32]. The Zn^{2+} -ion is not required for DNA-binding, but it is essential for DNA-hydrolysis [39].

In a recent paper [40] it was demonstrated that during the membrane translocation process the periplasmic extracts cleave ColE7 between K446 and R447 and only the nuclease domain (R447-K576) enters the cell. The R447E ColE7 mutant lost its cell killing activity due to failed inner membrane translocation, but K446E and N448A mutants retained it. However, it was shown in an *in vitro* assay that the R447E mutant of NColE7 (444-576) has only ~15% endonuclease activity of that of the wild-type NColE7. This difference was assumed to be the consequence of lower affinity towards DNA and not of the decrease in catalytic activity. Based on the crystal structure of VVn endonuclease with DNA it was proposed that the role of such a spatially close arginine residue might also be the stabilization of the enzyme – product complex [41].

The necessity of the N-terminal amino acids in NColE7 for the function of the C-terminal catalytic centre poses a possibility of an allosteric activation within the enzyme that would be a desired property for use in an artificial nuclease [42]. The N-terminal end of NColE7 forms a loop leaning near to the active centre, and the interactions between them might be decisive in control of the function. In this paper two N-terminally truncated derivatives of NColE7 (446-576): GST- Δ N25-NColE7 and GST- Δ N4-NColE7-C* (instead of the GST- Δ N4-NColE7 protein we studied its C-terminal mutant GST- Δ N4-NColE7-C* selected by bacterial cells; the sequences are defined later in Figure 1a) were expressed in *E. coli*. The proteins with and without the GST tag were purified for the studies of DNA and Zn^{2+} -binding activities. Gel mobility shift assay, synchrotron-radiation circular dichroism (SRCD), fluorescence spectroscopic and mass spectrometric experiments were carried out and complemented by bioinformatics, molecular dynamics and semiempirical quantum chemical calculations. The results will lead us to better understanding of the role of the N-terminal loop in the catalyzed reaction, as well as, its structural effects.

Materials and methods

Cloning, protein expression and purification

The genes of the mutant proteins were amplified by PCR from the pQE70 plasmid (a generous gift of Prof. K.-F. Chak, Institute of Biochemistry and Molecular Biology, National Yang Ming University, Taipei, Taiwan) by using the oligonucleotides Δ N4-NCoIE7-F: 5'-ggaattcccaggaaggcaacaggta-3' and Δ N25-NCoIE7-F: 5'-ggaattcgacttaggttctctgttcca-3' as forward and NCoIE7-R: 5'-gccgctcgagctatttacctcggtgaatatcaatatgc-3' as the reverse primer and inserted into the pGEX-6-P1 (GE Healthcare) vector between the EcoRI and XhoI restriction enzyme sites (underlined sequences). The insert DNA sequences contained a C-terminal stop codon (formatted to italic in the primer sequence). The plasmids encoding the mutant proteins with a glutathione S-transferase (GST) affinity tag at the N-terminus were cloned in *E. coli* DH10B or Mach1 (Invitrogen) cells and then transformed into *E. coli* BL21 (DE3), spread on LB/Amp plates and colonies were grown overnight at 37 °C. A small scale (4 ml) LB/Amp medium (Luria-Bertani medium supplemented with 100µg/ml ampicillin) was inoculated with a single colony and incubated overnight at 37 °C with shaking at 300 rpm. For large scale protein production, the small scale overnight cultures were transferred in 250 ml LB/Amp medium and bacteria were grown at 37 °C. The protein expression was induced by adding isopropyl- β -thiogalactoside (IPTG, 200 mg/ml) to the final concentration of 0.42 mM to the cultures at OD₆₀₀ 0.5-0.6. The shaking at 140 rpm was continued for 2-3 hours, until OD₆₀₀ increased up to 0.9-1.0. Cells were sedimented by centrifugation at 4 °C, 5000×g for 10 minutes, then resuspended in PBS buffer (1.4 M NaCl, 27 mM KCl, 100 mM Na₂HPO₄, 18 mM KH₂PO₄, pH 7.3). Pellets were disrupted by sonication and the debris was removed by centrifugation at 5000×g, 4 °C, 10 minutes. The supernatant, as well as the resuspended aggregates were analyzed by 12.5 % sodium dodecyl sulphate - polyacrylamide gel electrophoresis (SDS-PAGE). A significant amount of the desired protein was present in inclusion bodies. The GST based affinity purification step was only done with the soluble fractions. The protein solutions were loaded onto a 4 ml Glutathione Sepharose 4B affinity column (Amersham Bioscience). The column was washed with ice-cold PBS buffer to remove unbound material and then the bound fusion proteins were eluted with 15 mM reduced glutathione dissolved in PBS buffer, containing 0.1 % Triton X-100 non-ionic detergent. Following an SDS-PAGE analysis, the fractions containing the target protein were

pooled. Before the cleavage of the GST fusion tag, the excess of reduced glutathione was removed by dialysis against PBS buffer, applying 150-times dilution. For digestion 1 μl (2 units/ μl) of PreScission Protease (GE Healthcare) for each 100 μg of fusion protein was added. The solution was incubated at 5 °C for 4 hours. Following cleavage, batch purification was applied to remove the GST moiety and the PreScission Protease (siliconized tubes were used to prevent the resin to stick to the wall of the tubes): the Glutathione Sepharose 4B resin was incubated with the protein solution for 30 minutes at 4 °C with gentle rotation. The medium was sedimented by centrifuging at 500 \times g for 5 min and the supernatant was carefully transferred to new tubes. The efficiency of cleavage was checked by 17.5 % SDS-PAGE. The concentration was estimated from the gel, comparing the intensity of the protein band with the intensity of the standard low range (14-97 kDa) protein marker (BioRad) applied in known concentrations (i.e. 5 μl , 10 μl , 15 μl and 20 μl from a 100 ng/ μl stock solution). The large scale protein expression and purification procedure has been detailed elsewhere [43].

Electrospray ionization mass spectrometry (nano-ESI-MS)

Mass spectrometric (MS) measurements 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 10 – 20 μM in 100 mM ammonium acetate (Sigma) buffer. 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. A stock solution of 100 μM Zn(acetate)₂ (Sigma) was used to titrate the 20 μM $\Delta\text{N}25\text{-NCoIE7}$ protein solution. The recorded m/z data were deconvoluted using the MassLynxTM 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.

Gel mobility shift assay

A ~ 400 bp (bp) dsDNA was used as substrate for gel mobility shift assays. 50-100 ng of this DNA was used for each assay. The protein:dsDNA molar ratio varied between 0.5:1 and 50:1, the protein:Zn²⁺ molar ratio was 1:1, 50 mM Tris/HCl buffer, pH 7.4 or PBS buffer was used, and NaCl was adjusted to 50 mM final concentration. The solutions also contained 5 % glycerol. The reaction mixtures were incubated for 20-40 minutes at room temperature and then loaded onto 1% agarose gel and run at 30-50 V in TBE buffer (44.5 mM Tris, 1 mM EDTA, 44.5 mM boric acid, pH 7.3). Finally the gel was stained with ethidium-bromide solution with a concentration of 0.5 µg/ml for 30 minutes, followed by washing for several times with distilled water. The documentation was done using UV-light (Bioinstrument ATTO or UviDoc – Uvitec, Cambridge – gel documentation systems) at 312 nm.

SRCD spectroscopic measurements

The synchrotron radiation CD (SRCD) spectra of the proteins were recorded at the SRCD facility at the CD1 beamline [44] on the storage ring ASTRID at the Institute for Storage Ring Facilities (ISA), University of Aarhus, Denmark. The instrument was calibrated against camphor-sulfonic acid. All spectra were recorded with 1 nm steps and a dwell time of 3 s/step, using 10 µm quartz cells (SUPRASIL, Hellma GmbH, Germany), for the wavelength range of 175-350 nm. The proteins were dissolved in distilled water and the pH was adjusted by HCl and NaOH solutions. The protein concentrations ranged from 10-50 µM. From raw spectra the water baseline was subtracted.

Fluorimetry

The most popular fluorescent probes for Zn²⁺-ions in living organism contain quinoline-units, such as 6-Methoxy-(8-p-toluenesulfonamido)quinoline (TSQ) and its derivatives. These coordinate to the Zn²⁺-ion by four nitrogen donor atoms in the *bis*complex as illustrated in Fig. S1, but the protonation state of the complex changes with pH [45]. The fluorophore used in this study is an acid derivative of the above mentioned TSQ compound: TFLZn test (Sigma-Aldrich). It is soluble in water and shows high Zn²⁺-ion selectivity compared to other cations. Its affinity towards Zn²⁺-ions is high enough to bind the free biological metal ion ($K_D \approx 20$ µM), but is not high enough to extract Zn²⁺ coordinated in proteins. TFLZn shows little fluorescence in the absence of Zn²⁺. Upon formation of the *bis*complex, a 100-fold increase in

intensity is experienced. Maxima of the excitation and emission spectra are at 360 nm and 498 nm, respectively.

The measurements were carried out on a Hitachi F-4500 fluorescence spectrometer in a 1x1 cm pathlength quartz cell. The solutions were irradiated in a wavelength range between 340 and 420 nm, while the emission spectrum was recorded between 400 and 600 nm. TFLZn concentration was adjusted to 5 μM in all cases, mutant proteins were added in a concentration range between 0.37 and 4.8 μM . Zn^{2+} concentration ranged from 1.2 to 27.5 μM , EDTA concentration ranged from 1.2 to 30 μM and the DNA concentration (a 10 bp DNA string was considered to be a binding unit) ranged from 0.5 to 1.5 μM in the titration experiments.

Calculations

Initial conformation of NCoIE7 and the shortened mutants was taken from the structure of 1M08 [6]. This structure has a methionine (M446) instead of K446 at the N-terminus. The wild-type (WT) NCoIE7 calculations were made from the M446K mutant of the original pdb structure. All proteins had unprotected termini (ie. NH_3^+ and COO^- groups).

Molecular dynamics (MD) calculations were performed with GROMACS 4.05 [46,47], Gromos 53a6 [48] force field. The ionizable residues were charged according to the default $\text{p}K_a$ values at $\text{pH} = 7.2$, as no reason was found to alter the default $\text{p}K_a$ values by propKa 3.0 [49]. The protein was placed in a cubic box edge size of ca. 8 nm, and solvated by explicit SPC/E water model. About 16000 equilibrated water molecules were needed to fill the box that was neutralised due to positive charge of the protein by Cl^- ions replacing water molecules at the most positive parts of the box using GENION. Energy minimization was carried out using steepest descent method. 200 ps position restrained dynamics were performed in NVT ensemble to equilibrate the system (solvate and generate initial velocities with Maxwell distribution) including explicit water molecules. 20 ns productive MD simulations were performed in the NPT ensemble with periodic box conditions with the integration steps of 2 fs. The temperature was set to 300K and isotropic Berendsen p-coupling and T-coupling was used. For Coulomb interactions PME was applied with a 0.9 nm electrostatic and 1.6 nm van der Waals interactions cut-off, the dielectric constant was set to 1.0. The constraint algorithm LINCS was used. Trajectories were analyzed starting at 500 ps.

Semiempirical quantum chemical computations were performed by using MOPAC2009 [50] with PM6 method [51, 52]. Localized molecular orbitals were applied by

the MOZYME [53] model implemented in MOPAC2009. The solvation was considered by COSMO method [54] with the dielectric constant of 78.4. The geometry optimization was carried out by the quasi-Newton L-BFGS method after the initial minimization of the hydrogen positions. A gradient norm of ca. 2-3 kcal/mol was achieved while the heat of formation became essentially stationary.

Results and discussion

Cytotoxic effect of the mutant proteins

To study the necessity of the N-terminal end of the NCoIE7 protein for its activity first the genes of the truncated mutants were constructed and inserted into a pGEX-6P-1 vector. The Δ N4-NCoIE7 and Δ N25-NCoIE7 proteins were designed to delete the 4 and 25 N-terminal amino acids from the original NCoIE7 sequence (starting with K446 according to the CoIE7 protein original numbering). It is known that the native CoIE7 gene is toxic for the cells, due to the unwanted minor expression level of the protein during the cloning process [55]. Thus, the success of the cloning and expression of a CoIE7 mutant indicates that the given protein is not toxic for the cells. According to the PCR followed by agarose gel electrophoresis the genes were successfully inserted into the vector and cloned in either DH10B or Mach1 cells. The transformed BL21 (DE3) cells based on the change in OD₆₀₀ after the induction of the protein expression were grown in a similar manner as the cells expressing GST itself, unlike to those of the toxic variant of NCoIE7, where the cells started to die shortly after induction (Fig. 1b). The SDS-PAGE of the expressed proteins showed intense bands around the expected molecular mass (Fig. 1c). However, the DNA sequencing and the mass spectra of the proteins showed that while the GST- Δ N25-NCoIE7 sequence was correct, instead of the GST- Δ N4-NCoIE7 we had expressed a new mutant GST- Δ N4-NCoIE7-C* with a modified C-terminus – but containing all the amino acids necessary for the Zn²⁺-ion binding. This strongly suggests that the GST- Δ N4-NCoIE7 was cytotoxic. The sequences of the proteins after the GST cleavage are depicted in Fig. 1a.

Fig. 1

To check the effect of the C-terminal modification in GST- Δ N4-NCoIE7-C* we have also shown that the Δ N4-NCoIE7 mutant expressed from pET21a plasmid without any N-terminal tag but having the correct sequence at the C-terminus is non-toxic for the cells [43]. These results together proved that the N-terminal basic amino acids are necessary for the cell killing activity of the enzyme if it is overexpressed in bacterial cells. Looking at the available crystal structures of the NCoIE7 (Table 1) these amino acid residues with a special emphasis on R447 residue were observed in only few of them. In those few including this amino acid,

however, the R447 side-chain is situated close to the Zn^{2+} -ion in the active centre [5, 6, 10]. The two positively charged residues are bridged by a phosphate ion (Fig. 1d), which is most probably replaced by the scissile phosphodiester group of the DNA in the catalytically active complex [16]. In the NCoIE7-DNA crystals the R447 is mostly missing from the solved structure [16, 29, 32], but it is close to the phosphate backbone in the structure of a metal ion deficient mutant [57].

Table 1.

In view of the possible allosteric control, it is important to know, what is the function of the N-terminal residues with positively charged side-chains (one arginine and two lysines) and the role of the whole N-terminal chain - a component without autonomous secondary structure - regarding the activity of the NCoIE7 protein. Further in this manuscript we try to get closer to the solution of this problem by means of the investigation of the Zn^{2+} and DNA-binding abilities of the expressed mutant proteins.

Protein – Zn^{2+} interaction

Mass spectrometric investigations

Intact protein mass spectrometry was used to identify the truncated proteins and to further investigate their Zn^{2+} -ion binding abilities. Fig. 2a shows the mass spectrum of the purified $\Delta N4$ -NCoIE7-C* mutant recorded in the volatile ammonium acetate buffer without addition of Zn^{2+} -ions. The apparent mass of the main peak, i.e. 16251.9 Da, corresponded to the mass of the holo protein. This clearly demonstrates that the $\Delta N4$ -NCoIE7-C* mutant was purified in its Zn^{2+} -bound form. The multiply charged spectrum showed at almost all charge states the presence of a significant amount of acetato-complex as a result of a non-covalent interaction. Its amount increased with the decrease in the protein's charge state (see Fig. S2 for the m/z spectrum in ESI). Since the metal binding site consists of three histidine imidazole nitrogen ligands from the HNH motif, the presence of the acetate ligand, completing the tetrahedral coordination around Zn^{2+} -ion is expected here instead of the phosphate ion, occurring usually in the crystal structures.

Fig. 2

In contrast to $\Delta N4\text{-NCoIE7-C}^*$, Fig. 2b shows that the purified $\Delta N25\text{-NCoIE7}$ mutant did not contain Zn^{2+} -ions, and that it was not able to complete the metallation of the apo form even in the presence of 20-fold molar excess of Zn^{2+} -ions at pH 6.7 (pH of the ammonium acetate buffer used for MS measurements). The apparent mass of the apo protein (13124.8 Da) fits very well with the calculated theoretical mass. As a result of increasing amounts of Zn^{2+} -ions the molar ratio of the holo protein increased. In the presence of 10-fold metal ion excess the molar ratio of the apo and holo proteins is approximately 1:1, while 20-fold excess of Zn^{2+} -ions is required to achieve approximately 85% metallation. The K_D value calculated from the ratio of the ion signal intensities of the apo and holo proteins in the m/z spectra of the $11\times$ charged ion (see the series of the spectra in Fig. S3) was $74\pm 18\ \mu\text{M}$, assuming that no dissociation occurs during the transmission through the mass spectrometer and the metal ion binding to the protein does not alter the ionization efficiency of the non-covalent complex [59]. It shall, however, be noted that the above estimated constant would largely depend on the protonation state of the protein molecule. This means that the stability of the metal ion complex is lower than that published for the Zn^{2+} -binding of NCoIE9 nuclease (nanomolar K_D) and is similar to Ni^{2+} -binding of the same protein [33]. These data unambiguously indicate that the 21 N-terminal amino acids of the $\Delta N4\text{-NCoIE7-C}^*$ mutant play an important role in the metal binding in the HNH motif at the C-terminus of the protein. The amino acids of the N-terminal loop may e.g. affect the dynamics of the protein folding and promote the formation of the proper structure of the protein.

Fluorimetry

Fluorimetry can also be applied to monitor the Zn^{2+} -ion binding of proteins by probes being fluorescent in their zinc(II)-complexes [45], such as the TFLZn probe used by us. The maximal fluorescence intensity at 490 nm were monitored (Fig. S4). In the solutions containing the TFLZn probe and the truncated NCoIE7 proteins in 2:1 molar ratio different behaviour was observed for $\Delta N25\text{-NCoIE7}$ and $\Delta N4\text{-NCoIE7-C}^*$. There was no significant change in the fluorescence intensity of TFLZn in the presence of $\Delta N25\text{-NCoIE7}$, while the addition of an equivalent ($c(\text{Zn}^{2+}) = c(\text{P})$) of Zn^{2+} -ions caused a large increase in the intensity. At the same time increase of the fluorescence was observed upon addition of $\Delta N4\text{-NCoIE7-C}^*$ to the TFLZn solution. The resulting fluorescence intensity was in both cases significantly higher in the presence of the proteins (and metal ion), than in the $\text{Zn}^{2+} - \text{TFLZn}$ binary

system. This suggests that in agreement with the MS result $\Delta N25$ -NCoIE7 does not contain Zn^{2+} -ions, and that the proteins can not completely replace the TFLZn probe from the coordination sphere of the Zn^{2+} -ion. The latter can be explained, supposing that the proteins do not fill all the coordination sites around the metal ion (coordination occurs through the three histidine side-chains). Therefore, according to the thermodynamics of the system the formation of $Zn^{2+}(\text{protein})(\text{TFLZn})$ ternary complexes is also possible in which an enhancement of the fluorescence can be observed. The addition of DNA to protein containing solutions slightly decreased the fluorescence, probably due to the replacement of the dye from the ternary complex. The slightly larger extent of the change for $\Delta N4$ -NCoIE7- C^* protein points to its stronger DNA binding (see later).

SRCD spectroscopic results

In a chiral environment there is a difference between the absorption of the left and right circularly polarized light, and the plot of the difference in their absorption coefficients ($\Delta\epsilon = \epsilon_{\text{left}} - \epsilon_{\text{right}}$) vs. wavelength yields a characteristic circular dichroism (CD) spectrum of the sample. The relative position of chiral amide chromophores in proteins *i.e.* the secondary structure and its changes are responsible for this effect in the wavelength region of the UV light (180-250 nm). Since the synchrotron radiation circular dichroism (SRCD) spectroscopy provides an optimal and even flux of the UV light in a highly controlled manner it can be applied for accurate study of the solution structure and interactions of proteins [60].

The effect of metal ion binding on the structure of the mutant NCoIE7 proteins was investigated by monitoring the changes in their SRCD spectra on addition of Zn^{2+} ions and/or EDTA to their solutions, as described in the experimental part. Fig. 3a shows the spectra obtained for $\Delta N4$ -NCoIE7- C^* protein. As it can be seen, the addition of Zn^{2+} -ions did not affect the SRCD spectrum (not even at 5-fold Zn^{2+} excess – data not shown). This again suggests that the $\Delta N4$ -NCoIE7- C^* protein already includes a bonded metal ion. At the same time, an excess of EDTA caused a slight decrease in the intensity. This suggests that the removal of the Zn^{2+} ions by EDTA from the protein caused only a negligible change in the secondary structure composition of the protein suggesting that the structure is also stable without the metal ion.

Fig. 3

In similar experiments with $\Delta N25$ -NCoIE7 both the intensity and the shape of the SRCD spectra changed continuously (the spectral pattern becoming more similar to that of the $\Delta N4$ -NCoIE7-C* spectrum) upon gradual addition of up to 10 equivalents of Zn^{2+} -ions, while the extent of this change became negligible at higher metal ion excess. The addition of an excess of EDTA resulted in a similar spectrum to that recorded in the absence of metal ions (Fig. 3b). These results further show that the shorter protein binds Zn^{2+} -ions more weakly than $\Delta N4$ -NCoIE7-C*, which could result from the more extensive distortion of the metal ion binding site upon the deletion of the further 21 amino acids.

Protein – DNA interactions

SRCD spectroscopic results

SRCD was also applied to study the double-strand DNA (dsDNA) binding of the mutant proteins. The spectra recorded in the presence of dsDNA are presented in Fig. S5. Although the gel mobility shift experiments (see below) proved the DNA binding, the recorded spectra showed that the addition of DNA did not change the structure of $\Delta N4$ -NCoIE7-C* in solution. This is in line with the crystal structures of the NCoIE7 and NCoIE9 bound to dsDNA [16, 29, 31, 32, 57]. There was, however, a clear difference between the calculated and experimental spectra upon addition of DNA to $\Delta N25$ -NCoIE7 suggesting that the dsDNA binding induces a slight conformational change in the more flexible mutant, probably by the stabilization of the wild-type structure.

Gel mobility shift assays

To compare the DNA-binding ability of the truncated mutants a gel mobility shift experiment was conducted, in which increasing amounts of proteins were added to a 0.874 μM solution of a ~ 400 bp dsDNA sample (Fig. 4). For $\Delta N25$ -NCoIE7 approximately 10 times larger protein concentration than for $\Delta N4$ -NCoIE7-C* was applied to achieve substantial gel mobility shift, in agreement with its weaker DNA binding.

Fig. 4

Addition of Zn^{2+} -ions to the $\Delta N25$ -NCoIE7 solution containing 20-fold excess of protein (see Fig. 4a) did not cause change in the position of the band. Two reasons may account for this: (1) the protein already bound metal ions, however, this would be in contrast with our previous results, or (2) the binding of Zn^{2+} -ions is not necessary for DNA-binding – similarly as for NCoIE7. In agreement with this latter observation an excess of EDTA added to $\Delta N4$ -NCoIE7- C^* – DNA system did not cause any change in the position of the DNA band (Fig. 4a).

For $\Delta N4$ -NCoIE7- C^* an apparent stability constant was estimated based on the gel mobility shift assay. Simplifying conditions were introduced assuming 1:1 DNA binding site (10 bp DNA) - protein complex (P-DNA) formation and 100% complex formation at the saturation of the curve. In Fig. 4b the relative gel mobility shift vs. equilibrium protein concentration ($[P]$) was plotted. The latter was estimated as $[P] = c_P - [P-DNA]$, where c_P is the total concentration of the protein, and $[P-DNA]$ is the equilibrium concentration of the protein-DNA complex, which is proportional with the relative distance of the shifted band from the unbound DNA on the gel. $K_D = ([P] \times [DNA]) / [P-DNA]$, at the inflection point, where 50% of the DNA binding sites are occupied by the protein, *i.e.* $[DNA] = [P-DNA]$, and thus $[P] = K_D$, where K_D is the apparent dissociation constant related to the formation of protein-DNA complexes at each binding site. By the above considerations K_D was estimated to be $\sim 5.0 \mu M$ ($pK_D \sim 5.3$) for a $\Delta N4$ -NCoIE7- C^* -DNA binding site complex.

Molecular Dynamics Calculations

Structural changes of NCoIE7 and the mutant proteins (it shall be noted that in the calculations the native sequences were applied without any tags) in explicit SPC/E water were tracked by 22 ns molecular dynamics (MD) calculations. Fig. 5a describes the change in RMSD (root mean square deviation of backbone atoms in the molecule with respect to the reference structure at 500th ps) during the simulation. According to the RMSD diagram of NCoIE7 (446-576) simulation a relatively stable structure is formed after 2.5 ns of solvation, causing 0.2 nm difference, as compared to the start of the simulation. The structure slightly changes until 10 ns, and then it is fluctuating around 0.2 nm.

The structure of the $\Delta N4$ -NCoIE7 (450-576) mutant behaves similarly to the wild type NCoIE7, but the $\Delta N25$ -NCoIE7 (471-576) mutant goes through more serious changes reaching an RMSD of 0.35 nm by the end of simulation. The RMSD for both mutants

increased with time during the whole simulation in contrast to that of WT-NCole7. This means that shortening the protein caused remarkable changes in protein dynamics even if only 4 residues were cut at the N-terminus.

Fig. 5

Fig. 5b shows the average motion of each backbone atoms (RMSF, root mean square fluctuation) during the whole simulation. HNH-motif is at the C-terminal part of the protein, and such at the right side of the diagram. The intense peak at about the 550th residue corresponds to the loop that joins the two β -sheets of the HNH-motif. It is a functionally important part of the protein: the conserved residue N560 is located here, which is responsible for orienting the general base H545.

Significant difference can be observed between NCole7 and the shortened mutants. In two regions of the protein Δ N25-NCole7, *i.e.* the amino acid residues 485-487 and 511-515 show an increased motion. These parts of NCole7 are loops leaning approximately parallel to the original N-terminal part that is missing from Δ N25-NCole7. The α -helices in the neighborhood remained unchanged. The region including residues 530-535 also forms a loop at the N-terminal end of Δ N25-NCole7, and it is in interaction with the helix of the HNH-motif. Interestingly, these residues also show an increased RMSF in case of Δ N4-NCole7 which suggests that the deletion of the last four residues at the N-terminus has an influence on the dynamics of the middle part of the protein. Residues 547-560 form the loop between the β -strands of the HNH motif. Changes in the dynamics of this loop can strongly influence the function of the protein as mentioned above.

Fig. 5c shows the average structure of each mutants taken from the 20-25th ns region of simulation. The most obvious effect of shortening the N-terminus is the change in the orientation of the loop between the two β -strands of the HNH-motif (highlighted by a circle on the figure). In case of Δ N25-NCole7 the missing N-terminal loop caused an approach between the two neighbouring loops. Therefore, the loop in the HNH motif lost its original orientation. A smaller but not negligible movement can be also seen in Δ N4-NCole7, the HNH loop is shifted also in this case. As mentioned above, there are catalytically important residues here. The change of average orientation and flexibility of the loop between β -strands of HNH motif can be a reason of the decreased nuclease activity of the shortened proteins. The N-terminal loop can be considered as a structural spacer between the HNH loop and the

DNA binding loop of the protein. It is also worth mentioning, that the N-terminal loop remained in the unchanged position in Δ N4-NCoIE7 even without the positively charged amino acids deleted in this mutant. This has also been observed for the Δ N4-NCoIE7-C* mutant [43].

In a previous study of NCoIE7 [32] it was supposed that N-CoIE7 can bind the DNA substrate in two different manner: coordinating also a water molecule or without it. That is, the Zn^{2+} -ion may have a temporary 5th coordination site that can provide a general acid (assisting in the protonation of the leaving group) in the form of an induced coordinated water molecule. During the 25 ns simulations no such structure was found, there was no water molecule in the proximity of the metal ion. However, a change could be detected in the solvent distribution around the Zn^{2+} -ion in the mutants. The shorter the protein the looser the structure is, that allows more water molecules to get near or in the active center (Fig. S6).

Semiempirical Quantum Chemical Calculations

PM6/MOZYME/COSMO semiempirical quantum chemical calculations were performed to further investigate the fine changes in the active centre of the protein. Fig. 6 shows the active centre in the optimised structures of NCoIE7, Δ N4-NCoIE7 and Δ N25-NCoIE7. The proteins were superimposed with PyMOL [58] using the whole length of the corresponding sequences. The RMSD of the full-length backbone relative to the initial structure of calculations (PDB entry: 1M08) was 1.010 nm for the NCoIE7 (127 atoms fitted), 0.604 nm for Δ N4-NCoIE7 (109 atoms fitted) and 0.893 for Δ N25-NCoIE7 (100 atoms fitted). Aligning with the optimised structure of the NCoIE7 resulted in a RMSD of 0.527 nm (110 atoms) for Δ N4-NCoIE7 and 0.824 (89 atoms) for Δ N25-NCoIE7. The active centre of the Δ N4-NCoIE7 mutant is similar to that of the WT enzyme, as the orientation of histidine sidechains is almost identical (Fig. 6b). However, the small differences in the structures lead to different orientations of the phosphate and Zn^{2+} -ions. This is even more obvious in the case of Δ N25-NCoIE7. The changed geometry around Zn^{2+} and phosphate ions can be a reason of the decreased metal- and DNA-binding ability of the Δ N25-NCoIE7 mutant. This again points to the fact that the removal of the N-terminal part has significant effect on the structure of the C-terminal active centre.

Fig. 6

Conclusions

The necessity of the arginine residue at the N-terminus for the hydrolytic activity of NCoIE7 poses a possibility of a positive allosteric control in this protein. Mass spectrometry, SRCD and fluorescence spectroscopy, and agarose gel mobility shift assays provided information on the effect of the removal of N-terminal sequences on the Zn^{2+} -ion and DNA binding in $\Delta N4$ -NCoIE7-C* and $\Delta N25$ -NCoIE7 mutants. The longer protein bound both Zn^{2+} -ion and DNA stronger than the shorter counterpart due to the structural stabilization effect of the N-terminal amino acids. The C-terminal mutation in $\Delta N4$ -NCoIE7-C* might affect these properties, but our results here and in ref. 43 strongly suggest that the C-terminal flankig sequence does not participate in the metal ion or DNA binding. Molecular dynamics and semiempirical quantum chemical calculations performed in parallel showed that the absence of the N-terminal sequences resulted in significantly increased movement of the backbone atoms in regions of possible interactions with N-terminal loop: 485-487, 511-515 and 570-571 for $\Delta N25$ -NCoIE7, while 467-468, 530-535 and 570-571 residues for $\Delta N4$ -NCoIE7. The distortion of the active centre predicted by semiempirical quantum chemical calculations could also be the reason for weak Zn^{2+} -binding of $\Delta N25$ -NCoIE7. These results lead to a conclusion, that the N-terminal loop plays an important role in the positioning of the arginine residue for the control of the DNase activity. The question arose, whether this could be a common feature among the HNH family of endonucleases? Since the amino acid sequences of the bacterial colicins and pyocins display high similarity, the arginine is frequently found in similar position as in the NCoIE7. Also in the available crystal structures of different members of the HNH family, e.g. the SM endonuclease, VVn endonuclease or nuclease A we found arginine side-chains in the spatial vicinity of the catalytic centre (Fig. 7). The answer thus, seems to be positive. However, since the arginines are not always situated at the N-termini of the proteins, it is difficult to identify them without knowing the 3D structure. Therefore, a detailed bioinformatic study is foreseen in a future communication.

Fig. 7

Acknowledgements This work has received support through the Hungarian Science Foundation (OTKA-NKTH CK80850), TÁMOP-4.2.1/B-09/1/KONV-2010-0005 and TÁMOP-4.2.2/B-10/1-2010-0012. The computational resources at HPC of University Szeged

and financial support from the European Union Research Infrastructure Action FP7 (Integrated Activity on Synchrotron and Free Electron Laser Science, contract number FP7/2007-2013; N^o 226716) is also gratefully acknowledged. B.G. thanks JSPS, while I.N.S-J. and A.C. to Hungarian Scholarship Board for the fellowships provided.

References

- [1] Chak K-F, Kuo W-S, Lu F-M, James R (1991) *J General Microbiol* 137:91–100
- [2] Lin Y-H, Liao C-C, Liang P-H, Yuan HS, Chak K-F (2004) *Biochem Biophys Res Comm* 318:81–87
- [3] Liao C-C, Hsia K-C, Liu Y-W, Leng P-H, Yuan HS, Chak K-F (2001) *Biochem Biophys Res Comm* 284:556–562.
- [4] Cheng Y-S, Shi Z, Doudeva LG, Yang W-Z, Chak K-F, Yuan HS (2006) *J Mol Biol* 356:22–31
- [5] Sui M-J, Tsai L-C, Hsia K-C, Doudeva LG, Ku W-Y, Han GW, Yuan HS (2002) *Protein Sci* 11:2947–2957
- [6] Cheng Y-S, Hsia K-C, Doudeva LG, Chak K-F, Yuan HS (2002) *J Mol Biol* 324:227–236
- [7] Chak K-F, Safo MK, Ku W-Y, Hsieh S-Y, Yuan HS (1996) *Proc Natl Acad Sci* 93:6437–6442
- [8] Hsieh S-Y, Ko T-P, Tseng M-Y, Ku W-Y, Chak K-F, Yuan HS (1997) *EMBO J* 16:1444–1454
- [9] Dennis CA, Videler H, Paupit RA, Wallis R, James R, Moore GR, Kleanthous C (1998) *Biochem J* 333:183–191
- [10] Ko T-P, Liao C-C, Ku W-Y, Chak K-F, Yuan HS (1999) *Structure* 7:91–102
- [11] Kleanthous C, Walker D (2001) *Trends Biochem Sci* 26:624–631
- [12] Kolade OO, Carr SB, Kühlmann UC, Pommer A, Kleanthous C, Bouchcinsky CA, Hemmings AM (2002) *Biochimie* 84:439–446
- [13] Orłowski J, Bujnicki JM (2008) *Nucl Acids Res* 36:3552–3569
- [14] Eastberg JH, Eklund J, Monnat R, Stoddard BL (2007) *Biochemistry* 46:7215–7225
- [15] Mehta P, Katta K, Krishnaswamy S (2004) *Protein Sci* 13:295–300
- [16] Hsia K-C, Chak K-F, Liang P-H, Cheng Y-S, Ku W-Y, Yuan HS (2004) *Structure* 12:205–214
- [17] Michel-Briand Y, Baysse C (2002) *Biochimie* 84:499–510
- [18] Shen BW, Landthaler M, Shub DA, Stoddard BL (2004) *J Mol Biol* 342:43–56
- [19] Ghosh M, Meiss G, Pingoud A, London RE, Pedersen LC (2005) *J Biol Chem* 280:27990–27997
- [20] Kriukiene E, Lubiene J, Lagunavicius A, Lubys A (2005) *Biochim Biophys Acta* 1751:194–204

- [21] Saravanan M, Bujnicki JM, Cymerman IA, Rao DN, Nagaraja V (2004) *Nucl Acids Res* 32:6129–6135
- [22] Saravanan M, Vasu K, Ghosh S, Nagaraja V (2007) *J Biol Chem* 282:32320–32326
- [23] Cymerman IA, Obarska A, Skowronek KJ, Lubys A, Bujnicki MJM (2006) *Proteins: Struct Func Bioinf* 65:867–876
- [24] Jakubauskas A, Giedriene J, Bujnicki JM, Janulaitis A (2007) *J Mol Biol* 370:157–169
- [25] Sokolowska M, Czapinska H, Bochtler M (2009) *Nucl Acids Res* 37:3799–3810
- [26] Veluchamy A, Mary S, Acharya V, Mehta P, Deva T, Krishnaswamy S (2009) *Bioinform* 6:80–83
- [27] Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL, Bateman A (2008) *Nucl Acids Res* 36:D281–D288
- [28] Huang H, Yuan HS (2007) *J Mol Biol* 368:812–821
- [29] Wang Y-T, Yang W-J, Li C-L, Doudeva LG, Yuan HS (2007) *Nucl Acids Res* 35:584–594
- [30] Papadakos G, Wojdyla JA, Kleanthous C (2012) *Quart Rev Biophys* 45:57–103
- [31] Mate MJ, Kleanthous C (2004) *J Biol Chem* 279:34763–34769
- [32] Doudeva LG, Huang H, Hsia K-C, Shi Z, Li C-L, Shen Y, Cheng C-L, Yuan HS (2006) *Protein Sci* 15:269–280
- [33] Pommer AJ, Kuhlmann UC, Cooper A, Hemmings AM, Moore GR, James R, Kleanthous C (1999) *J Biol Chem* 274:27153–27160
- [34] Hannan JP, Whittaker SBM, Hemmings AM, James R, Kleanthous C, Moore GR (2000) *J Inorg Biochem* 79:365–370.
- [35] Keeble AH, Hemmings AM, James R, Moore GR, Kleanthous C (2002) *Biochemistry* 41:10234–10244.
- [36] van den Bremer ETJ, Jiskoot W, James R, Moore GR, Kleanthous C, Heck AJR, Maier CS (2002) *Protein Sci* 11:1738–1752
- [37] Hannan JP, Whittaker SB, Davy SL, Kuhlmann UC, Pommer AJ, Hemmings AM, James R, Kleanthous C, Moore GR (1999) *Protein Sci* 8:1711–1713
- [38] van den Bremer ETJ, Keeble AH, Visser AJWG, van Hoek A, Kleanthous C, Heck AJR, Jiskoot W (2004) *Biochemistry* 43:4347–4355
- [39] Ku W-Y, Liu Y-W, Hsu Y-C, Liao C-C, Liang P-H, Yuan HS, Chak K-F (2002) *Nucl Acids Res* 30:1670–1678
- [40] Shi Z, Chak K-F, Yuan HS (2005) *J Biol Chem* 280: 24663–24668

- [41] Li Ch-L, Hor L-I, Chang Z-F, Tsai L-Ch, Yang W-Z, Yuan HS (2003) *EMBO J* 22:4014–4025
- [42] Gyurcsik B, Czene A (2011) *Future Med Chem* 3:1935–1966
- [43] Tóth E, Czene A Gyurcsik B, Otten H, Poulsen J-CN, Larsen S, Christensen HEM, Nagata K (2012) *Acta Crystallogr Sect D* submitted for publication.
- [44] Limao-Vieira P, Giuliani A, Delwiche J, Parafita R, Mota R, Duflot D, Flament JP, Drage E, Cahillane P, Mason NJ, Hoffmann SV, Hubin-Franskin MJ (2006) *Chem Phys* 324:339–349
- [45] Fahrni CJ, O’Halloran TV (1999) *J Am Chem Soc* 121:11448–11458
- [46] Berendsen HJC, van der Spoel D, van Drunen R (1995) *Comp Phys Comm* 91:43-56
- [47] Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) *J Chem Theory Comput* 4:435–447
- [48] Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) *J Comput Chem* 25, 13:1656-1676
- [49] Olsson MHM, Søndergaard CR, Rotkowski M, Jensen JH (2011) *J Chem Theory Comput* 7:525-537
- [50] MOPAC2009, Stewart JJP (2008) Stewart Computational Chemistry, Colorado Springs, CO, USA, [HTTP://OpenMOPAC.net](http://OpenMOPAC.net)
- [51] Stewart JJP (2007) *J Mol Modeling* 13:1173–1213
- [52] Stewart JJP (2009) *J Mol Model* 15:765–805
- [53] Stewart JJP (1996) *Int J Quantum Chem* 58:133–146
- [54] Klamt A, Schüumann G (1993) *J Chem Soc Perkin Transactions* 2:799–805
- [55] Anthony LC, Suzuki H, Filutowicz M (2004) *J Microbiol Methods* 58:243–250
- [56] Levin KB, Dym O, Albeck S, Magdassi S, Keeble AH, Kleanthous C, Tawfik DS (2009) *Nat Struct Mol Biol* 16:1049–1055
- [57] Wang Y-T, Wright JD, Doudeva LG, Jhang H-C, Lim C, Yuan HS (2009) *J Am Chem Soc* 131:17345–17353
- [58] The PyMOL Molecular Graphics System, Version Open-Source PyMOL 0.99rc6
- [59] Jecklin MC, Schauer S, Dumelin CE, Zenobi R (2009) *J Mol Recognit* 22:319–329
- [60] Miles AJ, Wallace BA (2006) *Chem Soc Rev* 35:39–51

Table 1. Crystal structures of NCoIE7 containing different lengths of amino acid sequences.

PDB code / ref.	Mutation	Complex	Sequence in pdb file*	Reason for inactivity
1M08 / [6]	K446M	protein-Zn-PO ₄ **	446 MRNK-HRGK 576	-
1MZ8 / [5]	-	protein-Zn-PO ₄ -Im7	447 RNKP-IDIH 573	-
1PT3 / [16]	-	protein-8bpDNA	449 KPGK-HRGK 576	no metal ion
1ZNS / [32]	K443M/H545E	proten-Zn-12bpDNA	450 PGKA-DIHR 574	mutation
1ZNV / [32]	K443M/H545E	protein-Ni-PO ₄ -Im7	450 PGKA-HRGK 576	-
7CEI / [10]	-	protein-Zn-Im7	447 RNKP-IDIH 573	-
2IVH / [29]	H545Q	protein-Zn-18bpDNA	449 KPGK-IDIH 573	mutation
2JAZ / [28]	N560D	protein-Zn-PO ₄ -Im7	450 PGKA-HRGK 576	-
2JB0 / [28]	H573A	protein-Zn-Im7	449 KPGK-HIDI 572	-
2JBG / [28]	N560A	protein-Zn-SO ₄ -Im7	448 NKPG-HRGK 576	-
3GJN / [55]	H545A	protein-Zn-Im***	450 PGKA-HRGK 576	-
3GKL / [55]	H545A	protein-Zn-Im***	450 PGKA-HRGK 576	-
3FBD / [56]	D493Q	protein-18bpDNA	445 SKRN-HRGK 576	no metal ion

*All the proteins were expressed in the presence of the immunity protein. A general sequence of the NCoIE7 was MLDKES+446-576, with the exception of the one with PDB code 7CEI where an N-terminal hexahistidine tag in a form of MRGSHHHHHHGSES was attached to the 446-576 sequence. **Charges are omitted for simplicity in the table. ***Mutant Immunity proteins were applied in these experiments. The expression of the NCoIE7 was not described in detail in the original paper [55].

Figure legends

Fig. 1a The sequences of the NCoIE7 (from K446 to K576 according to the original CoIE7 numbering) and the deletion mutants: Δ N4-NCoIE7-C*, Δ N4-NCoIE7 and Δ N25-NCoIE7. The amino acids in red are fused to the N-terminus as a consequence of expression and purification from pGEX-6-P1 vector and cleavage by Prescission protease. The green amino acids indicate the result of the random mutation the C-terminus in Δ N4-NCoIE7-C*. All the blue residues are cut from Δ N25-NCoIE7 and only the dark blue residues from Δ N4-NCoIE7(-C*) and the HNH motif is in orange similarly to Fig. 1d. **b** Growth of the *E. coli* cells expressing different GST-protein variants after induction with IPTG as monitored by OD₆₀₀ measurements. GST protein itself was applied as a control protein without any nuclease activity. The uncertainty of the measurements has not been plotted to simplify the figure. The average error was considered to be ± 0.03 OD₆₀₀ units. **c** The SDS PAGE of the expressed proteins. (GST-NCoIE7' is a toxic variant of NCoIE7 – not detailed here.) **d** The structure of NCoIE7 (PDB id: 1MZ8) in complex with a phosphate ion. Among the N-terminal amino acids, R447 is the closest to the phosphate ion that is bridging it with the Zn²⁺-ion. Figure was created by PyMOL software [57].

Fig. 2a Mass spectrum of the Δ N4-NCoIE7-C* mutant. The main peak corresponds to the mass of the holo protein. The theoretical average mass of the apo protein is calculated to be 16188.1 Da, while the mass of the Zn²⁺-complex is 16253.5 Da. **b** Mass spectra of the Δ N25-NCoIE7 mutant in the presence of 1 to 20-fold molar excess of Zn²⁺-ions. The theoretical average mass of the apo protein is 13123.7 Da, while the mass of the Zn²⁺-complex is calculated to be 13189.1 Da.

Fig. 3 The comparison of the SRCD spectra recorded for **a** Δ N4-NCoIE7-C* (c = 36 μ M) and **b** Δ N25-NCoIE7 (c = 18 μ M) under various conditions. The spectra of the aqueous solutions of the proteins are in blue. The yellow curves belong to the systems where one equivalent of Zn²⁺-ions has been added to protein solutions. For Δ N25-NCoIE7 the SRCD spectrum recorded in the presence of ten equivalents of Zn²⁺-ions has also been plotted (orange curve), since the change here is more expressed than in the case of Δ N4-NCoIE7-C*. Finally, an excess of EDTA was added to the previous solutions and the spectra were recorded (green). In all cases the average of three measurements were plotted.

Fig. 4a Gel mobility shift assay for studying DNA-binding ability of Δ N25-NCole7 (left) and Δ N4-NCole7-C* (right). Lane 1 contains a 1000 base pair marker DNA; lane 2: 0.874 μ M dsDNA sample. In the following lanes equal amounts of the same DNA sample incubated for one hour with increasing amounts of mutant proteins in a constant volume of 10 μ l were loaded. The excess of the proteins was 10, 20, 30, 50, 80, 100 and 200 \times fold for Δ N25-NCole7 and 1, 2, 3, 5, 8, 10 and 20 \times fold for Δ N4-NCole7-C*. **b** Plot of the relative gel mobility shift of the Δ N4-NCole7-C* mutant (i.e. the normalized distance of the shifted band and the band of the unbound dsDNA; the saturation distance was taken as 1.0) vs. equilibrium protein concentration, [P]. At the inflection point [P] = K_D .

Fig. 5a The RMSD vs. time diagrams as a result of 22 ns molecular dynamics calculations in explicit SPC/E water. **b** The average motion of backbone atoms during the MD simulation in proteins. **c** The average structure of the proteins in the 20-25th ns of the simulations. The loop between the two beta-strands of HNH motif is highlighted by a circle. (WT NCole7 is in black, Δ N4-NCole7 in red and Δ N25-NCole7 in green in **5a**, **b** and **c** figures.)

Fig. 6 Superposition of the metal ion binding residues in the optimised structures of NCole7 (in blue), Δ N4-NCole7 (in yellow) and Δ N25-NCole7 (in red) with phosphate and zinc ions. The backbone atoms of the proteins were aligned with **a** 1M08 structure [6] and **b** with the optimised structure of the NCole7.

Fig. 7. The alignment of the HNH-motifs of NCole7 (red, PDB id: 1MZ8) and selected proteins belonging to the HNH superfamily. **a**) VVn endonuclease (PDB id: 1OUP) **b**) SM endonuclease (PDB id: 1G8T) **c**) nuclease *A* (PDB id: 1ZM8)

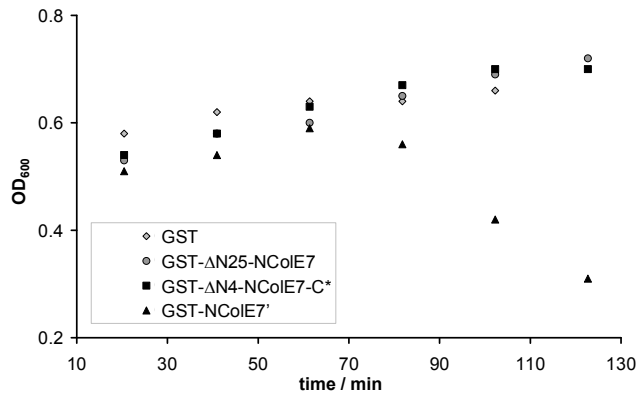
Fig. 1

a

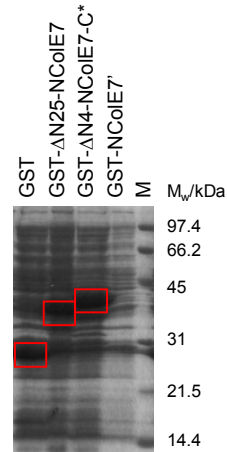
NCoIE7 **GPLGSPEF****KRNK****PGKATGKGPVNNKWLNNAGK**DLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
 ΔN4-NCoIE7-C* **GPLGSPEF**----PGKATGKGPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
 ΔN4-NCoIE7 **GPLGSPEF**----PGKATGKGPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
 ΔN25-NCoIE7 **GPLGSPEF**-----DLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL

 NCoIE7 SKQFSRNNNDRMKVGKAPKTRTQDV**SGKRTSFELHHEKPI****SQNGGVYDMDNISVVT****PKRHIDIHRGK**Stop-----
 ΔN4-NCoIE7-C* SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPI**SQNGGVYDMDNISVVT****PKRHIDIHQVNSSGRIVTD**Stop
 ΔN4-NCoIE7 SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPI**SQNGGVYDMDNISVVT****PKRHIDIHRGK**Stop-----
 ΔN25-NCoIE7 SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPI**SQNGGVYDMDNISVVT****PKRHIDIHRGK**Stop-----

b



c



d

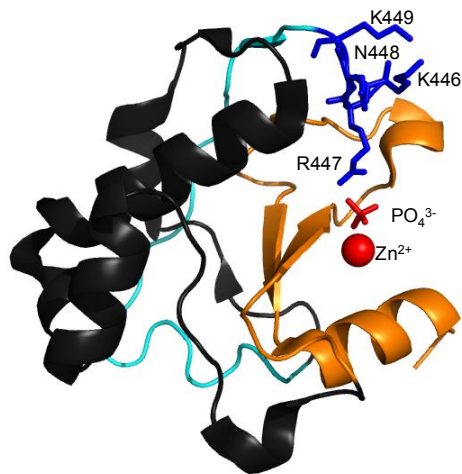
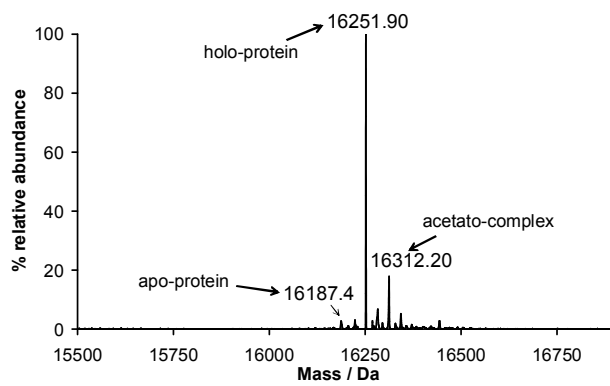


Fig. 2

a



b

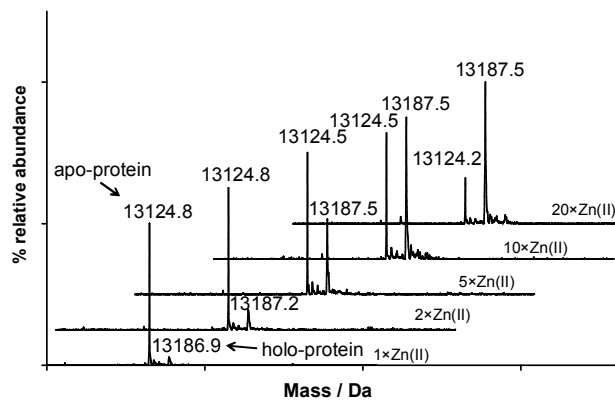
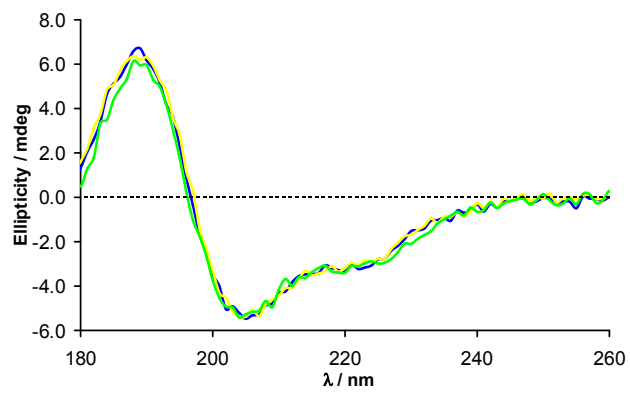


Fig. 3

a



b

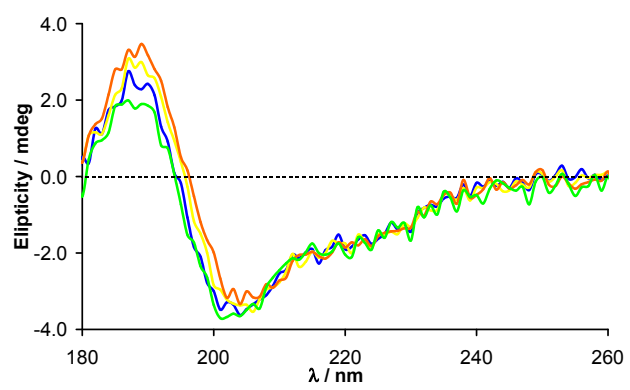
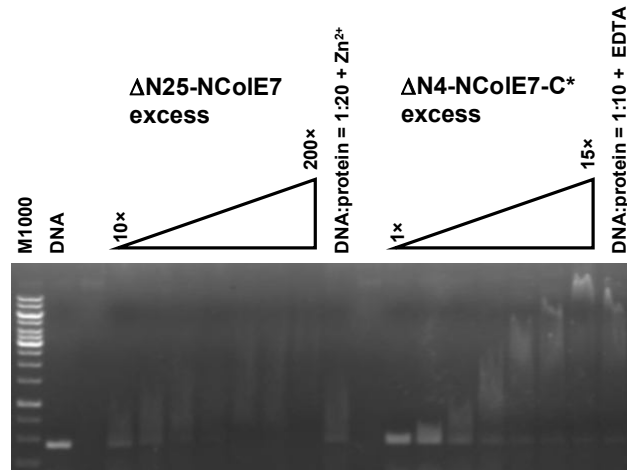


Fig. 4

a



b

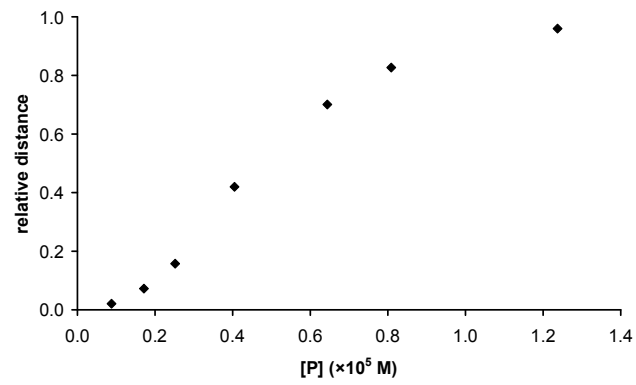
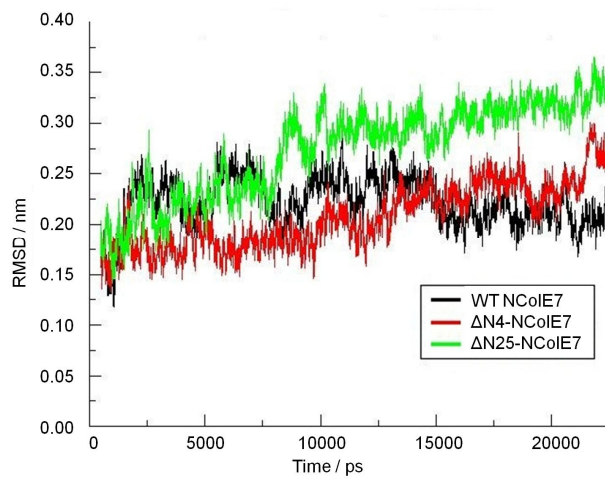
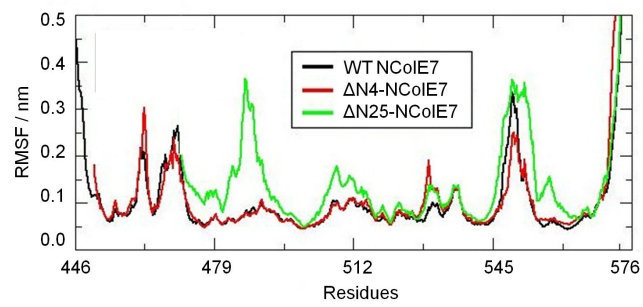


Fig. 5

a



b



c

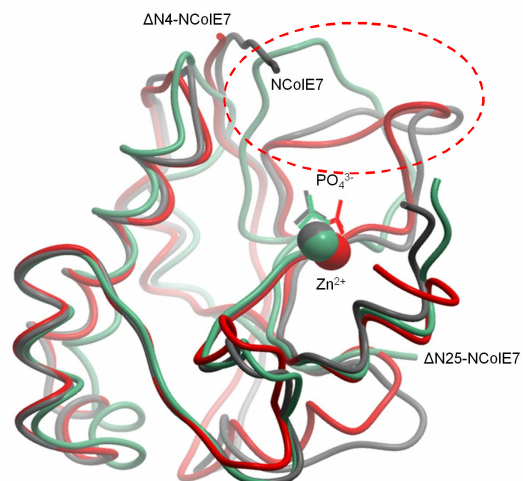
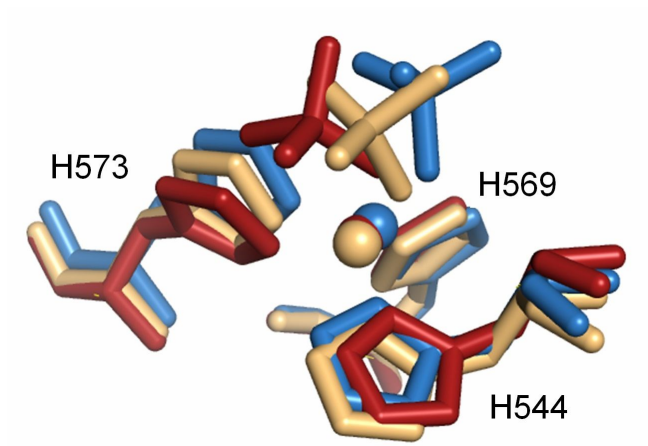


Fig. 6

a



b

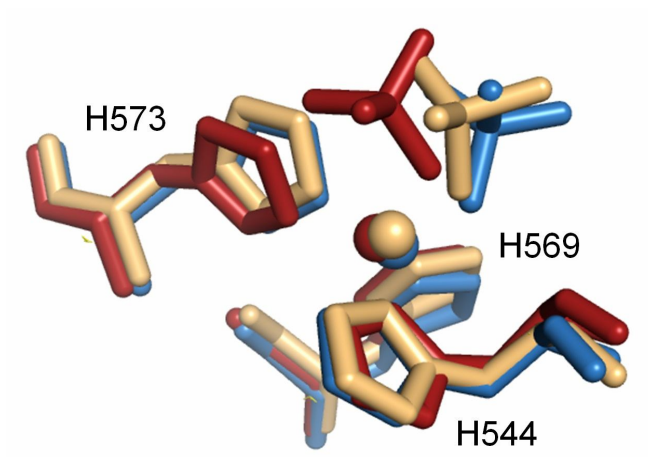
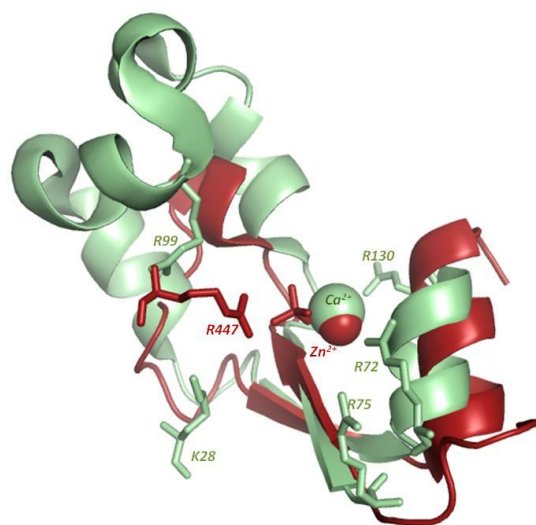
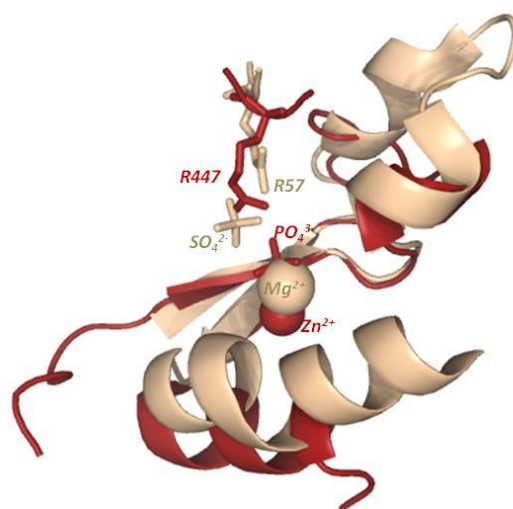


Fig. 7

a



b



c

