Electronic Supplementary Information for the manuscript:

A new insight into the zinc-dependent DNA-cleavage by the colicin E7 nuclease: a crystallographic and computational study

Anikó Czene, Eszter Tóth, Eszter Németh, Harm Otten, Jens-Christian N. Poulsen, Hans E.M. Christensen, Lubomír Rulíšek, Kyosuke Nagata, Sine Larsen* and Béla Gyurcsik*

S1. Protein setup for QM/MM calculations

The crystal structure of the mutant D493Q-NColE7 (residues 446-576) in complex with an 18 bp DNA (PDB code: 3FBD, Wang *et al.,* 2009) was used for setting up the QM/MM model. Residue D493 was restored in an orientation obtained by manual superposition with the corresponding parts of the 2IVH (Wang *et al.,* 2007) structure containing a H545Q mutant NColE7 in complex with a 18 bp DNA substrate and a Zn^{2+}-ion. The last five base pairs of the DNA molecule (not interacting with the protein in our model) and S445 were deleted. As in the 3FBD structure there was no Zn^{2+}-ion, it was positioned into the metal binding site of the HNH motif, using the Zn^{2+}-containing 1M08 structure as the template (Cheng *et al.*, 2002). The metal ion binding H544, H569 and H573 were protonated at the uncoordinated N^{ε}, N^{δ} and N^{δ} nitrogens, respectively, whereas H545 at the N^{ε} to be suitable to activate the nucleophilic water molecule (Shi *et al.,* 2005; Huang & Yuan, 2007, Eastberg *et al.,* 2007)

The protein-DNA complex was solvated by water molecules (keeping also those resolved in the crystal structure) in the sphere with a radius R = 35 Å from its centre (~ 4000 water molecules in total). An additional water molecule was inserted into the putative position for the catalytic water, between H545 and the scissile phosphate. Total of 42 K^+ and 32 Cl^- ions were added to neutralize the system and to achieve a salt concentration of 0.3 M.

The starting geometry for QM/MM calculations was obtained using *sander* module of an AMBER package. The minimization was carried out with heavy atoms fixed at their crystallographic position, whereas hydrogen atoms (added to the DNA-protein complex by the *Leap* module of Amber) and water molecules added to the model were allowed to relax. The oxygen atoms of catalytic and those water molecules resolved in the crystal structure were kept fixed. After initial minimization, 180 ps simulated annealing (T initial 353 K, T

final 0 K, no pressure scaling, NVT ensemble) with the same set of fixed atoms as described above was carried out. Finally, the resulting structure was subjected to an additional molecular mechanics minimization. No periodic boundary conditions were used. The water molecules were kept inside the sphere by a harmonic potential with a default force constant to prevent their dissociation in the course of the MD simulation (simulated annealing).

The initial QM/MM structure of the ΔN4-NColE7 was prepared using the above procedure after deletion of the KRNK amino acid string from the N-terminus of NColE7.

S2. Construction of the QM/MM system

The quantum system consisted of 169 atoms for NColE7 and 156 for ΔN4-NColE7 (Fig. S1). The Zn^{2+}-ion, the catalytic water molecule, the side chain of R447 (including only the positively charged guanidinium and the neighbouring -CH_2- group), the metal binding histidines and the catalytic H545, as well as, V555 (deemed important for positioning the H545 residue) was included in the model. In all cases the truncation involved "fairly inert" non-polar C-C bonds to prevent any artifact in the QM/MM calculations. Hydrogen-link approach was used to cap the dangling bonds in the QM part.

Beside the quantum region (denoted also as System 1), the standard approach in the QM/MM protocol is the division of the MM part into System 2 (surrounding of the quantum region) that is allowed to move in the QM/MM minimizations and System 3 (the rest of the protein) that is kept fixed. The cut-off for selection of System 2 was R = 6 Å (on per-residue basis), *i.e.* all residues containing any of their atoms 6 Å or less from the quantum system were included in the System 2. The System 2 consisted of 1439 atoms (54 residues, 8 DNA bases and 3 K^+ and 2 Cl^- ions) and 124 water molecules. The rest of the protein (System 3) consisted of 13719 atoms (75 residues, 16 DNA bases and 39 K^+ and 30 Cl^- ions and the remaining water molecules 3973).

S3. Description of the QM/MM procedure

All of the QM/MM calculations have been carried out with the ComQum program (Ryde, 1996; Ryde & Olsson, 2001). In the current version, it uses Turbomole 6.3.1. (Treutler & Ahlrichs 1995) for the QM part and AMBER 7 (Case *et al.,* 2002) with the Cornell force field (Cornell et al., 1995) for the MM part. In this approach, the protein and solvent are split into three subsystems: (1) The QM region (system 1) contains the most interesting atoms and

is relaxed by QM/MM forces. (2) System 2 consists of all of the residues within 6Å of any atom in system 1 and is relaxed by a full MM minimization in each step of the QM/MM geometry optimization. (3) Finally, system 3 contains the remaining part of the protein and surrounding solvent molecules and is kept fixed at the original (crystallographic) coordinates. In the quantum chemical calculations, the QM system is represented by a wave function, whereas all of the other atoms are represented by an array of partial point charges, one for each atom, taken from Amber libraries. Thereby, the polarization of the quantum chemical system by the surroundings is included in a self-consistent manner. In the MM calculations for the QM/MM forces and energies, all of the atoms are represented by the Amber force field. When there is a bond between systems 1 and 2 (a junction), the quantum region is truncated by hydrogen atoms, the positions of which are linearly related to the corresponding carbon atoms in the full system (the hydrogen-link approach) (Ryde, 1996; Reuter *et al.,* 2000). To avoid overpolarization of the quantum system, point charges on atoms in MM region bound to junction atoms are zeroed and the remaining charges on the truncated amino acid are adjusted to keep the fragment neutral. The total energy is calculated as

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM123}} - E_{\text{MM1}} \qquad (1)$$

Here, E_{QM} is the QM energy of the quantum system truncated by the hydrogen atoms, excluding the self-energy of the surrounding point charges. E_{MM1} is the MM energy of the quantum system, still truncated by hydrogen atoms, but without any electrostatic interactions. Finally, E_{MM123} is the classical energy of all the atoms in the system with normal atoms at the junctions and with the charges of the quantum system set to zero (to avoid double counting of the electrostatic interactions). By the use of this approach, which is similar to the one used in the Oniom method (Svensson *et al.,* 1996) errors caused by the truncation of the quantum system should cancel out. The calculated forces are the gradient of this energy, but owing to the differing junctions in the various calculations, they have to be corrected using the chain rule. The geometry optimizations were carried out in two steps. First, systems 2 and 3 were frozen and only the quantum system was optimized (this structure will be referred to as Protein_Fixed). Second, both systems 1 and 2 were allowed to relax. In the MM optimization of system 2, the charges on the quantum atoms were updated in each iteration of the QM/MM optimization. This optimization was performed with the looser convergence criterion of 10^{-4} a.u. for the change in energy and 10^{-2} a.u. for the maximum norm of the Cartesian gradient (0.26 and 50 kJ.mol^{-1} Å^{-1})). Then, system 2 was frozen again, and the geometry optimization

was continued with default convergence criteria (10^{-6} and 10^{-3} a.u.). The resulting structure is usually denoted Protein_Free. Such an approach was followed not only to get a feeling for the relaxation of the surrounding protein but also to minimize artifacts caused by the shortcomings of the MM force field (Ryde *et al.,* 2002; Titmuss et al., 2000).

References

Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Werner, P. K. & Kolman, PI. A. (2002) AMBER 7, University of California: San Francisco, CA

Cornell, W. D., Cieplak, P. I., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kolman, P. A. (1995) *J. Am. Chem. Soc.* **117**, 5179–5197.

Reuter, N., Dejaegere, A., Maigret, B. & Karplus, M. (2000) *J. Phys. Chem.* **104**, 1720–1735.

Ryde, U. & Olsson, M. H. M. (2001) *Int. J. Quantum Chem*. **81**, 335–347.

Ryde, U. (1996) *J. Comput.-Aided Mol. Des.* **10**, 153–164.

Ryde, U., Olsen, L. & Nilsson, K. (2002) *J. Comput. Chem.* **23**, 1058–1070.

Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S. & Morokuma, K. (1996) *J. Phys. Chem.* **100**, 19357–19363.

Titmuss, S. J., Cummins, P. L., Bliznyuk A. A., Rendell, A. PI. & Gready, J. E. (2000) *Chem. Phys. Lett.* **320**, 169–176.

Treutler, O. & Ahlrichs, R. (1995) *J. Chem. Phys*. **102**, 346–354.

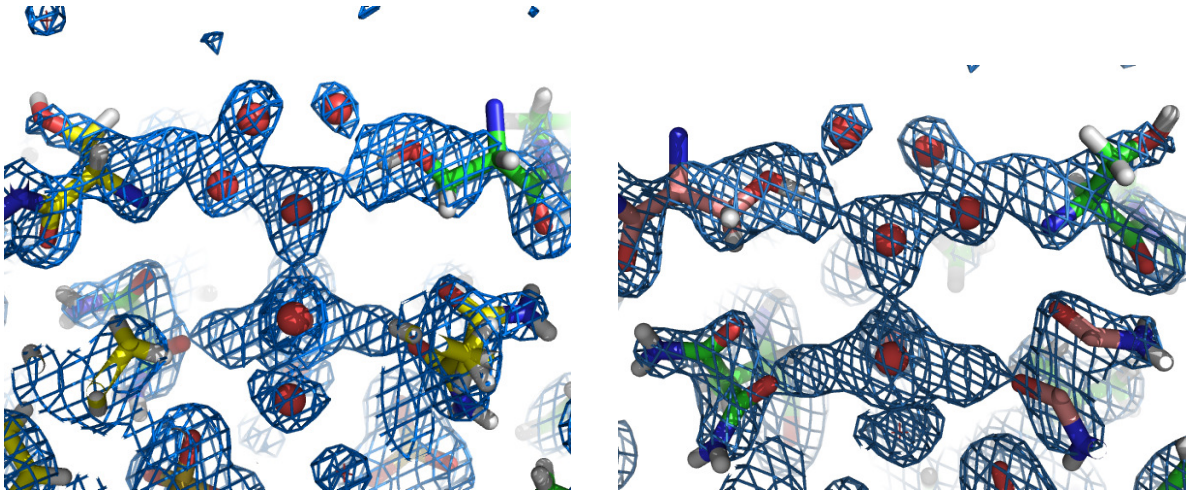


Figure S1. The crystal-packing visualization in the unit-cell with 35% solvent content along the 2-fold axis with density for overlapping molecules A and B with and a symmetry related molecules respectively. This remains the same in $P3_2$ or $P3_221$. The illustration sheds light on the crystallographic and respective NCS two-fold axis, where N-terminal density overlaps.

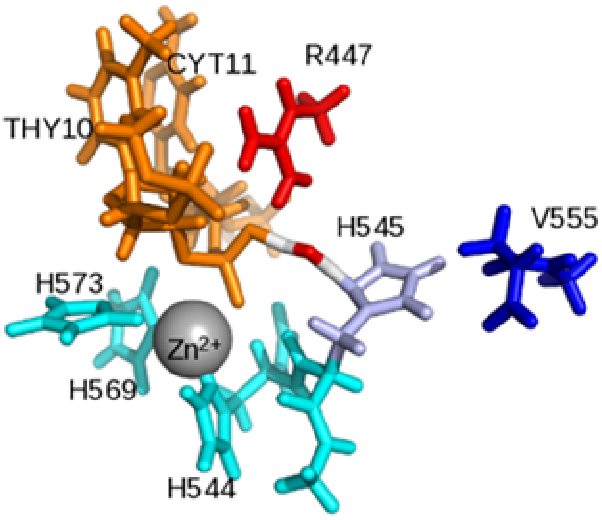
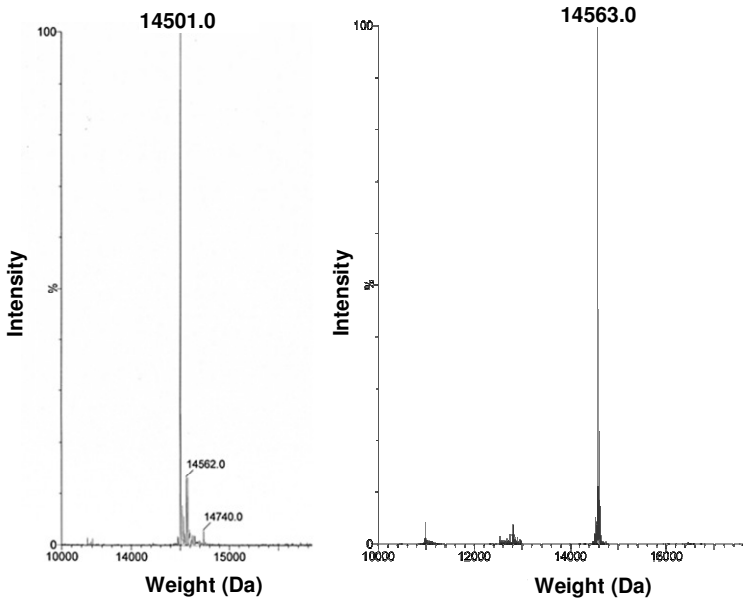


Figure S2. Quantum region (system 1) of the NColE7/DNA model, containing 169 atoms. At the dangling bonds hydrogen caps are used. The quantum region of ΔN4-NColE7/DNA (156 atoms) is the same, except for the missing R447 residue.



(a) (b)

Figure S3. Mass spectra of the (N/X)ΔN4-NColE7 protein expressed and purified from (a) minimal medium, (b) LB medium demonstrate the Zn^{2+} binding in the latter molecule.

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD

VKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLS

KLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAI

PQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSPEFPGKATGKGKPVNNK

WLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPELSKQFSRNNNDRMKVGK

APKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGKStop

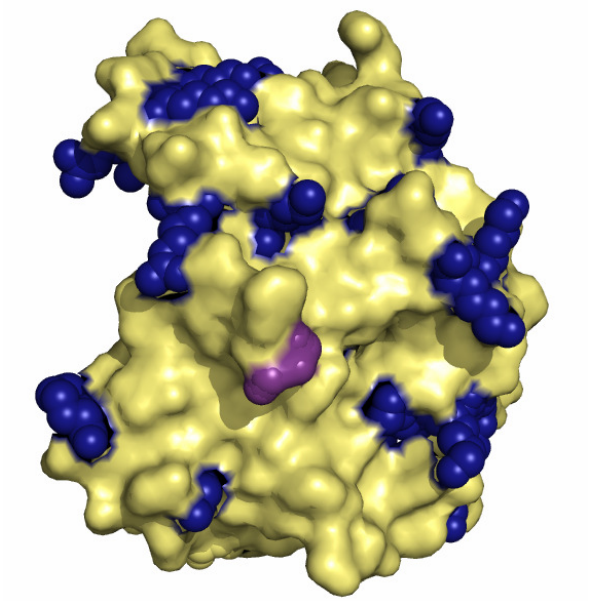
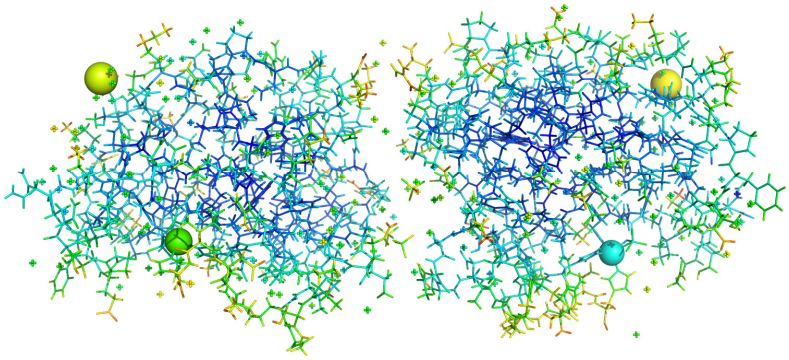
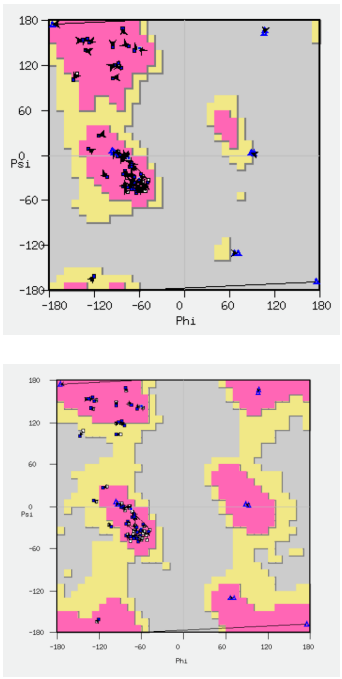


Figure S4. The upper panel shows the full amino acid sequence of the (GST)ΔN4-NColE7 mutant protein. The ΔN4-NColE7 sequence after the cleavage by the human rhinovirus C3 protease as described in Material and Methods is highlighted by green, while the amino acids with positively charged side chains in GST region by blue backgrounds. In the lower panel the crystal structure of the GST protein (PDB entry 1M99) is shown with the positively charged amino acids exposed at the surface. The C-terminus, *i.e.* the origin of the linker region is coloured in violet – such as in the upper panel.





(a) (b)

Figure S5. (a) Stick representation of the overall fold of ΔN4-NColE7-C*. Chlorine ions represented as yellow spheres and zinc ions in green and turquoise spheres, respectively. Atoms colored by atomic displacement factor, from blue to red (low to high): 5 - 116 $Å^2$, mean 25 $Å^2$. (b) Kleywegt plot of A *vs.* B molecule highlighting the differences in Ramachandran plots. Points are connected by arrows, which lengths relate to the difference between the two molecules.

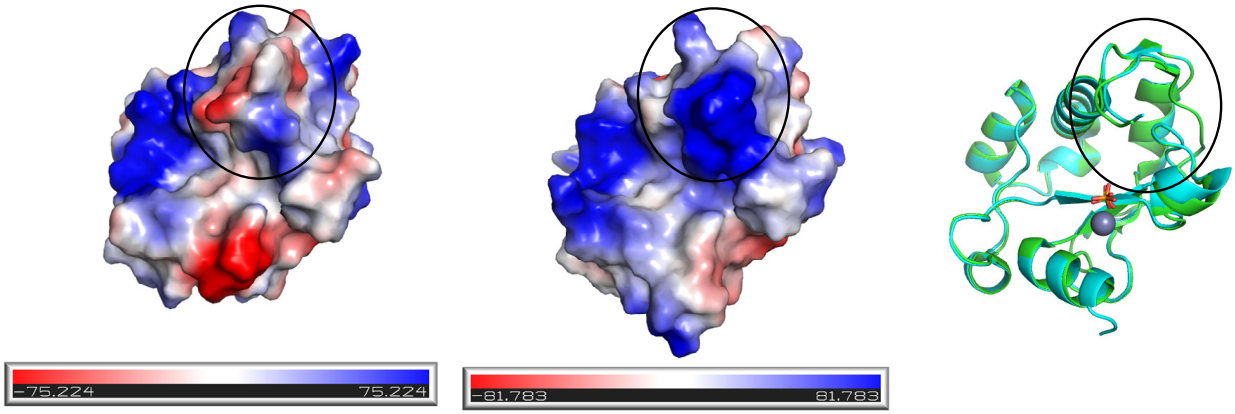


Figure S6. Vacuum electrostatics of (a) ΔN4-NColE7-C* (3ZFK) *vs.* (b) NColE7 (1MZ8) with KRNK at the N-terminus. The shorter N-terminus also has less positive charges and surface electrostatics change as shown within the circled area. Panel (c) shows the cartoon representation of the proteins for better understanding.

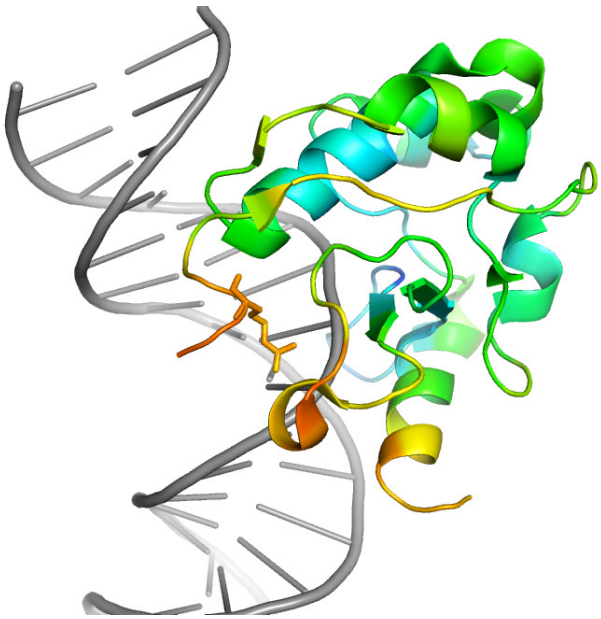


Figure S7. The cartoon of the NColE7 colored by temperature factors in the crystal structure of the NColE7/DNA complex, in the absence of Zn^{2+}-ion (PDB entry: 3FBD; Wang *et al.,* 2009) visualized in the Pymol program (DNA is colored in grey).