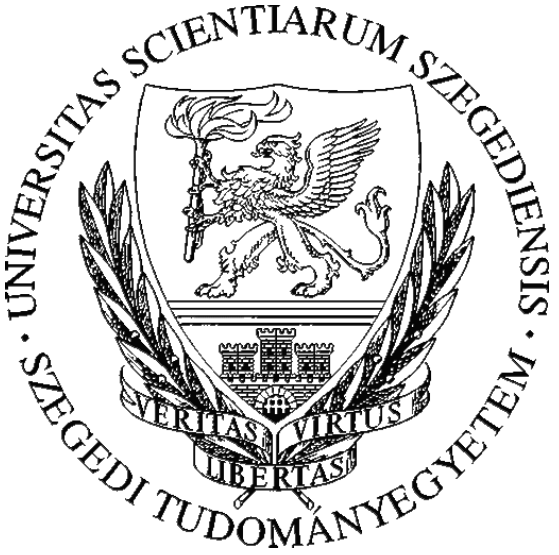


Ph.D. dissertation

ARTIFICIAL METALLONUCLEASES - MOLECULAR TOOLS FOR GENE THERAPY OF CANCER

HEBA ALAA ELDEEN HOSINY ABD ELHAMEED

Supervisor:
DR. BÉLA GYURCSIK
Associate professor

Doctoral School of Chemistry

Department of Inorganic and Analytical Chemistry

Faculty of Science and Informatics

University of Szeged, Hungary

2020

Table of Contents

List of Abbreviations

A549	human lung cancer cells
AAV	adeno-associated virus
AN	artificial nuclease
BL21(DE3), DH5α, Mach1	*E. coli* strains
Cas9	CRISPR-associated nuclease 9
CD	circular dichroism
CPP	cell-penetrating peptide
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	changeable CRISPR RNA
DLS	dynamic light scattering
DMEM	Dulbecco's modified eagle medium
DSB	double-strand break
E. coli	*Escherichia coli*
EGFR	epidermal growth factor receptor
ESI	electrospray ionization
BbsI, BsmBI, FokI, VspI	restriction endonucleases
FPLC	fast protein liquid chromatography
gRNA	guide RNA
HDR	homology-directed repair
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
KGNK	R447G mutant of the NColE7 protein
LB	lysogeny broth or commonly Luria-Bertani medium
LV	Lentivirus
MALDI	matrix-assisted laser desorption/ionization
MS	mass spectrometry
MTT	3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide
NColE7	nuclease domain of the Colicin E7
NHEJ	non-homologous end-joining
NMR	nuclear magnetic resonance spectroscopy
NTA	nitrilotriacetic acid
OD	optical density
PAM	protospacer adjacent motif
PCD	particle charge detector
PCR	polymerase chain reaction
pEGFP-N3, pET-21a, pX458	name of DNA plasmid
PEI	poly(ethylenimine)
RNP	Ribonucleoprotein
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEM	scanning electron microscope
SRCD	synchrotron radiation circular dichroism
dsDNA, ssDNA	double-strand and single strand DNA, respectively
TALEN	transcription activator-like effector nuclease
TCEP	tris(2-carboxyethyl)phosphine)
TEM	transmission electron microscopy
tracrRNA	trans-activating crRNA
WSLP	water-soluble lipopolymer
ZFN	zinc finger nuclease
ZFP	zinc finger protein
ΔN4-NColE7	nuclease domain of Colicin E7 lacking the four N-terminal amino acids

1. INTRODUCTION

All cancer cells contain multiple genetic mutations that allow them to grow progressively and exhibit the characteristics of malignancy. Therefore, targeting the cancer cell genome is an attractive approach of cancer therapy. The ability to redeem such cancer-associated mutations requires a reagent which should induce correction of genetic changes in a highly specific manner without off-target effects. The reagent would also require efficient delivery into all or nearly all cells of a tumor. In the last decades, novel genetic-editing technologies have been developed, based on artificial nucleases (ANs). These ANs have important potential clinical applications including the treatment of genetic diseases, viral infections, and cancer. These new classes of reagents, which can specifically target nucleotide sequences within cellular genomes, are of four major types: meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the clustered regulatory interspaced short palindromic repeat (CRISPR) and its associated nuclease 9 (Cas9) system. However, all of the above artificial nucleases have various levels of cytotoxicity.

Therefore, the goal of our research group is to develop a regulated zinc finger nuclease, which is based on Colicin E7 metallonuclease domain (NColE7) instead of the widely used FokI nuclease domain in the ZFNs and TALENs. Accordingly, new ANs should be developed, in which the various parts of the enzyme regulate the catalytic activity so that the DNA hydrolysis only occurs when the enzyme binds to its specific target site. In addition, any damage of the enzyme should cancel its function, rather than lead to non-specific cleavages.

On this path, the development of such an AN includes various approaches presented in this thesis. (i) The optimization of NColE7 based AN depends largely on understanding of the intramolecular regulation of the NColE7 nuclease within the native and modified variants, including the zinc finger – NColE7 ANs. (ii) The design of CRISPR/Cas9 ANs to target and knockout selected oncogenes promotes the selection of the most efficient DNA recognition sites for the new types of ANs. It also allows for comparison of different types of ANs. (iii) The optimization of gene delivery systems using high molecular weights poly(ethylenimine) (PEI) modified into a water-soluble lipopolymer (WSLP) increases the bioavailability and efficiency of the ANs.

The results of my PhD work are intended to contribute to the development of an artificial metallonulease with improved specificity and effectivity for potential gene therapy.

2. LITERATURE REVIEW

2.1. Cancer

Cancer threats the human persistently and is evaluated to cause one in seventh deaths over the world [1]. This illness is defined as the alteration of normal and healthy tissue into malignant and interfering tissue, which intensely divides and consumes the requirements of other cells [2]. Various mutations leading to cancer might be initiated by environmental factors like radiation or chemicals during the DNA replication in the normal cells [3], while approximately fifth of all the human cancers is related to viral infections [4,5]. Cancerous mutations can be classified into several groups according to gene function as (i) overexpressed oncogenes, (ii) nonfunctional tumor suppressors, (iii) modification of epigenetic controls and (iv) or genes that confer chemoresistance.

(i) Oncogenes: Proto-oncogenes are a group of genes that cause normal cells to become cancerous when they are mutated. The mutated version of a proto-oncogene is called an oncogene. Proto-oncogenes encode proteins that function to stimulate cell division, inhibit cell differentiation, and prevent cell death. Cancer cells typically exhibit increased production of these proteins, thus leading to the cell grows out of control. (**Figure 2.1**). Proto-oncogenes are e.g., the genes of the erythroblastic oncogene B (ErbB) family receptor tyrosine kinase enzymes. These proteins are structurally related to the epidermal growth factor receptor (EGFR). In humans, the ErbB family includes ErbB1 (EGFR), ErbB2, ErbB3 and ErbB4 [6]. As the consequence of gene mutation, these proteins are overexpressed mainly in breast cancer. *SRC* is a proto-oncogene coding for sarcoma (src) non receptor tyrosine kinase, which is mutated in numerous types of cancers of the colon, liver, lung, breast and the pancreas [7]. Additional types of oncogenes include the genes of the G-protein family (e.g. Rat Sarcoma; Ras), which are present in 20-30% of all human cancers [8] and intracellular protein kinases (e.g. Rapidly Accelerated Fibrosarcoma; Raf), with 20% of all human tumors displaying a mutated B-Raf gene mostly in renal cell carcinoma and melanoma [9]. Oncogenic nuclear transcription factors (e.g. myc, which is originally isolated from an avian myelocytomatosis virus) are amplified in epithelial ovarian, breast, gastric, pancreatic, colorectal and uterine cancers [10].

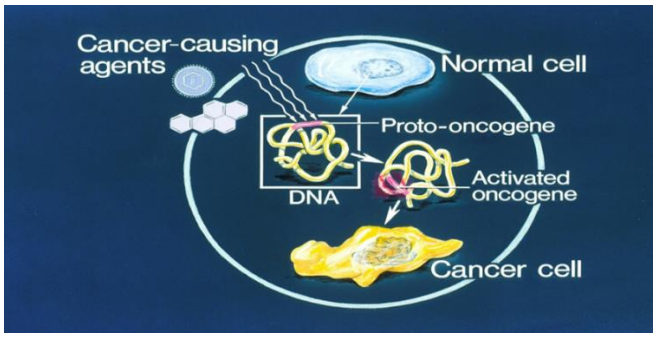


Figure 2.1. Illustration of how a normal cell is converted to a cancer cell when an oncogene becomes activated. This image was released by the National Cancer Institute, an agency part of the National Institutes of Health, with the ID 2351.

(ii) Tumors suppressors stop the cell division or initiate the apoptosis. The p53 tumor suppressor protein is the most known one mutated in about fifty percent of human cancers [11]. Breast cancer-1 (BRCA1) [12] and breast cancer-2 (BRCA2) proteins also represent tumor suppressors in breast and ovarian cancers, respectively. The ataxia-telangiectasia (ATM) tumor suppressor is mutated in some types of leukemias and lymphomas.

(iii) Epigenetic factors which regulate the gene expression are downregulated in cancer cells. Examples are DNA methyltransferases (DNMTs) and enzymes responsible for histone modifications, e.g., lysine-specific demethylase 1 (LSD1), enhancer of zeste homolog 2 (EZH2), and nuclear receptor binding SET domain protein 2 (NSD2) [13].

(iv) Chemoresistance genes: The ability of cancer cells to develop chemoresistance towards drugs is a major obstacle in many cancer therapies. Overexpression of e.g., ABC transporters leads to increased drug resistance. An example is the multi-drug resistance 1 (MDR1) gene, from which the P-glycoprotein is expressed [14]. Another efflux protein is the multidrug resistance-associated protein 2 (MRP2) [15]. There are enzymes that activate or deactivate the chemotherapies and decrease the amount of the drug available to target the cancerous cells, such as the glutathione-S-transferase (GST) enzyme, which inactivates cisplatin [16]. Another example is the superfamily of the cytochrome P450 monooxygenase heme enzymes metabolizing various drugs [17].

The aim of cancer treatment is to get rid of the cancer cells while keeping the normal cells unaffected. Surgery includes the removal of the solid tumors but it is ineffective towards the small and metastatic tumors [18]. The current therapeutic methods are based on chemotherapy or radiation therapy. Chemotherapy utilizes chemical drugs to disrupt the DNA replication to kill the rapidly multiplying cells. Oncogenic proteins are also major molecular

targets for anti-cancer drug design. However, these treatments have cytotoxic side effects because most of them target both normal and cancer cells [19].

Nowadays, gene therapy is a potentially effective and challenging therapeutic tool because all cancers have genetic mutations roots. Cancer gene therapy permits selective targeting and repair of cancer cells only. Gene therapy tools are still under progress, aiming at e.g., introducing tumor suppressor genes or knocking out oncogenes in the cancerous cells [20].

2.2. Genome editing

2.2.1. Definition and the mechanism of genome editing

In the last decades, genome editing as a novel technology has attracted much attention of the researchers and enabled to manipulate any gene sequence in plant, animal and human cells. A successful approach is based on artificial nucleases, which consist of a sequence-specific DNA binding domain cooperating with a non-specific DNA cleavage unit [21]. These ANs promise a potential to treat various human genetic diseases, such as e.g. haemophilia, sickle-cell anemia and several types of cancer. The therapeutic influence of genome editing is obtained by modifying accurately targeted specific DNA sequence inside cells by insertion, deletion, disruption, mutation, or replacement [22,23]. The first important step of genome editing is the double-strand break (DSB) at the DNA sequence of interest. This DSB is recognized in the cell and can be repaired by two pathways: non-homologous end-joining (NHEJ) or homology-directed repair (HDR) (**Figure 2.2**).

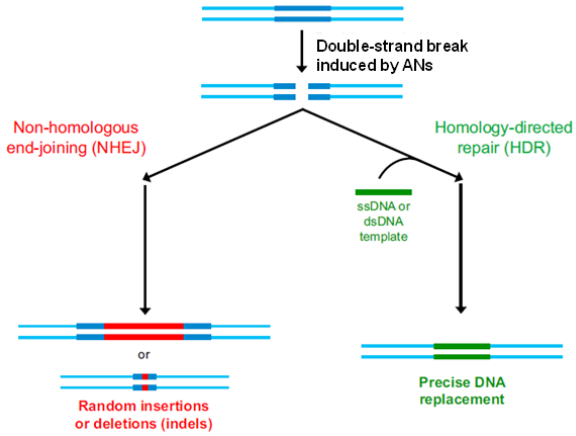


Figure 2.2. Genome editing is initiated by engineered nucleases through double-strand DNA breaks (DSBs) at specific site within the genome. These DSBs can be repaired by two pathways: non-homologous end-joining (NHEJ) left or homology-directed repair (HDR) right [24].

NHEJ results in small, random insertions / deletions (indels) at the cleavage site, which often leads to a non-functional protein product if the indels cause a reading frame shift within

the gene of the expressed protein. HDR catalyzes the DNA repair at the cleaved site by using a DNA template of intact sequence for the precise correction of the DNA. HDR repair mechanism is less effective than NHEJ [25].

2.3. Artificial nucleases

Programmable nucleases such as meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR) /CRISPR-associated nuclease 9 (Cas9) or often also called RNA-guided engineered nucleases (RGENs) are now widely used for genome editing in various cells, and possess a broad range of novel applications in biotechnology and medicine [26]. These engineered nucleases are 'molecular scissors' causing breaks in the DNA. ANs are categorized according to their mode of action to target a specific DNA sequence (**Figure 2.3**).

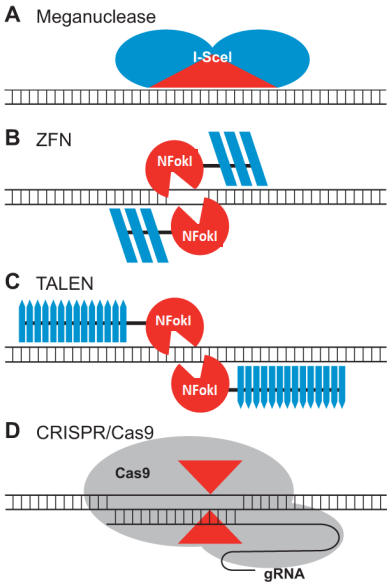


Figure 2.3. Schematic illustration of the four types of artificial nucleases. (A) The DNA cleavage domain of the I-SceI meganuclease is shown in red. (B) A ZFN dimer is represented in complex with the DNA. The two zinc fingers (blue) bind to the specific DNA target, while the nuclease domain of FokI (NFokI) (red) is the catalytic domain which dimerizes at the 5–7 bp spacer and cleaves the DNA within it. (C) A TALEN consists of NFokI as nuclease domain (red) and the DNA binding domains, shown in blue. The recognition sites are separated by a 15–20 bp spacer. (D) The CRISPR/Cas9 system binds to its specific DNA target by the help of the guide RNA (gRNA) through base pairing. The cleavage by Cas9 is represented by red arrowheads [27].

Meganucleases, ZFNs and TALENs use DNA-protein interactions in targeting specific DNA sequence for DNA cleavage, while CRISPR-Cas9 is directed by DNA-RNA interactions to bind the target gene. The DNA recognition of meganucleases is difficult to redesign. In the following, I will only discuss ZFNs, TALENs and RGENs.

2.3.1. Zinc finger nucleases (ZFNs)

Zinc finger NFokI nucleases represent the first generation of the modular enzymes for genome editing. They have been utilized to target and change specific gene sequences in numerous organisms such as bacteria, viruses, plants, insects, fish and mammals such as mice and pigs, as well as in cultured mammalian cells [28,29]. In humans, the C-C motif chemokine receptor 5 (*CCR5*) gene was one of the first ZFN targets. The product of this gene is a co-receptor for human immunodeficiency virus 1 (HIV-1) responsible for inducing acquired immunodeficiency syndrome (AIDS). ZFNs have been designed to target this gene in the T-cells isolated from HIV patients. The patients were injected with the gene edited cells. Hopeful initial results revealed increased T-cell levels, indicating that the virus was not able to destroy these cells[30].

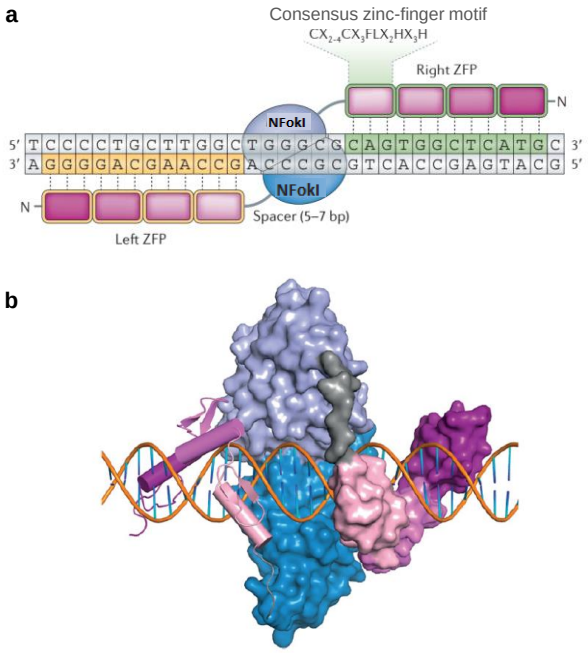


Figure 2.4. Structure of ZFNs. a, Schematic illustration of a zinc finger nuclease (ZFN) pair bound to specific DNA recognition sequence. Each ZFN is composed of a zinc-finger protein (ZFP) at the N-terminus and NFokI at C-terminus. In the consensus zinc-finger motif, X represents any amino acid. Target sequences of ZFN pairs are usually 18–36 bp in length. **b**, Model of ZFN pairs bound to DNA is shown. Zinc fingers are shown by pink ribbons (left) or by space-filling (right) representation. The grey region is the peptide linker between the DNA-binding and catalytic domains. The NFokI cleavage domains are shown in blue and purple using space-filling representations. DNA is represented as the sugar-phosphate backbone in orange and the bases in blue. The separation between ZF binding sites is 6 bp [31].

These artificial programmable nucleases are created by fusing an N-terminal zinc finger protein (ZFP) as a DNA binding domain to a C-terminal DNA cleavage domain of FokI restriction enzyme (NFokI) [32] (**Figure 2.4**). Zinc finger domains can be engineered to target

the desired specific DNA sequences, and this enables zinc-finger nucleases to target unique sequences within complex genomes.

ZFPs are composed of arrays of Cys_2His_2 (C2H2) zinc-fingers which are the most common types of DNA binding domains in the eukaryotes. Each zinc finger has about 30 amino acids, forming one α helix and two short β strands in the presence of the Zn^{2+} ion. Each finger binds to three adjacent nucleotides by the specific protein side chains within the major groove of DNA. The initial contacts are established by residues –1,3 and 6 related to the beginning of the α helix as shown in **Figure 2.5.** The DNA specificity of the zinc finger can be changed almost arbitrarily by varying the interacting amino acids. 3–6 zinc finger arrays are used to generate a ZFN subunit that binds to DNA sequences of 9–18 bp. A 18 base pair long DNA recognition sequence could increase the specificity even within the 68 billion bp of DNA being much larger than the human genome. The designed three or four finger ZFNs had successful applications. Increasing the number of the fingers increases the specificity but unfortunately decreases the possibility of finding a proper target site within the genome [33,21].

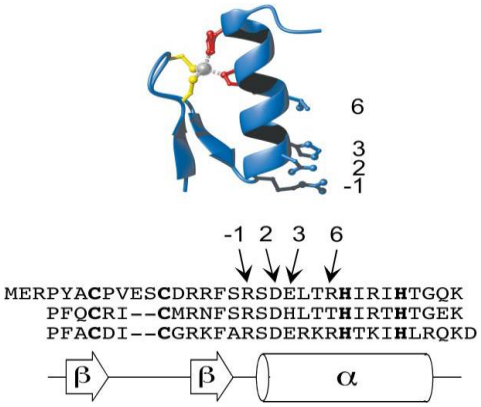


Figure 2.5. The Cys_2His_2 zinc finger motif and the amino acid sequences of finger motifs in Zif268 ZFP. (*Top*) A ribbon diagram of finger 2 from Zif268 including the two cysteine side chains (yellow) and two histidine side chains (red) that coordinate the zinc ion (silver sphere). The side chains of key residues at positions that make base contacts in the major groove of the DNA are indicated by the numbers identifying their position relative to the start of the recognition helix. (*Bottom*) The amino acid sequence alignment of the three fingers from Zif268 shows the conserved cysteines and histidines in boldface. Secondary structure elements are indicated at the bottom of the diagram [34].

The crystal structures of a ZFPs in their complexes with DNA indicated that zinc finger–DNA interactions are modular in nature; each zinc finger contacts almost individually with a triplet DNA sequence (PDB IDs: 1MEY, 2I13 and 1ZAA). Definitely, new ZFPs with preferred specificities can be modularly assembled from pre-distinguished zinc-finger units [35]. Zinc fingers prepared by this method are usually not strictly specific in their DNA targeting [36,37]. It still challenging to design ZFNs with high activity, specificity and low cytotoxicity using

widely available resources. Various computer programs are available for finding the possible ZFN target sites in any DNA sequence (**Table 2.1**).

Table 2.1. Resources for programmable nuclease design [31]

	Online resources for nuclease design	**Non-profit organization**	**Company**
ZFNs	• Genome-wide tag scanner for nuclease off-sites • The Segal Laboratory software site • ZFN target site algorithm for identifying sites compatible with the Lawson-Wolfe modular assembly system • Zinc-finger tools • ZiFiT Targeter software	• Addgene	• Sigma-Aldrich • ToolGen
TALENs	• E-TALEN • Genome engineering resources • Scoring algorithm for predicting TALE(N) activity • ToolGen TALEN Designer • ZiFiT Targeter software	• Addgene • TALEN library resource	• Cellectis Bioresearch • Life Technologies • ToolGen • Transposagen • Biopharmaceuticals
RGENs	• E-CRISP • Genome engineering resources • RGEN tools • ZiFiT Targeter software	• Addgene	• Life Technologies • Sigma-Aldrich • System Biosciences • ToolGen • Transposagen • Biopharmaceuticals

NFokI domains dimerize on the DNA and produce a double-strand break (DSB) with 5'overhangs [38]. Therefore, an active nuclease requires two A and B ZFN monomers. Each monomer must bind to its adjacent half-site. This binding mode duplicates the length of the recognized DNA site, which significantly increases ZFNs specificity. Nevertheless, both ZFNs can also form homodimers to cleave DNA when only A or B ZFNs bind to DNA causing undesirable off-target effects. The NFokI dimeric interface was genetically modified to allow only heterodimers [39,40], which greatly decreased the off-target effects of ZFNs in a test system. FokI domains with improved action were also produced by directed evolution.

2.3.2. Transcription activator like effector nucleases (TALENs)

Recently, TALEs as programmable DNA binding proteins were discovered [41,42]. TALEs are natural proteins derived from *Xanthomonas spp* plant pathogenic bacteria. TALEs penetrate the nucleus of the host cell, alter the gene to promote the bacterial infection [43]. TALEs structure is similar to that of ZFs but in contrast to zinc fingers, TALE repeat bind DNA

forward, in other words, the N-terminal repeat binds the 5’ base pair. A TALE protein is a tandem array of 33–35 amino acid repeat units. Each TALE repeat identifies a single deoxynucleotide in the major groove of the DNA [44]. The two amino acids at positions 12 and 13 determine the base pair specificity of a single repeat unit (**Figure 2.6**). These two amino acids are termed repeat variable diresidues (RVDs) [45]. Four different RVDs, namely Asn-Asn (NN), Asn-Ile (NI), His-Asp (HD) and Asn-Gly (NG) are most commonly used to recognize G, A, C, T deoxynucleotides, correspondingly. Only residue 13 binds DNA, while the side-chain of residue 12 folds back and stabilizes the contacts with other residues in the TALE protein module.

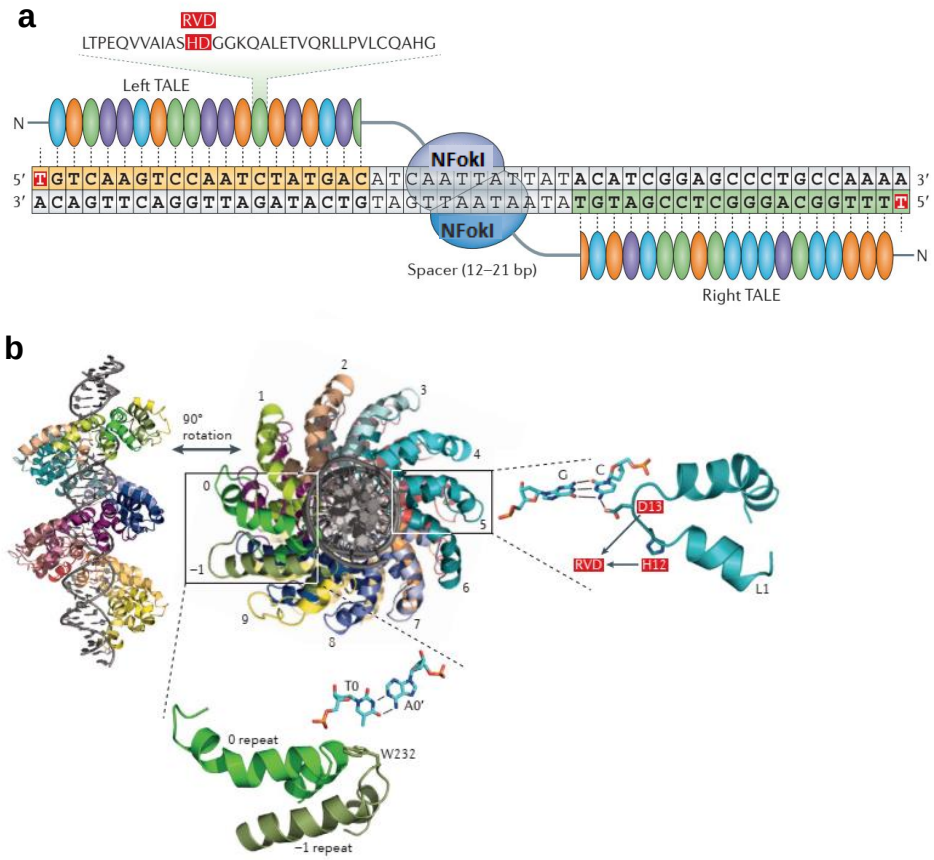


Figure 2.6. Structure of TALENs. a, Schematic representation of a TALEN pair. Each TALEN consists of transcription activator-like effectors (TALEs) at N-terminus and the NFokI nuclease domain at C-terminus. Each TALE repeat is composed of 33–35 amino acids and recognizes a single base pair through the amino acids at 12 and 13 positions, which is called the repeat variable diresidue (RVD, shown in red). Target sequences of TALEN pairs are typically 30–40 bp in length without the spacers. **b**, TALE–DNA co-crystal structure showing the interaction between the RVDs in TALE and the major groove of DNA. The amino-terminal repeats (represented as 0 and –1 in the box) contact 5′ T deoxynucleotide [44].

The one to one correspondence between the four RVD modules and the four nucleobases makes the design of new TALEs with the desired recognition sequence straightforward. This is an important advantage of TALEs over other types of DNA binding proteins, but the design of TALEs requires a T deoxynucleotide at the 5′ end of the target sequence which is recognized by two amino-terminal cryptic repeat folds. Thus, finding a target sequence with T at the 5′ end is always required, limiting the use of TALEs. Recently established TALE variants that distinguish other bases at the 5′ end would additionally increase the targetable sites [46,47]. Several online computer programs are available for the design of TALEs (**Table 2.1**).

TALENs are composed of a TALE protein (DNA binding domain) connected to the NFokI cleavage domain. TALENs have been applied for modification of endogenous genes in various species, including yeast[48], viruses[49], plants, nematodes, insects, frogs, fish and mammals such as mice, rats [50] and pigs, in addition to the cultured mammalian cells [51,21].

2.3.3. CRISPR/Cas RNA-guided nucleases

The newest tool of genome editing is known as clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. It is initially derived from the adaptive immune system of the bacterium *Streptococcus pyogenes* [52]. Since then, this system was found in 90% of archaea and 50% of bacteria. Bacteria use CRISPR/Cas system for the defense against the foreign viral DNA sequences (e.g. bacteriophage) [53]. CRISPR/Cas system is composed of a short RNA and Cas protein to target and cleave a specific DNA sequence, respectively. There are two classes of CRISPR/Cas [54]. Class 1 needs several Cas proteins but Class 2 only requires one Cas protein with many domains. The Class 2 CRISPR-Cas system is therefore, better for genome editing applications due to its simplicity [55].

CRISPR-Cas9 is a widely used example of Class 2. It includes a highly effective Cas9 enzyme. The Cas9 is an endonuclease protein with two nuclease domains: RuvC and HNH (**Figure 2.7A**). Thus, Cas9 acts as a pair of scissors to cleave DNA. CRISPR/Cas9 has two RNA components: a changeable CRISPR RNA (crRNA) and a fixed trans-activating crRNA (tracrRNA). The crRNA and the complementary tracrRNA form a double-strand RNA structure denoted as the guide RNA. These two RNAs can also be covalently linked together into a single guide RNA (sgRNA) (**Figure 2.7B**).

Each crRNA contains a conserved repeat sequence and a variable spacer of around 20 nucleotide length. This variable spacer should be complementary to the target DNA and it has to be selected so that a three base sequence, called PAM (protospacer adjacent motif) is present in the targeted DNA adjacent to it. The absence of the PAM sequence can influence the affinity

of the Cas9 enzyme towards the DNA/guide RNA complex. By the help of the sgRNA, Cas9 cuts the matched DNA [56]. The RuvC domain cuts non-complementary DNA strand, including the PAM sequence, while the HNH domain cuts the complementary DNA strand hybridized to the crRNA spacer region. The schematic of the molecular mechanism of DNA cleavage by the CRISPR-Cas9 system is shown in **(Figure 2.7 C)**.

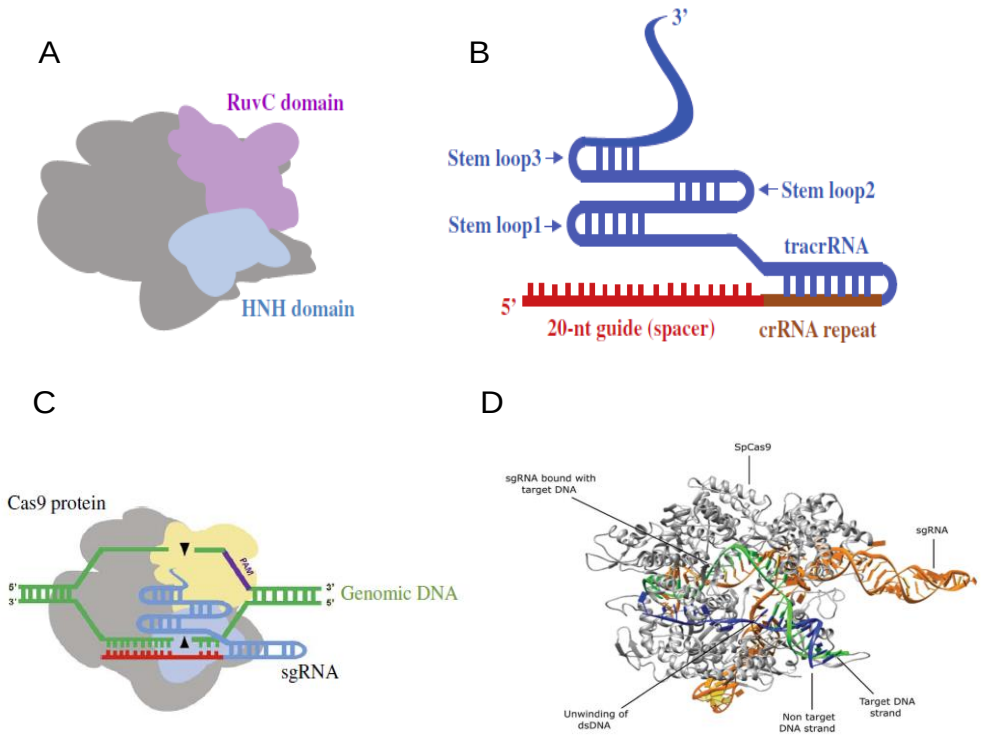


Figure 2.7. (A) Schematic diagram of the Cas9 protein consisting of two nuclease domains: the RuvC and the HNH domain. (B) The sgRNA consists of the tracrRNA and crRNA bound together covalently. The crRNA is composed of a 20bp protospacer element and additional nucleotides that are complementary to the tracrRNA. The tracrRNA hybridizes to the crRNA and binds to the Cas9 protein, forming the CRISPR-Cas9/sgRNA complex to edit genome. (C) Molecular mechanism of the CRISPR-Cas 9 system, Cas9 protein cleaves 3–4 deoxynucleotides upstream of the PAM site [57]. The RuvC domain cuts the non-complementary DNA strand, and the HNH domain cuts the complementary DNA strand. (D) Crystal structure of the catalytically active *Streptococcus pyogenes* Cas9 (PDB: 5F9R) in complex with the sgRNA and the dsDNA primed for target cleavage. Visualized with CHIMERA (UCSF) [58].

The targetable sites of this RNA guided nuclease are limited by the requirements of the PAM sequence, which should be NGG or NAG and the 5' G deoxynucleotide. G shall be added to the 5' end of the sgRNA if the initiation start site does not start with a G. The crystal structure of *Streptococcus pyogenes* CRISPR–Cas9 in complex with the single gRNA and its target double-stranded sequence is shown in **(Figure 2.7 D)**.

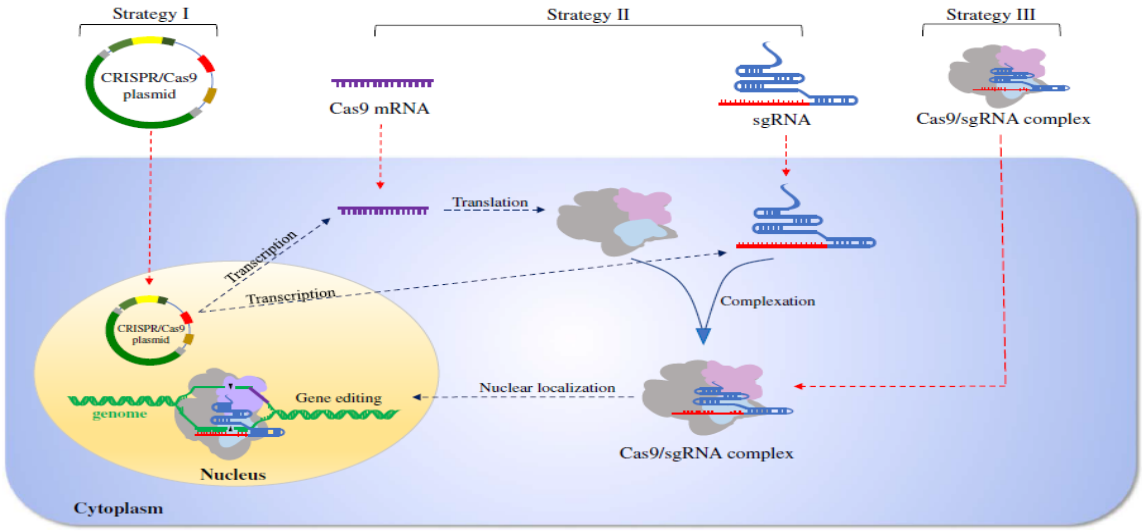


Figure 2.8. Various strategies to deliver the CRISPR-Cas9 system [57]: (i) as a plasmid-based CRISPR-Cas9 system encoding the Cas9 protein and sgRNA within the same vector, or (ii) as a mixture of Cas9 mRNA and sgRNA or (iii) direct delivery of the Cas9/sgRNA complex in the cells.

CRISPR-Cas9 can edit the genome by three different strategies (**Figure 2.8**). The most convenient strategy is using a CRISPR-Cas9 plasmid both expressing the Cas9 protein and transcribing the sgRNA [59]. Another approaches are delivering a mixture of the Cas9 mRNA [60] or Cas9 protein and the sgRNA [61]. The plasmid-based CRISPR/Cas9 system is a simple and convenient strategy to avoid the difficulty of the delivery of multiple components to the same cells and the plasmid-based CRISPR-Cas9 system is more stable than Cas9 mRNA with sgRNA. CRISPR-Cas9 is a promising tool in the treatment of genetic diseases by editing the mutated genes. RNA guided nucleases were already used to edit the genes in bacteria, eukaryota [59], zebrafish embryos [62], rats [63], non-human primates [60] and human pluripotent stem cells [64]. The current challenges focus on improving the specificity and delivery efficiency of the CRISPR-Cas9 system [57].

2.3.4. The NColE7 nuclease domain (structure, function, consequences of mutations and regulation)

Colicin E7 protein is a toxin produced naturally by *Escherichia coli* cells to protect themselves from the closely related bacteria, which contain the essential outer membrane receptors [65,66]. Colicin E7 consists of 567 amino acids including three functional domains: receptor binding (residues 3-308), membrane translocation (residues 288-446), and the toxic nuclease domain (residues 446-576). Under stress conditions, it binds to the receptor of the attacked cell and by the help of the translocation domain, the nuclease domain of the Colicin -

E7 (NColE7) enters the cytoplasm to kill the cell by non specific degradation of the chromosomal DNA. The host cell protects itself against the nuclease activity by co-expression of immunity protein (Im7), which blocks binding of NColE7 to the DNA [67].

NColE7 is a member of the HNH nuclease family [68]. The name of HNH originates from the three underlined amino acids residues in EXH<u>H</u>X_{14}<u>N</u>X_8<u>H</u>X_3H sequence. The HNH motif in NColE7 is the C-terminal catalytic part of the total 131 amino acids sequence. It binds to the minor groove of the DNA while the central helices of NColE7 bind to the major grove of the DNA **(Figure 2.9)**. The HNH motif forms a metal-binding site by three histidine side chains (H544, H569, and H573). Zn^{2+} is the divalent metal ion, which coordinates to these amino acids under physiological conditions in a tetrahedral coordination geometry. Water molecule occupies the fourth coordination site. The metal ion is essential for the nuclease activity of NColE7. Zn^{2+} coordinates to the oxygen of the scissile phosphodiester bond, replacing the water molecule. H545 activates the water molecule for the nucleophilic attack at the scissile phosphodiester bond [69].The DNA cleavage results in 5'-phosphate and 3'-OH ends [70].

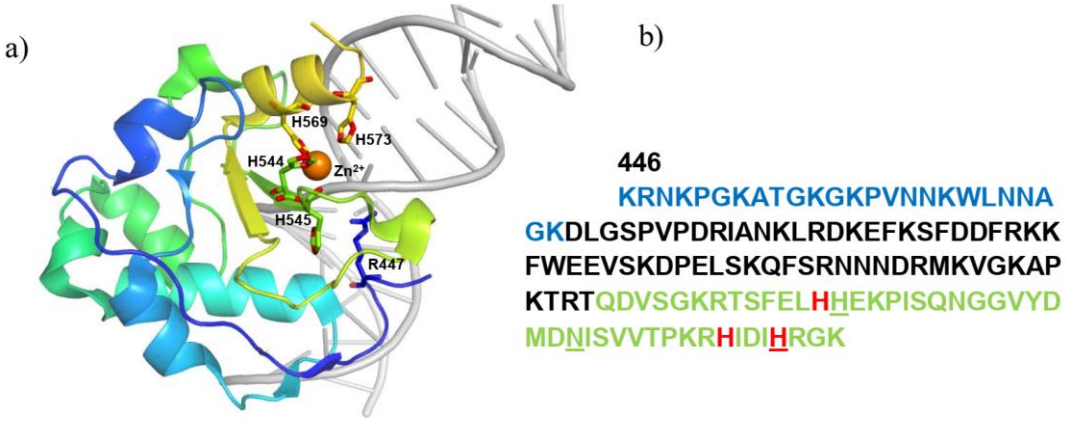


Figure 2.9. a, Crystal structure of NColE7 in complex with DNA (PDB: 3FBD). The orange sphere indicates the Zn^{2+}-ion, which was not included in this structure, and it was placed as a result of the alignment with the 1M08 structure (chain A). PyMol's rainbow coloring scheme was applied resulting in a color varied from blue (N-terminus) to the yellow (C-terminus). **b,** Sequence of NColE7, which is the C-terminal nuclease domain of the colicin E7 protein (446-576) The N-terminal loop is written in blue and the HNH motif in green except the metal binding residues in red. The underlined residues are the name-giving HNH residues [71].

The role of the N-terminus of NColE7 was studied in detail [72,73]. It activates the nuclease by its positively charged amino acid residues sterically close to the catalytic site. R447 is linked with Zn(II) ion through the phosphodiester group. R447 may also help the DNA

binding. Modifications or deletions in the N-terminal sequence affect the catalytic activity, and the W464A mutation destabilizes the NColE7 structure [74].

2.3.5. NColE7-based integrated zinc finger nucleases

Therapeutic applications of the ANs are still hampered by their off-target activity resulting in various levels of cytotoxicity [21,75]. One of the reasons might be the loss of the regulation of the catalytic activity, characteristic for the native enzymes. To overcome this hurdle, novel zinc finger ANs were designed in our laboratory, based on the NColE7 nuclease domain instead of the NFokI nuclease domain. Considering the properties of NColE7 with the emphasis of the need of the collaboration of its two termini in the catalytic action, the new ANs were expected to be more regulated.

NColE7 was divided into two parts: the C-terminal domain, which is the catalytic center and N-terminal domain, which serves as an activator. These two parts were fused to the opposite termini of a zinc finger protein as shown in **Figure 2.10**. In this kind of artificial metalonuclease, the HNH-motif at the C-terminus is expected to have nuclease activity only in the presence of the positively charged N-terminal domain offering a positive allosteric control [76].

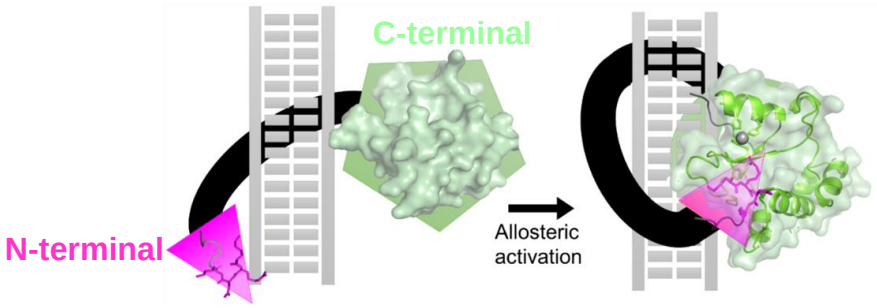


Figure 2.10. Schematic representation of the NColE7-based AN design strategy demonstrating the expected allosteric regulation. The inactivated C-terminal catalytic site is shown in green and the N-terminal activator is in pink. The specific DNA-binding domain (black) is a ZF protein that brings the above two parts together in space upon recognizing the target DNA sequence [77].

The new ANs consist of nuclease, DNA binding and regulator domains, rather than a simple chimera of two molecules in traditional ZFNs. Various series of NX-ZF-CY models were designed in which X is number of N-terminal amino acid residues and Y is the number of C-terminal amino acid residues of NColE7, ZF is a protein array of three fingers, the same published by Kim and Berg [78] (PDB code 1MEY). NColE7 and ZF domains were linked together by computer-designed peptide linkers. Three models of new ZFNs (N4–ZF–C45, N4–ZF–C105 and N46–ZF–C45 were established (**Figure 2.11**).

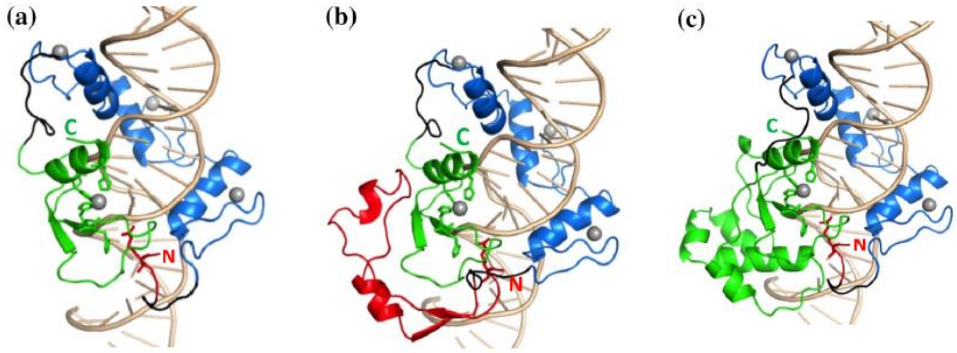


Figure 2.11. The designed ZFN structures. The N-terminal part of NColE7 is red, the C-terminal part is green, the ZF is blue, and the linkers are black. The grey spheres indicate Zn^{2+} ions. (a) N4–ZF–C45, (b) N46–ZF–C45, (c) N4–ZF–C105. The C- and N-termini of the ZFNs are marked by 'C' and 'N', respectively [79].

The three models have the allosteric regulator residues at their N-termini and the HNH motif at their C-termini such as the NColE7. These models were experimentally studied and have shown to exert slow but specific DNA cleavage. The reactivity of the new artificial nucleases followed the next order: N4–ZF–C105>N46–ZF–C45>N4–ZF–C45. These ZFN variants may have different activation mechanisms. The fusion of the C-terminal catalytic domain to the ZF domain could obviously increase its local abundance at the DNA, and the N terminal domain, fused to N-terminus of ZF domain acts as an activator for the catalytic domain. However, this kind of regulation of the catalytic domain was not effective in these nucleases. The expression and localization of NColE7-based ZFNs were studied in mammalian cells. The enzyme was detected in the cell nucleus, which is promising for future therapeutic applications in the human gene therapy [77], but further optimization is needed.

2.4. The delivery of programmable nucleases

The success of the selected genome editing technique largely depends on the effectivity of the delivery of the nuclease into the specific target cells. It is a challenge to safely and effectively deliver the genome editing tool into the mammalian cells. We briefly summarize the most common delivery strategies. The delivery systems can be classified into non-viral and viral vehicles as shown in **Figure 2.12**.

The non-viral methods include physical and chemical delivery. The most common physical methods are electroporation, microinjection, induced transduction, mechanical cell deformation, and hydrodynamic injection.

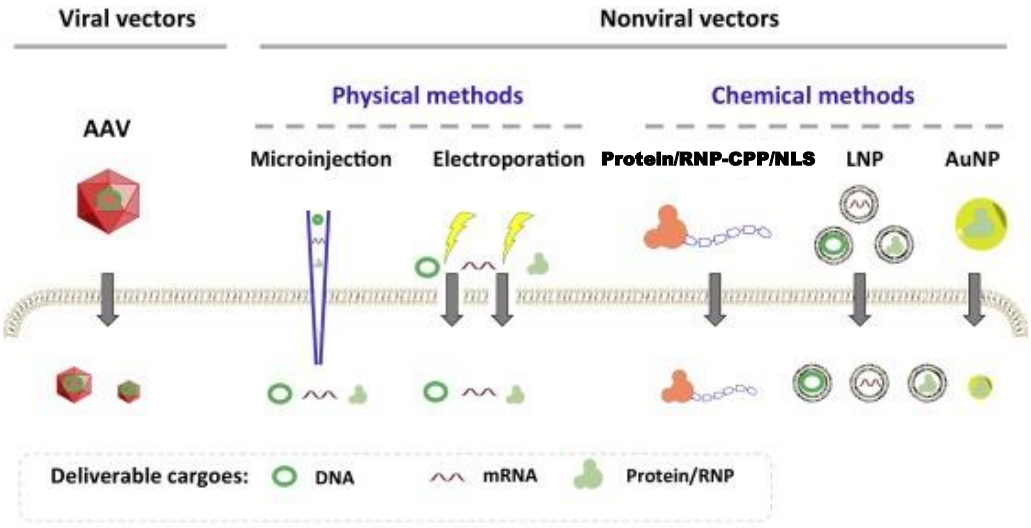


Figure 2.12. A summary of the types of delivery routes. The ANs may be delivered as a DNA or mRNA molecule encoding for the protein/RNA, or it may be delivered as a functional protein/ribonucleoprotein (RNP). Regardless of cargo format, the largest challenge lies in delivering the cargo across the cell membrane. A variety of viral and nonviral methods have been derived to achieve successful delivery across the cell membrane. AAV: adeno-associated virus; AuNP: gold nanoparticle; CPP: cell-penetrating peptide; LNP: lipid nanoparticle; NLS: nuclear localization sequence [80].

For chemical delivery lipid nanoparticles, polymer nanoparticles, cell penetrating peptides, DNA nanostructures, gold nanoparticles and further newly developed vehicles can be applied. Viral delivery systems include adeno-associated virus (AAV) and lentivirus (LV) vehicles.

2.4.1. Non-viral delivery methods

Non-viral delivery of programmable artificial nucleases has shown high efficiency because of its safety, large loading capacity and low cost production. Non-viral delivery systems have shown low level of the off-target effects [81]. These advantages make them attractive for delivery applications. The physical methods are widely applied in the laboratory but less used for the *in vivo* applications because they are based on physical modifications of the cells. The chemical vehicles became more promising for the *in vivo* applications.

2.4.1.1. Physical delivery methods

Physical delivery methods depend on making a simple pierce in the cell membrane for the direct injection of the cargo into the target cells.

a) Electroporation. Electroporation is a process of introducing nucleic acids or proteins into bacteria or other cells using a pulse of electricity to briefly open the pores in the cell membranes. The disadvantages of electroporation are that plasmid DNA is only inserted into around 0.01%

of the target cells and it also leads to high level of cell death. Nevertheless, electroporation has newly been used to deliver CRISPR-Cas9 based plasmid to cancer cells, embryonic stem cells and CD34+ stem cells [82,83]. Editing efficiencies of 87% and 94% were achieved in induced pluripotent stem cells and Jurkat T cells, respectively via the electroporation of RNPs, but lower efficiency in induced pluripotent stem cells (20% and 32%) and Jurkat T cells (63% and 42%) was achieved using plasmid-based CRISPR-Cas9 system and the Cas9 mRNA/sgRNA, respectively [84].

b) Microinjection. Microinjection is a technique to deliver the DNA or protein into a single cell by direct injection using glass micropipette. Microinjection of CRISPR-Cas9 was used to edit genes in different types of cells [85,86]. Although microinjection is very simple and accurate it requires a high level of manual skills not to induce cell damage. Another disadvantage is that only a single cell can be targeted in each injection, so this method is only suitable for a limited number of cells.

c) Induced transduction by osmocytosis and propanebetaine (iTOP). Proteins and other compounds can be delivered into the cells by iTOP method by using hyperosmolar buffer consisting of sodium chloride and propanebetaine (a transduction compound). iTOP method is effective for intracellular delivery of the Cas9 protein and sgRNAs separately, or direct delivery of RNPs. But iTOP has limited use because it needs high salt concentration for dissolving the proteins, which is not suitable for *in vivo* applications [57].

d) Mechanical cell deformation. Mechanical deformation induces the disruption of the cellular membrane to uptake the cargos through the passive diffusion. Mechanical cell deformation is an effective delivery with low death rate [87]. A microfluidic device with optimized physical shrinkage was used successfully to deliver a plasmid-based CRISPR-Cas9 system *in vitro* and achieved 90% knockout efficiency of EGF in MDA-MB-231 (breast cancer cells) [88], but it is limited to *in vitro* use only.

e) Hydrodynamic injection. Hydrodynamic pressure is induced by rapid and sudden injection of a high volume of nucleic acid solution. This pressure leads to the temporary formation of pores in the membrane allowing the entry of the nucleic acids into the cells [89]. This method is simple and effective [90].

2.4.1.2. Chemical delivery methods

Chemical systems are in general nanomeric complexes, more widely used than physical methods. They include compaction of negatively charged nucleic acid by mostly cationic

particles. These nanomeric complexes can enter the cells easily and release the nucleic acids [91]. Chemical delivery systems have low toxicity and antigenicity compared to viral vectors.

a) Lipid nanoparticles. Lipid nanoparticles are common and effective delivery systems for the nucleic acids. The negative nucleic acids are encapsulated or complexed with the positively charged lipid nanoparticles through electrostatic attractions. The lipid vector protects the nucleic acids from the nuclease degradation and immunological response. It can be modified with polyethylene glycol (PEG) to increase its lifetime inside the blood circulation system and can be modified also by specific antigen to target the specific tissue.

Cationic systems consist of either a single synthetic cationic amphiphile (cytofectin), such as DOTAP, DOTMA, DOSPA, DMRIE and DC-cholesterol or a combination of a cationic amphiphile and a neutral lipid, such as DOPE, DSPC and cholesterol (**Figure 2.13**). These neutral helper lipids enhance the transfection efficiency through destabilization of the endosomal membrane to facilitate lipid exchange and membrane fusion between lipoplexes and endosomal membrane leading to higher gene expression [92,93].

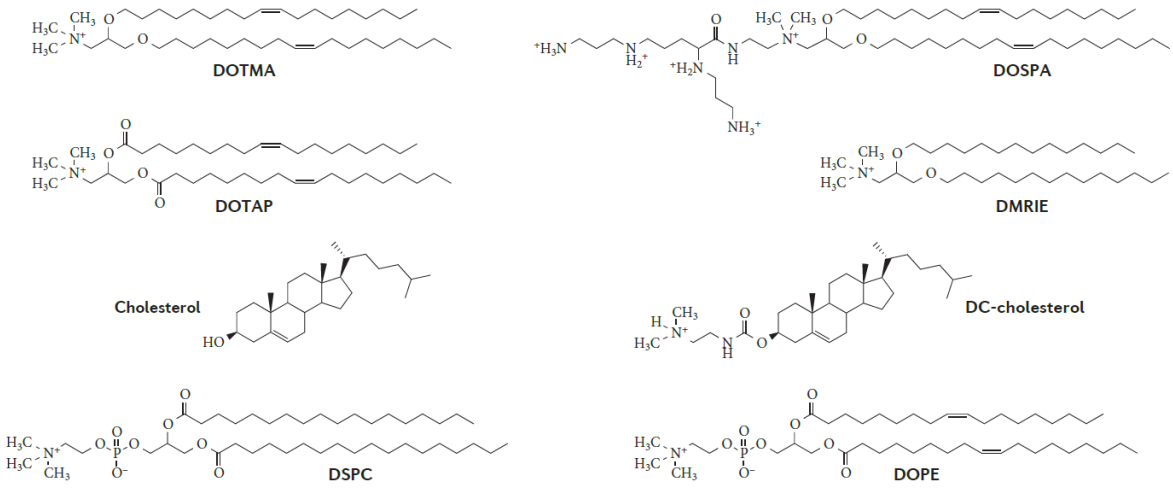


Figure 2.13. Chemical structures of cationic and neutral lipids. Liposomal formulations used for DNA delivery typically include a mixture of a neutral lipid and a cationic lipid. Cationic lipids (such as DOTMA, DOSPA, DOTAP, DMRIE and DC-cholesterol) have an active role in DNA binding and transfection. They are characterized structurally by a cationic head group, a hydrophobic tail and a linker region. Neutral lipids (such as the phospholipids DSPC and DOPE, and the membrane component cholesterol) function as 'helper lipids' to further enhance nanoparticle stability and overall transfection efficacy.

The cationic lipid nanoparticles were utilized as *in vitro* delivery vectors for ZFNs [94] and have been used in several therapeutic applications of the CRISPR/Cas9 [95]. Lipofectamine is the most common commercial optimized delivery system. Genome editing efficiency using lipid nanoparticles can more than 70% in HEK cells. These delivery systems have some

drawbacks such as the poor stability and rapid clearance, as well as inflammatory or anti-inflammatory responses [96].

b) Polymer nanoparticles

Cationic polymers have high DNA condensing ability. There are several types of cationic polymers such as poly(ethylenimine) (PEI), poly(L-lysine) (PLL), poly[2-(dimethylamino) ethyl methacrylate] (PDMAEMA), and polyamidoamine (PAMAM) dendrimers (**Figure 2.14**).

PEI is one of the leading reagents because of the high concentration of positively charged ammonium groups interacting strongly with the negatively charged DNA. It forms stable and condensed DNA/PEI complexes and protects the genetic material from the enzymatic degradation. PEI is synthesized in linear and branched form. There is a relationship between the length of the polymer, gene delivery efficiency, and toxicity. As the length of the polymer increases, so its efficiency increases because the net positive charge of it also increases and it is able to bind DNA strongly [97,98]. Recently branched PEI was used to deliver plasmid-based CRISPR-Cas9 [99].

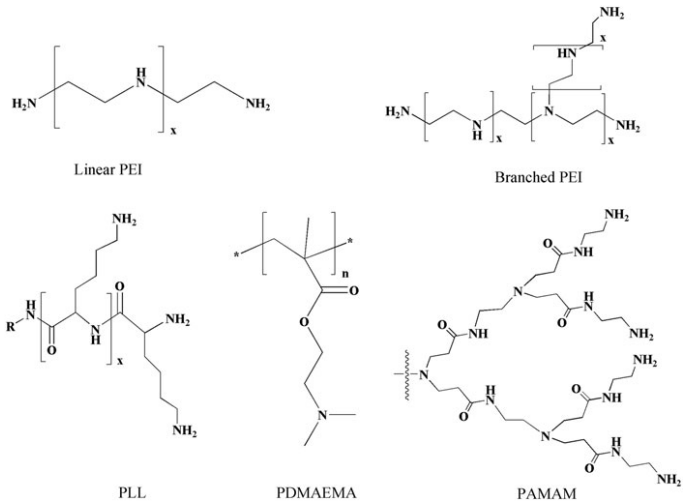


Figure 2.14. Chemical structures of polymeric vectors. PEI: poly(ethylenimine); PLL: poly(L-lysine); PDMAEMA: poly[2-(dimethylamino) ethyl methacrylate]; PAMAM: polyamidoamine [100].

Modification of PEI by polymers and/or ligands has recently been investigated to overcome the increasing cytotoxicity with increasing molecular weight and aggregation in the case of high ionic strength [101,102]. Changing its physicochemical properties is aimed to enhance biocompatibility and transfection efficiency [103,104].

c) Cell-penetrating peptide (CPP). CPPs typically consist of 5–30 amino acids including several arginine and lysine residues, which are positively charged at the phsiological pH. They

can cross the cellular membrane by translocation or penetration [105]. Cas9 protein was covalently attached to a CPP and the sgRNA conjugated with another CPP by electrostatic interactions for *in vitro* expreiments. The following disadvantages can limit the applications of CPPs: (i) the conjugated protein can be degraded by the proteases; (ii) CPP cannot target specific cells; (iii) this method needs chemical conjugation [106].

d) DNA nanostructure. DNA nanostructures have several applications in medicine, such as imaging and drug delivery. DNA nanostructure is a sphere like DNA structure, formed by base pairing of small DNA sequences. It needs a high amount of DNA. DNA nanostructures called nanoclew are being developed by rolling circle replication (RCR). PEI can be added to this nanoclew to be positively charged to improve the endosomal escape and cellular uptake. Nanoclew has been used in 2015 to deliver the Cas9 protein and sgRNA complex and it achieved a mutation efficiency around 36% [107]. The main disadvantages of the nanoclews are their complicated assembly and the possibility of immunological response.

e) Gold nanoparticles (AuNP). AuNP has been used as a new delivery vector for Cas9/sgRNA ribonucleoprotein complexes (RNPs), achieving around 90% transfection and 30% genome editing efficiency [108]. It is a very promising AN delivery tool, but it is limited for the *in vivo* application due to the toxicity at applied concentrations.

2.4.2. Viral delivery methods

The viral delivery vectors can be used to deliver nucleic acids for gene therapy purposes with high delivery efficacy but they have many limitations such as carcinogenesis, immunogenicity, limited DNA packaging size and difficulty in production [109].

a) Adeno-associated virus (AAV). The genome of AAV is composed of a single DNA strand about 4.7 kb. It is widely used for the gene therapy because it does not cause diseases in the humans, has moderate immunogenicity and has a wide range of serotypes so it can infect various cells [110]. The first promising gene therapy trial with AAV was in 2012 for treatment of lipase deficiency [111]. AAV was used to deliver ZFNs. The transfection efficiency of AAV/ZFN particles was investigated *in vitro* and *in vivo* [112].

It has been used also for delivery of CRISPR/Cas9. The sequence encoding the *Streptococcus pyogenes* Cas9 protein and sgRNA is approximately 4.2 kb.The challenge for AAVmediated CRISPR-Cas9 delivery is the packaging limit of AAV (approximately 4.5 kb) [113]. For this reason, CRISPR-Cas9 system was delivered by alternative ways: (i) AAV was used to deliver the sgRNA only into cells modified to express Cas9 protein by using the

microinjection method [114]; (ii) another method was the packaging of the Cas9 and sgRNA separately into two AAV particles but this is quite complex procedure [115].

b) Lentivirus (LV). Lentivirus is a widely used vector for gene therapy due to its high infection efficiency. It can infect even the non-dividing cells, has mild immunogenicity and can be used both for the *in vitro* and *in vivo* applications. The size of LV is approximately 80-100 nm, being better for large insertions and suitable for the CRISPR/Cas9 delivery [110]. An advanced type of LV called integrase-defective LV was utilized in ZFN-based genome editing in human stem cells [116].

Viral particles with the packaged Cas9 and sgRNA were used to infect HEK293T cells. One plasmid expressed the proteins and enzymes which are necessary for viral particle formation and another plasmid encoded the CRISPR/Cas9 system [117]. Promising results were obtained in the chronic hepatitis B virus (HBV) treatment by the lentivirus-based CRISPR/Cas9 system [118]. It was utilized also for development of a mouse model with myeloid leukemia, targeting genome regions of herpes simplex virus-1 and to mutate genes in primary immune cells of mouse [119].

3. AIMS AND OBJECTIVES

The aim of the research conducted in the Bioinorganic Chemistry Research Group of the Chemistry Institute at University of Szeged is to develop an effective artificial nuclease as a promising molecular tool for gene therapy. I have joined this research team in 2016. At that time the research has focused on the design of a novel zinc finger artificial nuclease based on the NColE7 nuclease domain instead of the NFokI nuclease domain. Therefore, my work is aimed at optimization of the newly designed ZFNs for targeting oncogenes in cancer cells, as well as at comparison of these artificial nucleases with the currently widely used CRISPR/Cas9 system. Furthermore, we decided to develop a novel drug delivery system to increase the efficiency of the designed artificial nucleases by their delivery into the target cells.

The summary of my aims is shown in **Figure 3.1** which explains my interdisciplinary research including chemistry and biology. According to this scheme, my research consists of three main steps: (i) optimizing of a promising artificial nuclease, (ii) development of an effective drug delivery system and (iii) studying of the behavior of these systems in cancer cells.

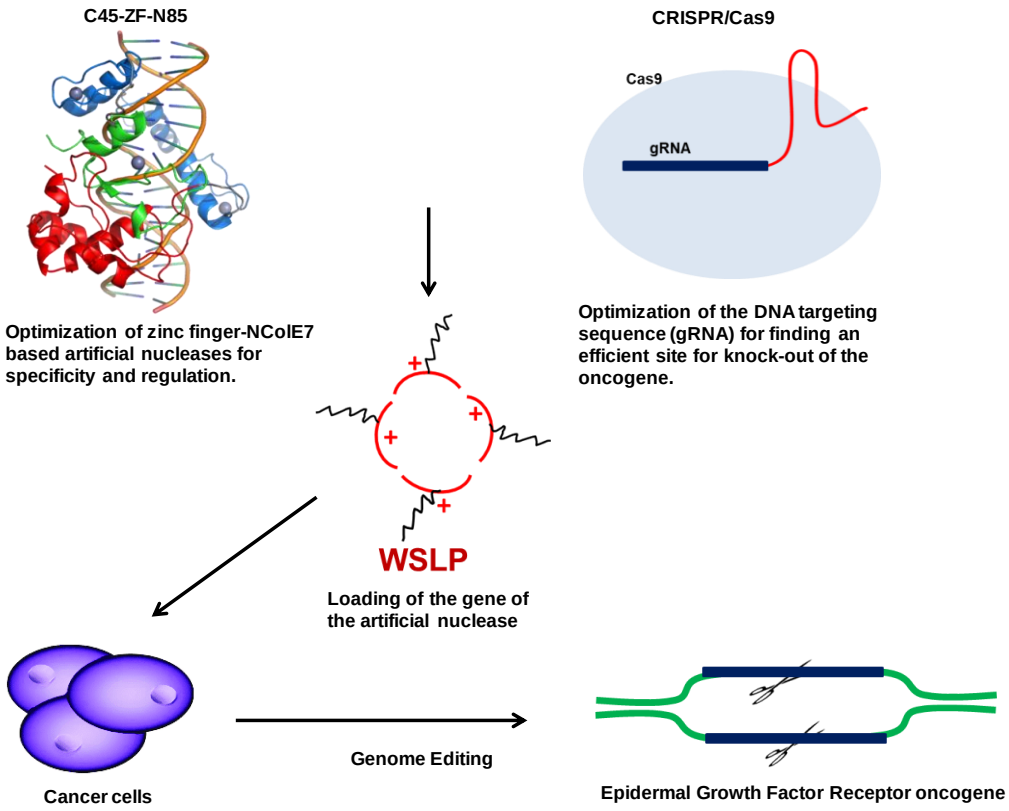


Figure 3.1. The schematic demonstration of the aims of my PhD research project.

During my PhD studies I focus on the following objectives:

1- Construction of a new protein expression system for effective purification of artificial nucleases

The production of proteins in high yield and purity is usually difficult and requires multistep chromatographic methods often leading to a significant loss of target proteins. In our laboratory, the synthesis of proteins with precisely determined sequences is very important because the computer design of artificial nucleases requires the reproduction of the protein sequence. Therefore, it is a topical issue to develop purification procedures for the designed proteins with precisely determined sequences. A strategy was elaborated by means of which the designed artificial nucleases can expressed and purified by immobilized metal ion affinity chromatography, but without any additional remaining amino acids at the termini of the protein after removal of the hexahistidine affinity tag. This system can also be applied for the expression of purification of native proteins and can be used with various C-terminal affinity tags.

2- Improving of the newly designed NColE7-based zinc finger nuclease

There is a chance that the metal ion binding hexahistidine affinity tag may influence and thus, modulate the catalytic activity of the NColE7 metalloenzyme, used as the catalytic center of the new artificial nuclease. Therefore, we decided to explore this effect for development of a novel regulation mechanism. The regulation may also be achieved in the novel zinc finger nucleases by inactivating the nuclease domain itself in the absence of a specific DNA target sequence. To better understand the regulation mechanisms, and to further optimize the artificial nucleases, the expression and purification of the promising AN variant (C45-ZF-N85), its mutants, as well as 6×His tagged NColE7 mutants was foreseen followed by the study of their solution structure by circular dichroism spectroscopy and their specificity and/or regulation by *in vitro* catalytic experiments.

3- Design of CRISPR/Cas9 artificial nucleases for oncogene targeting

I planned to apply the CRISPR/Cas9 system to target epidermal growth factor receptor (EGFR) oncogene in A549 human lung cancer cells. Designing various recognition sequences within the gene would allow for detection of the most efficient artificial nuclease target sites, which can also be targeted by the new NColE7-based ANs upon redesign of the ZF domain. This would also allow for comparison of the different ANs.

4- Development of a novel drug delivery system for artificial nucleases

The success of the selected genome editing technique is influenced by the effectivity of the delivery system used to carry the nuclease to the specific target cells to induce its therapeutic effect. Improvement of the delivery and specificity of the ZFNs and CRISPR-Cas9 system could help to overcome current challenges, and promote the artificial nuclease technology to be used therapeutically. Chemical delivery methods are being considered to be the most promising. For this purpose, we decided to synthesize a water-soluble lipopolymer from a high molecular weight poly(ethylenimine) by cholesterylation, to characterize its DNA loading properties and study its gene delivery i.e., transfection efficiency in cancer cells.

4. METHODS

4.1. Recombinant protein production

In this work, proteins were produced from different recombinant DNA elements for various experiments. Proteins produced with the recombinant DNA technology are the so-called recombinant proteins. The first step includes the genetic recombination to bring together genetic material from multiple sources thereby, creating recombinant DNA sequences that are naturally not found in the genome, and the DNA cloning. Then, the production of recombinant proteins follows, including the protein expression from the recombinant DNA in cells, the protein purification by a specific method and quality control of the purified proteins.

4.1.1. Recombinant DNA technology

Recombinant DNA technology allows DNA of interest to be produced through artificial means. It also involves specific mutations of the DNA sequence by polymerase chain reaction (PCR), insertion of the mutated DNA sequence into a specific carrier plasmid DNA, and producing copies of it within the bacterial cells as shown in **Figure 4.1**. Each step is explained below.

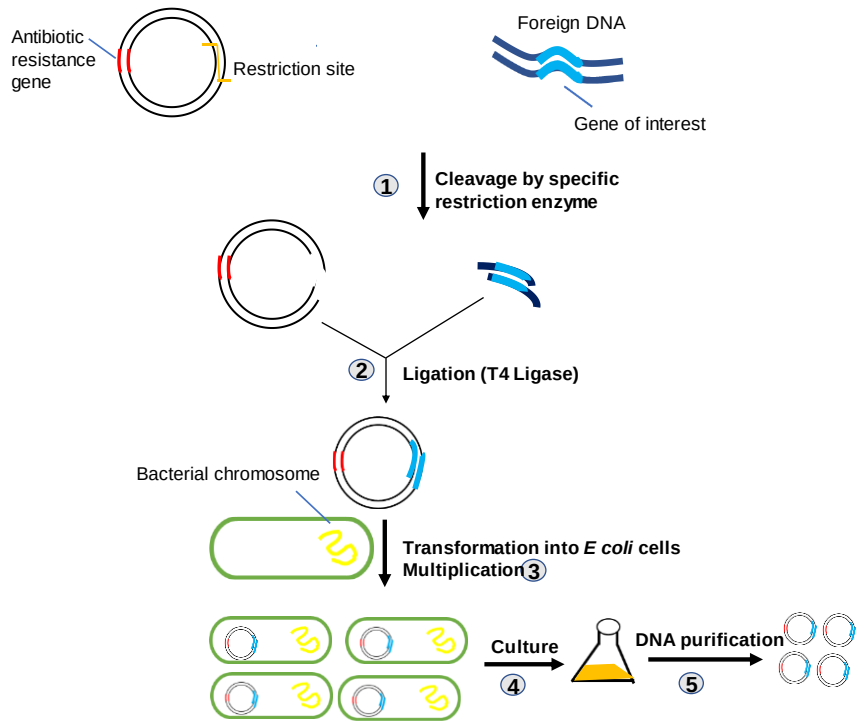


Figure 4.1. Schematic illustration of recombinant DNA production.

4.1.1.1. Polymerase chain reaction (PCR)

For the production of specific proteins or their modified versions, first their genes are obtained and amplified by polymerase chain reaction (PCR). This method is used also to insert restriction sites into the target DNA sequence, so that it can be digested by a specific restriction endonuclease There are five main ingredients of the PCR reaction mixture:

- DNA template.
- Two (forward and reverse) primers, which are short oligonucleotides essential for initiating the polymerase enzyme to start the amplification step, by hybridizing to the single-strand template.
- The mixture of the four deoxynucleoside triphosphates (dNTPs), which are the building blocks of DNA (dATP, dTTP, dGTP, and dCTP)
- Heat-stable DNA polymerase enzyme (usually Taq polymerase from *Thermus aquaticus*).
- A buffer, that also contains Mg^{2+} to keep the reaction in the optimal condition.

PCR involves periodic heating and cooling of the reaction mixture by a thermal cycler instrument. Three main steps of a cycle are: denaturation, annealing, and extension (**Figure 4.2)**. First, the PCR reaction mixture is heated to 94-95°C for 15-30 seconds to denature the dsDNA template into two ssDNA strands, which act as templates for the formation of new DNA strands. Then the temperature is decreased to a suitable value for the primers to bind to their complementary sequences on the single-strand DNA – usually between 50 and 60°C for 10-30 seconds. This temperature is adjusted considering the melting temperature of the primers used.

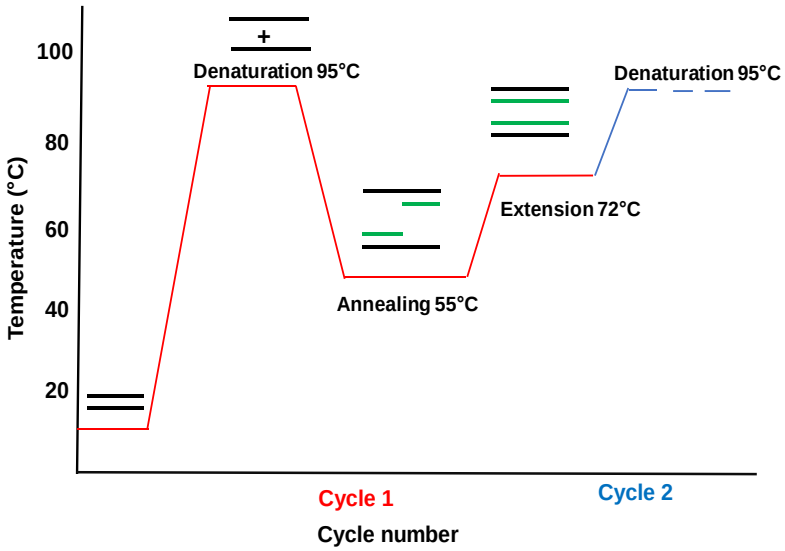


Figure 4.2. Principle and major steps of the Polymerase Chain Reaction (PCR).

The forward and reverse primers allow the two complementary single strands to be extended from their 5' end toward the 3' end, when the temperature is increased again to 72°C, suitable for Taq DNA polymerase enzyme function. So it can attach to the primer and add the nucleotides to the growing strand at its 3' end, resulting in a new double-strand DNA molecule. The duration of this step depends on the length of the DNA to be amplified. Copying of a 1 kbp DNA takes ~1 minute. These three steps of the thermal cycling are repeated ~30 times. Since the new DNA molecules behave as templates in the subsequent cycles, the amplification is exponential. The PCR product resulted from this reaction is checked by the gel electrophoresis. The successfully amplified PCR product can be used for cloning into a carrier circular DNA molecule.

4.1.1.2. Digestion and ligation processes

Specific restriction sites are designed within the primer sequences and are added to the termini of the gene of the target protein through the PCR. The resulted gene can be inserted into a specific carrier, so-called plasmid DNA (pDNA). The plasmid is a circular double-strand DNA molecule, originally isolated from the bacterial cells. It consists of a few thousands of DNA base pairs. These plasmids have been modified for the purpose of molecular cloning in the laboratory. Both the target gene and the plasmid are digested by a restriction endonuclease at a specific sequence to get double-strand DNAs with sticky ends (short, single-strand overhangs). Using the same restriction enzyme these sticky ends will be complementary to each other.

The cleaved target gene is ligated into the linearized plasmid by the ligase enzyme, which covalently joins the phosphate backbone of DNAs with blunt or compatible sticky ends. This ligation process results in a recombinant DNA molecule containing the gene of the target protein. This ligation reaction needs the optimization of the temperature for bringing the DNA ends together, and then ligate them. We applied incubation for 1 hr at 16°C and then lowering the temperature to 4°C overnight. The plasmid and insert genes should have an optimal molar ratio (1/6) in the reaction for effective ligation.

4.1.1.3. Transformation of bacteria

The ligated plasmid is introduced into harmless *E. coli* bacteria used in labs, in a process so-called transformation. This is usually done by a heat shock treatment. The plasmid contains an antibiotic resistance gene that enables the bacterial cells to survive in the presence of a specific antibiotic, so only the bacteria containing the plasmid can grow without any

contamination. A bacterial cell will grow into a small colony carrying a single type of plasmid. Not all the colonies contain the modified plasmid, because in some cases the ligation fails. So the plasmid DNA is purified from several colonies. This involves the culturing of the bacteria in a suitable liquid medium, the bacterial cell lysis, DNA extraction and precipitation by ethanol. The plasmids are checked for the successful ligation by different methods like PCR, agarose gel electrophoresis, restriction enzyme digestion, and DNA sequencing.

4.1.2. Protein expression and purification

Several types of expression vectors (plasmids) are available for production of recombinant proteins. The pET-21a commercial plasmid (**Figure 4.3**) and its modified versions were used in this study.

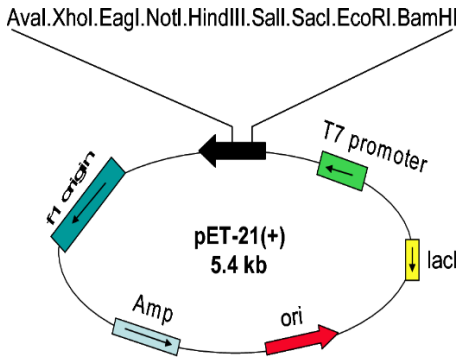


Figure 4.3. The schematic representation of the pET-21a plasmid. The cloning site contains unique restriction sites. The ampicillin resistance gene allows for the selection of bacteria against the antibiotic.

The pET-21a plasmid contains a lac (lactose) operon which is one of the most common systems used for the protein expression control, see the map of pET-21a plasmid in **Appendix A.1**. The lac repressor protein (LacI) senses the presence of lactose, which is the inducer of lactose operon as it triggers the transcription of the lac operon to express the protein in BL21(DE3) *E. coli* strain. Isopropyl β-D-1-thiogalactopyranoside (IPTG) is a structural mimic of lactose used as an inducer for the protein expression. It binds to the lac repressor then activates the transcription of the gene in the lac operon. IPTG reagent is used instead of the lactose because the lactose is degraded by β-galactosidase enzyme during the bacterial metabolic pathway but the IPTG is not, since it contains sulfur atom preventing the enzymatic hydrolysis. Therefore, the concentration of IPTG remains constant and the protein expression will not be decreased or inhibited resulting in overexpression of the target protein. IPTG is added when the bacterial growth in the LB medium causes an optical density (OD) of ~ 0.6–0.8.

The pET-21a plasmid contains DNA sequence coding for six histidine residues fused to the C-terminus of the gene of the recombinant protein, resulting in the expression of the protein with a poly-His tag. These His-tagged proteins are purified by immobilized metal ion affinity chromatography (IMAC). Both metal affinity and cation exchange chromatography were applied in this work.

4.1.2.1. Immobilized metal affinity chromatography (IMAC)

IMAC is a rapid purification method of the proteins with oligohistidine affinity tags. The desired metal ion used in the IMAC can be immobilized by different supports, such as magnetic particles or beaded agarose derivatized by chelating groups. These supports are used for binding and purification of proteins. Nitrilotriacetic acid (NTA) and iminodiacetic acid (IDA) are common chelators used as ligands. NTA-agarose resin loaded with the desired metal ion (Ni^{2+}) was used in this work. The resin is packed into a suitable column and serves as the stationary phase in the fast protein liquid chromatography (FPLC), while the resin can be used for batch purification, as well. The oligohistidine tags can bind to the IMAC resin in neutral buffer conditions. The binding and washing buffers consist of Tris-HCl with a low concentration of imidazole (competing ligand) to prevent the non-specific protein binding. The His-tagged protein is eluted from the IMAC column by using high imidazole concentration.

In this work, IMAC Sepharose 6 Fast Flow column was used. It consists of highly cross-linked 6% spherical agarose and has a high protein binding capacity. The binding capacity is metal ion- and protein dependent. The medium is easy to pack and use, and its high flow properties make it excellent for scaling up.

4.1.2.2. Cation exchange chromatography

Cation exchange chromatography is used to separate molecules having net positive surface charges on a negatively charged ion exchange resin. The proteins and enzymes interacting with DNA, such as the zinc finger nucleases designed in our laboratory are positively charged. Thus, cation exchange chromatography was used in this work. All ion exchange chromatographies rely on electrostatic interactions between the resin functional groups and proteins of interest. The conditions of cation exchange chromatography should be adjusted to best suit the protein of interest, and the ion exchange resin chosen. The pH and ionic strength greatly affect protein binding to the resin.

To perform the protein purification by this method, first the column is equilibrated with a suitable buffer. After the protein sample is loaded onto the column, it is washed to remove the

unbound protein molecules and impurities. Finally, the target protein is eluted from the column by gradient elution changing the buffer pH or the ionic strength. The gradient elution is optimized for the best elution conditions. Once the elution profile of the protein of interest has been established and it is known at what ionic strength and pH a protein elutes, a step elution can be used to speed the purification process. After all remaining protein has been eluted from the resin, the column is equilibrated in low ionic strength buffer.

In this work, a Sepharose SP Fast Flow column was used. In the Fast Flow ion exchanger columns 90 µm agarose beads serve as a matrix, that is highly crosslinked (6%) providing high physical and chemical stability and can be used with high flow rates in a wide pH-range.

4.2. Characterization of proteins

4.2.1. SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is widely used for the separation of proteins based on their molecular weights. The polyacrylamide gel consists of acrylamide and N,N'-methylenebisacrylamide which form a highly crosslinked matrix. The gel polymerization is initiated by a free radical mechanism: sulfate radicals are provided by ammonium persulfate. N,N,N′,N′-Tetramethylethylenediamine (TEMED) stabilizes the radicals for the polymerization reaction.

The protein samples are denatured before loading to the gel by a buffer including 2-mercaptoethanol (ME) or dithiothreitol (DTT) and SDS and heating this mixture at temperatures > 90°C for a few minutes. ME or DTT reduces the disulfide bridges that might be present between the cysteine residues. SDS denatures and linearizes the protein and provides a net negative charge proportional to the M_w by adsorbing to the hydrophobic parts of the protein. Specific buffers are used to conduct the electric current between the cathode and the anode through the gel. The denatured samples are loaded onto the gel, which is placed in electrophoresis buffer. A voltage around ~100 V is applied. The proteins sorted by the size are visualized by staining with Coomassie Brilliant Blue dye, and destaining by using a solution composed of methanol and glacial acetic acid. The molecular weight and the concentration of the protein band can be estimated by using a mixture of a known size proteins as a marker.

4.2.2. Mass spectrometry of metalloproteins

Mass spectrometry (MS) is an analytical technique that reveals the molecular mass by ionizing the chemical compounds and sorting them based on their mass-to-charge ratio (m/z). The so-called soft ionization techniques like ESI (electrospray ionization) and MALDI (matrix-

assisted laser desorption/ionization) are used for the study of biological macromolecules [120]. The intact molecular ions are generated in the ionization chamber and transferred to the mass analyzer, which sorts the ions according to their m/z value. The separated ions are then passed to the detector systems to measure their relative amounts, and the results are shown as a mass spectrum [121].

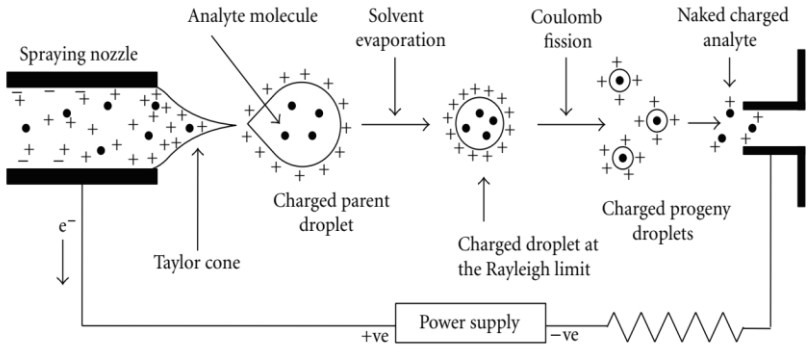


Figure 4.4. Schematic representation of the electrospray ionization process.

In an ESI ion source, a high voltage is applied to a steel capillary to generate charged droplets containing the analyte molecules. When solvent evaporates, charge accumulates until the Rayleigh limit is reached and the droplet undergoes Coulomb fission into smaller droplets. This process continues until the naked charged analyte ions enter the mass spectrometer (**Figure 4.4.**). The actual voltage required depends on the needle's inner diameter (i.d.) as well as the solvents which make up the solution. Typical nanospray instruments use fused silica or glass capillaries with orifice inner diameter < 10 µm from which solution normally flows at rates < 100 nL/min. Micro-electrospray devices use larger capillaries (10 µm < i.d. < 100 µm) with slightly higher flow rates (100 nL/min < flow rate < 500 nL/min).

The ESI process produces multiply charged ions of the same protein molecule due to protonation of the basic sites (in positive ion-mode) or deprotonation of acidic sites (in negative ion-mode) denoted as $(M+zH)^{Z+}$and $(M–zH)^{Z–}$, respectively [122]. An example of a multiply charged spectrum is given in **Figure 4.5.**

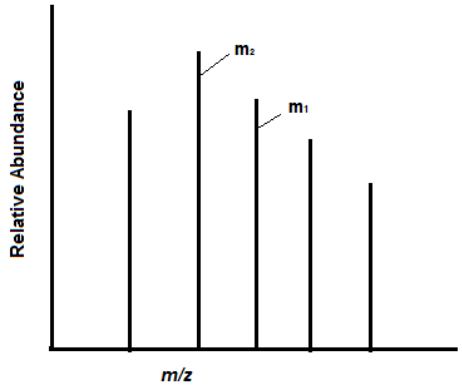


Figure 4.5. A hypothetical multiply charged spectrum.

From this spectrum the molecular mass of the analyte molecule can be mathematically determined using a simple algorithm for two adjacent peaks differing by one charge state:

$$m_1 = \frac{M - z_1 M_{H^+}}{z_1} \quad (4.1)$$

$$m_2 = \frac{M - z_2 M_{H^+}}{z_2} = \frac{M - (z_1 + 1) M_{H^+}}{z_1 + 1} \quad (4.2)$$

where m_1 and m_2 are m/z values of two adjacent peaks in the spectrum with the charge states $z_2 = z_1+1$. The molecular mass of an analyte is M and M_{H^+} is the mass of a proton. Solving the **Equations 4.1** and **4.2** for z_1 yields **Equation 4.3** for negative ion mode:

$$z_1 = \frac{m_{2+} M_{H^+}}{m_2 - m_1} \quad (4.3)$$

Substitution of this value into equations 4.1 or 4.2 provides the value for the molecular mass of an analyte, M. This process can then be repeated for different pairs of adjacent charge states from the spectrum, and hereafter the molecular mass is determined by averaging the obtained values. Using this algorithm the m/z for the multiply charged ion is converted to the mass of the zero-charged molecule [123]. This common algorithm assumes that all charge present on an ion is due to an excess of protons in positive ion-mode or a deficit of protons in a negative ion-mode. This assumption is true for the apo-proteins while for the metalloproteins additional considerations need to be taken into account [124]. Metalloproteins are interesting targets for mass spectrometric analysis because the stoichiometry and oxidation state of the metal or metal-cluster can be obtained, even though the information that is sought are determined by the ability of protein molecules to form multiple-charged ions.

In this study, intact protein analysis was performed by mass spectrometry (MS) on an Orbitrap EliteTM (Thermo Scientific) Hybrid Ion Trap-Orbitrap Mass Spectrometer coupled with a TriVersa NanoMate (Advion) chip-based electrospray ion (ESI) source. Prior to ESI-MS measurements, we performed buffer exchange on the protein to 10 mM NH_4HCO_3 pH 8.0 buffer

using Amicon Ultra 0.5 mL centrifugal filters with 3 kDa molecular weight cut-off. All the masses were measured in the Orbitrap mass analyzer in positive ion mode with 120 k resolution (R = 120,000 at 400 m/z). Protein intact masses (single protonated forms) were determined by deconvolution using the Freestyle 1.6 software tool (Thermo Scientific).

4.2.3. Study of protein solution structure by circular dichroism spectroscopy

Circular Dichroism (CD) spectroscopy is a technique used for the analysis of the conformation, structural stability and interaction of chiral macromolecules (proteins, DNA) in various conditions (temperature, ionic strength, and presence of interacting agents). [125-127]. The difference in absorption of the left and right circularly polarized light components by the sample is measured. Using the Beer-Lambert law the concentration of the sample can be obtained (**Equation 4.4**).

$$\Delta A = A_L - A_R = \varepsilon_L lc - \varepsilon_R lc = \Delta\varepsilon lc \qquad (4.4)$$

Where L is left and R is right, ε is the molar absorption (an intrinsic property of the molecule) in $M^{-1}cm^{-1}$, l is the optical path-length in cm, and c is concentration in M. The difference in absorption results in an elliptically polarized light, characterized by the ellipticity (θ), in milidegrees (**Figure 4.6)**.

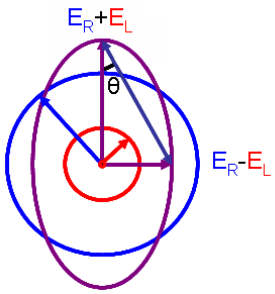


Figure 4.6. Elliptically polarized light (violet) is composed of unequal contributions of right (blue) and left (red) circular polarized light.

Elipticity and the absorption difference are related by **Equation 4.5**.

$$\Delta A = \frac{\theta}{32980} \qquad (4.5)$$

In the case of proteins, the results are usually expressed as the mean residue molar ellipticity, [θ] (**Equation 4.6**).

$$[\theta] = \frac{\theta\ 100\ M_w}{c\ l\ n} \qquad (4.6)$$

Where θ is the ellipticity in degrees, l is the optical path in cm, c is the concentration in g/ml, M_w is the molecular weight and n is the number of amino acid residues in the protein. The mean residue molar ellipticity [θ] is given in $deg{\cdot}cm^2{\cdot}dmol^{-1}$. CD investigates the secondary

structure of proteins through the induced asymmetry of the peptide chromophores in the 170–250 nm wavelength region. Hypothetical proteins with all α-helix, all β-sheet or a random coil show characteristic far-UV CD spectra (**Figure 4.7**). Using this information several algorithms have been developed to determine the secondary structure fractions of proteins from their CD spectra.

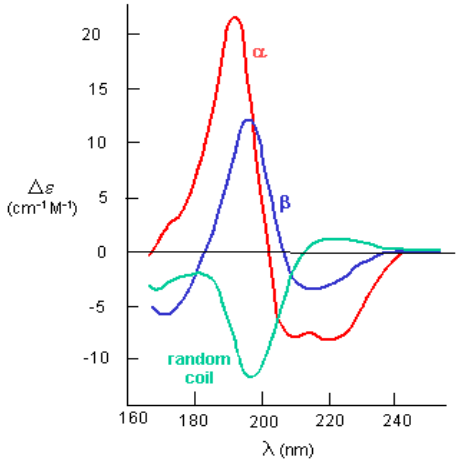


Figure 4.7. The CD spectra characteristic for alpha-helix, beta-sheet, and random coil. Each of the three basic secondary structures of a polypeptide chain (α-helix, β-sheet and random coil) show a distinct CD spectrum [128].

In our experiments, circular dichroism spectra were recorded both on a Jasco J-815 spectropolarimeter or at the CD1 beamline of the storage ring ASTRID at the Institute for Storage Ring Facilities (ISA), University of Aarhus, Denmark [129,130]. Camphor-sulfonic acid served as a calibration material for the instrument. All spectra were recorded using 1 nm steps and a dwell time of 2 s, using quartz cells (SUPRASIL, Hellma GmbH, Germany) with a 0.1 mm optical length, in the wavelength range of 180–300 nm. Prior to the measurements, the protein solutions were dialyzed or ultrafiltrated against a 10 mM HEPES, pH 8.2 buffer. Spectra were baseline-corrected by using the spectrum of water or buffer.

4.2.4. Determination of protein concentration

The measurement of protein concentration in an aqueous sample is essential for further investigations, applications ranging from circular dichroism spectroscopic analysis to enzymatic studies and biopharmaceutical analysis. Spectrophotometric protein quantitation assays rapidly determine the concentration of proteins either relative to a standard e.g., Lowry, Bradford, Comassie blue staining, biuret methods, or using estimated molar absorption coefficients [131,132]. In our lab, common UV absorption spectroscopic and Nanodrop

instruments are used to measure the protein concentration. The simplest and most direct assay method for protein concentration determination in solution is to measure the UV absorbance at 280 nm. Amino acids containing aromatic side chains (i.e., tyrosine, tryptophan and phenylalanine) exhibit strong UV-light absorption. Therefore, proteins and peptides absorb UV-light in proportion to their aromatic amino acid content and total concentration. The concentration is calculated using Beer's law. The molar absorbance (ε) is estimated based on the amino acid sequence using Expasy's ProtParam tool (http://web.expasy.org/protparam/).

Protein concentration can be detected also at wavelength 205 nm by using short path length cells, in the case of a sample of small concentration (< μM). A large number of peptide bonds within proteins make such measurement more sensitive and display less protein-to-protein variability than the measurements at 280 nm. However, this method is very sensitive to interferences from impurities and buffer components. Therefore, the protein sample has to be very pure and the composition of the buffer used for the reference measurement has to be precisely defined.

4.3. Catalytic activity assays

Catalytic experiments were performed to study the nuclease activity. Most commonly a plasmid DNA was applied as a substrate. The plasmid DNA has various forms with different mobility properties in an agarose gel. The double-strand DNA cleavage converts the supercoiled form into linear form, while the single-strand cleavage results in an open circular form. The catalytic reactions were performed under many different conditions, varying the concentration of the substrate and the enzyme, metal-to-protein molar ratios, ionic strength, time, etc. The reaction mixtures were incubated at 37°C and the catalytic reaction was stopped by adding SDS to a 1% (m/V) final concentration. Then the DNA cleavage was checked by agarose gel electrophoresis.

4.3.1. Agarose gel electrophoresis

Agarose gel electrophoresis is a suitable analytical method for separating DNA fragments of different sizes ranging from 100 bp to 25 kb. DNA fragments smaller than 100 bp are more effectively separated using polyacrylamide gel electrophoresis. DNA is loaded into pre-cast wells in the gel and electric current applied. Due to the phosphate backbone of the DNA (and RNA) molecule is negatively charged, thus DNA fragments will migrate toward the positively charged anode.

In this work, the catalytic activity of the nucleases in the DNA cleavage experiments was monitored by 1% (m/V) agarose gel electrophoresis. The DNA was visualized by ethidium bromide. Electrophoresis was performed in TAE buffer consisting of 40 mM Tris, 20 mM acetic acid, and 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0, using Bio-Rad Wide Mini Sub CellVR GT at 7 V/cm. Gene Ruler 1kb Plus DNA Ladder (Thermo Scientific) or Lambda DNA/EcoRI marker (Thermo Scientific) was applied for comparison as shown in **Figure 4.8**.

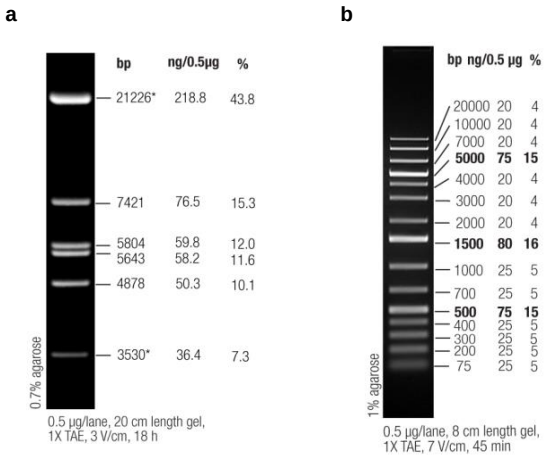


Figure 4.8. Agarose gel electrophoresis picture of a, Lambda DNA/EcoRI Marker and b, GeneRuler 1 kb Plus DNA Ladder (https://www.thermofisher.com/).

4.4. Synthesis and characterization of the water-soluble lipopolymer

This work was done in collaboration with Interdisciplinary Excellence Centre, Department of Physical Chemistry and Materials Science, and MTA-SZTE Biomimetic Systems Research Group, Department of Medical Chemistry, Faculty of Medicine (University of Szeged).

4.4.1. Synthesis and purification of the water-soluble lipopolymer

The applied materials were: Poly(ethyleneimine) (PEI, 30% (m/V) in water was purchased from Tokyo Chemicals), trimethylamine ($(C_2H_5)_3N$, 99%), methylene chloride (CH_2Cl_2, anhydrous, ≥99.8%), cholesteryl chloroformate ($C_{28}H_{45}ClO_2$, 95%), diethyl ether ($(CH_3CH_2)_2O$, anhydrous, ≥ 99.7%) were purchased from Sigma and acetone (CH_3COCH_3, ≥ 99%) was manufactured by Molar. The synthesis and the purification of WSLP were carried out based on the description in the literature [133]. Briefly: 3 g of PEI (10 mL PEI solution) was mixed with 10 mL ice-cold methylene chloride and 100 µL trimethylamine. Then 1 g of cholesteryl chloroformate was added to 5 mL chilled methylene chloride and this solution was slowly mixed with the PEI solution. The mixture was stirred at 0 °C using 800 rpm for 12 h

resulting in a pale yellow solid product. First, the product was completely dried by rotary evaporation (IKA, RV 3 evaporator) then the powder was dissolved into 50 mL of 0.1 M HCl and extraction using 50 mL methylene chloride was repeated three times. For the removal of the large aggregates, vacuum filtration was applied by standard P3 glass and paper filters with 16-40 µm and 3-5 µm pore size, respectively. The filtered liquid was concentrated by rotary evaporation (50 °C, 2 h) then the WSLP was precipitated by acetone and the filtration was repeated as described above. Finally, the solid WSLP has been obtained by freeze drying (Christ Alpha, 1 2 LD dryer) for 1 day.

4.4.2. Particle charge detection

The charge of the initial PEI and the synthesized WSLP was determined at pH = 7.4 using a PCD-04 type of Mütek Particle Charge Detector. Namely, 10 mL of 0.01 M sodium dodecyl sulfate (SDS) solution in the PCD cell was titrated separately with 25.2 mL of 0.3 mg L^{-1} PEI and 26 mL of 0.6085 mg L^{-1} WSLP in 300 and 1000 µL steps, respectively. The streaming potential was recorded 30 s after each step. The measurements were analyzed and the charge neutralization point was appointed based on the sigmoidal shape of the charge titration curve. The possible extent of the PEI protonation, the cholesteryl substitution, and the presumable molecular weight of the synthesized WSLP were also calculated.

4.5. Characterization of WSLP/pDNA complexes

After WSLP was synthesized, plasmid DNA (pDNA) was mixed with it at different ratios to obtain WSLP/pDNA complexes. Condensation of the pDNA into WSLP was detected by agarose gel electrophoresis. The size, zeta potential, and morphological properties of free WSLP and WSLP/pDNA were studied by the following methods.

4.5.1. Dynamic light scattering

Dynamic light scattering (DLS) is a physical technique that can be used to detect the size distribution of polymers in solution or suspension of small particles [134]. DLS can be used for emulsions, micelles, polymers, proteins, nanoparticles or colloids. The basic principle is simply based on that the sample is illuminated by a laser beam and the fluctuations of the scattered light are determined at a known scattering angle θ by a fast photon detector.

In this study, the hydrodynamic diameter and the ζ-potential of the WSLP itself and the WSLP/pDNA complexes were determined by dynamic and static light scattering (DLS and SLS, respectively) using a Malvern Zetasizer Nano ZS ZEN 4003 apparatus equipped with a

He-Ne laser (λ = 633 nm) at 25 ± 0.1 °C. The refractive index gradient was obtained by a Mettler Toledo Excellence RM50 refractometer.

The average molecular weight of the initial PEI was determined by DLS measurements based on the Debye plot (KC/R as a function of the C concentration) using the Zimm equation **(Equation 4.7)**. The slope of the fitting provided the second virial coefficient ($2A_2$) and the intercept yielded the reciprocal of the molecular weight (1/Mw).

$$\frac{KC}{R} = \frac{1}{M_{wP}} + 2A_2C \qquad (4.7)$$

Where K is a constant depending on the sample dn/dc, c is the sample concentration in g mL^{-1}, R is the Rayleigh ratio (the ratio of scattered light intensity to incident light intensity). P is an angular dependent term (by default, assumed to be 1 for small molecules < 15 nm radius).

4.5.2. Electron microscopy

Scanning electron microscope (SEM), utilizes focused electron beam of low energy for studying the surfaces of solid objects. Electrons in the electron gun are released from the cathode and accelerated to the energy 1 – 50 keV by the anode. The electron beam is compressed by the condenser lenses (one or two). The magnetic field, created by the scan coils deflects the electron beam back and forth. The electron beam, centered by the objective lens to very fine spot (1-5 nm), scans the surface of the sample in a raster pattern. Primary electrons interact with the sample surface atoms, causing emissions of the secondary electrons that are determined, producing the image (**Figure 4.9 a**). The backscattered electrons are used to contrast the sample parts with different chemical compositions. In this work, the scanning electron microscopy (SEM) images were recorded by a Hitachi S-4700 Field Emission Scanning Electron Microscope using 20 kV accelerating voltage.

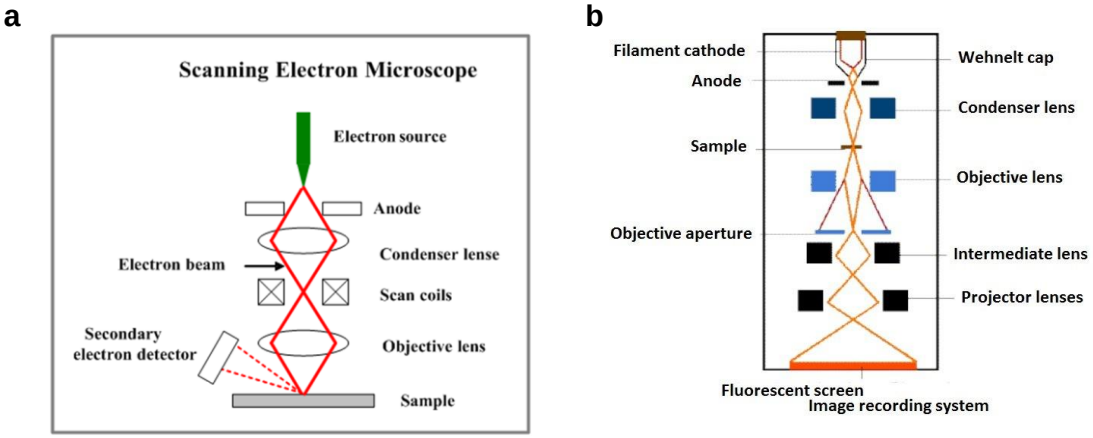


Figure 4.9. A schematic diagram of **a,** scanning electron microscope (SEM) [135] and **b,** transmission electron microscope (TEM) [136].

Transmission electron microscopy (TEM) is a technique in which a beam of electrons is transmitted through a specimen to generate an image. The specimen is most often an ultrathin section less than 100 nm thick or a suspension on a grid. An image is produced from the interaction of the electrons with the sample as the beam is transmitted through the specimen. The basic principle is analogous to that of a light microscope but electron beams are used instead of light beams. The electrons are produced in an electron gun composed of a filament, an anode, and a Wehnelt cap1. The filament acts as a hot cathode in which a wire flows electric current and sets the electrons free. Those electrons are accelerated by a high potential (~100 kV). Lenses act as a coil device focusing the beam in order to obtain the magnification of the image and examine sensitive details on the nanoscale. All this happens in a vacuum to avoid collisions with the air molecules. TEM is equipped with a camera through which the transmitted beam passes and generates the image. A fluorescent screen is used to adjust the electron path through TEM and to perform a rough search for the sample structures (**Figure 4.9 b**). A Jeol JEM-1400plus equipment (Japan) with 120 keV accelerating voltage was applied and the samples were dropped on the surface of formvar-coated grid. The pictures were analyzed by ImageJ software.

4.6. Experiments with cancer cell lines

The experiments performed to detect the cytotoxicity, biocompatibility, and transfection efficiency of the water-soluble lipopolymer and the various plasmids were carried out in collaboration with the Department of Biochemistry and Molecular Biology.

4.6.1. Cell culture materials

A549 human lung cancer cells and HeLa cervical adenocarcinoma cells were obtained from the American Type Culture Collection (ATCC, USA). Cells were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), L-glutamine (2 mM) (all purchased from Lonza), penicillin (100 IU/mL) and streptomycin (50 μg/mL) (SERVA) in a humidified atmosphere containing 5% CO_2 at 37°C. The stock solutions were freshly prepared using MQ (Millipore, Milli-Q Integral3) ultrapure water (18.2 MΩ·cm at 25 °C) in every case. The pEGFP-N3 and the pSpCas9(BB)-2A-GFP (pX458) plasmids were purchased from Addgene (see detailed information in **Appendix A.2, A.3**. The Qiagen Plasmid Mini Kit was purchased from Qiagen.

The MTT reagent (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) was obtained from SERVA. Lipofectamine™ 3000 transfection reagent was obtained from Thermo Fisher Scientific.

4.6.2. MTT assay

The MTT assay is a sensitive, accurate, and quantitative colorimetric assay that detects viability, proliferation, and activation of cells [137]. The assay depends on the ability of mitochondrial dehydrogenase enzymes in viable cells to convert the yellow water-soluble substrate 3-(4,5-dimethylthiazol-2-y1)-2,5-diphenyl tetrazolium bromide (MTT) into purple formazan crystals that are insoluble in water (**Figure 4.10**) [138].

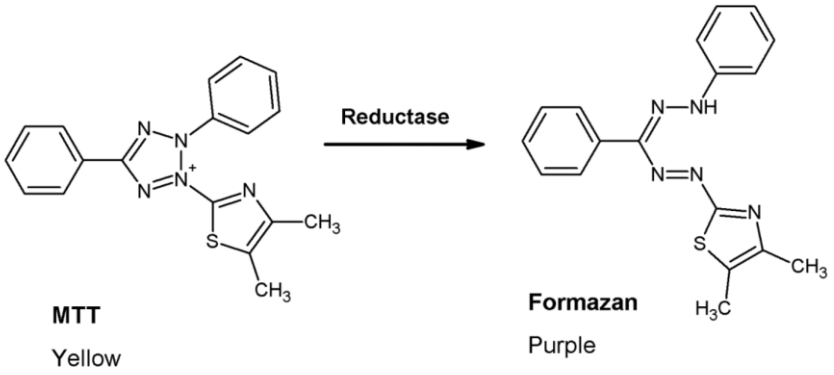


Figure 4.10. Conversion of MTT to formazan by NADH-dependent reductases.

The insoluble formazan crystals are solubilized and the resulting color intensity is quantified by measuring absorbance in the range of 500-600 nm by using a multi-well spectrophotometer. The darker the solution, the greater the number of viable and metabolically active cells.

In this work, the cytotoxicity of WSLP/pDNA complexes prepared according to the mass and N/P ratios was investigated by MTT assay. A549 and HeLa cells were seeded at 5000 cells/well density in 96-well plates and incubated for 24 h at 37 °C to reach 70% confluence. Then the medium was removed and the WSLP, WSLP/pDNA complexes were added to the cells in serum-free DMEM. Following the incubation at 37 °C for 24 h and 48 h, the medium was removed and the cells were washed with phosphate buffered saline (1×PBS, pH 7.4). Cell viability was assessed by adding the MTT reagent prepared in serum free DMEM to a final concentration of 0.5 mg mL^{-1}. After incubation (1 h at 37 °C), the media was removed and 100 µL DMSO (Serva Electrophoresis GmbH) was added to dissolve the formazan crystals. The absorbance was measured at 570 nm using Synergy HTX plate reader (BioTek). Data were analyzed by GraphPad Prism 7 software (GraphPad Software), where untreated cells served as the control.

4.6.3. Internalization of plasmid DNA into the cancer cells

Transfection is a process by which nucleic acids are deliberately introduced into eukaryotic cells. Transfection can be carried out using different delivery systems mentioned previously in section 2.4. During our work commonly A549 and HeLa cells were seeded in 6-well tissue culture plates at a density of 2×10^5 cells per well in 10% FBS-containing DMEM and were left to grow. After the cells reached 70% confluency, the medium was replaced with serum-free DMEM and the cells were transfected with WSLP/pDNA complexes at DNA concentration of 2.5 μg per well. After 24 h incubation at 37°C, the medium was replaced by a DMEM medium supplemented with 10% FBS, L-glutamine, penicillin, and streptomycin in concentrations described above. The cells were incubated for an additional 48 h, then were trypsinized, washed, and resuspended in PBS. The transfection efficiency was detected by flow cytometry using FACSCalibur™ platform via quantification of EGFP fluorescence. The analysis of the FACS data was performed with Flow Jo and Graphpad softwares.

4.6.4. Fluorescence microscopy

A fluorescence microscope can selectively detect the distribution of fluorescent molecules [139]. The image acquisition process with the fluorescence microscope is shown in **Figure 4.11**. The light from the source passes through the excitation filter, to select the appropriate range of wavelengths. The dichroic mirror focuses the light on the specimen. The light absorbed by the fluorophores allows them to emit a lower-energy radiation. The emitted light passes through the dichroic mirror and the emission filter to reach the detector [140].

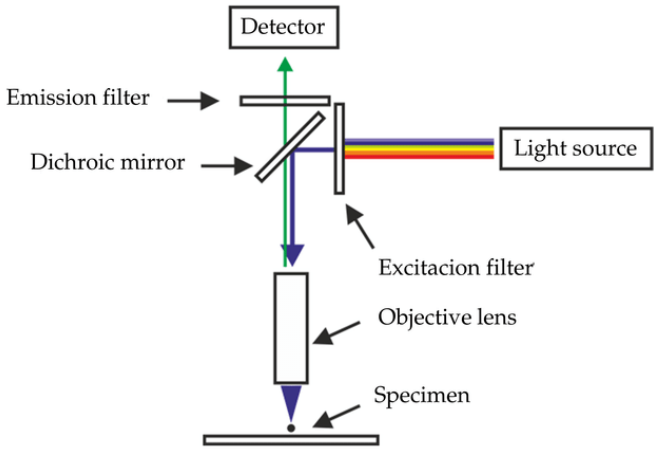


Figure 4.11. The scheme of the fluorescence microscope [141].

The samples can be expressed fluorescent proteins or molecules labelled with fluorescent stains. Hoechst blue fluorescent dyes are supravital minor groove–binding DNA stains with AT selectivity (**Figure 4.12**) [142].

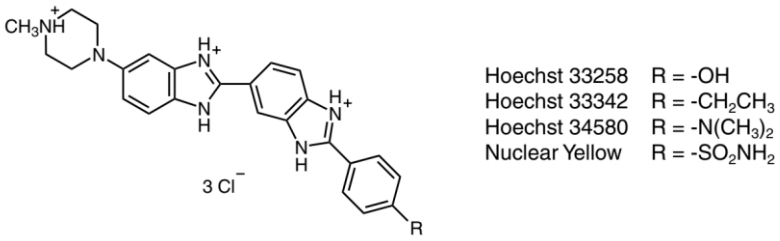


Figure 4.12. Structure of Hoechst dyes (from Handbooks of Invitrogen detection technologies).

We have assessed the transfection process by fluorescence microscopy as well. For this A549 cells were seeded on glass coverslips (VWR) placed into the wells of a 24-well plate and were left to grow. When the cells reached about 70% confluency, the culture medium was replaced to serum-free DMEM and the cells were transfected with WSLP/pEGFP-N3 and with PEI/pEGFP-N3 complexes at 10/1 N/P ratio with DNA concentration of 0.75 μg per well. After 24 h the cells were washed with PBS and were left to grow for 48 hours in FBS-containing culture medium. Cells were visualized by OLYMPUS BX51 fluorescence microscope and the photos were taken by Olympus DP70 camera (Olympus, Tokyo, Japan) with or without staining with 3.25 μM Hoechst 33342 (Sigma-Aldrich) dye for 30 minutes.

5. RESULTS AND DISCUSSION

5.1 Design of a new DNA construct for affinity based purification of ΔN4-NColE7

The computer design of artificial nucleases relies on precise reproduction of the designed sequence during the synthesis of the protein. However, production of such proteins in high yield and purity is usually difficult and requires multistep chromatographic methods often leading to a significant loss of target proteins. Therefore, it is a topical issue to develop purification procedures for the designed and similarly, for the native proteins with precisely determined sequences [143-147].

Affinity chromatography is a popular approach, allowing for easy protein purification. The affinity tag may be a short amino acid sequence or a protein fused to either the C- or N-terminus of the target recombinant protein. Such tags facilitate the purification and detection of the target protein, and are removed by a specific protease after the purification procedure. Unfortunately, the tag removal is not always a straightforward procedure.

The search for ideal affinity tags and tag-removal methods is necessitated in protein chemistry as exemplified by recent reviews [148-158]. Various methods have been elaborated to cleave and remove the affinity tags after getting rid of impurities. Specific protein-based cleavage enzymes, however, display several disadvantages for example, shorter or longer amino acids strings will remain at the termini of the purified protein after the enzymatic cleavage. These may disturb the folding and/or the function of the target protein. To avoid such consequences, we have modified an existing protein expression vector: pET-21a. The C-terminal 6×His affinity tag can be completely removed from the proteins purified from this new DNA construct. The new DNA vector is an advantageous tool for protein purification, when the complete removal of the affinity (or other) tags, without any remaining amino acid residue is essential.

5.1.1. Construction of the new cloning site into the pET-21a (Amp^R) DNA vector

BsmBI (Esp3I) restriction endonuclease recognizes 5’-gagacg-3’ DNA sequence and cleaves the DNA outside the recognition region. Including such restriction site in the cloning region of a plasmid provides an opportunity to eliminate this recognition sequence during the insertion of a gene into the DNA carrier. For this reason, we redesigned the cloning site of the pET-21a plasmid (Novagen) as it is schematically depicted in **Figure 5.1**. This construct allows to insert the target gene precisely between the 5’-atg-3’ start codon included in the NdeI recognition site, and the SRHS code which will serve the removal of the affinity tag.

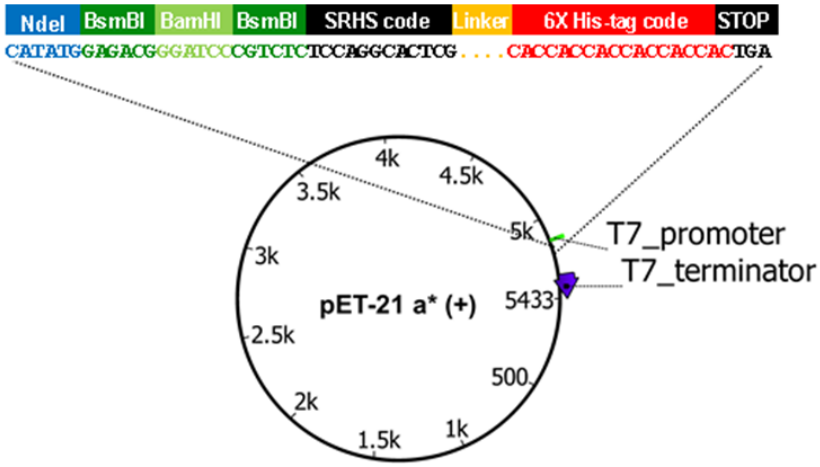


Figure 5.1. Schematic representation of pET-21a*-SRHS plasmid with the redesigned cloning site.

The original pET-21a sequence contains two additional BsmBI restriction sites as shown in **(Figure 5.2 a)**, so first, the knockout of these restriction sites was necessary. Both sites were knocked out using a modified QuikChange protocol (QCM, Stratagene). We used KOD-FX-DNA polymerase in PCR, because this enzyme precisely amplifies long DNA sequences. The following primers were applied for the knockout modifications: BsmBI-1F 5'-gtcactggtgaaaagaaaaaccacc-3'; BsmBI-1R 5'-acgggcaacagctgattgcc-3' and BsmBI-2F 5'-gacggtcacagcttgtctgtaag-3'; BsmBI-2R 5'-accgggagctgcatgtgtca-3'. The success of modifications was checked by the sequencing reactions with the designed Seq-BsmBI-1F: 5'-aatgcgcgccattaccga-3 or Seq-BsmBI-2F: 5'-ggcatcagtgaccaaacag-3' primers. One of these BsmBI sites was located within the ampicillin resistance gene **(Figure 5.2 a)**, so the mutation of one base pair within the restriction site had to be carried out without affecting the ampicillin resistance gene properties. The success of this procedure was verified by transformation of bacterial cells, and subsequently growing them on LB(Amp+) plates.

A specific double-strand oligonucleotide was designed and inserted into the new cloning site using the original NdeI and SalI restriction enzymes. This sequence contains the two new BsmBI sites for cloning the target gene into the mutated pET-21a. This is followed by a sequence encoding for SRHX, where X can be any amino acid and it is encoded by the nnn codon in the DNA sequence.

a

gtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccaggccagccacgtttctg
cgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgct
gattggcgttgccacctccagtctggccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtg
gtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtgggctgatcattaact
atccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaa
cagtattatttttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggccca
ttaagttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattcagccgatagcggaacgggaaggcgact
ggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgatcagatggcgct
gggcgc***aatgcgcgccattaccga***gtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagctcatgttat
atcccgccgttaaccaccatcaaacaggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtga
agggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattc
attaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtaagttagctcactcattaggcaccggg
atctcgaccgatgcccttgagagccttcaacccagtcagctccttccggtgggcgcggggcatgactatcgtcgccgcacttatgactgtcttc
tttatcatgcaactcgtaggacaggtgccggcagcgctctgggtcattttcggcgaggaccgctttcgctggagcgcgacgatgatcggcctgt
cgcttgcggtattcggaatcttgcacgccctcgctcaagccttcgtcactggtcccgccaccaaacgtttcggcgagaagcaggccattatcgc
cggcatggcggccccacgggtgcgcatgatcgtgctcctgtcgttgaggacccggctaggctggcggggttgccttactggttagcagaatgaa
tcaccgatacgcgagcgaacgtgaagcgactgctgctgcaaaacgtctgcgacctgagcaacaacatgaatggtcttcggtttccgtgtttcgt
aaagtctggaaacgcggaagtcagcgccctgcaccattatgttccggatctgcatcgcaggatgctgctggctaccctgtggaacacctacatc
tgtattaacgaagcgctggcattgaccctgagtgatttttctctggtcccgccgcatccataccgccagttgtttaccctcacaacgttccagt
aaccgggcatgttcatcatcagtaacccgtatcgtgagcatcctctctcgtttcatcggtatcattacccccatgaacagaaatcccccttaca
cg***ga******ggcatcagtgaccaaacag***gaaaaaaccgcccttaacatggcccgctttatcagaagccagacattaacgcttctggagaaactcaacga
gctggacgcggatgaacaggcagacatctgtgaatcgcttcacgaccacgctgatgagctttaccgcagctgcctcgcgcgtttcggtgatgac
ggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaag

b

5′-tatg**GAGACG**ggatcc**CGTCTC**tccaggcacnnngaattcgagctcg-3′

3′- a**cCTCTGC**cctagg**GCAGAG**aggtccgtgnnncttaagctcgagcagct-5′

NdeI *BsmBI* BamHI *BsmBI* EcoRI SacI

5′-catatg*gagacg*ggatcc*cgtctc*tccaggcactcggaattcgagctc

M E T G S R L S R H S E F E L

Ni(II) cleavage site

SalI HindIII NotI XhoI

gtcgacaagcttgcggccgcactcgagcaccaccaccaccaccactga-3′

V D K L A A A L E H H H H H H -

6×His-tag Stop

Figure 5.2. Construction of pET-21a*-SRHS plasmid. **a,** Part of the pET-21a vector with the two highlighted BsmBI sites (red background) in the original sequence. The ampicillin resistance gene is highlighted by green background, while the designed sequencing primers to check the success of BsmBI knockout experiments are denoted by bold italic. **b,** The sequence of the double-strand oligonucleotide used to establish the new cloning site in the carrier DNA. The redesigned cloning site includes all the restriction enzyme recognition sites of the pET-21a vector and two additionally inserted BsmBI recognition sites (the arrows pointing downwards show the BsmBI cleavages outside the recognition sites). Cloning the gene of the target protein between the two BsmBI sites allows the expression from the atg start codon (underlined) within the NdeI restriction site.

Thus, by cloning this designed DNA sequence into the BsmBI knockout pET-21a plasmid, we have got a library of new plasmids, from which we isolated those containing the codes of SRHS and SRHH amino acids sequences. As it is well-known from the literature, the SRHX protein sequences can be hydrolyzed at the N-terminus of this site using Ni(II) ions [160]. Therefore, the SRHX sequence serves as the hidden stop codon, which is activated only in its translated form and ideally after the protein purification procedure. The SRHS sequence is followed by the hexahistidine code through a linker sequence containing the restriction enzyme recognition sites of the original pET-21a vector **(Figure 5.2 b)**. The modified plasmid with SRHS selected for the further experiments, was denoted as pET-21a*-SRHS. The SRHS sequence within the expressed protein from this new DNA construct can be cleaved by Ni(II) and thus, the C-terminal fusion sequence including the 6×His-tag can be removed.

5.1.2 Expression and purification of ΔN4-NColE7-6×His by nickel-affinity resin

To verify the suitability of the new plasmid for the purpose of protein purification without any remaining amino acid sequence, first, we started with a well studied protein in our laboratory, which is the ΔN4-NColE7 protein – the nuclease domain of Colicin E7 bacterial toxin lacking four N-terminal amino acids. The template plasmid pQE70-NColE7/Im7 [159] was kindly provided by prof. Kin-Fu Chak (Institute of Biochemistry, National Yang-Ming University, Taipei, Taiwan, Republic of China). The DNA segment, encoding the ΔN4-NColE7 sequence was modified to include BsmBI restriction site at its termini by PCR using the following oligonucleotide primers:

B-DN4 – forward: 5′-attcgtctccatatgccagggaaggcaa-3′ and

B-ColC – reverse: 5′-attcgtctccctggatttacctcggtgaatatcaatatg-3′.

Both the pET-21a*-SRHS plasmid and the PCR product were digested by BsmBI restriction endonuclease **(Figure 5.3 a)**. The ligated pET-21a*-ΔN4-NColE7-SRHS plasmid was applied to transform competent DH5α *E. coli* cells for DNA amplification. Several colonies were selected and cultured in separate LB (amp+) solutions for plasmid purification. The resulted new plasmids should contain the target sequence between the NdeI restriction site and SRHS sequence. We have checked this by BamHI cleavage. This restriction enzyme site is inserted between the to BsmBI restriction site, and therefore, it is also eliminated upon successful cloning. Finally, the DNA sequencing proved the precise ligation.

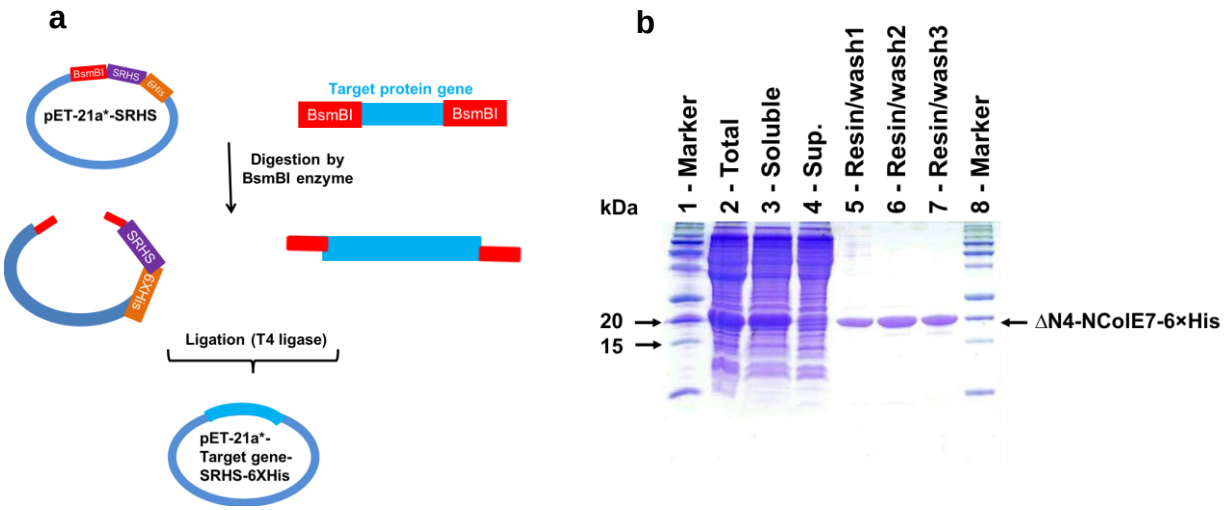


Figure 5.3. a) Scheme of the gene cloning technology by using the new plasmid. The target gene is amplified by PCR, then both the plasmid and the target genes were digested by BsmBI at 37 °C for two hours. The DNA fragments were ligated by T4 ligase to form the recombinant DNA. **b)** SDS PAGE analysis of the batch Ni(II)-affinity purification procedure. Lane 2: 20 µl aliquot of the total protein fraction obtained after resuspending the pellet in 10 mL 1×binding buffer. Lane 3: 20 µl aliquot of the soluble protein fraction, i.e., from the supernatant after the centrifugation of total protein fraction. Lane 4: 20 µl aliquot of the supernatant after binding to the Ni-NTA resin. Lane 5 (Wash1): 1 µl aliquot of the resin after washing with 4×2 bed volume wash buffer (20 mM Tris pH 7.9; 60 mM Imidazole; 500 mM NaCl; 0.1% Triton X-100. Lane 6 (Wash2): 1 µl aliquot of the resin after washing with 2×2 bed volume wash buffer with an increased imidazole concentration of 100 mM. Lane 7 (Wash3): 1 µl aliquot of the resin after washing with 2×2 bed volume of 100 mM pH 8.2 HEPES buffer.

The ΔN4-NColE7-6×His protein was expressed in BL21 (DE3) *E. coli* cells using the pET-21a*-ΔN4-NColE7-SRHS plasmid and then purified by batch type nickel-affinity chromatography by using a Novagen His·Bind® resin as described in the experimental section. ΔN4-NColE7-6×His protein was separated from other bacterial proteins by its specific binding to the resin through the 6×His affinity tag. The success of protein expression and purification was analyzed by SDS-PAGE **(Figure 5.3 b)**. The overexpressed target protein could quantitatively be solubilized in the binding buffer as shown in lane 3. The binding efficiency of the 6×His-tag was demonstrated by the lack of the target protein in the supernatant (lane 4) solution after the treatment with the nickel-affinity resin. Wash buffer (500 mM NaCl, 20 mM Tris-HCl pH 7.9) containing 60 mM imidazole was used to remove all the impurities from the resin and a clear band on the gel close to 20 kDa size marker band was assigned to the ΔN4-NColE7-6×His protein – calculated molecular weight is 17220 Da. The 6×His tag was removed from the C-terminus of this protein by Ni(II)-induced cleavage reaction, set up directly on the protein-loaded resin. The details of the Ni(II)-induced cleavage are parts of another Ph.D. topic.

This purification strategy provided the metalloprotein with a native protein sequence, avoiding any remnant amino acid residues after the removal of the affinity tag, while overcoming several above listed disadvantages of the proteases. This one-step purification method by using the new plasmid resulted in ΔN4-NColE7 protein samples of high purity with an average yield of ~3.0 mg protein from 1 L culture (~4.0 g of wet cells). This yield is higher compared to the previous purification method in our lab, which was 2.5 mg per liter of culture most probably because the toxicity of ΔN4-NColE7 protein was decreased by the C-terminal 6×His tag.

The new DNA construct designed by us, can be used with various tags at the C-terminus for affinity chromatographic applications. As a result of the protein expression from this vector, a C-terminal hexahistidine peptide tag is fused to the target protein [160]. The cytotoxicity of ΔN4-NColE7, attributed to the non-specific DNA cleavage within the cells during the protein expression, was decreased by using the C-terminal 6×His tag. Furthermore, it was found that the new vector could also carry the gene of NColE7 and R447G-NColE7 (KGNK) mutant successfully in DH5α bacterial cells, while this was not possible with the original pET-21a vector without the simultaneous cloning of the gene of the inhibitory protein Im7 [157,161]. I will discuss these experiments in the following chapter.

5.2 Modulation of the catalytic activity of NColE7 by the 6×His tag

The oligohistidine sequence in the C-terminal affinity tag introduced by the new protein expression DNA construct described in the previous chapter, may modulate the nuclease activity of the nuclease domain of the Colicin E7 metalloprotein (NColE7) via both non-specific and coordinative interactions. This highlights the possibility of tailoring the properties of a selected metalloprotein by applying an oligohistidine fusion sequence. To investigate this phenomenon we have synthesized the genes of the NColE7-6×His protein (**Scheme 5.1**) and its selected mutants and whenever it was possible, we also purified the protein, for further characterization of its solution structure, metal ion-binding and catalytic activity.

NColE7-6×His KRNKPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWE
EVSKDPELSKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGK
SRHSEFELVDKLAAALEHHHHHH

Scheme 5.1. Scheme of the amino acid sequence of the NColE7-6×His protein. The arginine (R447) amino acid highlighted by the grey background close to the N-terminus was mutated to G (R447G mutation) in the KGNK-6×His protein. H544, H569 and H573 metal ion-binding histidines are underlined in the HNH motif (red characters), and the general base H545 histidine is underlined and also highlighted by the grey background. The fusion sequence is in italic, in which the SRHS nickel(II)-sensitive motif and the hexahistidine function is underlined. The numbering of the amino acids is based on the original Colicin E7 amino acid sequence.

5.2.1. DNA cloning and cytotoxicity experiments reflecting the catalytic activity of the nucleases

The NColE7 protein is highly cytotoxic due to the non-specific DNA cleavage within the bacterial cells. The minor leakage of the protein expression prevents even the gene cloning procedure in pET-21a without the parallel expression of the inhibitory protein. pET-21a*-NColE7-SRHS and pET-21a*-KGNK-SRHS plasmids were constructed similarly such as pET-21a*-ΔN4-NColE7-SRHS by using the pET-21a-NColE7/Im7 plasmid containing the genes of NColE7 and the Im7 immunity proteins, as a template to amplify NColE7 and KGNK genes. The B-ColNNN, forward (5′-attcgtctccatatgaaacggaataagccag-3′) and B-ColC reverse (5′-attcgtctcctggatttacctcggtgaatatcaatatg-3′) oligonucleotide primers were used for NColE7 amplification and the B-KGNK-N* forward primer (5′-attcgtctccatatgaaacggaataagccag-3′) instead of B-ColNNN for KGNK (the explanation of the protein sequences is shown in **Scheme 5.1**.

After inserting the genes of NColE7 and KGNK proteins into the pET-21a*-SRHS vector and transforming the bacterial cells, bacterial colonies appeared on the LB(amp+) agar

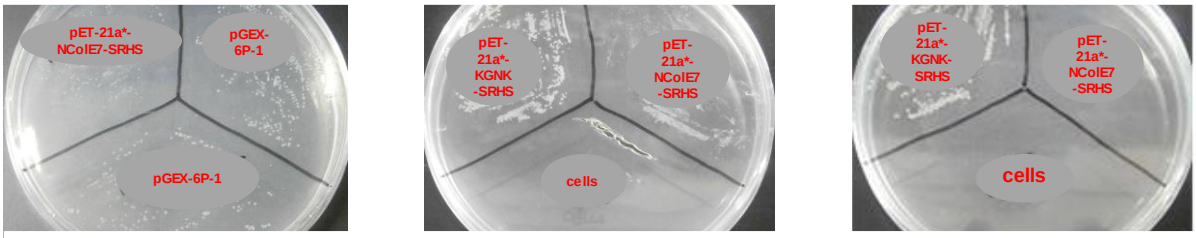
plates. This suggests that the modified 6×His-tagged proteins show decreased catalytic activity compared to those with native sequences. However, the absence of the Im7 gene during the cloning procedure enhances the selection of plasmids including erroneous genes, which results in expressing non-toxic proteins. Therefore, we have checked the nucleotide sequences of the cloned genes in randomly selected bacterial colonies. Among these, we could indeed identify the pET-21a*-NColE7-SRHS plasmid with the error-free gene, coding the expected NColE7-6×His protein, and thus, supporting the decreased catalytic activity of the tagged protein. Nevertheless, the sequencing experiments also revealed several pET-21a*-NColE7-SRHS plasmid preparations containing wrong, randomly mutated NColE7 genes, from which modified proteins with decreased or no toxicity could be expressed. The NColE7 nuclease sequence suffered from either random point mutations or random mutations causing reading frame shifts (**Table 5.1**). It is worth mentioning that some of the mutations occurred at the same locations as detected in a previous article [79], such as the point mutation at D557 or frameshift at …IDIHRGK… sequence, the latter mutation affecting the H573 residue, essential for the metal ion binding. Ont he other hand, the sequences of the verified pET-21a*-KGNK-SRHS plasmids were in agreement with the expected sequence. Thus, the cytotoxicity of the KGNK-6×His protein seems to be negligible compared to that of NColE7-6×His.

Table 5.1. Mutations found in the gene of NColE7 in the DNA cloning experiments.

Types of mutations	Mutation site
Single point mutation	S504L
	D557G
Reading frame shift	…IDIHRGK → …MIFTEVN…
	…PGKA… → …ATLStop
	…RKKFWE… → …VRNSGK…
	…FDDFR… → …LMILStop

After the identification of the pET-21a*-NColE7-SRHS and pET-21a*-KGNK-SRHS plasmids containing the expected sequences, we have applied these DNA constructs for the transformation of various bacterial cells in order to check their toxic effect reflecting the catalytic activity of the expressed metallonucleases. In case of Mach1 and DH5α cells optimized for DNA cloning, we observed the appearance of bacterial colonies, although the bacterial growth was partially inhibited (**Figure 5.4**). Mach1 colonies were significantly smaller than the DH5α colonies using pET-21a*-NColE7-SRHS. The colony size of the Mach1 cells

transformed with pET-21a*-NColE7-SRHS was significantly smaller than that of the pET-21a*-KGNK-SRHS containing cells. It is shown that the effect of the NColE7-6×His protein expression leakage is already lethal to BL21(DE3) cells, but bacterial colonies still appeared after the transformation experiments with pET-21a*-KGNK-SRHS 6×His plasmid. These results support that the catalytic activity of the 6×His tagged enzymes decreased in comparison with the native sequences. The slightly more pronounced protein expression leakage and/or sensitivity in Mach1 could even differentiate between the NColE7-6×His and KGNK-6×His sequences, in agreement with the higher activity of the former enzyme.



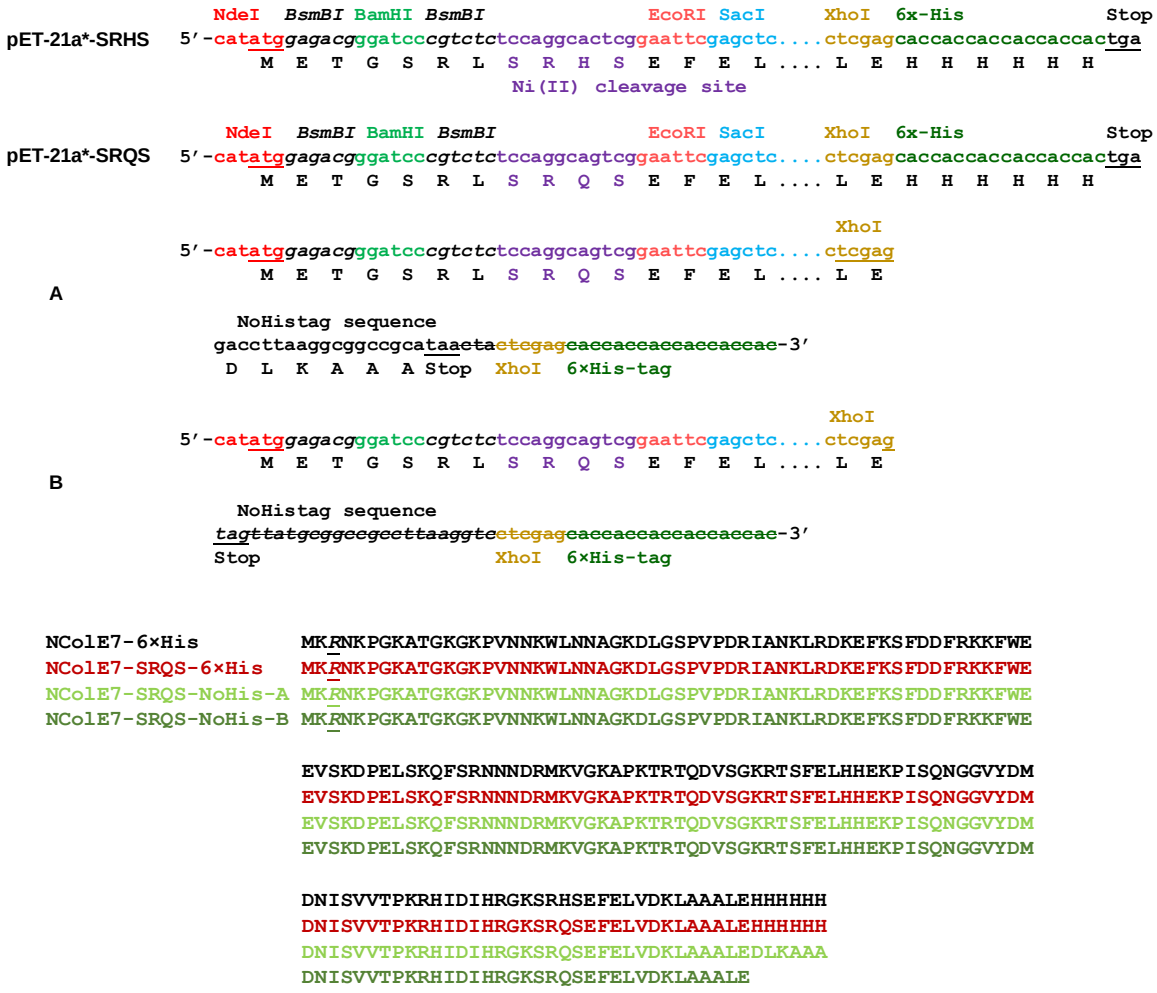
DH5α DNA | Mach1 DNA | BL21 Protein

Figure 5.4. Formation of bacterial colonies on the LB(amp+) agar plates. DH5α and Mach1 cells are optimized for DNA cloning, while BL21(DE3) is a cell for protein expression.

The Zn^{2+} ion in the active center of NColE7 has an essential role in the catalytic mechanism. The scissile phosphodiester group of the substrate can bind to the metal ion at the fourth coordination site, replacing the coordinated water molecule. Any additional histidine within the C-terminal tag can influence this event. To monitor the possible role of these histidines we have carried out mutation experiments. First, we changed the code of histidine within the SRHS sequence to the code of glutamine (Q) resulting in pET-21a*-NColE7-SRQS and pET-21a*-KGNK-SRQS plasmids. The sequencing results revealed that the mutation of histidine of SRHS was successful in both plasmids, but we could also detect random mutations in some of the pET-21a*-NColE7-SRQS plasmids. Thus, the cytotoxicity of these plasmids (through the minor protein expression) in Mach1 and DH5α cells did not show significant change as compared to the pET-21a*-NColE7-SRHS and pET-21a*-KGNK-SRHS plasmids.

The hexahistidine tag is expected to inhibit the catalytic reaction when it binds to the catalytic Zn^{2+} ion. To verify this hypothesis, we modified the genes of the NColE7-SRQS and KGNK-SRQS proteins by inserting a designed short dsDNA sequence (NoHistag) using the XhoI restriction endonuclease site between the codes of the SRQS and 6×His motifs. This short dsDNA can be inserted in two different orientations during the ligation experiment. For this

reason, it was designed to include a stop codon independently of the orientation. Thus, it prevents the expression of the 6×His sequence. Therefore, we expected an increased catalytic activity and cytotoxicity of the new plasmids. The two different proteins are expressed with C-terminal linker sequences of different lengths (**Scheme 5.2**) without bearing strongly coordinating amino acid side-chains (His or Cys).



Scheme 5.2. Top: The nucleotide sequence of the modified cloning region of pET-21a*-SRHS vector showing the mutations that led to SRHS → SRQS amino acid sequence modification, as well as the results (A and B) of the insertion of the NoHistag short dsDNA in two different orientations into the XhoI restriction enzyme recognition site. Note that the scored sequences after the newly established stop codons are not expressed. Information about the further restriction sites is also included in the scheme. **Bottom:** The sequences of the expressed NColE7-type proteins based on the DNA constructs above. The KGNK protein sequences can be read by exchanging the underlined R to G.

In the transformation experiments we still obtained colonies, which were verified for the successful modifications of the plasmids. To this end, we inserted a unique AflII restriction endonuclease cleavage site (5'-cttaag-3') into the NoHistag dsDNA sequence, so that the

successfully ligated oligonucleotide plasmid constructs could be cleaved by this enzyme. The plasmid digestion experiments by AflII revealed more than 50% success of the insertion of the oligonucleotide. The modified plasmids were subjected to sequencing.

We expected an increased catalytic activity and cytotoxicity of the new plasmids, able coding for proteins with less coordinating histidine side-chains in the C-terminal tag. **Figure 5.5** shows the comparison of the effect of pET-21a*-Nucl-SRHS, pET-21a*-Nucl-SRQS and pET-21a*-Nucl-SRQS-NoHistag plasmids (where Nucl is either NColE7 or KGNK) on the bacterial colony formation with BL21 cells optimized for protein expression. All the NColE7 containing plasmids killed the cells, but for all KGNK plasmids colony formation was observed. The inhibitory effect could still be observed, which decreased the toxicity to an extent that is enough for protein-expressing bacteria harboring the KGNK-type plasmids to survive, but not for those transformed by the more active NColE7-type plasmids.

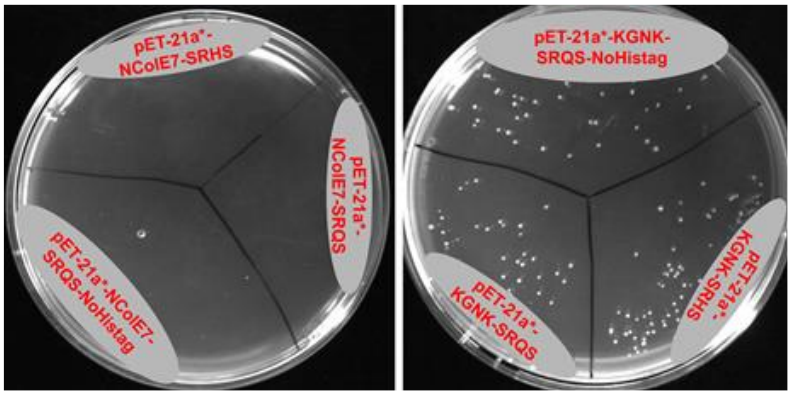


Figure 5.5. Inhibition effect of pET-21a*-Nucl-SRHS, pET-21a*-Nucl-SRQS and pET-21a*-Nucl-SRQS-NoHistag plasmids (where Nucl is NColE7 – **left** – or KGNK – **right**) on BL21(DE3) bacterial colony formation with and without His affinity tag. There is no colony formation when Nucl = NColE7, but a few bacterial colonies appeared when Nucl = KGNK.

This result suggests that the C-terminal tag (even without bearing strongly coordinating amino acid residues) can decrease the catalytic activity of KGNK through a non-specific steric effect, likely involving non-covalent interactions with the active site. Thus, inhibition most probably occurs through a complex mechanism, involving the combination of the steric and coordinative effects. We purified the KGNK-6×His protein as described in the experimental section for more detailed *in vitro* catalytic investigation to learn more about the effect of His-tag on the enzyme function.

5.2.2. Purification and characterization of KGNK-6×His protein

The KGNK-6×His protein fractions were analysed by SDS-PAGE **(Figure 5.6)**. The soluble protein fraction contains the overexpressed target protein at around ~ 20 kDa (lane 2). The calculated average molecular mass is 17779.03 Da. The binding efficiency of the 6×His-tagged protein to the nickel-loaded IMAC resin was indicated by the absence of the target protein in the supernatant (lane 4) following the treatment of the protein solution with the resin.

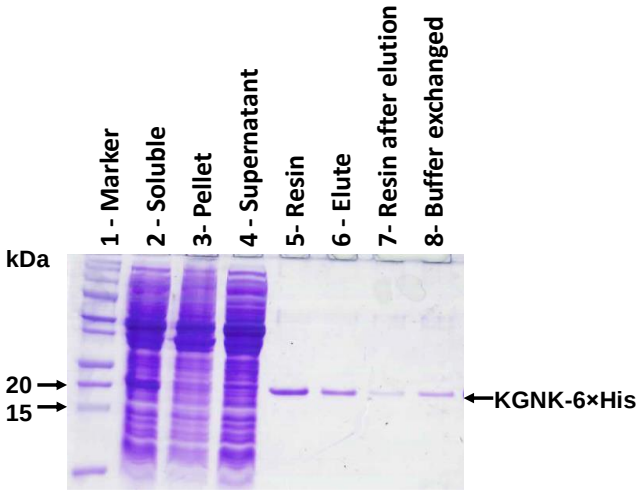


Figure 5.6. The SDS PAGE analysis of the batch Ni(II)-affinity purification of the KGNK-6×His protein is visualized on the gel photo above. **Lane 1**: protein marker. **Lane 2**: 20 µL sample of the supernatant after the sonication and centrifugation of the resuspended cells. **Lane 3**: 20 µL sample obtained after sonicating, centrifuging of the cells and resuspending the pellet in 5 mL of 1×binding buffer – this fraction represents the insoluble protein fraction lacking the target protein. **Lane 4**: 20 µL aliquot of the supernatant after mixing the soluble protein fraction with the Ni-NTA resin and centrifugation. **Lane 5**: 0.38 µL aliquot of the resin after washing with 5×2 bed volume of wash buffer (20 mM Tris pH 7.9; 60 mM imidazole; 500 mM NaCl). **Lane 6**: 0.3 µl aliquot of the protein eluted from the resin by elution buffer (20 mM Tris-HCl, pH 7.9, 0.5 M NaCl, 1 M imidazole). **Lane 7:** 1 µL aliquot of the resin after the protein elution. **Lane 8:** 0.23 µL aliquot of the protein solution after buffer exchange.

All the impurities were removed from the resin by subsequent washing steps, and a clear band on the gel was assigned to the KGNK-6×His protein (lane 5). The protein was eluted from the resin by the elution buffer (lane 6) and then it was buffer-exchanged with 20 mM HEPES pH 7.7 prior to catalytic experiments.

It is worth mentioning that the yield of the pure KGNK-6×His protein was ~ 0.3 mg from 1 L culture, *i.e.* from ~ 4.0 g of wet cells which is at least 10 times less than the yield we could achieve with e.g. the ΔN4-NColE7-6×His protein previously. This reflects that the 6×His attachment couldn't completely inhibit the nuclease activity of the KGNK protein and the expressed protein is slightly toxic to the cells.

CD spectroscopy was applied to monitor the protein folding after purification and to estimate the fractions of secondary structure elements. Evaluation of the CD spectrum of the KGNK-6×His protein by the BeStSel software [162] revealed that the relative α-helical content is ~14% **(Figure 5.7)**. As expected, this is slightly less than the α-helical content of the Zn^{2+}-loaded KGNK protein calculated by BeStSel: ~16% [73,144]. The reason for this is that KGNK-6×His is longer by 16 amino acids due to its C-terminal fusion tag, displaying most probably non-helical structure. This result suggests that the protein is in its functionally folded form.

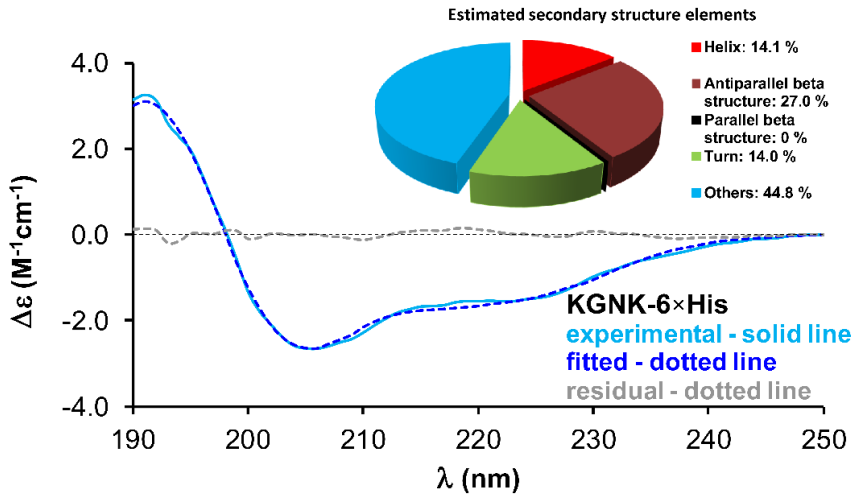


Figure 5.7. Fitting of the far UV CD spectra of the KGNK-6×His protein between 190 and 250 nm to determine the secondary structure fractions by the BeStSel program. The result of the secondary structure estimation is shown in the inset.

ICP-MS measurements with the same batch of the purified KGNK-6×His revealed that the protein was obtained in its Zn^{2+}-loaded form containing approximately one equivalent of metal ion (36.99 μM Zn^{2+} for 30 μM protein). The Zn^{2+} ions in NColE7 and NColE9 are bound extremely tightly in the HNH catalytic center, characterized by an apparent dissociation constant $K_d \sim 10^{-9}$ value [163,164]. Therefore, we propose that this should be the primary metal ion-binding site in KGNK-6×His, as well.

5.2.3. DNA cleavage experiments with the KGNK-6×His protein

The cleavage of a randomly selected pUC-EGFP plasmid by the KGNK-6×His enzyme was monitored by agarose gel electrophoresis in the presence of various amounts of Zn^{2+} ions. A pUC-EGFP plasmid DNAs served as the template. The transitions between its supercoiled (sc), open circular (oc) and linear (lin) forms was followed by agarose gel electrophoresis.

Subsequent double strand cleavages fractionate the linear DNA into small fragments of various sizes resulting in a smear on the agarose gel.

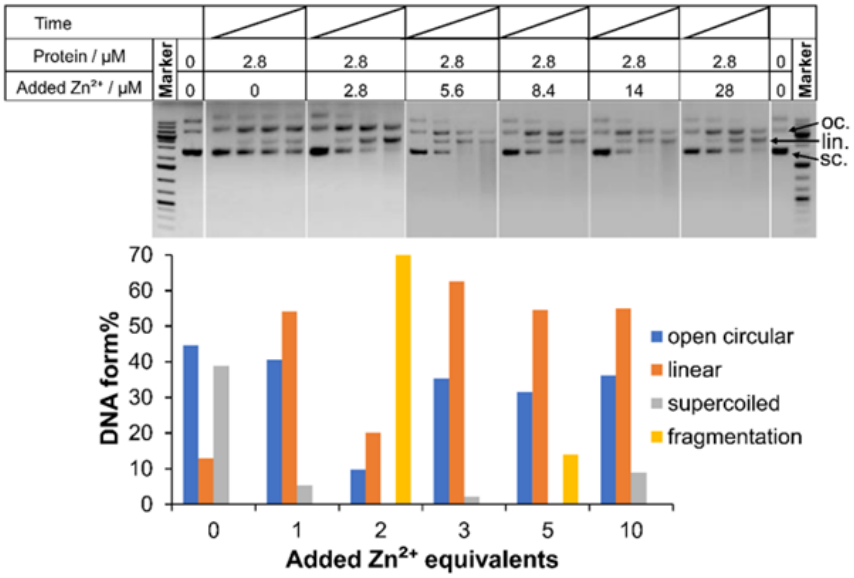


Figure 5.8. Top: Agarose gel electrophoresis images showing the cleavage of pUC-EGFP (74 μM for nucleotide pairs) by 2.8 μM KGNK-6×His protein containing increasing amounts of Zn^{2+} ions. Each column in one experiment represents the catalytic activity after 0, 15, 60, 120 min, from left to right. Gene Ruler 1kb Plus DNA ladder was loaded as a reference, and the untreated pUC-EGFP plasmid for negative control. **Bottom:** Evaluation of the gel pictures by ImageJ program [165] for quantitation of the various forms of DNA after 120 min.

The purified KGNK-6×His protein has a low nuclease activity. The band of the supercoiled form of the plasmid starts to decrease in intensity, while the band related to the linear form also appears within 2 hours of incubation (**Figure 5.8**). Since the active center of the purified protein contains Zn^{2+} ions, this low activity is attributed to the inhibitory effect of the affinity tag. By adding Zn^{2+} ions to the reaction mixture in increasing amounts we could observe a significant increase in the catalytic activity. The maximum of the DNA cleavage activity was achieved between ~ 1:3 and 1:5 protein-to-Zn^{2+} molar ratio. Under these conditions, the band of the supercoiled form completely disappears at 60 min; even the band related to the open circular form shows a decreased intensity and smearing of the band of the linear form occurs as a result of the fragmentation due to efficient catalytic process. After a further increase of the metal ion concentration, at 1:10 protein-to-Zn^{2+} molar ratio, the catalytic activity starts to decline again.

These results are in agreement with the observation that the catalytic activity of the KGNK-6×His protein is decreased compared to the KGNK protein (see e.g. the results of the cloning experiments for comparison), due to the inhibition effect of the C-terminal tag.

However, the increase of the DNA cleavage ability of the enzyme upon increasing metal ion-to-protein ratio suggests that the added metal ion can compete for the 6×His sequence, which originally may bind to the free site of the catalytic Zn^{2+} ion (**Figure 5.9 a**) accounting for the coordinative inhibition. Thus, by binding to the 6×His sequence, the added metal ion enhances the catalytic activity.

By saturating the histidines outside the catalytic center with one or two Zn^{2+} ions, H545 may also be metallated on further addition of metal ion, preventing the generation of the OH^- nucleophile by the side-chain of this histidine residue. The higher is the metal ion excess the higher is the probability of this type of coordination. This would cause the decrease of the catalytic activity under such conditions. It can not be excluded that the second or third metal ion, which most probably binds to the hexahistidine sequence, may also coordinate to H545, but the extensive formation of such complex would result in inhibition of the enzyme instead of the activation as observed in the catalytic experiments.

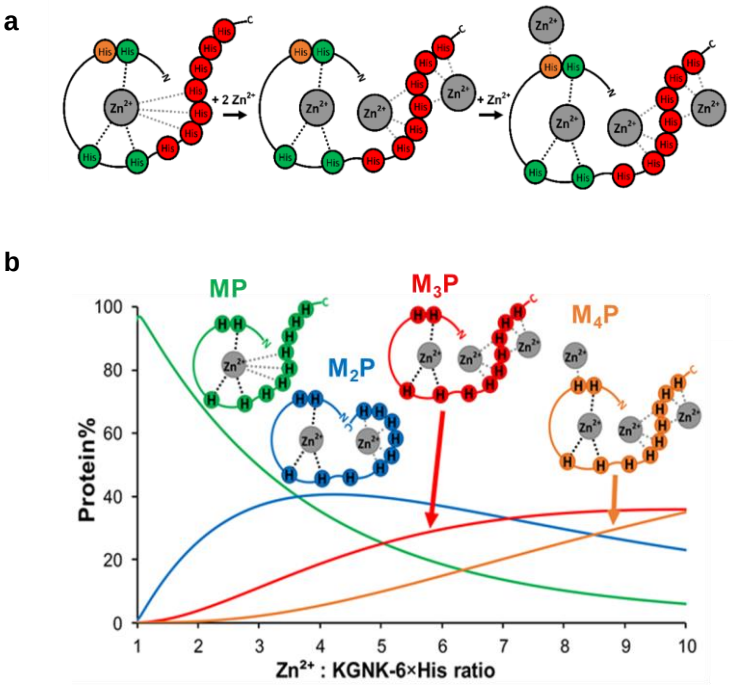


Figure 5.9. a, Coordination of histidine residues inside the KGNK-6×His to an increasing number of Zn^{2+} ions in parallel with the increase of the Zn^{2+} : protein molar ratio. The histidines, coordinating to Zn^{2+} ion inside the nuclease active center, are green. H545 is orange, while other histidines outside the native NColE7 sequence are marked red. **b,** Calculated species distribution diagram of the Zn^{2+} – KGNK-6×His (P) system based on the estimated apparent stability constants related to the various suggested metal ion binding sites (c_P = 2.8 μM).

We attempted to simulate these complex formation processes by estimating and refining the apparent stabilities of the Zn_xP species to obtain the species distribution best describing the catalytic experiments. Up to four bound metal ions were included considering the number of

binding sites potentially offering stability, which is high enough to enable the complex formation in micromolar concentration range. The log*K'* values obtained in this way are ~9.0 for ZnP; ~5.3 for ~Zn_2P; ~4.9 for Zn_3P and ~4.7 for Zn_4P. The first value is in agreement with those published for the Zn^{2+} affinity towards the HNH active center [163,164]. The values for Zn_2P and Zn_3P are consistent with Zn^{2+} coordination towards the oligohistidine peptides binding one or two metal ions [166-168]. The hypothetical distribution diagram **(Figure 5.9 b)** shows that the formation of Zn_4P species becomes significant at >5 fold Zn^{2+} excess. Since under such conditions the inhibition of the DNA hydrolysis can be observed in accordance with the literature data on NColE7 itself [169], we suppose that the coordination site in Zn_4P is most probably the H545 residue. Taking this into account, the stability related to the binding of the fourth metal ion is fairly high. An explanation of this could be a stability enhancing effect of further amino acid side-chains available for metal ion binding together with H545.

Multiple explanations can account for the observations so far: (i) The cooperative action of the N- and C-termini is required for NColE7 to hydrolyze DNA. This kind of intramolecular allosteric activation mechanism may be sterically hindered by the C-terminal protein tag. This phenomenon would allow the fine tuning of the enzyme by optimizing the length of the affinity tag. The tagged enzyme can also be activated by the tag removal from the protein via an optimally positioned hydrolytically sensitive amino acid sequence either by using a specific protease or by Ni(II)-induced cleavage. (ii) The metal ion containing active center at the C-terminus of NColE7 is close enough to the hexahistidine sequence to allow the coordination of this motif to the free site of the catalytic Zn^{2+} ion. This event would also inhibit the nuclease action, as it prevents the metal ion coordination to the scissile phosphodiester group, which is an essential step of the DNA cleavage. In such a mechanism of inhibition the enzyme could be reactivated by the addition of metal ions, which can effectively compete with Zn(II) for the hexahistidine binding site, but do not replace Zn(II) in the active center. (iii) Furthermore, the hexahistidine tag may also form a ternary complex with Zn^{2+} ions including the H545 residue, which is responsible for the formation of the nucleophilic attacking hydroxide in the catalytic reaction.

5.2.4. Mass spectrometric analysis of the metal ion binding properties of the KGNK-6×His

ESI-MS experiments were carried out to verify the feasibility of the hypothesis described in **Figure 5.9 a** concerning the formation of mono and oligonuclear complexes of KGNK-6×His. The experiments were executed by gradually increasing the concentration of the

metal ion up to ~ 4 equivalents of Zn^{2+} relative to the protein. The results of the measurements are presented in **Figure 5.10**. As the ICP-MS measurements indicated, the purified KGNK-6×His protein binds approximately one equivalent of Zn^{2+}. The mass spectrum without adding additional metal ion (**Figure 5.10 *a***), shows the presence of the ZnP monocomplex as the major species, while the Zn_2P complex containing two metal ions also appears.

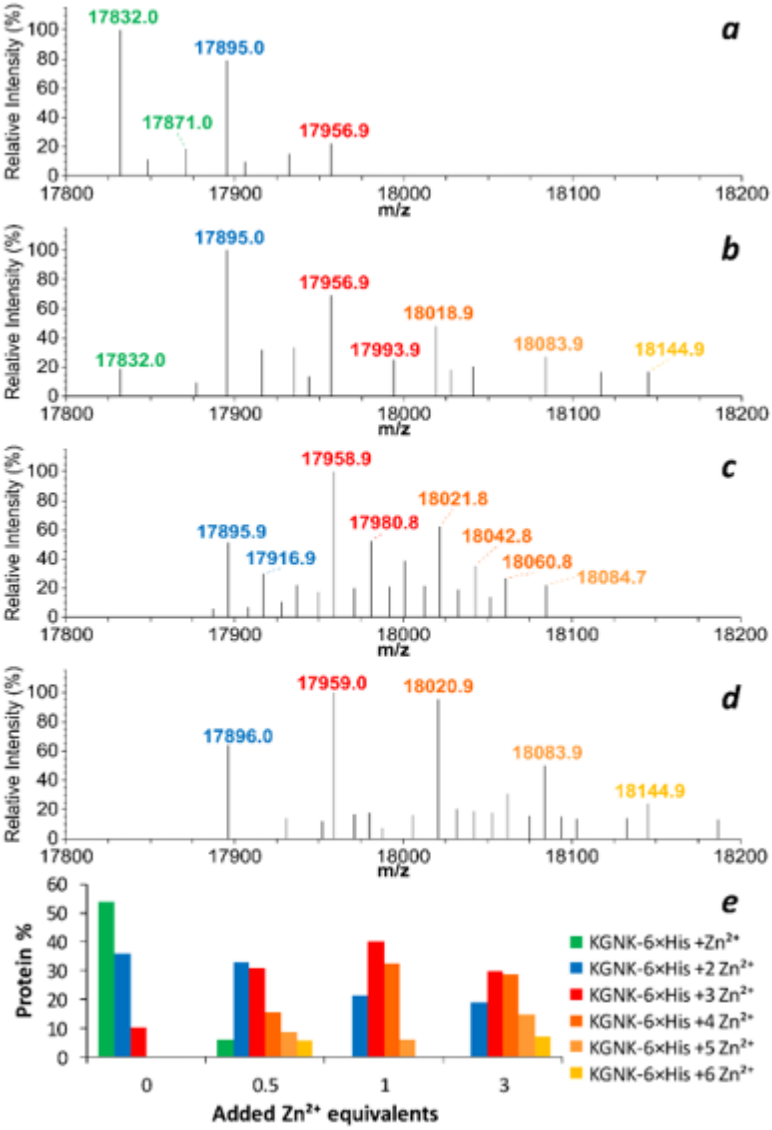


Figure 5.10. Deconvoluted monoisotopic MH^+ m/z spectra of the KGNK-6×His protein solution containing increasing amounts of Zn^{2+}. **(*a*)** The protein solution with approximately 1 equivalent of Zn^{2+}. **(*b*)**, **(*c*)** and **(*d*)** after the addition of further 0.5, 1.0 and 3.0 equivalents of Zn^{2+}, respectively. Panel **(*e*)** shows the estimated relative abundance of the various metal complexes of the protein as a function of the added Zn^{2+} equivalents. Green colour refers to the ZnP (Zn^{2+} : protein = 1 : 1) monocomplex; Zn_2P is marked with blue, Zn_3P with red; while the orangish colours refer to the further Zn^{2+} adducts containing more metal ions (Zn_4P-Zn_6P).

The mass spectrometric experiments revealed that the KGNK-6×His protein is able to bind more metal ions. Up to 6 bound Zn^{2+} ions have been detected by ESI-MS. Based on these

measurements the relative amounts of the complexes containing various number of metal ions were estimated and plotted versus the added Zn^{2+} equivalents (**Figure 5.10** ***e***). For this, we identified the major peaks of the mass spectra in **Figure 5.10** ***a-d*** and assigned them to the potential Zn^{2+}–KGNK-6×His species, which could be present in the solution (from the monocomplex till the six Zn^{2+} containing adduct). The list of the peaks considered for the calculation of the relative abundance diagram is shown in **Table 5.2**.

Table 5.2. List of the deconvoluted monoisotopic MH^+ ions, which were used to obtain the estimated relative abundance diagram in Fig. 4*e*. **(a)** column refers to the purified protein solution containing approximately 1 equivalent of Zn^{2+} ions. Data in **(b), (c)** and **(d)** columns refer to protein solutions with additional 0.5, 1.0 and 3.0 equivalents of Zn^{2+} ions, respectively. The differences between the calculated and measured masses are due to the deconvolution process. In some cases Na^+ and K^+ adducts were observed, which were also taken into account during the calculations.

Species	$MH^+_{(mono)}$ calc:	$MH^+_{(mono)}$ measured:			
		(a)	(b)	(c)	(d)
KGNK-6×His +Zn^{2+} -$1H^+$	17832.08	17832.00	17832.02		
-$1H^+$ + K^+	17871.05	17870.95			
KGNK-6×His +2 Zn^{2+} -$2H^+$	17895.00	17894.97	17895.00		
-$1H^+$	17896.01			17895.92	17895.96
-$3H^+$ + Na^+	17916.99			17916.92	
KGNK-6×His +3 Zn^{2+} -$2H^+$	17956.92	17956.94	17956.95		
-$4H^+$	17958.93			17958.87	17958.95
-$3H^+$ + Na^+	17980.91			17980.85	
-$6H^+$ + K^+	17993.94		17993.87		
KGNK-6×His +4 Zn^{2+} -$3H^+$	18021.85			18021.80	
-$4H^+$	18020.85				18020.91
-$6H^+$	18018.83		18018.93		
-$5H^+$ + Na^+	18042.83			18042.78	
-$3H^+$ + K^+	18060.75			18060.82	
KGNK-6×His +5 Zn^{2+} -$4H^+$	18084.78			18084.75	
-$5H^+$	18083.77		18083.91		18083.88
KGNK-6×His +6 Zn^{2+} -$8H^+$	18144.67		18144.93		18144.88

We summed up the intensities of the peaks belonging to the same species at each titration state and then plotted the percentage of these complexes in **Figure 5.10** ***e***. The maximal percentage of the Zn_2P complex of KGNK-6×His, expected to be catalytically most active, is observed in the presence of ~1.0–2.0 equivalents of Zn^{2+}. Compared to the catalytic activity of the KGNK-6×His as a function of the Zn^{2+} concentration, this would infer that the Zn_3P species

could still be active, while the increasing metal ion concentration results in further coordinated Zn^{2+} ions and inhibition of the nuclease action. Although some of the detected complexes may not form in significant amounts in dilute aqueous solutions and under the conditions applied in the catalytic experiments, these results support the considerations made in **Figure 5.9** and the clearer interpretation of the catalytic properties of the enzyme, too.

5.2.5. Influence of the His tag on the structural integrity of KGNK-6×His

Temperature dependent SRCD measurement series were carried out with the KGNK-6×His protein in the presence and absence of metal ion with the aim to study the thermal stability of the enzyme. Earlier it was suggested that the removal of the metal ion decreases thermal stability of NColE7 [170]. The results shown in **Figure 5.11** indicate that the changes in the structure of the KGNK-6×His enzyme do not occur in a single step.

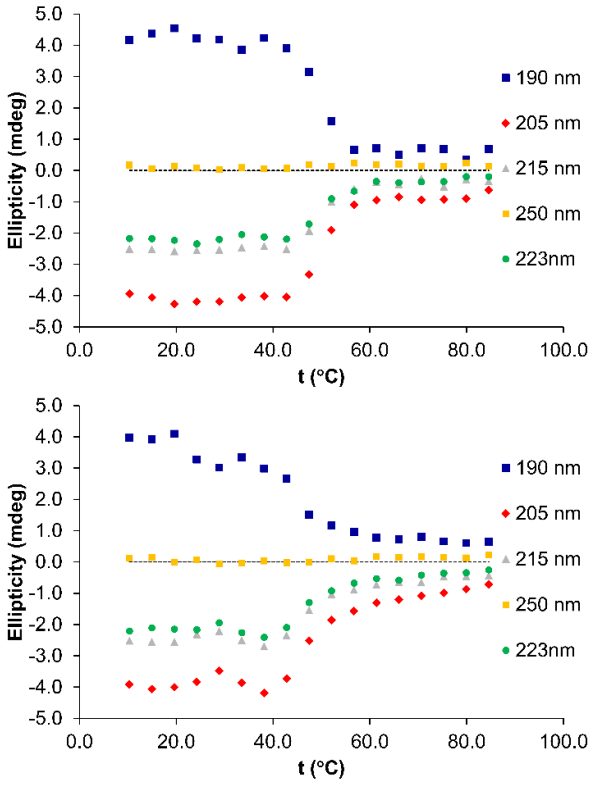


Figure 5.11. Temperature dependent SRCD spectra of the Zn^{2+}-loaded (top) and apo (bottom) forms of the KGNK-6×His protein.

The first conformational change occurred between 25–35 °C and this was significantly affected by the removal of the metal ion from the enzyme. Then the main melting process occurred at around 49 °C. This probably reflects the conformational change of the central helical regions of the protein, which is not significantly influenced by the metal ion. The curves

recorded in the presence of metal ion could be satisfactorily fitted with a single process with a melting point around 50 °C.

The melting process at low temperature is most probably related to the HNH motif, which might be slightly more flexible and thus, more sensitive to the increasing temperature in the absence of the metal ion. A plausible explanation may be that the C-terminal fusion tag destabilized the interaction between the two termini of NColE7, which is responsible for the stabilization of the structure of the preformed catalytic site [171]. This effect is, however, influenced by the metal ion, suggesting that the coordinated His tag has a negligible disturbing effect on the structure of the protein. This also supports the above observations, that the complex effect of the 6×His sequence involves steric hindrance, but it also inhibits the enzyme activity by binding to the metal ion in the catalytic center.

5.3 DNA targeting by artificial nucleases

One of the aims of my PhD study is to design an artificial nuclease that can efficiently target a selected (proto-)oncogenic DNA sequence in human genome. For the optimal artificial nuclease design, a suitable DNA target sequence has to be selected, which is essential for the protein expression from an oncogene in a cancerous disease. Targeting of the selected sequence is straightforward with the CRISPR/Cas9 artificial nuclease system, as its DNA recognition is defined by the oligonucleotide sequence of the guide RNA, that hybridizes to the target DNA sequence. In the experiments described in this chapter, I have studied the ability of the CRISPR/Cas9 system to target selected DNA sequences within the epidermal growth factor receptor (EGFR) oncogene in A549 human lung cancer cells.

5.3.1. Construction and optimization of CRISPR/Cas9 system for targeting the epidermal growth factor receptor (EGFR, ErbB1) oncogene

Inside the CRISPR/Cas9 complex, the gRNA must be properly formulated to target the correct region of genomic DNA. The gRNA binds to the corresponding DNA sequence allowing the Cas9 enzyme to cleave the genome close to the PAM site, as described in more detail in the literature review, section 2.3.3. The pX458 plasmid was chosen for our study because it has already been successfully applied earlier [59]. It provides ampicillin resistance as a selectable marker in *E. coli* and offers green fluorescent protein (GFP) expression for tracking in mammalian cells. A map of pX458 is shown in **Appendix A.3**. This plasmid allows for expression of Cas9 and for the transcription of the CRISPR single guide RNA (sgRNA) simultaneously.

The sgRNA sequences hybridizing to the DNA target sequences within the epidermal growth factor receptor (EGFR) oncogene (Homo sapiens transcript variant 1, mRNA) in A549 cancer cells were designed by Guide Design software (http://crispr.mit.edu/) as shown in **Table 5.3**. These exons to be targeted were chosen due to their role in the EGFR oncogene function [172,173]. I also selected one targeted site on exon 17 based on the literature, since it demonstrated high genome editing efficiency to be used as a positive control [174]. The crispr.mit.edu database compares the possible target 20 base pair (bp) DNA sequences to the rest of the human genome to determine their frequency of appearance. The sequences are then given a score from 0-100%. The higher the score the less often the target is found elsewhere, meaning that using this sequence in the guide RNA would lead to very few off-target interactions. Sequences with a score over 85% were saved for use in the experiments. The sgRNA Exon 21* (score 67%) was chosen to verify the effect of the score parameter on the

CRISPR/Cas9 activity later. The position of the selected sequences in the gene of the EGFR is shown in **Appendix A.4**.

The gRNAs include an additional G base at the 5' termini of the selected sequences and the forward and reverse sequences have CACC and CAAA 5' overhangs, respectively. The remaining 20 nucleotides hybridize with a 20 nt target sequence upstream of the PAM sequence in the oncogene. The forward and reverse gRNA sequences were annealed together as described in **Appendix A.5**.

Table 5.3. List of the designed sequences of CRISPR single guide RNA for EGFR oncogene targeting. Exon 19M was selected from the literature [172], the target of this sgRNA is located adjacent to a PAM sequence only in mutant EGFR, but not in the normal gene. Thus, Cas9 activity is only expected to occur at the mutant locus.

Single guide RNA name	Single guide RNA sequence and overhang	Score percentage
Exon 1-CGG	5′-caccgCTGCGCTCTGCCCGGCGAGT-3′ 3′- cGACGCGAGACGGGCCGCTCAcaaa-5′	90%
Exon 1-GGG	5′-caccgTGCGCTCTGCCCGGCGAGTC-3′ 3′- cACGCGAGACGGGCCGCTCAGcaaa-5′	90%
Exon 17	5′-caccgAGATCCCGTCCATCGCCACT-3′ 3′- cTCTAGGGCAGGTAGCGGTGAcaaa-5′	89%
Exon 19	5′- caccgTAAAATTCCCGTCGCTATCA-3′ 3′- cATTTTAAGGGCAGCGATAGTcaaa-5′	94%
Exon 19M	5′-caccgATTAAGAGAAGCAACATCTC-3′ 3′- cTAATTCTCTTCGTTGTAGAGcaaa-5′	n.d.
Exon 20	5′-caccgAGCCTACGTGATGGCCAGCG-3′ 3′- cTCGGATGCACTACCGGTCGCcaaa-5′	86%
Exon 21*	5′-caccgCAAGATCACAGATTTTGGGC-3′ 3′- cGTTCTAGTGTCTAAAACCCGcaaa-5′	67%
Exon 21**	5′-caccgATACTTGGAGGACCGTCGCT-3′ 3′- cTATGAACCTCCTGGCAGCGAcaaa-5′	97%

The pX458 plasmid was digested by BbsI restriction endonuclease as shown in **Figure 5.12**.

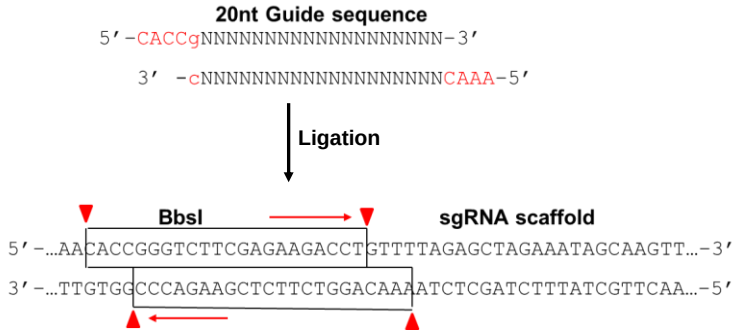


Figure 5.12. Schematic representation of cloning of the guide sequences into the cloning site of the pX458 plasmid.

The BbsI recognition sequence is located within the cloning site for sgRNA scaffold, as it is shown in detail in the literature review (**Section 2.3.3**). The success of the pX458 digestion was checked by 0.7% agarose gel electrophoresis: a single band (linear form) appeared at the expected size of about 10 kbp (lane 3 in F**igure 5.13**).

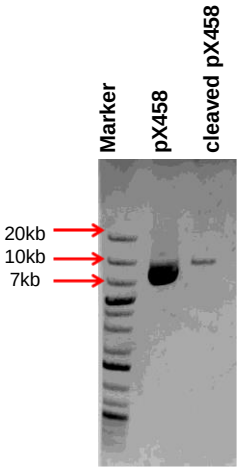


Figure 5.13. Agarose gel electrophoresis of digested pX458 plasmid.

The hybridized gRNA oligonucleotides were ligated with the cleaved plasmid by T4 ligase, then the resulted plasmids were applied to transform DH5α competent *E. coli* for multiplication purpose. The purified plasmids were sequenced by using pX-F primer (5'-gagggcctatttcccatgattccttc-3') to confirm the successful insertion.

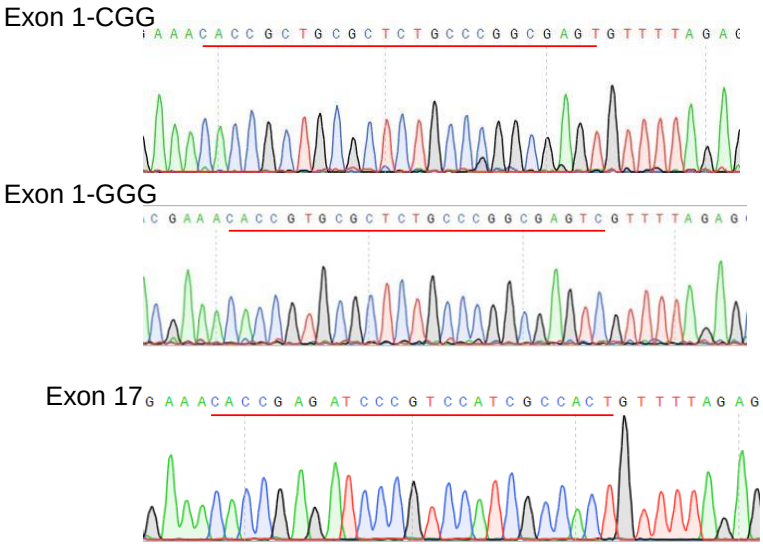


Figure 5.14. Selected electropherograms of the sequenced plasmids. Underlined regions illustrate the presence of CRISPR single guide RNA sequences successfully cloned into pX458.

Based on the results shown in **Figure 5.14**, the CRISPR/Cas9 constructs (pX458-E1-CGG, pX458-E1-GGG and pX458-E17-GGG) have been successfully generated and sequenced within plasmids that contain the chimeric sgRNA, Cas9 gene, and GFP gene.

5.3.2. Screening of CRISPR/Cas9 system on mammalian cell culture

5.3.2.1.Transfection of the cancer cells

A549 human lung cancer cells were selected for the transfection experiments because they overexpress EGFR. Targeting the EGFR oncogene should influence the amount of the expressed protein. The transfection efficiency, i.e. the percentage of cells that are transfected compared to the entire population, was checked by flow cytometry. In this experiment, the cells expressing the GFP are detected. According to the results, the transfection efficiency by pX458 plasmid which contains the guide RNA reached up to 68% within few optimization steps (**Figure 5.15**). This is a reasonably high value to enable the CRISPR/Cas9 system to successfully perform its genome editing function

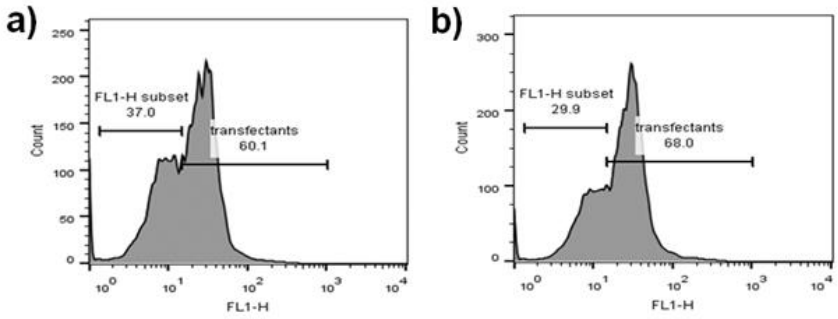


Figure 5.15. Measurement of GFP expression by flow cytometry in A549 cells transfected by pX458 plasmid using Lipofectamine 3000 transfecting agent. **a)** transfection efficiency of the original pX458 and **b)** transfection efficiency with the pX458-E1-GGG after inserting the guide RNA.

5.3.2.2. Genomic DNA extraction and sequencing

The genomic DNA was extracted from the transfected cells, as well as from the control cells including the non-transfected cells after incubation for four and seven days. For this, we used DNeasy Blood and tissue kit (Qiagen). The direct sequencing of the targeted DNA segments to check the EGFR oncogene knockout efficiency of the applied CRISPR/Cas9 systems was not successful, because this DNA preparation did not provide enough template. Therefore, the target sites were amplified in PCR using Dream Taq-DNA polymerase and specifically designed primer pairs for the different targeted sites **(Appendix A.6)**. These primer pairs were complementary to the upstream and downstream sequences around the target

sequences. The PCR products were sequenced by Sanger sequencing to determine whether the targeted gene was edited.

CodonCode Aligner software was applied for sequence data analysis [175], and the TIDE (Tracking of Indels by Decomposition) tool [176] was applied for the analysis of CRISPR/Cas9 activity. TIDE determines the spectrum and frequency of targeted mutations generated in a pool of cells by e.g. the CRISPR/Cas9 system. The analysis is based on the input of CRISPR single guide RNA sequence, control sample sequence and transfected sample sequence (using *.ab1 extension). The web tool reports the identity of the detected indels and their frequencies.

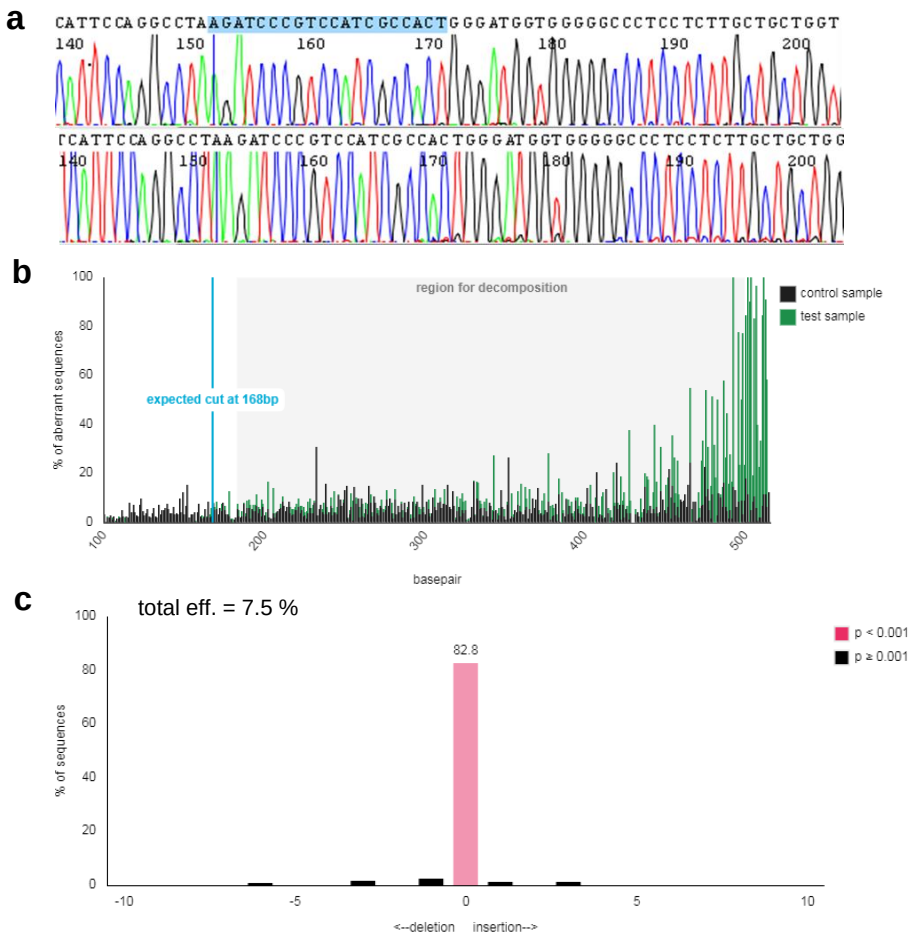


Figure 5.16. a, CodonCode Aligner sequences of EGFR demonstrating CRISPR/Cas9 activity after seven days transfection. Top sequence represents control sequence without gene knockout. Highlighted region shows the region complementary to CRISPR single guide RNA. Bottom sequence shows a slight increase of the background noise, due to the low extent of the DNA modification. **b,** TIDE data demonstrate the CRISPR/Cas9 activity by showing level of decomposition and aberration in sequences. Dark green sample represents the control sequence showing no decomposition. Light green represents test sequence illustrating slight increase in decomposition **c,** TIDE data show the probability of insertions or deletions, which correlates to mutation efficiency observed to be 7.5%. 82.8% represents the wild type DNA without any mutation.

According to the sequencing results of the DNA from the control and the transfected samples, there is a tiny increase in the noise beyond the complementary region of single guide RNA due to the effect of mutations in this portion of the sequence. TIDE analysis detected differences between the control and CRISPR/Cas9-treated sequences. It revealed a mutation efficiency of 7.5% for sgRNA targeting exon 17 after seven days incubation period with the CRISPR/Cas9 as shown in **Figure 5.16**. This validates the specificity of the designed guides, however, the efficiency of this system is low and requires further optimization is required in the future.

5.3.2.3. RNA extraction and qRT-PCR analysis

Total RNA was extracted from the cells transfected with CRISPR/Cas9 plasmid targeting exons 1 and 17 of the EGFR gene and with the empty CRISPR/Cas9 plasmid (mock), as well as from the untransfected control cells **(Appendix A.7)**. Quantitative real-time PCR (qRT-PCR) was applied to measure the amount of the transcribed RNA and thus, to detect the knockout level of the target gene in A549 cells. Various sets of designed primers were used for exons 1 and 17 in addition to the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) encoding the positive control sequence as a reference gene (**Table 5.4**). The expression level of EGFR as the target gene was normalized to an average of GAPDH gene.

Table 5.4. Details of the oligonucleotide sequences used for qPCR analysis.

Gene	Forward and Reverse primers
GAPDH	GTCTCCTCTGACTTCAACAGCG ACCACCCTGTTGCTGTAGCCAA
EGFR-Exon 1	AGTCGGGCTCTGGAGGAAA AGCTCCTTCAGTCCGGTTTT
EGFR-Exon 1*	AAGGATCCTGACTCCGTCCAGTAT TTGGATCCTGTAAATTTCTCATGGGC
EGFR-Exon 17	TGCCATCCAAACTGCACC TGCGCTTCCGAACGATGT

According to the results, the CRISPR/Cas9 system targeting EGFR exon 17 induced 20% decrease in EGFR mRNA expression level after four days incubation period (**Figure 5.17**). The level of EGFR gene expression was most probably decreased due to the small insertions or deletions (indels) as detected in DNA sequencing experiments (**Figure 5.16**). These mutations within the target sites may induce the shift of the reading frame and/or premature stop codons, leading to the expression of a truncated protein of target gene or

preventing gene expression [177]. Promising preliminary results have been obtained with other targets, as well, but these experiments need further optimization.

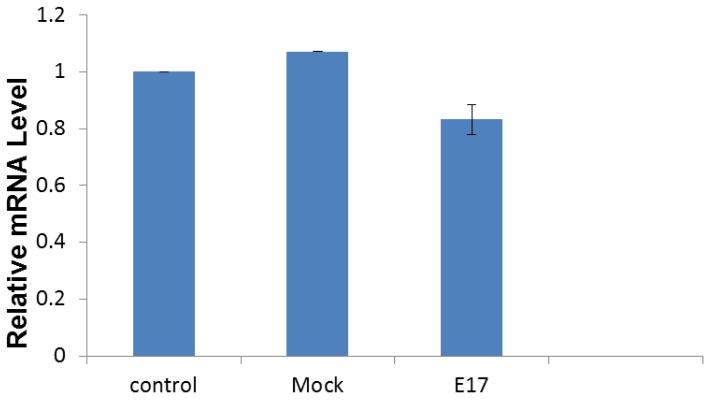


Figure 5.17. EGFR knockout efficiency was determined by qRT-PCR in A549 cells transfected by CRISPR/Cas9 target the EGFR oncogene following the recommended transfection protocol. EGFR mRNA expression levels were measured four days post-transfection. Values are normalized to untreated sample. Data are presented as means ± SD.

5.4. Delivery of the artificial nucleases into the mammalian cells

The success of the selected genome editing technique is influenced by the effectivity of the delivery system used to carry the nuclease to the specific target cells. By improvement of the delivery and specificity of the ZFNs and CRISPR-Cas9 system, the current challenges could be overcomed, promoting this technology to be used therapeutically. Chemical delivery methods are being considered to be the most promising. For this reason, a novel cationic water-soluble lipopolymer (WSLP) was applied to encapsulate the pDNA for transfection in this work. The water-soluble cationic lipopolymer (WSLP) was obtained by the modification of high molecular weight (HMW) branched poly(ethylenimine) (PEI) with cholesteryl chloroformate. It was characterized and assessed for better cellular membrane permeability. This part of the work was conducted in strong collaboration with two research groups. Researchers Dr. Edit Csapó and Dr. Anita Ditta Ungor provided facilities and helped with the measurements on the nanoparticle characterization, while Dr. Monika Kiricsi, Nóra Igaz and Mohana Krishna Gopisetty helped with the experiments using cancer cell lines.

5.4.1. Synthesis and characterization of water-soluble lipopolymer

WSLP was prepared from PEI and cholesteryl chloroformate as described in the experimental section. The reaction is schematically depicted in **Figure 5.18**. Due to the balanced ratio between the hydrophilic PEI and lipophilic cholesterol, the prepared WSLP is well soluble in water. The critical micelle concentration (CMC) of the WSLP was determined to be 0.35 ± 0.03 mg/mL by using dye solubilization method as it is described in more detail in the **Appendix A.9**.

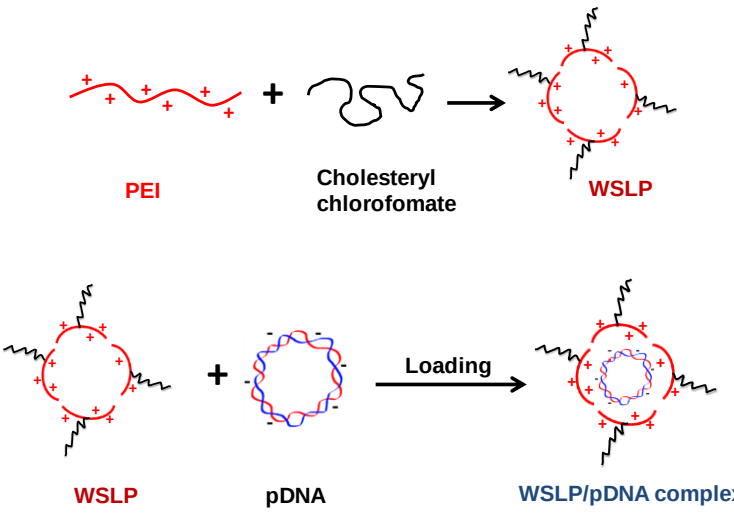


Figure 5.18. Schematic representation of the WSLP synthesis and the complex formation with the pDNA.

5.4.1.1. Particle charge detection

The average molecular weight of PEI was not provided and therefore, we had to determine it. For the average molecular mass of PEI M_w ~ 139±1 kDa was obtained from DLS measurements (it can be found in the **Appendix A.8**), however, the same method was not suitable for the determination of the molecular mass of WSLP, because of the artefacts arising from the light scattering of some aggregates. The WSLPs obtained by cholesterylation of low molecular weight PEI are usually characterized by NMR and MALDI TOF MS in terms of the molecular weight and cholesterylation extent [104,133,178-184]. These methods perform well with small and medium size molecules, but become of limited use for compounds with the molecular weight with several hundreds of kDa. The broadening effect of the NMR bands due to the slow molecular tumbling can already be experienced with the WSLP having M_w > 10 kDa [105,185-186] while according to our knowledge no MS data are published for high molecular weight PEI or WSLP. Therefore, we were looking for a readily available simple method for checking the cholesterylation extent. Since cholesterylation or the modification in general, may affect the specific charge of the polymer both through the reaction with the amine groups and through increasing the molecular weight, such measurements could provide useful information, independently of the size of the polymer. The specific charge of the polymers was determined by the titration of sodium dodecyl sulfate (SDS) by PEI and WSLP, and checking the charge neutralization points. As it can be seen in **Figure 5.19**, the negative streaming potential continuously increased by the increase of the amount of the polymers. Based on the measured data, the charge compensation occurred upon adding of 6.5 mg PEI or 14.8 mg WSLP into the particle charge detector (PCD) cell. Knowing the applied amount of SDS, the specific charges of PEI and WSLP were calculated, to be q(PEI) = 15.38 mmol g^{-1} and q(WSLP) = 6.76 mmol g^{-1}, respectively. Using these data and the molecular weight of the initial PEI, the protonated fraction of PEI can be estimated. The theoretically maximal specific charge of the applied PEI polymer is 23.22 mmol g^{-1}, suggesting that ~ 66% of the nitrogens are protonated in PEI under the measurement conditions. This is in very good agreement with the suggested values published in the literature for branched PEI molecules determined from acid-base titrations [187].

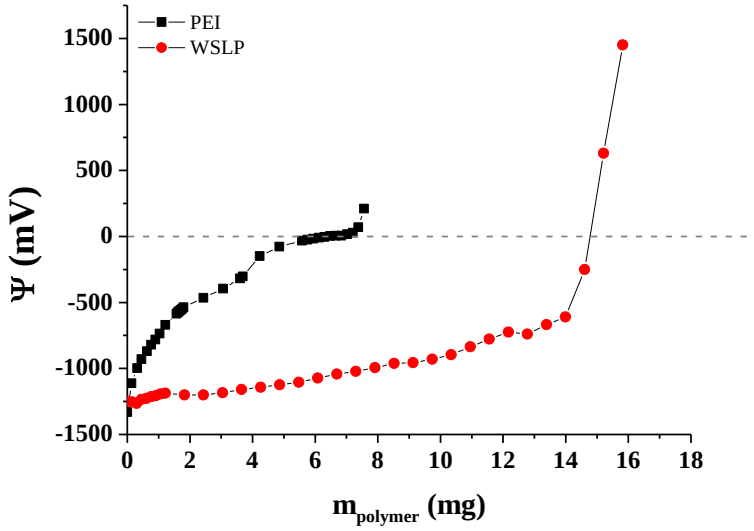


Figure 5.19. Titration of SDS with PEI or WSLP solutions for the detection of the charge neutralization point in a charge particle detection experiment, as described in the experimental section.

It can also be concluded that the specific charge has decreased in the cholesterylation step. The ratio of q(WSLP)/q(PEI) is ~ 0.44, meaning that the WSLP has 56% less specific charge than PEI. Since the specific charge (q) depends on the molecular weight and the number of charges in a single molecule (n^+)the decrease of the specific charge may occur both by the increase of M_w and decrease of the number of the protonated nitrogens according to **Equation 5.1**.

$$q\left[\frac{mmol}{g}\right] = \frac{1}{M_w} n^+ 1000 \qquad (5.1)$$

The cholesterylation occurs at more reactive primary amine groups which have to be deprotonated under the reaction conditions, but they might be protonated in a buffer solution at pH ~ 7. An initial PEI molecule with M_w of 139 kDa contains 3227 nitrogens, out of which ~ 2138 are protonated based on the measured specific charge. The increase of the M_w may play major role in the decrease of the specific charge of WSLP compared to PEI, if the number of protons at the nitrogens does not change as the consequence of the modification. In this case the increased M_w can be calculated from Equation 1, based on the specific charge value of the WSLP. This calculation yields ~ 316.5 kDa as the new Mw. This molecular weight is the sum of the original 139 kDa of the PEI and the mass increase due to the cholesterylation, i.e:

316492 = 139000 + x(449.11 – 35.45). Here x is the number of the conjugated cholesteryl groups per a PEI polymer, which can be easily calculated to be x ~ 429. This reveals that approximately 13% of the nitrogens have been cholesterylated. On the other hand, we also may carry out the calculation assuming that the charge on the new WSLP molecule is decreased by the number of the cholesteryl groups attached to the nitrogens of PEI. In this case **Equation 5.2** will be used for the calculation of x in its following form:

$$6.76 = \frac{1}{139000+x(449.11-35.45)}(2138-x)1000 \qquad (5.2)$$

From this, x ~ 316 is obtained. Thus, ~ 10% is suggested for the possible fraction of the cholesterylated nitrogens. The molecular weight of the synthesized copolymer is ~ 269.7 kDa by this assumption. The real number of the attached cholesteryl groups might be between the above two values assuming that no further deprotonation of the WSLP is caused by the modification. This calculation provides closer estimates for the possible domains of important parameters such as the molecular weight and the charge content of the newly prepared polymer.

5.4.2. pDNA condensation efficiency of the water-soluble lipopolymer

WSLP/pDNA complexes are usually prepared by adjusting the mass or N/P ratios. It has to be mentioned that the two types of preparation can not be directly compared with each other. The N/P ratio is the ratio of the number of nitrogen atoms of WSLP and the number of the phosphorous atoms of plasmid DNA. The number of nitrogen atoms is obtained by dividing the molecular weight of the polymer with the molecular weight of the monomer (there is one N atom in a monomer) and P atoms are calculated as 2000×kb of the plasmid DNA. For this, the knowledge of the molecular weight, i.e. the extent of the modification of PEI, as well as the protonation state of the polymer is essential. These parameters are usually very difficult to determine precisely.

We optimized the loading efficiency of WSLP using pEGFP-N3 and pX458 plasmids within both the mass and the N/P ratios ranging from 1.25 to 20.0 as it is described in more detail in the **Appendix A.10**. DNA encapsulating ability of WSLP was monitored by agarose gel electrophoresis. The results on N/P ratio series are shown in **Figure 5.20**.

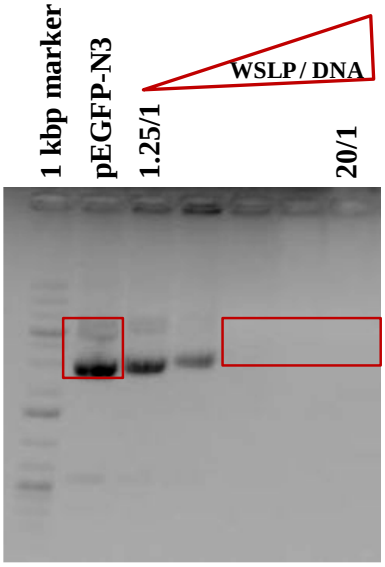


Figure 5.20. Agarose gel electrophoresis picture of naked pEGFP-N3 plasmid and WSLP/ pEGFP-N3 complexes formed at N/P ratios 1.25, 2.5, 5, 10 and 20 in lanes 3–7, respectively.

On an increase of the N/P ratio, free DNA could not be detected on agarose gel when the value of N/P ratio reached 5/1, confirming the effectiveness of condensation by the polymer. This is due to the fact that positively charged head groups of WSLP interact strongly with the negatively charged phosphate ions in the backbone of plasmid DNA independently of its nucleotide sequence. Nevertheless, the specific charges (ratio of the charge of anion or particle for its mass) put the case more clearly, since the specific charge of the DNA which should be 3.24 mmol g^{-1}, can directly be compared with that of the polymer. At 5/1 N/P ratio the reaction mixture was constructed by mixing 1.00 μg of pDNA with 0.76 μg of WSLP. Based on the specific charges this corresponds to a positive/negative charge ratio of 5.14/3.24, i.e. the total charge of the WSLP/pDNA system is positive, while at N/P = 2.5/1 is still negative. This is in good agreement with the fact that the band of the DNA can be detected at N/P = 2.5/1 ratio on the agarose gel, although its mobility is less than that of the free DNA. At the same time the DNA band at N/P = 5/1 ratio does not appear on the gel. At the same time, the mass ratio of 5/1 would correspond to a positive/negative charge ratio of 33.8/3.24, which means that this particle is already highly positively charged. Thus, by measuring the specific charge of the polymers the preparations using various types of ratios become comparable with each other.

5.4.3. Characterization of WSLP/DNA complexes

Preparations of WSLP/pEGFP-N3 and WSLP/pX458 complexes with various composition ratios were also investigated by DLS to determine the particle size and ζ-potential. The loading of plasmids into the polymer with defined colloidal and surface properties plays a major role in controlling stability, biodistribution, and intracellular fate of formulated plasmids. In contrast to the large size of naked DNA [185] complex formation between WSLP and plasmid DNA resulted is small size particles. For WSLP/pEGFP-N3 complexes, the mean particle size was in the range of 53.1 to 268.6 nm depending on the N/P ratio. ζ-potential of WSLP/DNA complexes increased with the increase of the charge ratio. At the N/P ratio of 2.5/1, it was –22.4 mV, but increased to +15.3 mV when formulated at N/P ratio of 20/1 as shown in **Table 5.5**. The size and ζ potential values for WSLP/pX458 complexes differ from the WSLP/pEGFP-N3 complexes as shown in **Table 5.6.**

Table 5.5. The average size and ζ -potential values of WSLP/pEGFP-N3 plasmid complexes at different N/P ratios.

N/P ratio	d (nm)	ζ-potential (mV)
1.25/1	53.1 ± 9.5	-22.4 ± 0.4
2.5/1	93.2 ± 9.2	-25.3 ± 0.2
5/1	138.8 ± 24.6	-20.3 ± 1.1
10/1	142.5 ± 14.8	+1.9 ± 0.2
15/1	163.8 ± 39.2	+12.3 ± 0.5
20/1	268.6 ± 19.2	+15.3 ± 1.3
WSLP	36.7 ± 3.5	+61.6 ± 0.4

Table 5.6. The average size and ζ -potential values of WSLP/pX458 plasmid complexes at different N/P ratios.

N/P ratio	d (nm)	ζ-potential (mV)
2.5/1	61.0 ± 4.7	-19.7 ± 4.9
5/1	63.1 ± 7.6	-3.2 ± 3.3
10/1	133.2 ± 5.2	+23.1 ± 4.5
20/1	131.9 ± 3.5	+28.9 ± 1.7

The WSLP/pEGFP-N3 complexes were characterized by SEM and TEM. The recorded images are shown in **Figure 5.21**. Both SEM and TEM detected the particle sizes being in good agreement with the values determined by DLS. The morphology of the WSLP/pEGFP-N3 complex (5/1 mass ratio preparations) is shown in **Figure 5.21A, B**. The significant increase in size of the pDNA/WSLP nanoparticles (> 100 nm; **Figure 5.21D**) in comparison with the WSLP itself (< 50 nm; **Figure 5.21C**) is clearly demonstrated by TEM, supporting the successful condensation of the negatively charged pDNA by the cationic lipopolymer. Also, the pure WSLP in most cases showed fibrous structure on solid surfaces (e.g. copper grid) as can be also seen on the **Figure 5.21C**, while the morphology became nearly spherical-like upon complexation with DNA. The image shown in **Figure 5.21D** suggests a core-shell-like architecture of the pDNA/WSLP complex, which is assumed based on the different electron diffractions of the two components, due to their significantly distinct molecular weight. However, we could not unequivocally confirm such structure upon analyzing more particles (**Figure 5.21E** and **5.21F**).

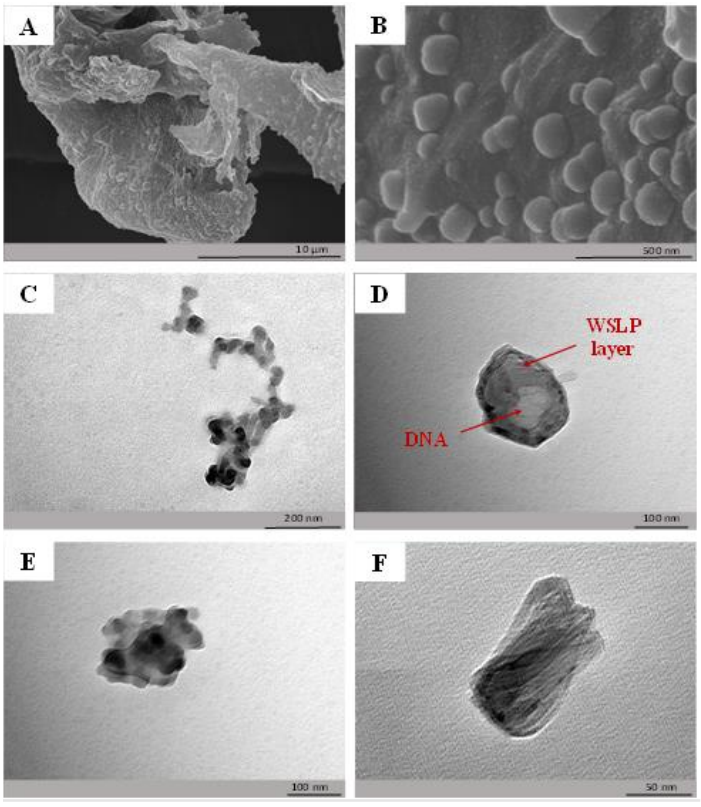


Figure 5.21. A and B: SEM images of the WSLP/pEGFP-N3 complex using 5/1 mass ratio preparations under different magnifications. C: TEM image of the WSLP itself. D, E and F: TEM images of selected WSLP/pEGFP-N3 nanoparticles from the 10/1 NP ratio preparation.

5.4.4. Cytotoxicity assay

The cytotoxicity of WSLP and WSLP/pEGFP-N3 complexes was determined on A549 and HeLa cell lines after 24 h and 48 h incubation at 37 °C by MTT assay as shown in **Figure 5.22.**

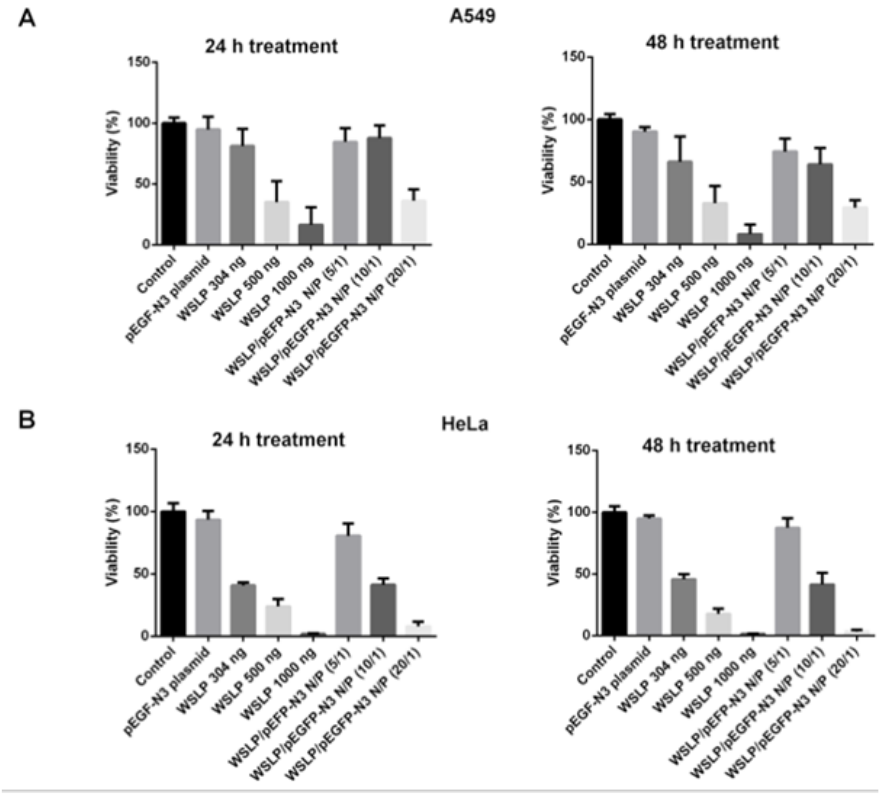


Figure 5.22. *In vitro* cytotoxicity analysis of naked DNA, WSLP and WSLP/pEGFP-N3 at various N/P ratios on (A) A549 and (B) HeLa cells using MTT assay. Viability of the naked DNA, WSLP and WSLP/pEGFP-N3 treated cells was normalized to the untreated controls.

According to the results the polymer is slightly toxic to the cells at 304 ng/well WSLP concentration. However, cell viability decreased significantly upon increase of the WSLP concentration, which indicates a remarkable dose-dependence of cytotoxicity. We have also checked the cytotoxicity of the WSLP/pEGFP-N3 preparations as a function of the ratio of the two components.

The average cell viabilities upon treatment with WSLP/pEGFP-N3 of increasing N/P ratios from 5 to 20, decreased from ~84% to ~36% for A549 cells, and from ~80% to ~8%, for HeLa cells after 24 h. It is important to note that upon association of the WSLP with the pDNA, the toxic effect did not increase compared to the WSLP itself. Cytotoxicity was reasonably low even at 10/1 N/P ratio for A549 cells, nevertheless it depended remarkably on the type of the

applied cell. The microscopic images of the cells treated with WSLP/pEGFP-N3 complexes show that the majority of the cells are destroyed by the excess of the WSLP **(Appendix A.11)**.

We have also compared the toxic effect of the WSLP preparations with the starting HMW PEI, as well as with the commercial standard transfection reagent Lipofectamine 3000. The results are included in **Figure 5.23**, showing that PEI is extremely toxic to the cells, while upon cholesterylation, the resulted WSLP performs similarly to Lipofectamine.

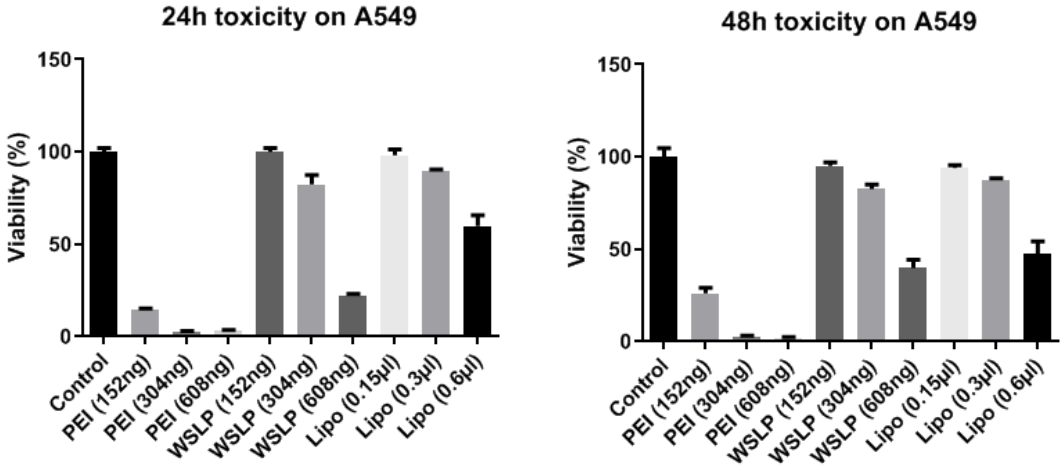


Figure 5.23. Comparison of the cytotoxic effect of WSLP with the high molecular weight PEI starting material and the standard commercial transfection agent Lipofectamine™ 3000 in *in vitro* cytotoxicity analysis of naked on A549 cells at various amounts of the reagents using MTT assay. Viability of the treated cells was normalized to the untreated control. While the numbers shown for PEI and WSLP relate to the mass in ng of the substances per well, for Lipofectamine the 0.3 µL is the amount providing the optimal concentration as suggested by the supplier.

5.4.5. *In vitro* transfection

The transfection efficiency of WSLP/pEGFP-N3 complexes at various N/P ratios was determined in A549 and HeLa cells as shown in **Figure 5.24**. According to the results of flow cytometry, the most efficient transfection was achieved at 20/1 N/P ratio in both cell lines. This can be surprising, since the 20/1 N/P ratio preparation caused increased cytotoxicity in the MTT assays compared to the preparations with lower N/P ratios. This effect is often reported in the literature [133]. It may be explained by the fact that the stably transfected cells may efficiently proliferate during the incubation after the transfection procedure. Indeed, the transfection efficiency reached 36.5% in the case of A549 cells and 64.3% in the case of HeLa cells. These values render this type of WSLP promising for its applicability in *in vivo* transfection experiments after the optimization of the cytotoxic side-effect.

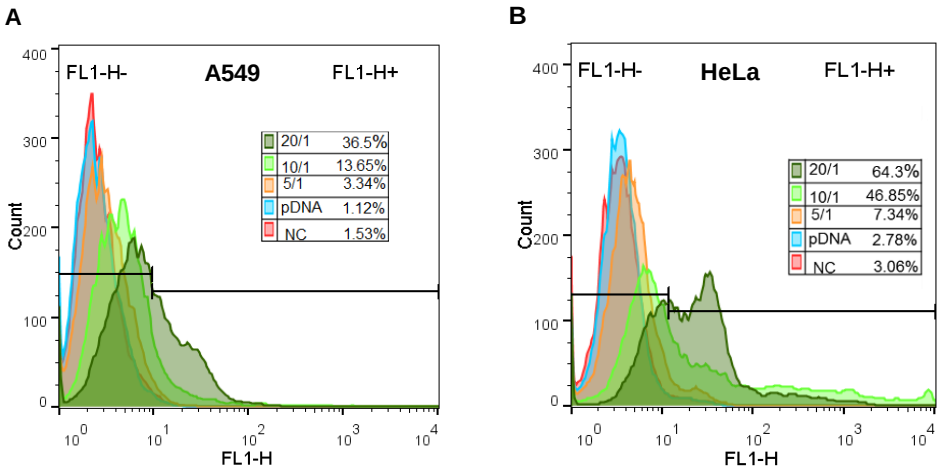


Figure 5.24. Flow cytometric analysis of *in vitro* transfection efficiency. **A**) A549 and **B**) HeLa cells were transfected with WSLP/pEGFP-N3 complexes of various N/P ratios.

As another proof of the successful gene delivery using WSLP/DNA complexes is the direct visualization of the expressed enhanced green fluorescent protein (EGFP) in A549 cells was confirmed using an inverted fluorescent microscope.

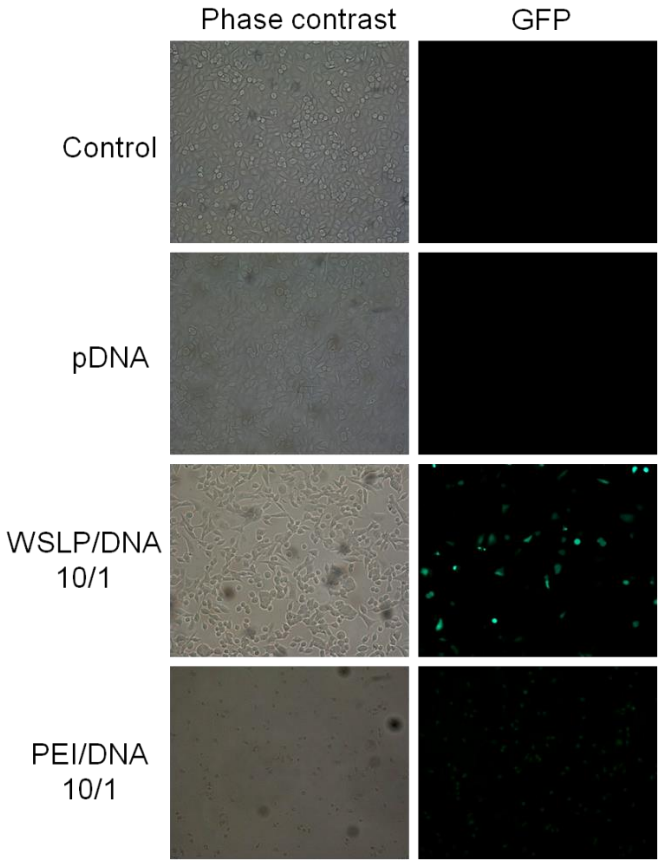


Figure 5.25. Fluorescence microscopic images of A549 cells transfected by WSLP/pEGFP-N3 complexes at 10/1 N/P ratio after 48 h incubation time. PEI was used for comparison. The scale bar represents 500 μm and it is same for all images in this figure.

Figure 5.25 clearly indicates that GFP expression is more intense in cells transfected with the WSLP/pDNA than with PEI/pDNA complexes at 10/1 N/P ratio. The fluorescence is more diffuse when PEI was applied instead of WSLP, which can be easily explained by the phase contrast images, showing that PEI is much more toxic than WSLP.

GFP expression was also observed using confocal microscopy after transfected A549 cellular nuclei were stained with Hoechst fluorescent dye. The results revealed that more GFP is expressed using WSLP as transfection agent than with PEI **(Appendix A.13)**. In other words, the WSLP/DNA complexes showed higher intracellular distribution than PEI/DNA.

5.4.6. Assessment of WSLP/pX458 complexes activity by in vitro gene knockout assay

After the high transfection efficiency was achieved by using the WSLP carrier, the genome editing ability of the designed CRISPR/Cas9 system was monitored, demonstrating the biological activity of WSLP/pX458 complexes at 10/1 and 20/1 N/P ratios. The genomic DNA was extracted from the transfected A549 cells, sequenced and analysed as described previously in 5.3.2.2 section. A mutation efficiency of 8% and 6% for sgRNA targeting Exon 17 was obtained after seven days incubation period with WSLP/pX458 10/1 and 20/1 respectively as shown in **Appendix A.14**. These results are comparable to the mutation efficiency obtained using the commercial Lipofectamine as transfecting agent. This demonstrates the feasibility of applying the cationic WSLP in combination with genome editing materials, which could broaden the application of nanomaterials in gene therapy. These experiments still need optimization for more effective cancer gene therapy.

5.5. Development of specific and regulated NColE7-based zinc finger artificial nuclease

To decrease the probability of the off-target activity of the artificial nucleases, novel zinc finger ANs were designed in our laboratory, based on the NColE7 nuclease domain instead of the NFokI nuclease domain. NColE7 was selected as the nuclease domain in the new ANs because of the knowledge available in the literature and from previous experiments in our laboratory, as discussed in the literature review 2.3.4. During the AN design, the nuclease domain NColE7 was divided into two parts and then joined by a zinc finger protein, which recognizes a selected specific DNA sequence [79]. In the prototype ANs, the ZF is zinc finger protein containing three finger units, identical to that in the crystal structure with the PDB Id: 1MEY. The two segments of NColE7 are expected to fit together into a functional unit upon DNA binding. The regulatory N-terminal segment then activates the catalytic C-termial sequence. This is a positive intramolecular allosteric activation process. As mentioned earlier, the NX-ZF-CY (where NX is the N-terminal segment and CY is the C-terminal segment of NColE7 with X and Y referring to the number of residues involved, respectively) type ANs could achieve specificity and moderate activity, but the regulation was only slightly functional [77]. Therefore, our goal was to develop a regulated zinc finger nuclease in which the nuclease domain itself is inactivated in the absence of a specific DNA target sequence.

We chose the CY–ZF–NX model for further optimization, in which the N-terminus of the ZF is close to the C-terminus of NColE7, while the C-terminus of the ZF is close to the N-terminus of NColE7. Thus short linkers can be designed to directly bind the termini of the proteins to form a continuous circular sequence in the model. By deleting a selected amino acid from the NColE7 sequence the circular AN can be linearized. In the resulting CY–ZF–NX proteins X + Y is always 130, one amino acid less than the total number of amino acids in NColE7. C123–ZF–N7 was computationally analysed by molecular dynamics (MD) simulations and found to be promising in terms of catalytic activity, and the structural robustness [79].

5.5.1. Design of new zinc finger nuclease variants

The amino acid to be deleted from the NColE7 part can in fact be selected arbitrarily. Along this path genes of several new ZFN variants were established by Dr. Eszter Németh as CY-ZF-NX, together with their CY-ZF pairs in which the N-terminal activator domains have been deleted (**Figure 5.26**).

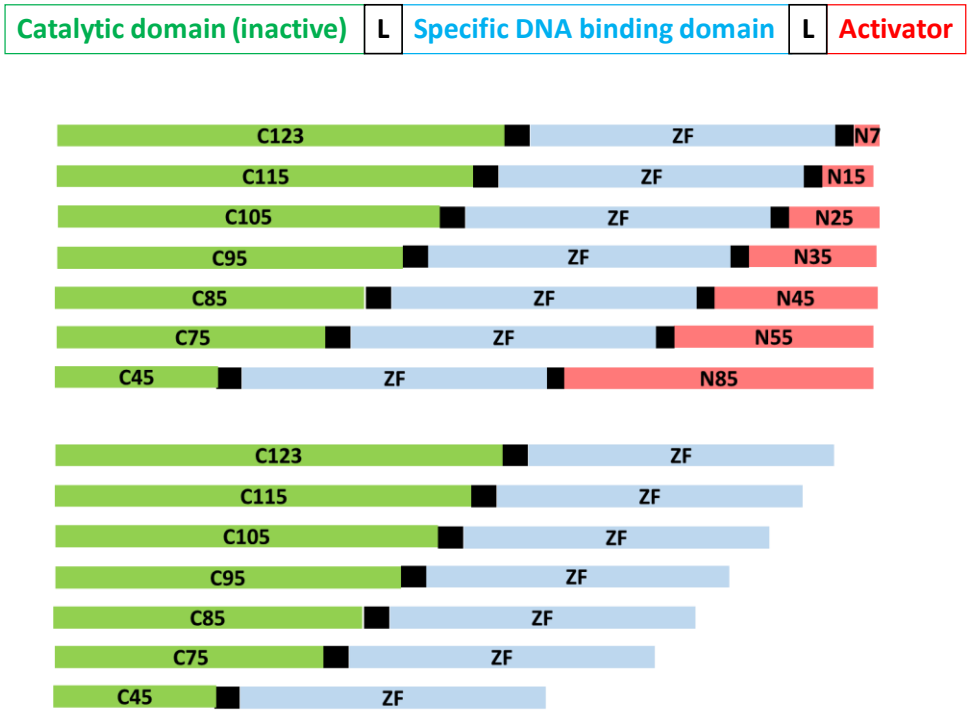


Figure 5.26. Schematic representation of the CY-ZF-NX ANs and their truncated pairs. The C- and N-terminal residues of NColE7 are shown in green and red, respectively; the ZF array is in blue, and the short linker sequences joining these units are in black.

Protein expression experiments were performed in BL21 cells for comparison of the selected CY–ZF–NX and CY–ZF protein pairs to detect their different levels of the toxicity. OD_{600} of the bacterial cultures containing transformed BL21 cells was registered after the protein expression was induced by IPTG at different incubation times (**Figure 5.27**).

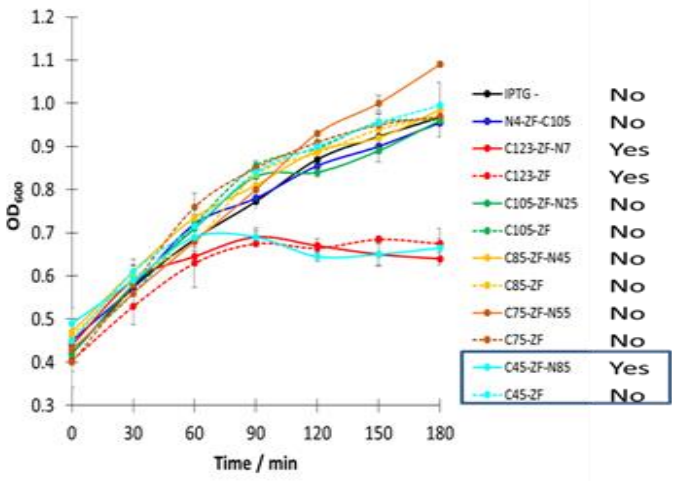


Figure 5.27. *In vitro* catalytic experiment of NX–ZF–CY, CY-ZF-NX and CY-ZF ANs inside BL21 (DE3) *E. coli* cells. The blue box refers to the most promising AN pair.

The bacterial growth was not significantly influenced by most of the proteins expressed, including control experiments, as well, while it was found that both the C123-ZF-N7 and C123-ZF proteins were toxic indicating that the allosteric control is not functional in these ANs. However, C45-ZF-N85 was only active in the presence of N85 activator domain suggesting that this variant is under allosteric regulation.

5.5.2. Proteins used in this work

The cytotoxicity of NColE7 is based on its non-specific DNA cleavage, i.e. its nuclease activity. However, we expected that the new ANs will cleave the DNA in a specific way. Therefore, we have first isolated the DNA from the BL21(DE3) cells treated by the various CY–ZF–NX variants. The results are shown in **Figure 5.28**.

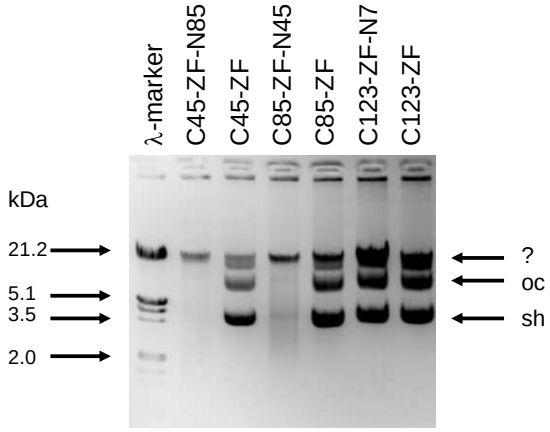


Figure 5.28. Agarose gel electrophoresis picture of the DNA preparations from BL21 (DE3) *E. coli* cells transformed by plasmids encoding for various CY–ZF–NX ANs.

The gel electrophoresis picture clearly shows the difference between the nuclease behavior of the C45–ZF–N85 and its C45–ZF truncated pair. The bands of the superhelical (sh) and open circular (oc) plasmid DNA are clearly visible in the presence of the truncated protein, but disappeared due to their digestion by C45–ZF–N85. This suggests non-specific DNA cleavage by the designed AN. However, the large size DNA, which might be the genomic DNA extracted from the bacteria, was not fragmented is the same way. This might be explained by the competitive effect of the *E.coli* DNA binding proteins preventing non-specific cleavages. Surprisingly, similar effect was observed with the C85–ZF–N45 and C85–ZF AN pairs, which was not observed in the cytotoxicity experiment. The C123–ZF–N7 and its C123–ZF pair did

not show significant DNA cleavage, although both of these ANs were cytotoxic. Thus, the cytotoxicity seems to be influenced by multiple factors.

Since the C45–ZF–N85 vs. C45–ZF AN pair appeared to be different both in the cytotoxicity and in the DNA cleavage experiments, my following experiments focused on a detailed investigation of the C45–ZF–N85 protein.

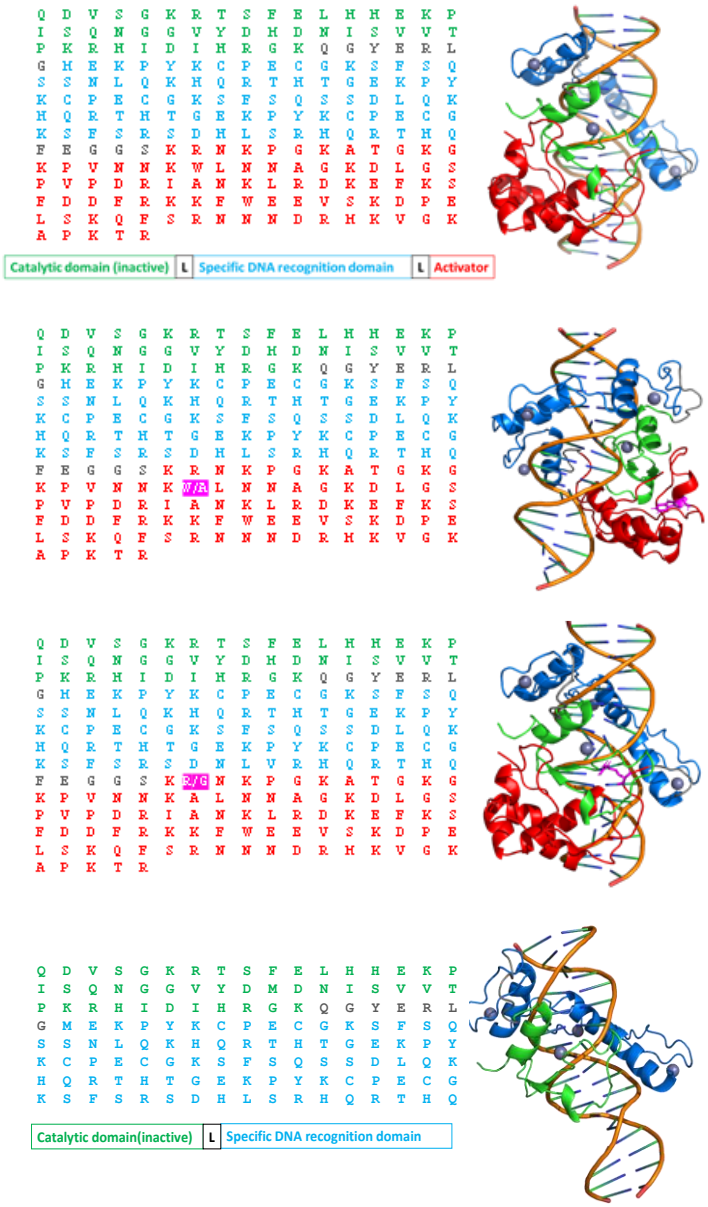


Figure 5.29. Amino acid sequences and computer models showing of the DNA binding of C45-ZF-N85 protein and its mutants.

To understand more about the function of N85 part in the allosteric activation of the new AN variants, two further mutations were performed within this region. (i) In C45-ZF-N85 W160A mutant (abbreviated by W/A) a tryptophan was mutated to alanine. It has previously

been demonstrated that the corresponding W464A mutation leads to collapse of the secondary structure of NColE7 because this amino acid contributes to the proper folding and stabilizes the structure of the metal-binding HNH motif due to its extensive hydrophobic interactions [171]. However, the functional fold of the W464A NColE7 mutant is induced by binding of the DNA substrate. This phenomenon can be explored in the C45-ZF-N85 W160A mutant AN, as well. (ii) In C45-ZF-N85 R143G variant (abbreviated by R/G) an arginine was changed to glycine, influencing the catalytic activity of the enzyme [186]. The R447G mutation in NColE7 affects the number of the positively charged residues in the regulatory region and thereby also the interaction with DNA. Therefore, this mutation in C45-ZF-N85 is expected to decrease its non-specific cleavage. Amino acid sequences and computer models of C45-ZF-N85 and its mutants, including also the truncated C45-ZF protein are shown in **Figure 5.29**. The results of the experiments concerning the solution structure and the nuclease activity of these proteins, where these investigations were possible, are presented in the following.

In order to study the allosteric regulation of the C45-ZF-N85 NColE7-based artificial nuclease and to decrease the nonspecific DNA cleavage, the catalytic experiment was performed in transformed BL21 cells, similarly as described above. The DNA was purified from these bacterial cells and analysed by agarose gel electrophoresis **(Figure 5.30)**. The results showed the nonspecific cleavage properties of C45-ZF-N85 in which the plasmid DNA is fragmented, while in the case of its mutants, the plasmid bands remained visible suggesting the decrease of the non-specific cleavage ability.

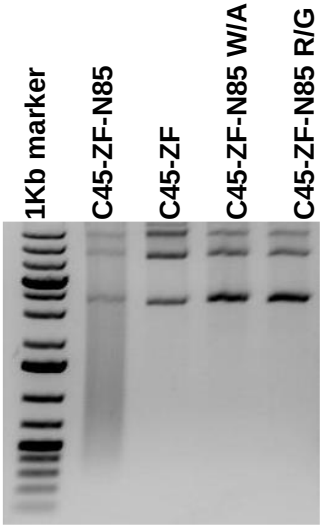


Figure 5.30. Purified plasmids containing genes of C45-ZF-N85 and its mutants, run on 1 % (m/V) agarose gel. Gene Ruler 1kb Plus DNA Ladder served as the reference.

5.5.3. Expression and purification of the designed proteins

The new AN proteins were expressed and purified to study their solution structure, interaction with DNA and Zn^{2+} ions, and their catalytic activity *in vitro*. The new pET-21a*-SRHS plasmid (introduced in chapter 5.1) seemed to be useful for the purification of the artificial nuclease as a 6×His tagged protein by nickel-affinity chromatography, without any remaining amino acids after the Ni(II)-induced cleavage of the affinity tag. However, previously it has been shown that the 1MEY ZF protein used in the prototype ANs can itself be cleaved by Ni(II) ions, since it contains two SXH sequence motifs. Therefore, a mutation of the ZF motif according to the suggested protocol [160] was performed without changing its DNA binding specificity, prior to the further experiments. The modified ZF is denoted as ZF# in the following. The genes of C45-ZF#-N85 and C45-ZF# were then cloned into the pET-21a*-SRHS plasmid using the BsmBI restriction enzyme. Based on the sequenced plasmids, the 6×His-tagged proteins were overexpressed in BL21 (DE3) cells and purified on Ni(II)-NTA column. After this step, the protein samples were checked by SDS PAGE, showing the success of the affinity chromatography (**Figure 5.31**).

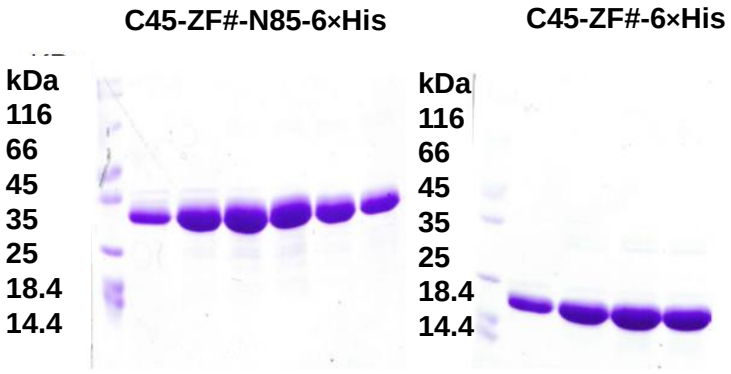


Figure 5.31. The 12.5% SDS PAGE analysis of the purified C45-ZF#-N85-6×His and C45-ZF# -6×His proteins (collected fractions from nickel affinity chromatography).

Due to the strong aggregation tendency of the ANs most of the protein precipitated and aggregated during the Ni(II)-induced cleavage of the 6×His tag. For this reason, this purification strategy was not successful in the case of these proteins. The process needs further optimization e.g. a reducing agent containing buffer. In parallel we have also carried out the expression and purification of the ANs in their designed forms.

C45-ZF-N85, C45-ZF, the W/A the R/G mutant proteins were expressed in BL21 (DE3) cells and purified by a multistep cation exchange chromatography by Eniko Hermann and

myself [188]. The success of the expression and purification of the proteins was verified by SDS PAGE as shown in **Figure 5.32**.

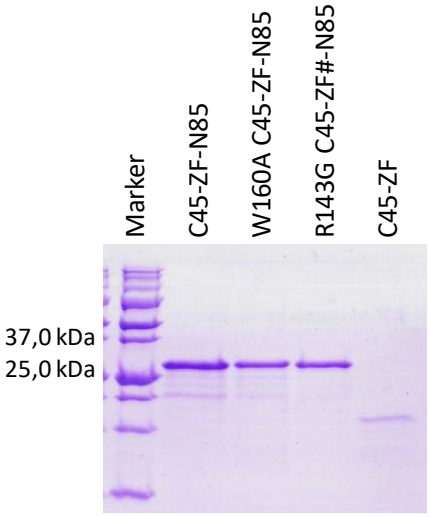


Figure 5.32. The 12.5% SDS PAGE analysis of the purified C45-ZF-N85 protein and its mutants.

To characterize the purified proteins their solution structure was investigated by CD spectroscopy and their secondary structures were calculated from CD data using BeStSel program [162]. The results of the calculations obtained from the spectral region between 190 and 250 nm are shown in **Table 5.7**.

Table 5.7. The percentage of each secondary structural element in the tested proteins, calculated using the BeStSel program from their CD spectra.

Protein Name	α-Helix	β-sheet	Turn	Others
C45-ZF-N85	17.5%	26.9%	14.1%	41.5%
C45-ZF	17.3%	25.7%	15.3%	41.6%
C45-ZF-N85 W160A	13.1%	31.3%	15.7%	39.9%
C45-ZF-N85 R143G	18.5%	21.1%	14.1%	46.3%

The data obtained for the C45-ZF-N85 and the R/G mutant proteins show that this fold is similar to that expected from the data of the independent ZF and NColE7 proteins. This may be surprising, since the functional structure is expected to be stabilized only in the presence of the DNA substrate. However, the CD data analysis does not reflect the tertiary structure. On

the other hand, the W/A mutant displays smaller fraction of the helical parts in agreement with the effect of this mutation in NColE7. The concentration of the C45-ZF protein and thus, the intensity of the CD spectrum was very low (~ 2 μM) and therefore, the calculated values are considered to be ambiguous.

We have also applied ESI-MS to determine the molecular weight of the purified proteins. The measurements with C45-ZF-N85 under acidic conditions revealed a M_w value, which is in good agreement with the expected one: 26057.0 Da (monoisotopic mass). It has to be mentioned that on the addition of ~ 5 equivalents of Zn^{2+} ions the protein could bind three equivalents of metal ion, and the catalytic center could only be saturated by the metal ion upon adding high excess of the Zn^{2+} ions.

5.5.4. Plasmid cleavage activity of the C45-ZF-N85 artificial nuclease

The catalytic activity of the purified C45-ZF-N85 and C45-ZF proteins was studied by using a nonspecific plasmid (pGEX-6P-1), which does not contain the zinc finger recognition site and a specific plasmid (pGEX-6P-1-Z0, abbreviated by Z0 in the following) that contains two recognition sites of the artificial nuclease opposite to each other **(Figure 5.33)**.

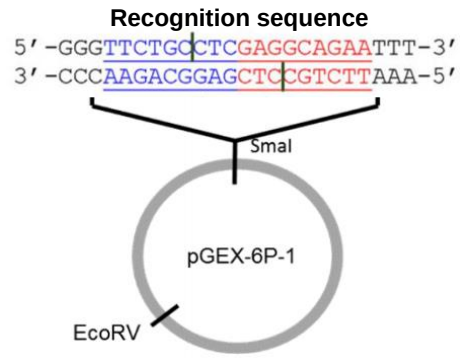


Figure 5.33. Scheme of pGEX-6P-1-Z0 plasmid in which two 1MEY recognition sites (shown in blue and red) are inserted into the SmaI restriction site of the original pGEX-6P-1 plasmid. The two green lines represent the expected double strand cleavage resulting in a linear plasmid with sticky ends.

The Z0 plasmid was cleaved into open circular form by C45-ZF-N85 after 24 h but C45-ZF did not cleave the DNA. This shows that the C45 catalytic motif can only function in the presence of an activator sequence, i.e. N85 is essential for the catalytic activity, indeed. On the other hand, the non-specific pGEX-6P-1 plasmid was not cleaved by either C45-ZF-N85 or C45-ZF. This proves the specificity of the artificial nuclease **(Figure 5.34)**.

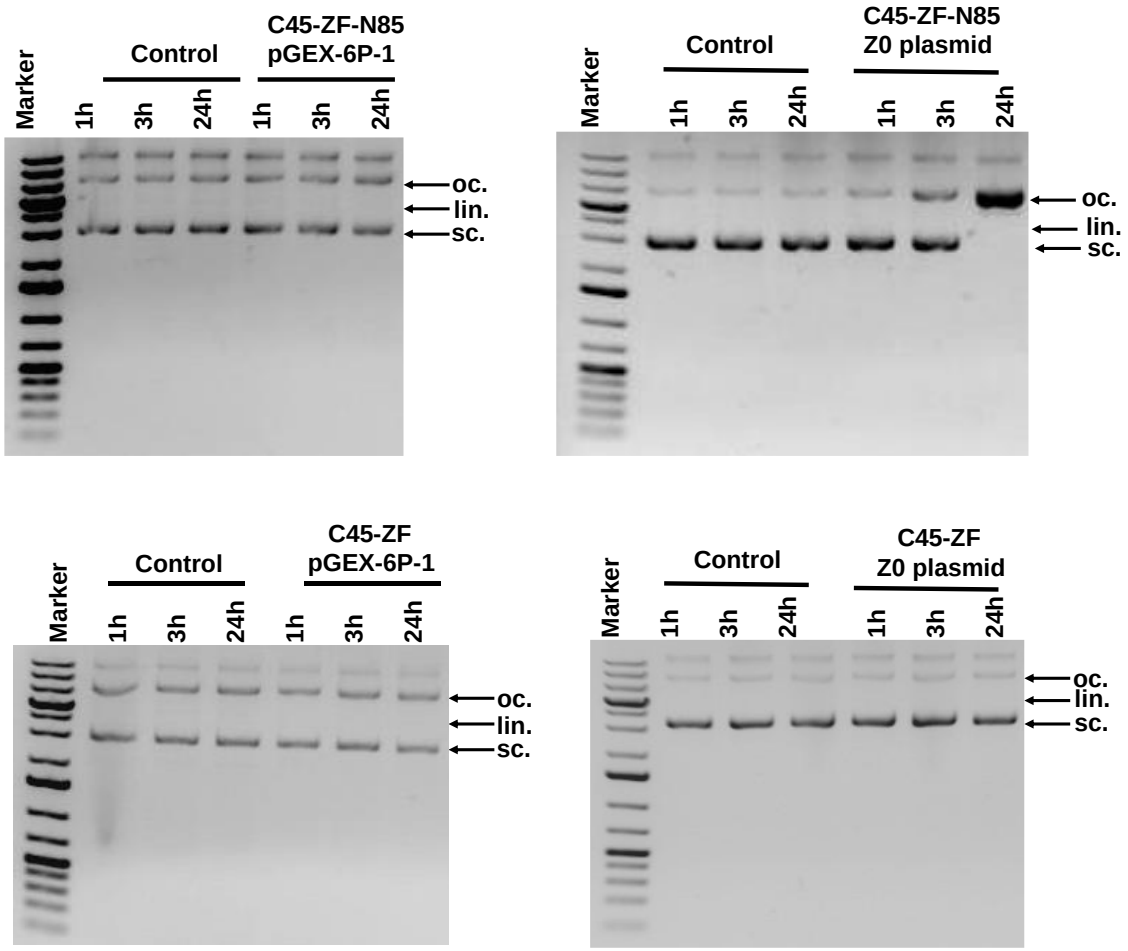


Figure 5.34. Cleavage of the pGEX-6P-1 and pGEX-6P-1-Z0 plasmids (12 nM) by 200nM protein. 200 nM $ZnSO_4$ and 150 mM NaCl was adjusted in 20 mM Tris pH 7.9. The reaction mixtures were incubated up to 24 hours at 37 °C. The reactions were stopped by 1% (m/V) SDS and freezing, and subsequently run on 1 % (m/V) agarose gel. The untreated plasmids served as negative controls. Gene Ruler 1kb Plus DNA ladder served as the reference.

The results shown in **Figure 5.34** are the outcome of optimization experiments to adjust the ionic strength, Zn(II) and protein concentration, required for the specific and regulated nuclease activity. Finally, the concentrations of NaCl and Zn^{2+} ions concentrations were fixed at 150 mM and 200 nM, respectively. Double stand cleavage was expected, in which the superhelical form of the plasmid is converted to linear form but only single strand cleavage happened. This may be explained by the strong specific DNA binding of the AN protein sterically inhibiting the binding of a second protein molecule and the occurrence of cleavage on the opposite strand of DNA, under the measurement conditions. The above experiments demonstrated that it was possible to establish the reaction conditions under which the specificity and regulation of C45-ZF-N85 could be proved.

The cleavage of pGEX-6P-1-Z0 plasmid by the C45-ZF-N85 was monitored by agarose gel electrophoresis in the presence of various amounts of Zn^{2+} ions **(Figure 5.35)**. The protein is completely inactive in the absence of added Zn^{2+} ions. This suggests that the C45-ZF-N85 was purified without, or with a minimal amount of zinc(II), bound in the catalytic center. The maximum of the DNA cleavage activity was achieved at 1:1 protein-to-Zn^{2+} molar ratio. Under these conditions, the supercoiled DNA changed to open circular form at 24 h. After a further increase of the metal ion concentration, at 1:2 protein-to-Zn^{2+} molar ratio, the catalytic activity is completely inhibited.

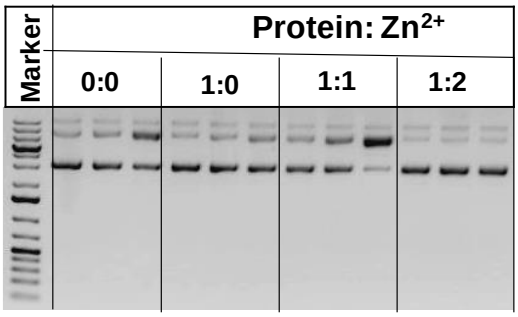


Figure 5.35. Agarose gel electrophoresis images showing the cleavage of pGEX-6P-1-Z0 (12 nM) by 200 nM C45-ZF-N85 protein containing increasing amounts of Zn^{2+} ions. Each column in one experiment represents the catalytic activity after 1, 3, 24 h, from left to right. Gene Ruler 1kb Plus DNA Ladder was loaded as a reference, and the untreated pGEX-6P-1-Z0 plasmid as the negative control.

Our previous studies showed that the CD spectra of apo NColE7 shifted slightly upon binding the Zn^{2+} ion, as a consequence of minor changes in the solution structure [73] and the addition of Zn^{2+} ions to N4-ZF-C105 AN resulted in a spectral shift, similar to that observed for NColE7 itself [77]. The above cataytic results using C45-ZF-N85 are in agreement with these observations. First the zinc binding to the AN lacking the metal ion in the active center increases its catalytic activity, while the metal ion binding to the catalytic histidine decreases the nuclease activity. At 1:1 protein-to-Zn^{2+} molar ratio, the catalytic center of C45-ZF-N85 is saturated with the Zn^{2+} then H545 may also be metallated on further addition of metal ion, preventing the generation of the OH^- nucleophile by the sidechain of this histidine residue. The higher is the metal ion excess the higher is the probability of this type of coordination as shown previously [189].

The Zn^{2+} ion is not required for DNA binding, but it is essential for DNA hydrolysis in NColE7 [171]. During the catalytic reaction, the Zn^{2+} ion in the active center binds the oxygen donor atoms of the DNA scissile phosphodiester group. The positively charged amino acids at

the N-terminus are sterically close to the catalytic site. R447 is bridged with the Zn^{2+} ion by the scissile phosphodiester group [80]. The metal ion, having a free coordination site, has essential multiple roles in DNA cleavage since it stabilizes the active centre, polarizes the P–O bond for nucleophilic attack. In addition it is suggested to stabilize the phosphoanion transition state and the leaving sugar alcoholate group. The attacking nucleophilic OH^- in NColE7 is supposed to be generated by the most conserved histidine residue of the HNH motif which does coordinate to the Zn^{2+} ion.

In case of incomplete saturation of the zinc finger units by the metal ion, oxidation of the thiol groups into disulfide bridges may easily occur. This will most probably result in inactive protein. The cleavage experiment of the pGEX-6P-1-Z0 plasmid was performed by C45-ZF-N85 under reducing conditions, as well, to verify the integrity of the ANs. C45-ZF-N85 was treated by 100 equivalents of tris(2-carboxyethyl)phosphine (TCEP) as a reducing reagent. It was expected that by reducing the eventual disulfide bonds within the AN protein, lacking some of the zinc ions, its catalytic activity will increase. The catalytic activity of the TCEP treated-protein was monitored in the presence of varying Zn^{2+} ion equivalents **(Figure 5.36)**.

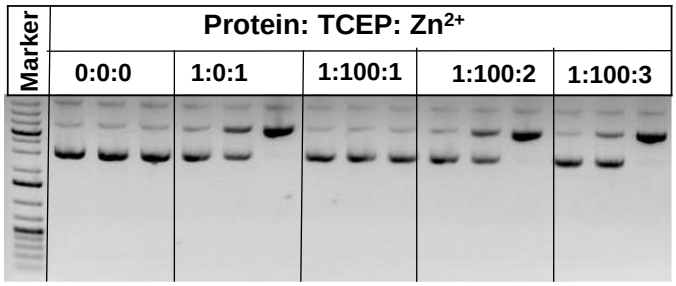


Figure 5.36. Agarose gel electrophoresis images showing the cleavage of pGEX-6P-1-Z0 (12 nM) by 200 nM C45-ZF-N85 protein and treated by 100 equivalents of TCEP in the presence of increasing amounts of Zn^{2+} ions. Each column in one experiment represents the catalytic activity after 1, 3, 24 h, from left to right.

The untreated protein in the presence of 1 equivalent Zn^{2+} served as a positive control. The DNA cleavage was inhibited completely at 1:100:1 protein to TCEP to Zn^{2+} ratio. This may be due to the combined effect of the decrease of the zinc(II) amount available for the catalytic center: (i) the reduction of any disulfide bonds results in metal ion binding by the tiolate donors (ii) the TCEP excess can bind metal ion. The catalytic activity increased again by adding 2 and 3 equivalents Zn^{2+} ion to the reaction mixture. Although the catalytic activity was expected to increase more than the positive control upon increase of the active catalyst concentration by increasing the Zn^{2+} ion amounts, it did not happen within the conditions of the

present experiment series. This is most probably attributed to the possibility of the multiple metal ion binding modes in this complicated system.

5.5.5. Double strand cleavage in short oligonucleotides

As mentioned above, we could not achieve double strand DNA cleavage by using the pGEX-6P-1-Z0 plasmid as the substrate. Therefore, a short dsDNA (ZV, 37bp) was designed for the catalytic experiments to confirm the double-strand cleavage by C45-ZF-N85. The short DNA contains two ZF recognition sequences separated by the restriction site of VspI restriction enzyme **(Figure 5.37 a)**. The recognition site of VspI was added to use this enzyme as a positive control for the double-strand DNA cleavage. Another 34 bp dsDNA (S0), which does not contain the recognition site of C45-ZF-N85 protein and the VspI restriction site was applied as a negative control. (Forward strand: 5'–ctagtttgctgaactggggtcacatagattaata-3' and reverse strand: 3'-gatcaaacgacttgaccccagtgtatctaattat-5'). Both ZV and S0 dsDNAs were synthesized as complementary forward and reverse ssDNA strands and hybridized by PCR through increasing the temperature to 95 ºC and cooling to 4 ºC step by step.

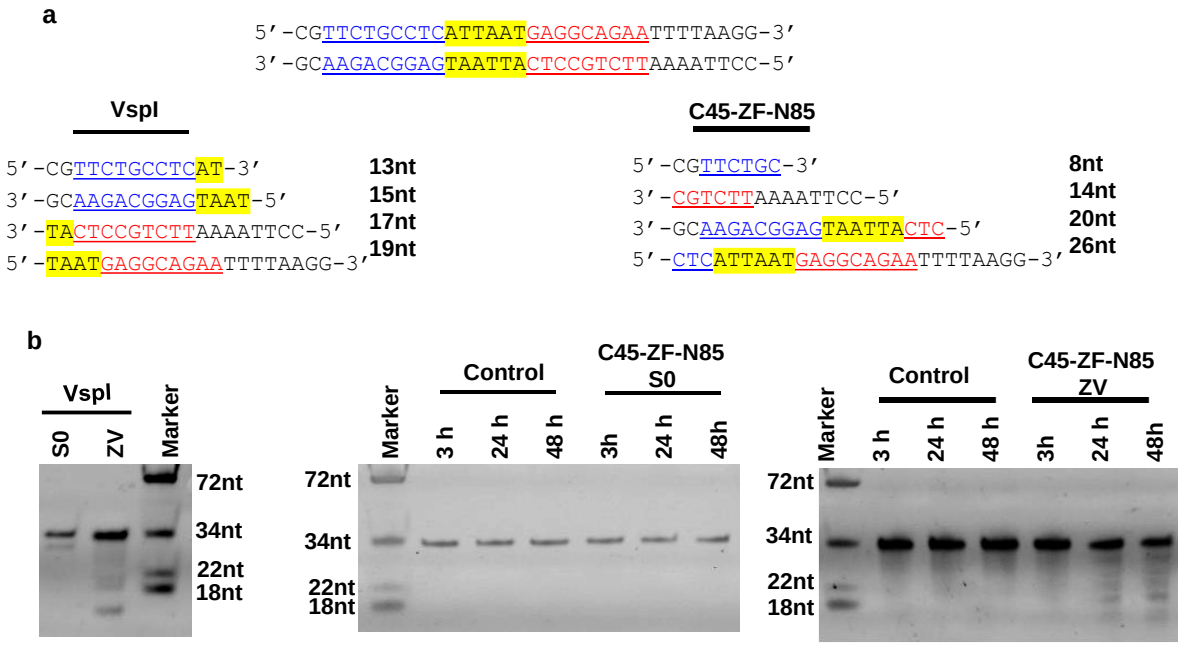


Figure 5.37. a, Illustration of ZV short oligonucleotide which consists of two 1MEY recognition sites (shown in blue and red) and separated by VspI restriction site (yellow highlighted) and the expected fragment products which result from the cleavage experiment. **b,** UREA-PAGE (15%) analysis of the cleavage reaction products of oligonucleotides by VspI and C45-ZF-N85. The reaction mixtures containing 1000 nM substrate, 400 nM protein, 400 nM $ZnSO_4$, 100 mM NaCl and 20 mM Tris pH 7.9 were incubated for up to 48 h at 37 ºC. The reaction was stopped by formamide and freezing at -20 °C. Mixture of ssDNAs with different nucleotides (nt) size (72, 34, 22, and 18) was used as a reference. The digestion reaction of ZV and S0 was performed also by VspI restriction endonuclease at 37 °C for two hours as a positive control.

The cleavage of S0 and ZV was monitored by urea PAGE gel, in which ZV was expected to appear as four single strand fragments upon the single strand cleavages on both strands of the dsDNA. The resulting fragmentation pattern by VspI and C45-ZF-N85 is expected to be similar, as shown in **Figure 5.37 b**. After 24 h ZV DNA was cleaved but S0 DNA was not affected by both the restriction endonuclease and the C45-ZF-N85 AN. This catalytic activity experiment proved that double strand cleavage may be obtained by C45-ZF-N85 within a suitably designed cleavage sequence. The above results showed that activation of the C-terminal HNH nuclease motif was possible by fusion of the N85 N-terminal segment to the ZF in integrated ANs. Further improvement requires optimization of the molecular design.

5.5.6. Mutation of the C45-ZF-N85 and C45-ZF nucleases for targeting ErbB1(EGFR) oncogene

To prove the structural variability of the newly designed AN, we introduced changes in the zinc finger domain to target the EGFR gene at the same DNA sequence as it was selected for the CRISPR/Cas9 system designed in chapter 5.3. Each recognition site consists of three fingers. Mutations the DNA sequence were performed in a stepwise manner by a modified QuikChange PCR protocol using a different sets of primers for each zinc finger unit. The success of the mutations was confirmed by DNA sequencing. The initial check of the catalytic activity was carried out in transformed BL21 (DE3) cells, similarly as described in section 5.5.2.

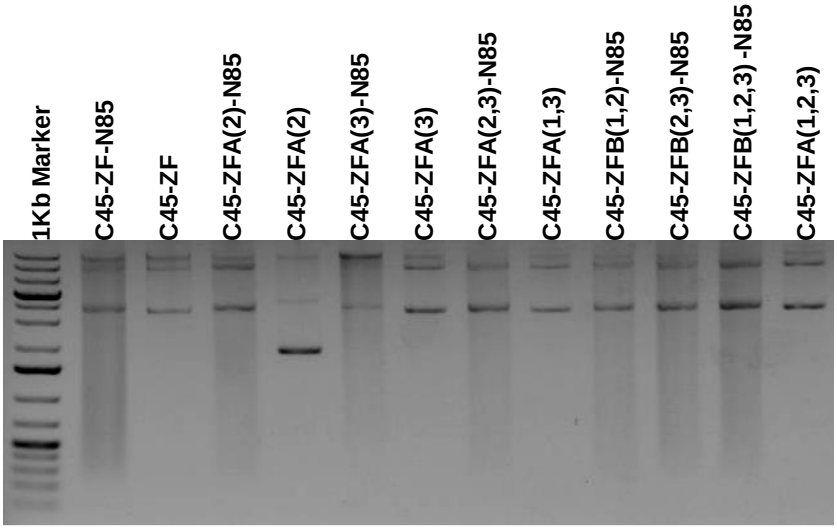


Figure 5.38. Agarose gel electrophoresis picture of the DNA preparations from BL21 (DE3) *E. coli* cells transformed by plasmids encoding for various C45-ZF*-N85 and C45-ZF* with mutated zinc fingers to target ErbB1. Plasmids encoding C45-ZF-N85 and C45-ZF were used as a positive control. Gene Ruler 1kb Plus DNA ladder served as the reference. Lane 5 most probably contains a wrong plasmid.

The available full size (C45-ZF*-N85) and truncated (C45-ZF*) protein versions were compared to each other in this experiment. The applied plasmids contain the gene of ANs with one, two or three modified zinc fingers for one of the two recognition sites (A and B) positioned oppositely within the DNA targeting region. The DNA was purified from the transformed bacterial cells and analysed by agarose gel electrophoresis **(Figure 5.38).** The results showed the nonspecific cleavage properties of the full size proteins against the plasmid DNA similarly to the prototype C45-ZF-N85. In contrary, the plasmids containing the genes of the truncated C45-ZF* proteins remained intact. These results suggest the robustness of the designed AN.

6. SUMMARY AND FUTURE OUTLOOK

My PhD research work is a part of the research project conducted in the Bioinorganic Chemistry Research Group in the Chemistry Institute of University of Szeged, aiming to develop effective artificial nucleases as promising molecular tools for gene therapy. Such nucleases need to be safe in terms of avoiding off-target cleavages. Therefore, novel zinc finger ANs were designed based on the NColE7 nuclease domain. The HNH active center of NColE7 enzyme functions under intramolecular allosteric activation, providing the basis for a new approach towards integrated artificial nucleases. In the resulting constructs the allosteric activation is expected to occur upon the specific DNA recognition by the zinc finger domain. The optimization of such regulated ANs can lead to gene editing enzymes applicable in parctice.

Therefore, in my work I investigated the newly designed ANs to achieve deeper understanding about the mechanism of this new nuclease. This knowledge can be utilized further in the AN design to achieve better regulation and specificity. The robustness of the design is also verified by the variability of the zinc finger domain, to target various DNA sequences, beside the prototype model. I used CRISPR/Cas9 artificial nuclease for knocking out oncogenes in order to identify suitable target DNA sequences for artificial nucleases, which can be applied in gene therapy of cancer, and to compare them with the currently often used CRISPR/Cas9 artificial nucleases. I also developed a novel drug delivery system to increase the efficiency of the ANs.

In the first part of my work, I designed a new DNA construct (pET-21a*-SRHS) for affinity purification strategy of proteins with a native sequence. This DNA construct allows for the removal of the affinity tag avoiding any remnant amino acid residues, while overcoming several disadvantages of the proteases: sensitivity, high price and extra amino acids remaining after the protease cleavage. Such DNA carrier can provide a feasible purification pathway for those proteins, where the sequence is precisely determined e.g. by the computer design. The synthesis of the ANs relies on precise reproduction of the designed sequence. The new DNA construct can be used with various tags at the C-terminus for affinity chromatographic applications.

I observed that the C-terminal 6×His tag of the new construct causes an inhibitory effect by which the catalytic activity of the NColE7 or other metalloenzymes can be modulated in a multiple manner without changing the core enzyme sequence. In the light of this the genes of NColE7 protein and its R447G (KGNK) mutant were successfully cloned into a modified pET-21a*-SRHS DNA vector fusing the 6×His affinity tag to the proteins upon expression. The

cloning into pET-21a would be otherwise not be possible in the absence of the gene of the Im7 inhibitory protein. The solution secondary structure of KGNK 6×His protein was studied by CD technique and the CD spectrum evaluated by the BeStSel software revealed that its relative α-helical content is ~14%. This result suggested that the protein was obtained in its functionally folded form. ESI-MS was applied to validate the molecular mass and to monitor metal ion binding property of the KGNK-6×His protein. Accordingly, this protein was able to bind up to six Zn^{2+} ions under the measurement conditions. Based on ESI-MS measurements, the relative amounts of the complexes containing various numbers of metal ions were estimated and plotted versus the added Zn^{2+} equivalents. The catalytic activity experiments at increasing Zn^{2+} ion to protein concentration ratio revealed that the DNA cleavage ability of the 6×His tagged enzyme was first enhanced by an increase of the metal ion concentration, while high excess of Zn^{2+} ions caused a lower rate of the DNA cleavage. Based on the comparison of the MS and catalytic experiments I concluded that the initial inhibition by the 6×His tag was at least partially caused by its binding to the free coordination site of the Zn^{2+} ion in the active center. Saturation of the tag by metal ions leads to increase of the activity, as the original coordination sphere of the catalytic Zn^{2+} ion was restored. Upon further increase of Zn^{2+} amounts the catalytic histidine is also inhibited by the metal ion excess. Mutations within the C terminal tag revealed that the inhibitory effect of the 6 × His fusion tag on the nuclease activity is a complex process via both coordinative and non-specific steric interactions.

New ZFNs were designed by Dr. Eszter Németh as CY-ZF-NX and CY-ZF in which the CY catalytic center and NX activator sequences of NColE7 were fused these to the opposite termini of a zinc finger protein, which is the DNA recognition domain. We found that among the various constructs with X+Y = 130, C45-ZF-N85 variant is under intramolecular allosteric regulation. C45-ZF-N85, C45-ZF, as well as W/A and R/G mutants of C45-ZF-N85 protein were successfully expressed and purified by cation exchange chromatography. The catalytic activity was also studied by using nonspecific and specific DNA (pGEX plasmid and Z0 plasmid). Z0 plasmid contains the recognition site of the artificial nuclease and is cleaved into the open circular form by C45-ZF-N85. The truncated version C45-ZF, however, did not cleave the DNA. By this, we could prove that the catalytic domain cannot work in the absence of the activator and C45-ZF-N85 is a very promising artificial nuclease enzyme.

The variability of the recognition sequence of the above ANs was also checked. For this, first the CRISPR/Cas9 system was used in my PhD study to detect which DNA target sequence could be suitable for the practical application of the new zinc finger nuclease. New guide RNA sequences were designed to target lung cancer through EGFR oncogene by CRISPR/Cas9

system. The new CRISPR/Cas9 plasmids were used for transfection of A549 lung cancer cells and mutation efficiencies around 7.5% were revealed after seven days incubation period by the sequencing of the target region of the genomic DNA. Further improvement of the gene editing efficiency is needed. The redesign of the zinc-finger part of the new ANs has been initiated through several steps of single finger mutations. Comparison of the full length and truncated versions proved the robustness of the AN design.

A water-soluble cationic lipopolymer (WSLP) was synthesized as a novel drug delivery system as a candidate to improve the efficiency of the artificial nuclease action in cells. WSLP was prepared by the modification of high molecular weight branched poly(ethylenimine) (PEI) with cholesteryl chloroformate, then it was characterized and assessed for DNA loading and cellular membrane permeability. The cholesterylation was efficient, in average ~ 10% of the PEI nitrogens were reacted. This kind of modification of a cytotoxic high molecular weight PEI proved to be promising, since it retained the positive charges for strong interaction with DNA, while decreased the toxicity of the preparation. pEGFP-N3 and a CRISPR/Cas9 (pX458) carrier plasmid DNA both expressing the GFP fluorescent protein for detectability of transfection were efficiently condensed by the new water soluble lipopolymer. A549 and HeLa cells were efficiently transfected under the applied conditions. WSLP/pX458 complexes could achieve ~ 8% mutation efficiency in A549 cancer cells at 10/1 N/P ratio.

The results of my PhD work demonstrate how the specificity and effectivity of artificial metallonuleases can be improved by regulating the enzymatic function and by developing effective drug delivery system. In the future, we plan to better understand the gene editing of the artificial nucleases by using cell sorting technique to separate the transfected cells. In this way, the properties of the newly designed enzymes can be compared with others, e.g. with the CRISPR/Cas9 system. In my hope, the results of my PhD studies are expected to contribute to the development of better molecular tools towards cancer gene therapy.

7. REFERENCES

1. American Cancer Society, “Cancer Facts & Figures 2016,” Cancer Facts Fig. 2016 2016: 1–9

2. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: The next generation. Cell 144: 646–674

3. White MK, Khalili K (2016) CRISPR/Cas9 and cancer targets: future possibilities and present challenges. Oncotarget 7: 12305–12317

4. Zur Hausen H (2007) Infections causing human cancer. Wiley & Sons

5. White MK, Pagano JS, Khalili K (2014) Viruses and human cancers: A long road of discovery of molecular paradigms. Clin. Microbiol 27: 463–481

6. Vennström B, Michael Bishop J (1982) Isolation and characterization of chicken DNA homologous to the two putative oncogenes of avian erythroblastosis virus. Cell 28: 135–143

7. Parker RC, Varmus HE, Bishop JM (1981) Cellular homologue (c-src) of the transforming gene of Rous sarcoma virus: isolation, mapping, and transcriptional analysis of c-src and flanking regions. Proc. Natl. Acad. Sci 78: 5842–5846

8. Chang EH, Gonda MA, Ellis RW, Scolnick EM, Lowy DR (1982) Human genome contains four genes homologous to transforming genes of Harvey and Kirsten murine sarcoma viruses. Proc. Natl. Acad. Sci 79: 4848–4852

9. Kozak C, Gunnell MA, Rapp UR (1984) A new oncogene, c-raf, is located on mouse chromosome. J. Virol 49: 297–299

10. Vennstrom B, Sheiness D, Zabielski J, Bishop JM (1982) Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain. J. Virol 42: 773–779

11. Levine AJ (1997) P53, the Cellular Gatekeeper for Growth and Division. Cell 88: 323–331

12. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684-1689

13. Gnyszka A, JASTRZĘBSKI Z, Flis S (2013) DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. Anticancer Res 33:2989-2996

14. Reeve JG, Rabbitts PH, Twentyman PR (1989) Amplification and expression of mdr1 gene in a multidrug resistant variant of small cell lung cancer cell line NCI-H69. Br. J. Cancer 60: 339–342

15. Peaston AE, Gardaneh M, Franco AV., Hocker JE, Murphy KM, Farnsworth ML, Catchpoole DR, Haber M, Norris MD, Lock RB, Marshall GM (2001) MRP1 gene expression level regulates the death and differentiation response of neuroblastoma cells. Br. J. Cancer 85: 1564-1571

16. Ishikawa T, Ali-Osman F (1993) Glutathione-associated cis-diamminedichloroplatinum(II) metabolism and ATP-dependent efflux from leukemia cells. Molecular characterization of glutathione-platinum complex and its biological significance. J. Biol. Chem 268: 20116–20125

17. Mcdonnell AM, Dang CH (2013) Basic review of the cytochrome p450 system. J Adv Pract Oncol 4:263–268

18. Lindsay-Mosher N, Su C (2016) A Toolkit for Targeting Cancer. Gene Ther 30: 20–27

19. Wang H, Sun W (2017) CRISPR-mediated targeting of HER2 inhibits cell proliferation through a dominant negative mutation. Cancer Lett 385:137–143

20. Bhattacharjee R, Das Purkayastha K, Adapa D, Choudhury A (2017) CRISPR/Cas9 genome editing system in the diagnosis and treatment of cancer. J. RNAi Gene Silenc13:585–591

21. Gaj T, Gersbach CA, Barbas CF (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31: 397–405

22. Sun W, Gu Z (2017) Tailoring non-viral delivery vehicles for transporting genome-editing tools. Sci. China Mater 60: 511–515

23. Hotta A, Yamanaka S (2015) From Genomics to Gene Therapy: Induced Pluripotent Stem Cells Meet Genome Editing. Annu. Rev. Genet 49: 47–70

24. Hu JH, Davis KM, Liu DR (2016) Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. Cell Chem. Biol 23:57–73

25. Gupta RM, Musunuru K (2014) Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. J. Clin. Invest 124: 4154–4161

26. Wang M, Glass ZA, Xu Q (2017) Non-viral delivery of genome-editing nucleases for gene therapy. Gene Ther 24: 144–150

27. Voytas DF, Gao C (2014) Precision genome engineering and agriculture: opportunities and regulatory challenges. PLoS Biol 12: e1001877

28. Perez-Pinera P, Ousterout D, Gersbach C (2012) Advances in targeted genome editing. Curr. Opin. Chem 16: 268–277

29. Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics 188:773–782

30. Tebas P et al (2014) Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. N. Engl. J. Med 370: 901–910

31. Kim H, Kim JS (2014) A guide to genome engineering with programmable nucleases. Nat. Rev. Genet 15: 321–334

32. Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci 93:1156–1160

33. Doyon Y, Vo TD, Mendel MC, Greenberg SG, Wang J, Xia DF, Miller JC, Urnov FD, Gregory BD, Holmes MC (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat methods 8: 74–79

34. Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 Å: A model system for understanding zinc finger-DNA interactions. Structure 4: 1171–1180

35. Joung JK, Voytas DF, Cathomen T (2010) Reply to 'Successful genome editing with modularly assembled zinc finger nucleases. Nat. Methods 7:91–92

36. Sander J, Dahlborg E, Goodwin M (2010) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nature 8: 67–69

37. Bhakta MS, Henry IM, Ousterout DG, Das KT, Lockwood SH, Meckler JF, Wallen MC, Zykovich A, Yu Y, Leo H, Xu L, Gersbach CA, Segal DJ (2013) Highly active zinc-finger nucleases by extended modular assembly. Genome Res 23:530-538

38. Bitinaite J, Wah DA, Aggarwal AK Schildkraut I (1998) FokI dimerization is required for DNA cleavage. Biochemistry 95:10570–10575

39. Szczepek M, Brondani V, Büchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat. Biotechnol 25: 786–793

40. Guo J, Gaj T, Barbas CF (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J. Mol. Biol 400: 96–107

41. Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat. Rev. Mol. Cell. Biol 14: 49-55

42. Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501–1501

43. Li L, Atef A, Piatek A, Ali Z, Piatek, M, Aouida M, Sharakuu A, Mahjoub A, Wang G, Khan S, Fedoroff NV, Zhu J, Mahfouz MM (2013) Characterization and DNA-binding specificities of Ralstonia TAL-like effectors. Mol plant 6: 1318-1330

44. Mak AN, Bradley P, Cernadas R, Bogdanove AJ, Stoddard BL (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. Science 335:716–719

45. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326: 1509-1512

46. Doyle EL, Hummel AW, Demorest ZL, Starker CG, Voytas DF, Bradley P, Bogdanove AJ (2013) TAL effector specificity for base 0 of the DNA target is altered in a complex, effector-and assay-dependent manner by substitutions for the tryptophan in cryptic repeat–1. PloS one 8: e82120

47. Lamb BM, Mercer AC, Barbas CF (2013) Directed evolution of the TALE N-terminal domain for recognition of all 5′ bases. Nucleic Acids Res 41: 9779–9785

48. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186: 757-761

49. Bloom K, Ely A, Mussolino C, Cathomen T, Arbuthnot P (213) Inactivation of Hepatitis B virus replication in cultured cells and in vivo with engineered transcription activator-like effector nucleases. Mol. Ther 21: 1889–1897

50. Li W, Teng F, Li T, Zhou Q (2013) Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. Nat. Biotechnol 31: 684–686

51. Tesson L, Usal C, Ménoret S, Leung E, Niles BJ, Remy S, Santiago Y, Vincent AI, Meng X, Zhang L, Gregory PD, Anegon I, Cost GJ (2011) Knockout rats generated by embryo microinjection of TALENs. Nat. Biotechnol 29: 695-696

52. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res 39: 9275–9282

53. Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. Annu. Rev. Genet 45: 273–297

54. Makarova KS et al (2015) An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol 13: 722–736

55. Van Der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol 12: 479–492

56. Mei Y, Wang Y, Chen H, Sun ZS, Da Ju X (2016) Recent Progress in CRISPR/Cas9 Technology. J. Genet. Genomics 43: 63–75

57. Liu C, Zhang L, Liu H, Cheng K (2017) Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications. J. Control. Release 266: 17–26

58. Singh A, Chakraborty D, Maiti S (2016) CRISPR/Cas9: a historical and chemical biology perspective of targeted genome engineering. Chem. Soc. Rev 45:6666–6684

59. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. Nat. Protoc 8: 2281–2308

60. Niu Y et al., (2014) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell 156: 836–843

61. Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen Z, Liu DR (2015) Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat. Biotechnol 33: 73

62. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JRJ, Joung K (2013) Efficient in vivo genome editing using RNA-guided nucleases. Nat. Biotechnol 31: 227–229

63. Sung Y, Kim J, Kim H, Lee J, Jeon J (2014) Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res 24: 125–131

64. Ding Q, Regan S, Xia Y, Oostrom L, Cowan C, Musunuru K (2013) Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell stem 12: 393–394

65. Chak KF, Kuo WS, Lu F, James R (2009) Cloning and characterization of the ColE7 plasmid. J. Gen. Microbiol 137:91–100

66. Lin YH, Liao CC, Liang PH, Yuan HS, Chak KF (2004) Involvement of colicin in the limited protection of the colicin producing cells against bacteriophage. Biochem. Biophys. Res. Commun 318: 81–87

67. Hsia KC, Chak KF, Liang PH, Cheng YS, Ku WY, Yuan HS (2004) DNA Binding and Degradation by the HNH Protein ColE7. Structure 12:205–214

68. Ko TP, Liao CC, Ku WY, Chak KF, Yuan HS (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. Structure 7:91–102

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

70. Pommer AJ, Cal S, Keeble AH, Walker D, Evans SJ, Kuhlmann UC, Cooper A, Connolly BA, Hemmings AM, Moore GR, James R, Kleanthous C (2001) Mechanism and cleavage specificity of the HNH endonuclease colicin E9. J. Mol. Biol 314: 735-749

71. Németh E (2014) Redesign of the Colicin E7 Nuclease domain into a controlled and specific artificial enzyme. Thesis, University of Szeged.

72. Shi Z, Chak KF, Yuan HS (2005) Identification of an essential cleavage site in ColE7 required for import and killing of cells. J. Biol. Chem 280: 24663–24668

73. Németh E, Körtvélyesi T, Thulstrup PW, Christensen HE, Kožíšek M, Nagata K, Czene A, Gyurcsik B (2014) Fine tuning of the catalytic activity of colicin E7 nuclease domain by systematic N-terminal mutations. Protein Sci 23: 1113-1122

74. 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. JBIC, J. Biol. Inorg. Chem 18:309-321

75. Maggio I Gonçalves MAFV (2015) Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol 33: 280–291

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

77. Németh E, Asaka MN, Kato K, Fábián Z, Oostenbrink C, Christensen HE, Nagata K, Gyurcsik B (2018) Chemical Approach to Biological Safety: Molecular-Level Control of an Integrated Zinc Finger Nuclease. ChemBioChem 19:66-75

78. Kim CA, Berg JM (1996) A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. Nat. Struct. Biol 3: 940–945

79. Németh E, Schilli GK, Nagy G, Hasenhindl C, Gyurcsik B, Oostenbrink C (2014) Design of a colicin E7 based chimeric zinc-finger nuclease. J. Comput. Aided. Mol. Des 28: 841–850

80. Ain QU, Chung JY, Kim YH (2015) Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. J. Control. Release 20:120–127

81. Mout R, Ray M, Lee YW, Scaletti F, Rotello VM (2017) In vivo delivery of CRISPR/ Cas9 for therapeutic gene editing: progress and challenges. Bioconjug. Chem 28: 880–884

82. Yang H, Wang H, Shivalila CS Cheng AW, Shi L, Jaenisch R (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154: 1370-1379

83. Mandal PK, et al., (2014) Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell Stem Cell 15:643–652

84. Liang X, Potter J, Kumar S, Zou Y, Quintanilla R, Sridharan M, Carte J, Chen W, Roark N, Ranganathan S, Ravinder N, Chesnut JD (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J. Biotechnol 208:44-53

85. Yan Q, Zhang Q, Yang H, Zou Q, Tang C, Fan N, Lai L (2014) Generation of multi-gene knockout rabbits using the Cas9/gRNA system. Cell regeneration 3: 12

86. Hruscha A, Schmid B (2015) Generation of Zebrafish models by CRISPR/Cas9 genome editing. Neuronal Cell Death 1254 :341–350

87. Worthen C, Rittié L, Fisher G (2017) Mechanical deformation of cultured cells with hydrogels. Methods Mol. Biol 1627:245–251

88. Han X, Liu Z, chan Jo M, Zhang K, Li Y, Zeng Z, Li N, Zu Y, Qin L (2015) CRISPR-Cas9 delivery to hard-to-transfect cells via membrane deformation. Sci. Adv 1: e1500454

89. Suda T, Liu D (2015) Hydrodynamic delivery. Adv. Genet 89:89–111

90. Liu F, Song Y, Liu D (1999) Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. Gene Ther 6: 1258–1266

91. Scherer F, Anton M, Schillinger U, Henke J, Bergemann C, Krüger A, Gänsbacher B, Plank C (2002) Magnetofection: Enhancing and targeting gene delivery by magnetic force in vitro and in vivo. Gene Ther 9:102-109

92. Chen C, Han D, Cai C, Tang X (2010) An overview of liposome lyophilization and its future potential. J Control Release 142:299-311

93. Stenekes RJ, Loebis AE, Fernandes CM, Crommelin DJ, Hennink WE (2000) Controlled release of liposomes from biodegradable dextran microspheres: A novel delivery concept. Pharm Res 17:690-695

94. Perez EE, Wang JB Miller JC Jouvenot Y Kim KA Liu O, Wang N, Lee G, Bartsevich VV, Lee Y, et al. (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biothechnol 26:808–817

95. Cong L, et al., (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339: 819–823

96. Lonez C, Vandenbranden M, Ruysschaert JM (2008) Cationic liposomal lipids: from gene carriers to cell signaling. Prog. Lipid Res 47:340–347

97. Goyal P, Goyal K, Kumar SG, Singh A, Katare OP, Mishra DN (2005) Liposomal drug delivery systems—clinical applications. Acta Pharm 55:1-25

98. Pradhan P, Giri J, Rieken F, Koch C, Mykhaylyk O, Döblinger M, et al. (2010) Targeted temperature sensitive magnetic liposomes for thermo chemotherapy. J Control Release 1:108-121

99. Ryu N, Kim MA, Park D, Lee B, Kim YR, Kim KH, et al. (2018) Effective PEI-mediated delivery of CRISPR-Cas9 complex for targeted gene therapy. Nanomedicin 14: 2095-2102

100. Li L, He ZY, Wei XW, Gao GP, Wei YQ (2015) Challenges in CRISPR/CAS9 delivery: potential roles of nonviral vectors. Hum. Gene Ther 26: 452-462

101. Jiang D, Salem AK (2012) Optimized dextran–polyethylenimine conjugates are efficient non-viral vectors with reduced cytotoxicity when used in serum containing environments. Int. J. Pharm 427: 71-79

102. Wang J, Dou B, Bao Y (2014) Efficient targeted pDNA/siRNA delivery with folate–low-molecular-weight polyethyleneimine–modified pullulan as non-viral carrier. Mater. Sci. Eng., C 34:98-109

103. Lv P, Zhou C, Zhao Y, Liao X, Yang B (2017) Modified-epsilon-polylysine-grafted-PEI-β-cyclodextrin supramolecular carrier for gene delivery. Carbohydr. Polym 168:103-111

104. Monajati M, Tavakoli S, Abolmaali SS, Yousefi G, Tamaddon A (2018) Effect of PEGylation on assembly morphology and cellular uptake of poly ethyleneimine-cholesterol conjugates for delivery of sorafenib tosylate in hepatocellular carcinoma. BioImpacts BI 8: 241–252

105. Kurrikoff K, Gestin M, Langel U (2016) Recent in vivo advances in cell-penetrating peptide-assisted drug delivery. Expert Opin. Drug Deliv 13:373–387

106. Suresh B, Ramakrishna S, Kim H (2017) Cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA for genome editing, Methods Mol. Biol. 1507:81–94

107. Sun W, Ji W, Hall JM, Hu Q, Wang C, Beisel CL, Gu Z (2015) Self-assembled DNA nanoclews for the efficient delivery of CRISPR–Cas9 for genome editing. Angew. Chem., Int. Ed 54:12029-12033

108. Mout R, Ray M, Yesilbag Tonga G, Lee YW, Tay T, Sasaki K, Rotello VM (2017) Direct cytosolic delivery of CRISPR/Cas9-ribonucleoprotein for efficient gene editing. ACS nano 11: 2452-2458

109. Shim G, Kim D, Park GT, Jin H, Suh SK, Oh YK (2017) Therapeutic gene editing: delivery and regulatory perspectives. Acta Pharmacol. Sin 38: 738–753

110. Lino CA, Harper JC, Carney JP Timlin JA (2018) Delivering CRISPR: a review of the challenges and approaches. Drug delivery 25: 1234-1257

111. Yla-Herttuala S, Endgame: glybera finally recommended for approval as the first gene therapy drug in the European Union, Mol. Ther 20: 1831–1832

112. Wang L, Li F, Dang L, Liang C, Wang C, He B et al. (2016) In vivo delivery systems for therapeutic genome editing. Int. J. Mol. Sci 17: 626

113. Wu Z, Yang H, Colosi P (2010) Effect of genome size on AAV vector packaging. Mol Ther 18:80–86.

114. Carroll KJ, Makarewich CA, McAnally J, et al. (2016) A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. Proc Natl Acad Sci USA 113:338–343

115. Hung SS, Chrysostomou V, Li F, et al. (2016). AAV-mediated CRISPR/Cas gene editing of retinal cells in vivo. Invest Ophthalmol Vis Sci 57:3470–3476

116. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov, FD, Galli C, Gregory PD, et al. (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat. Biotechnol 25:1298–1306

117. Schnell T, Foley P, Wirth M, Munch J, Uberla K (2000) Development of a self-inactivating, minimal lentivirus vector based on simian immunodeficiency virus. Hum. Gene Ther. 11:439– 447

118. Dong C, Qu L, Wang H, Wei L, Dong Y, Xiong S (2015) Targeting hepatitis B virus cccDNA by CRISPR/Cas9 nuclease efficiently inhibits viral replication. Antiviral Res 118:110-117

119. Roehm PC, Shekarabi M, Wollebo HS, Bellizzi A, He L, Salkind J, Khalili K (2016) Inhibition of HSV-1 replication by gene editing strategy. Sci. Rep 6:23146

120. Glish GL, Vachet RW (2003) The basics of mass spectrometry in the twenty-first century. Nat. Rev. Drug Discov 2:140-150

121. Banerjee S, Mazumdar S (2012) Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. Int J Anal Chem 2012:1-40

122. Watson JT, Sparkman OD (2007) Introduction to Mass Spectrometry. 4th Edition: John Wiley & Sons.

123. Johnson KA, Verhagen MFJM, Brereton PS, Adams MWW, Amster IJ (2000) Probing the stochiometry and oxidation states of metal centers in iron-sulfur proteins using electrospray FTICR mass spectrometry. Anal. Chem 72:1410-1418

124. Martic M (2011) Ferredoxins containing Heterometallic Clusters – Synthesis and Characterization. Thesis, Technical University of Denmark.

125. Corrêa DH, Ramos CH (2009) The use of circular dichroism spectroscopy to study protein folding, form and function. Afr. J. Biochem 3: 164-173

126. Woody RW (1995) Circular dichroism. Method. Enzymol 246:34-71

127. Woody RW, Koslowski A (2002) Recent developments in the electronic spectroscopy of amides and alpha-helical polypeptides. Biophys. Chem 101:535-551

128. Jakubowski H (2003) Biochemistry online: An approach based on chemical logic. College of St. Benedict, St. John's University.

129. Miles AJ, Hoffmann SV, Tao Y, Janes RW, Wallace BA (2007) Spectroscopy 21:245–255

130. Miles AJ, Janes RW, Brown A, Clarke DT, Sutherland JC, Tao Y, Wallace BA, Hoffmann SV (2008) J. Synchrotron Radiat 15:420–422

131. Noble JE, Bailey MJ (2009) Quantitation of protein. In Methods in enzymology. Academic Press 463: 73-95

132. Olson BJ, Markwell J (2007) Assays for determination of protein concentration. Curr Protoc Protein Sci 48: 3-4

133. Han SO, Mahato RI, Kim SW (2001) Water-soluble lipopolymer for gene delivery. Bioconjug. Chem 12: 337-345

134. Berne BJ, Pecora R (2000) Dynamic Light Scattering: with applications to chemistry, biology, and physics. Courier Corporation

135. http://www.substech.com/dokuwiki/lib/exe/detail.php?id=scanning_electron_microscop e&cache=cache&media=electron_microscope.png Last updated on 2012; Accessed on 07/04/2014.

136. Neumeier J (2015) Photophysics of Graphene Quantum Dots. Thesis, Max Planck Institute of Quantum Optics

137. Mosmann T (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity tests. J. Immunol. Methods 65:55–63

138. Fotakis G, Timbrell JA (2006) In vitro cytotoxicity assays: comparison of LDH, neutral red. MTT and protein assay in hepatoma cell lines following exposure to cadmium chloride. Toxicol. Lett 160:171–177

139. Lakowicz JR (1999) Principles of Fluorescence Spectroscopy. Kluwer Academic /Plenum Publishers, 2nd edition

140. Petty HR (2007) Fluorescence microscopy: Established and emerging methods, experimental strategies, and applications in immunology. Microsc Res Tech 70:687–709

141. Majtner T (2015) Texture-Based Image Description in Fluorescence Microscopy. Thesis, University of Southern Denmark

142. Latt SA, Stetten G (1976) Spectral studies on 33258 Hoechst and related bisbenzimidazole dyes useful for fluorescent detection of deoxyribonucleic acid synthesis. J. Histochem. Cytochem 24: 24–33

143. Peroza EA, Freisinger E (2008) Tris is a non-innocent buffer during intein-mediated protein cleavage. Protein Expr. Purif 57:217–225

144. Németh E, Balogh RK, Borsos K, Czene A, Thulstrup PW, Gyurcsik B (2016) Intrinsic protein disorder could be overlooked in cocrystallization conditions - an SRCD case study. Protein Sci 25:1977–1988

145. Goh HC, Sobota RM, Ghadessy FJ, Nirantar S (2016) Going native: complete removal of protein purification affinity tags by simple modification of existing tags and proteases. Protein Expr. Purif 129:18–24

146. Frey S, Görlich D (2014) A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. J. Chromatogr. A 1337:95–105

147. Cheriyan M, Chan SH, Perler F (2014) Traceless splicing enabled by substrate-induced activation of the Nostoc punctiforme Npu DnaE intein after mutation of a catalytic cysteine to serine. J. Mol. Biol 426: 4018–4029

148. Kimple ME, Brill AL, Pasker RL (2013) Overview of affinity tags for protein purification. Curr. Protein Pept. Sci 73:608–616.

149. Waugh DS (2011) An overview of enzymatic reagents for the removal of affinity tags. Protein Expr. Purif 80: 283–293

150. Zhao X, Li G, Liang S (2013) Several affinity tags commonly used in chromatographic Purification. J. Anal. Methods Chem 2013

151. Schmidt T, Skerra A (2015) The Strep-tag system for one-step affinity purification of proteins from mammalian cell culture in Affinity Chromatography Methods Mol. Biol 1286:83–95

152. Yadav DK, Yadav N, Yadav S, Haque S, Tuteja N (2016) An insight into fusion technology aiding efficient recombinant protein production for functional proteomics. Arch. Biochem. Biophys 612:57–77

153. Kosobokova EN, Skrypnik KA, Kosorukov VS (2016) Overview of fusion tags for recombinant proteins. Biochemistry (Mosc.) 81: 187–200

154. Pina AS, Lowe CR, Roque ACA (2014) Challenges and opportunities in the purification of recombinant tagged proteins. Biotechnol. Adv 32:366–381

155. Young CL, Britton ZT, Robinson AS (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. Biotechnol. J 7:620–634

156. Wood DW (2014) New trends and affinity tag designs for recombinant protein purification. Curr. Opin. Plant Biol 26:54–61

157. Block H, Maertens B, Spriestersbach A, Kubicek J, Schäfer F (2014) Proteolytic affinity tag cleavage. Methods Enzymol 559:71–97

158. Arnau J, Lauritzen C, Petersen GE, Pedersen J (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr. Purif 48:1–13

159. Sui MJ, Tsai LC, Hsia KC, Doudeva LG, Ku WY, Han GW, Yuan HS (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

160. Belczyk-Ciesielska A, Csipak B, Hajdu B, Sparavier A, Asaka MN, Nagata K, Gyurcsik B, Bal W (2018) Nickel(II)-promoted specific hydrolysis of zinc finger proteins. Metallomics 10:1089–1098

161. Sambrook J, Russel DW (2006) Condensed Protocols from Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

162. Micsonai A, Wien F, Kernya L, Lee YH, Goto Y, Réfrégiers M, Kardos J (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc. Natl. Acad. Sci 112: E3095–E3103.

163. 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

164. Doudeva LG, Huang H, Hsia KC, Shi Z, Li CL, Shen Y, Cheng YS, Yuan HS (2006) Crystal structural analysis and metal-dependent stability and activity studies of the ColE7 endonuclease domain in complex with DNA/Zn2+ or inhibitor/Ni^{2+} Protein Sci 15:269–280

165. Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. Biophotonics Int 11: 36-42

166. Watły J, Hecel A, Rowinska-Zyrek M, Kozłowski H (2018) Impact of histidine spacing on modified polyhistidine tag–Metal ion interactions. Inorg. Chim. Acta 472: 119–126

167. Jakab NI, Lőrincz O, Jancsó A, Gajda T, Gyurcsik B (2008) Approaching the minimal metal ion binding peptide for structural and functional metalloenzyme mimicking. Dalton Trans. 48:6987–6995.

168. Jancsó A, Kolozsi A, Gyurcsik B, Nagy NV, Gajda T (2009) Probing the Cu2+ and Zn2+ binding affinity of histidine-rich glycoprotein. J. Inorg. Biochem 103: 1634–1643

169. Ku WY, Liu YW, Hsu YC, Liao CC, Liang PH, Yuan HS, Chak KF (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. Nucl. Acids Res 30:1670–1678

170. Lee Y, Lim C (2008) Physical basis of structural and catalytic Zn-binding sites in proteins. J. Mol. Biol. 379: 545–553

171. Németh E, Kožíšek M, Schilli GK, Gyurcsik B (2015) Preorganization of the catalytic Zn^{2+}-binding site in the HNH nuclease motif—A solution study. J. Inorg. Biochem 151:143–149

172. Oppel F, Schürmann M, Goon P, Albers AE, Sudhoff H (2018) Specific targeting of oncogenes using CRISPR technology. Cancer Res 78:5506-5512

173. Tang H, Shrager JB (2016) CRISPR/Cas-mediated genome editing to treat EGFR-mutant lung cancer: a personalized molecular surgical therapy. EMBO Mol. Med 8: 83-85

174. Huang K et al (2017) The CRISPR/Cas9 system targeting EGFR exon 17 abrogates NF-κB activation via epigenetic modulation of UBXN1 in EGFRwt/vIII glioma cells. Cancer Lett 388:269-280.

175. Richterich P (2004) CodonCode Aligner Version 1.2 Released. Genet Med 6:162–163

176. Brinkman EK, Chen T, Amendola M, van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res 42:e168.

177. Shalem O, Sanjana NE, Zhang F (2015) High-throughput functional genomics using CRISPR–Cas9. Nat. Rev. Genet 16: 299-311

178. Furgeson DY, Chan WS, Yockman JW, Kim SW (2003) Modified linear polyethylenimine− cholesterol conjugates for DNA complexation. Bioconjugate Chem 14: 840-847

179. Wang DA, Narang AS, Kotb M, Gaber AO, Miller DD, Kim SW, Mahato RI (2002) Novel branched poly (Ethylenimine)− cholesterol water-soluble lipopolymers for gene delivery. Biomacromolecules 3: 1197-1207

180. Mahato RI, Lee M, Han SO, Maheshwari A, Kim SW (2001) Intratumoral delivery of p2CMVmIL-12 using water-soluble lipopolymers. Mol. Ther 4:130-138

181. Bhattacharya S, Bajaj A, (2008) Fluorescence and thermotropic studies of the interactions of PEI-cholesterol based PEI-chol lipopolymers with dipalmitoyl phosphatidylcholine membranes. Biochim. Biophys. Acta (BBA)-Biomembranes 1778: 2225-2233

182. Bajaj A, Kondaiah P, Bhattacharya S (2008) Synthesis and gene transfection efficacies of PEI− cholesterol-based lipopolymers. Bioconjug. Chem 19: 1640-1651

183. Holycross DR, Chai M (2013) Comprehensive NMR studies of the structures and properties of PEI polymers. Macromo 46: 6891-6897

184. Concellón A, Hernández-Ainsa S, Barberá J, Romero P, Serrano JL, Marcos M (2018) Proton conductive ionic liquid crystalline poly (ethyleneimine) polymers functionalized with oxadiazole. RSC adv 8: 37700-37706

185. Lim YB, Han SO, Kong HU, Lee Y, Park JS, Jeong B, Kim SW (2000) Biodegradable polyester, poly [α-(4-aminobutyl)-L-glycolic acid], as a non-toxic gene carrier. Pharm. Res 17:811-816

186. Németh E, Körtvélyesi T, Kožíšek M, Thulstrup PW, Christensen HE, Asaka MN, Nagata K, Gyurcsik B (2014) Substrate binding activates the designed triple mutant of the colicin E7 metallonuclease. J Biol Inorg Chem 19: 1295-1303

187. Kobayashi S, Hiroishi K, Tokunoh M, Saegusa T (1987) Chelating properties of linear and branched poly (ethylenimines). Macromol 20: 1496-1500

188. E. Hermann (2019) Development of novel compined zinc finger nucleases. MSc Thesis, University of Szeged.

189. Abd Elhameed HAH, Hajdu B, Jancsó A, Kéri A, Galbács G, Hunyadi-Gulyás E, Gyurcsik B (2020) Modulation of the catalytic activity of a metallonuclease by tagging with oligohistidine. J. Inorg. Biochem 206:111013

8. ACKNOWLEDGEMENTS

First of all, thank God, Allah, the beneficent, the merciful for helping me to accomplish this work.

I would like to express my sincere gratitude to my supervisor Dr. Béla Gyurcsik for his continuous help, support and motivation, as well as for his effective help in applications and publishing papers. His guidance helped me in all the time of research and writing of this thesis.

I am thankful to Dr. Gábor Galbács and Dr. Tamás Kiss, as the current and the former Head of Department, for allowing me to carry out my PhD work in the department.

I would like to thank our collaborators for their help in my research and their excellent laboratories and for the professional help that I got during my research visits.

- Dr. Gábor Galbács, Albert Kéri, University of Szeged, ICP MPS
- Dr. Éva Hunyadi-Gulyás, Biological Research Centre, ESI-MS
- Dr. Edit Csapó Dr. Ditta Ungor, University of Szeged, preparation and characterization of nanoparticles
- Dr. Mónika Kiricsi, Nóra Igaz, Mohana Krishna Gopisetty, University of Szeged, Cell culture experiments

I am grateful to all the members of the Bioinorganic Research Group of the Chemistry Institute at University of Szeged for their help and the nice working atmosphere. I thank to Balint Hajdu and Zeyad Nafaee PhD students for reading the fisrt version of my thesis.

I would like to thank my family and friends for their support, patience, and inspiration that I was given during my study.

Last but not the least, I would like to acknowledge the financial support provided by the Hungarian National Research, Development and Innovation Office (NKFIH K_16/120130, GINOP 2.3.2-15-2016-00038), COST Action CA15126 ARBRE MOBIEU, Stipendium Hungaricum and Cultural Affairs & Mission Sector in Egypt.

9. PUBLICATION LIST

Identification number in the Hungarian Collection of Scientific Publications (MTMT): 10069615

Publications related to the dissertation

1. **H.A.H. Abd Elhameed**, B. Hajdu, R.K. Balogh, E. Hermann, W. Bal, E. Hunyadi-Gulyás, B. Gyurcsik: Purification of proteins with native terminal sequences using a Ni(II)- cleavable C-terminal hexahistidine affinity tag.
   Protein Expr. Purif., 159, 53–59 (2019)
   IF= 1.291

2. **H.A.H. Abd Elhameed**, B. Hajdu, A. Jancsó, A. Kéri, G. Galbács, E. Hunyadi-Gulyás, B. Gyurcsik: Modulation of the catalytic activity of a metallonuclease by tagging with oligohistidine.
   J. Inorg. Biochem., 206, 111013 (2020)
   IF= 3.2

3. **H.A.H. Abd Elhameed**, D. Ungor, N. Igaz, M.K. Gopisetty, M. Kiricsi, E.Csapó, B. Gyurcsik: High molecular weight PEI-based water-soluble lipopolymer for transfection Of Cancer cells.
   Macromolecular Bioscience, (2020); doi.org/10.1002/mabi.202000040
   IF= 2.895

ΣIF = 7.386

Presentations at international conferences

1. **H.A.H. Abd Elhameed**, B. Hajdu, Z. Fabian, E. Hermann, W. Bal, B. Gyurcsik: Affinity protein purification resulting in protein sequence without remaining amino acid residues. Talking molecules: the network that shape the living world. 2018 plenary meeting of the Association of Resources for Biophysical Research in Europe-Molecular BioPhysics in Europe (ARBRE-MOBIEU) COST Action CA 15126, Warsaw, Poland, 19-21 March, (2018)-poster

2. **H.A.H. Abd Elhameed**, B. Hajdu, Z. Fabian, E. Hermann, W. Bal, B. Gyurcsik: Affinity protein purification resulting in protein sequence without remaining amino acid residues. Biotechnology and Research conference, Rome, Italy, 25-27 April, (2018)-poster

3. **H.A.H. Abd Elhameed**, Bálint Hajdu, Enikő Hermann, Mohana Krishna Goppisetty, Mónika Kiricsi, Ditta Ungor, Edit Csapó, Wojciech Bal, Béla Gyurcsik: Metal ions as regulatory elements of artificial nucleases. ISMEC2019, International Symposium on Metal Complexes, Hajdszoboszló, Hungary, 11-14 June, (2019)-lecture
4. **H.A.H. Abd Elhameed**, Béla Gyurcsik, Mohana Krishna Goppisetty, Mónika Kiricsi, Ditta Ungor, Edit Csapó: Artificial Nucleases-Molecular Tools for Gene Therapy. The 2nd International Conference “Plant Genome Editing & Genome Engineering, Vienna, Austria, 5-6 July, (2019)-lecture

5. **H.A.H. Abd Elhameed**, Bálint Hajdu, Mohana Krishna Goppisetty, Mónika Kiricsi, Ditta Ungor, Edit Csapó, Béla Gyurcsik: 6×HIS tag modulates the catalytic activity of NColE7 nuclease. Living Molecules: towards Integrative Biophysics of the cell. 2020plenary meeting of the Association of Resources for Biophysical Research in Europe-Molecular BioPhysics in Europe (ARBRE-MOBIEU), Prague, Czech Republic, 24-26 February, (2020)-poster

Coauthor at international conferences

1. Z. Fábián, B. Hajdu, E. Hermann, **H.A.H. Abd Elhameed**, W. Bal, B. Gyurcsik: Affinity protein purification resulting in protein sequence without remaining amino acid residues. ISMEC2018, International Symposium on Metal Complexes, Florence, Italy 3-7 June, (2018)-poster

2. B. Gyurcsik, Z. Fábián, E. Hermann, E. Németh, B. Hajdu, R.K. Balogh, **H.A.H. Abd Elhameed**, C. Oostenbrink, K. Nagata: Development of novel zinc finger-based artificial nucleases. 43rd International Conference on Coordination Chemistry (ICCC2018), Sendai, Japan 30 July - 4 August, (2018)

3. B. Gyurcsik, B. Hajdu, Z. Fábián, E. Hermann, E. Németh, R.K. Balogh, **H. A.H. Abd Elhameed**, C. Oostenbrink, K. Nagata: Multiple allosteric control in novel zinc finger-based artificial nucleases. 14th European Biological Inorganic Chemistry Conference (EuroBIC 14), Birmingham, UK, 26-30 August, (2018)

4. B. Gyurcsik, B. Hajdu, E. Hermann, R.K. Balogh, **H.A.H. Abd Elhameed**: Intramolecular allosteric control of NColE7 metallonuclease based on the specificprotease action of nickel (II) ions. Molecular Biophysics: ABC of the puzzle of Life, ARBRE-MOBIEU Plenary
Meeting. March 18-20, 2019, Zagreb, Croatia.

5. **H.A.H. Abd Elhameed**, B. Hajdu, B. Gyurcsik: Modulation of catalytic activity of the NColE7 metallonuclease. ISMEC2019, International Symposium on Metal Complexes, 11-14 June 2019, Debrecen, Hungary.

6. B. Gyurcsik, B. Hajdu, **H.A.H. Abd Elhameed**, W. Bal, K. Nagata: Metal ions as regulatory elements of artificial DNA cleaving enzymes. 15th International Symposium on Applied Bioinorganic Chemistry (ISABC15), June 2-5, 2019, Nara, Japan.

7. E. Hermann, **H.A.H. Abd Elhameed**, E. Németh, R. Csáki, B. Hajdu, B. Gyurcsik: Purification and characterization of the C45-ZF-N85 artificial zinc-finger nuclease and its mutants. Progressive Trends In Coordination, Bioinorganic and Applied Inorganic Chemistry, XXVII. International Conference on Coordination and Bioinorganic Chemistry
(XXVII. ICCBIC), June 2-7, 2019, Smolenice, Slovakia.

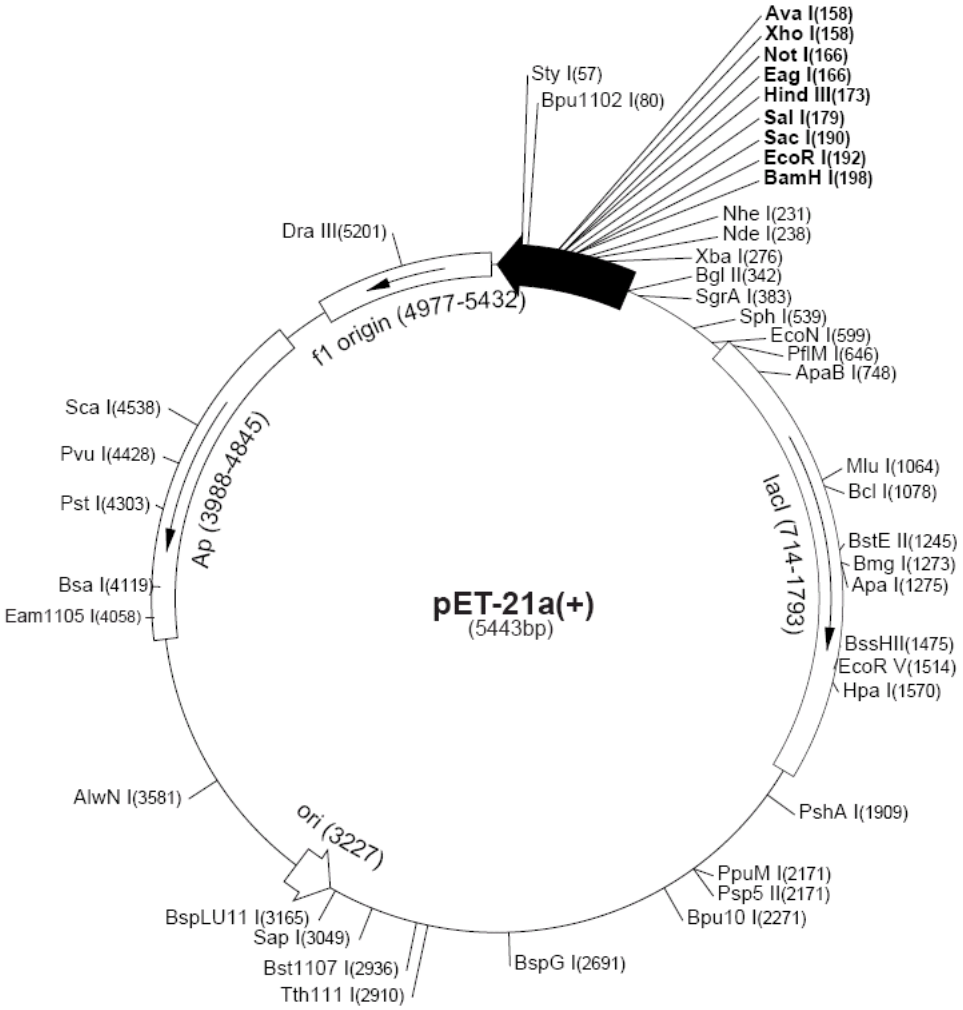
8. B. Hajdu, **H.A.H. Abd Elhameed**, E. Hermann, R.K. Balogh, É. Hunyadi-Gulyás, K. Kato, K. Nagata, W. Bal, B. Gyurcsik: Applications of Ni(II)-induced peptide bond cleavage. 19th International Conference on Biological Inorganic Chemistry (ICBIC-19), August 11-16 2019, Interlaken, Switzerland.

9. B. Gyurcsik, B. Hajdu, **H.A.H Abd Elhameed**, W. Bal, K. Nagata: Metal ions as regulatory elements of artificial DNA cleaving enzymes. Serbian Biochemical Society, Ninth Conference with international participation, University of Belgrade – Kolarac Endowment 14-16.11.2019. Belgrade, Serbia.

10. APPENDIX

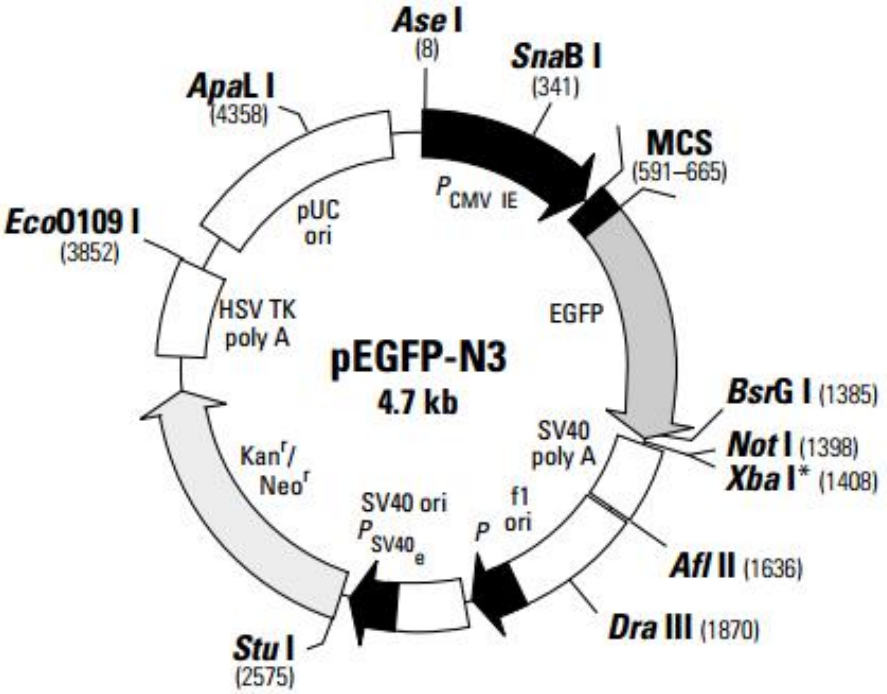
A. 1. pET-21a plasmid

Plasmid pET-21a was purchased from Novagen and used in our laboratory for DNA cloning and protein expression. It provides ampicillin resistance as a selectable marker in *E. Coli* and contains C-terminal His•Tag sequence for the purpose of protein purification. Schematic representation of plasmid pET-21a from Addgene was shown below.



A.2. pEGFP-N3 plasmid

pEGFP-N3 plasmid from BD Bioscience Clontech was used as a reference for loading it into the WSLP, it provides kanamycin resistance as a selectable marker in *E. Coli*. It can be used simply to express EGFP in the mammalian cell lines of interest as a transfection marker. A map of pEGFP-N3 is shown below.



A.3. pX458 (pSpCas9(BB)-2A-GFP plasmid

Plasmid pX458 (pSpCas9(BB)-2A-GFP from Addgene) was chosen for our study as it provides ampicillin resistance as a selectable marker in *E. coli* and offers green fluorescent protein (GFP) expression for tracking in mammalian cells. A map of pX458 is shown below. This plasmid encodes for the expression of a CRISPR single guide RNA (sgRNA), Cas 9 protein and 2A-EGFP simultaneously. It contains U6 promoter which controls the transcription of the guide RNAs by RNA polymerase III in cells.

A.4. DNA sequence of Exon1 and Exon17 for the epidermal growth factor receptor (EGFR, ErbB1) oncogene (Homo sapiens transcript variant 1)

The target genes selected to design the gRNA are included below.

Exon 1

CCCCGGCGCAGCGCGGCCGCAGCAGCCTCCGCCCCCCGCACGGTGTGAGCGCCCGACGCGGC
CGAGGCGGCCGGAGTCCCGAGCTAGCCCCGGCGGCCGCCGCCGCCCAGACCGGACGACAGGC
CACCTCGTCGGCGTCCGCCCGAGTCCCCGCCTCGCCGCCAACGCCACAACCACCGCGCACGG
CCCCCTGACTCCGTCCAGTATTGATCGGGAGAGCCGGAGCGAGCTCTTCGGGGAGCAGCGAT
GCGACCCTCCGGGACGGCCGGGGCAGCGCTCCTGGCGCTGCTGGCTGCGCTCTGCCCGGCGA
GTCGGGCTCTGGAGGAAAAGAAAG

Exon 17

GGCCTAAGATCCCGTCCATCGCCACTGGGATGGTGGGGGCCCTCCTCTTGCTGCTGGTGGTG
GCCCTGGGGATCGGCCTCTTCATGCGAAGGCGCCACATCGTTCGGAAGCGCACGCTGCGGAG
GCTGCTGCAGGAGAGGGAG

A.5. Annealing of gRNA oligonucleotides:

The forward and reverse gRNA sequences were annealed together by using the thermal cycler as follows (95°C 2 min then decrease the temperature gradually to 25°C with a gradient -1°C / min; 70 cycles).

Reaction mixture	**Volume**
Forward oligonucleotide (100 µM)	1 µl
Reverse oligonucleotide (100 µM)	1 µl
MilliQ H2O	8 µl
Total	10 µl

A. 6. Genomic DNA extraction and sequencing

The genomic DNA was extracted from the transfected cells to sequence the targeted DNA segments to check the EGFR oncogene knock-out efficiency of the applied CRISPR/Cas9 systems. The primers used to amplify the targeted regions located on exon 17 in the PCR reaction are listed below. The PCR products were loaded to 1% agarose gel to check the primer specificity then the gel bands were sliced and the PCR products purified by GeneJET Gel Extraction Kit (Thermo Scientific). The purified PCR products were sequenced by Sanger sequencing using the forward primer of the primer set used for each reaction.

Table A2 . Sequence of the primers used in the PCR

Name	Sequence (5' – 3')
E17-FD	GGAAACGTTGCCTTAGAAG
E17-RD	GCTTTGTGGACATTGAAGG

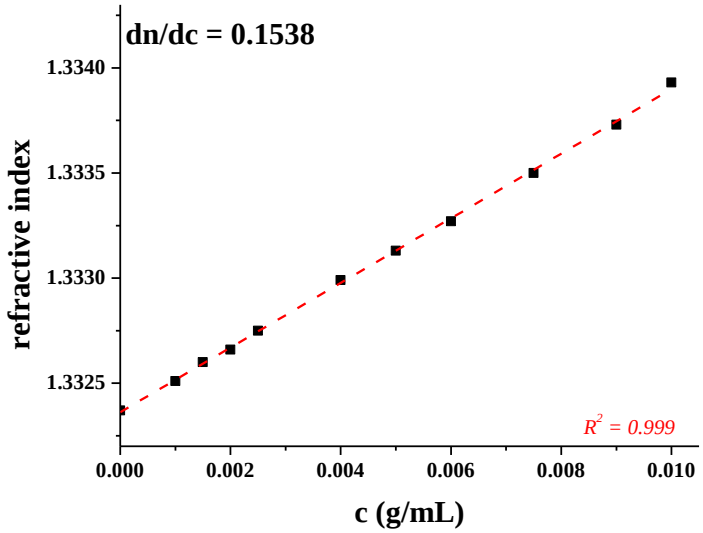
A.7. RNA extraction and SYBR-based qRT-PCR analysis

Total RNA was extracted from A549 cells after 72 h incubation by using RNeasy Plus Mini Kit (QIAGEN, GERMANY). One microgram of RNA was reverse transcribed to cDNA by using RT Kit (Taqman Reverse Transcription Reagents, Applied biosystem). The reverse transcription parameters were as follows: 25°C for 10 minutes, 37°C for 30 minutes, 95°C for 5 minutes and cooling to 4°C indefinitely. qRT-PCR was performed using Thermo Scientific™ PikoReal™ Real-Time PCR System with SYBR Green Supermix (Bio-Rad, USA). 200ng and 200nM of the cDNA and primers respectively were used in the reaction mixture. The PCR conditions used, 95°C, 7 min, followed by 40 cycles of 95°C, 15sec, 58°C,1min, data acquisition and 20°C,10 sec. EGFR expression was analyzed compared to the control by using delta delta Ct method**.** The housekeeping gene GAPDH was used for normalizing gene expression.

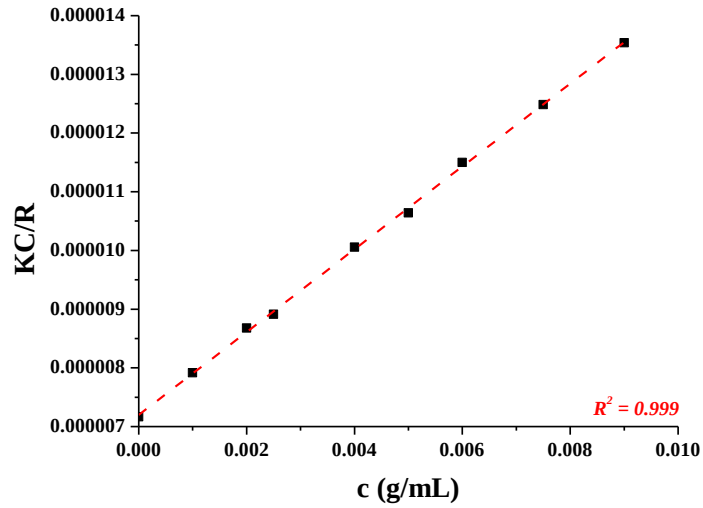
A.8. Determination of the molecular weight of PEI

The average molecular weight of PEI was not provided and therefore, we had to determine it. For the average molecular mass of PEI M_w ~ 139±1 kDa was obtained from DLS measurements.

A

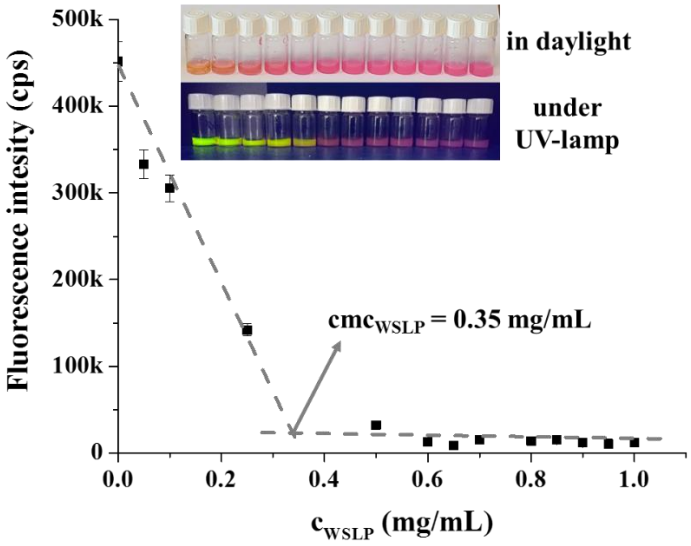


B



A.9. Determination of critical micelle concentration

WSLP is able to behave like a surfactant and forms micelles in aqueous solutions thanks to its amphiphilic chemical structure. For the determination of CMC, representing the concentration, above which the micelle formation occurs. The CMC of WSLP was determined by dye solubilization method using fluorescence spectroscopy (λ_{ex} = 490 nm, λ_{em} = 540 nm), with the photos of samples in daylight and under UV-lamp. Fluorescent Eosin red indicator was used. The fluorescence of the dyes is rather sensitive to the chemical environment. The increase of the WSLP concentration caused a decrease of the Eosin fluorescence intensity. WSLP behaves as a quencher of the dye. Based on the measured data, it can be established that the polymer applied at higher concentration than CMC caused complete fluorescence quenching in contrast to the conventional monocationic surfactants, where the fluorescence of the dye starts to increase above of CMC. The rapid decrease in the fluorescence intensity indicates the development of micelles, while the intensity does not change further upon completion of this process. The CMC of 0.35 ± 0.03 mg/mL was obtained by this method for the WSLP.



A.10. Details of WSLP/pEGFP-N3 complex preparation

From pEGFP-N3 plasmid (stock concentration = 493.9 ng μL^{-1}) ~ 1 µg amount was used for each sample preparation. Fresh samples of WSLP/pDNA were prepared according to either mass or N/P ratio.

(1) According to the mass ratio (the ratio between m(WSLP) and m(DNA) in µg):

Components	**2.5/1 ratio**	**5/1 ratio**	**10/1 ratio**	**20/1 ratio**
pEGFP-N3 (1 µg)	2 µl	2 µl	2 µl	2 µl
WSLP (1.0 mg mL^{-1})	2.5 µl (2.5 µg)	5µl (5 µg)	10 µl (10 µg)	20µl (20 µg)
Water	45.5 µl	43 µl	38 µl	28 µl
Total Volume	50 µl	50 µl	50 µl	50 µl

(2) According to WSLP–DNA N/P ratio:

2.5/1: 76 ng/200 ng/well
5/1: 152 ng/200 ng/well
10/1: 304 ng/200 ng/well
20/1: 608 ng/200 ng/well

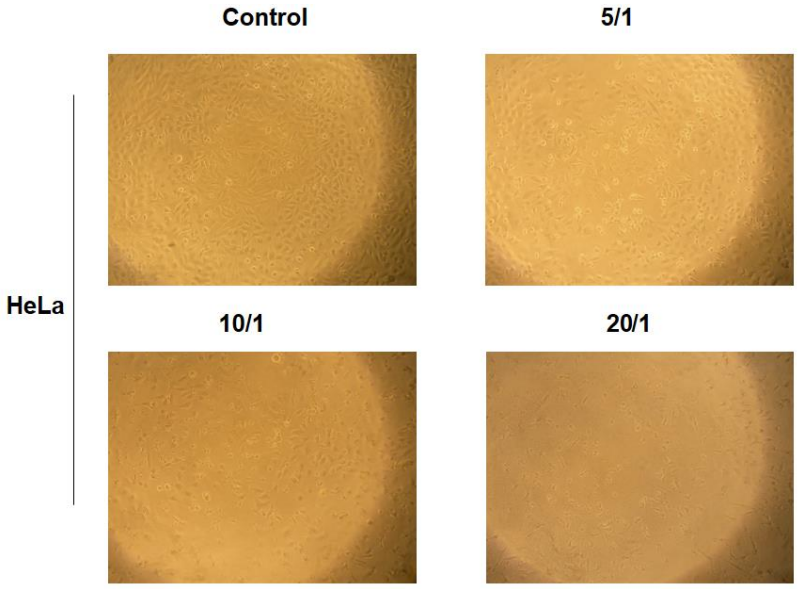
Components	**2.5/1 ratio**	**5/1 ratio**	**10/1 ratio**	**20/1 ratio**
pEGFP-N3 (1 µg)	2 µl	2 µl	2 µl	2 µl
WSLP (1.0 mg mL^{-1})	0.38 µl (0.38 µg)	0.76 µl (0.76 µg)	1.52µl (1.52 µg)	3.04µl (3.04 µg)
Water	47.62 µl	47.24µl	46.48µl	44.96 µl
Total Volume	50 µl	50 µl	50 µl	50 µl

The mixtures were incubated for 20 minutes at room temperature and then stored at 4 °C. For MTT assays 10 µl per well of these samples were applied so that the final amount of the DNA was 200 ng per well. Preparation of 2.5/1 ratio was not applied for the MTT assay.

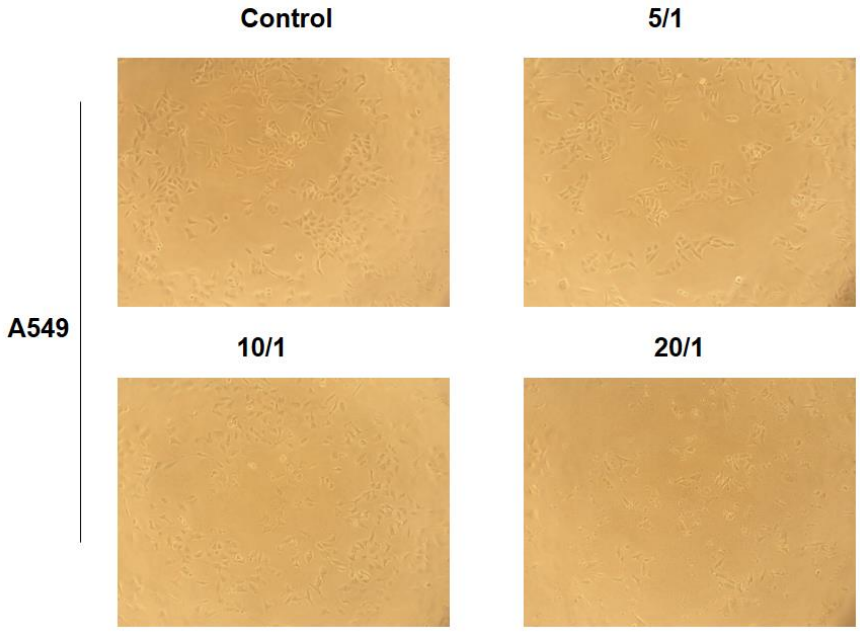
A.11. Microscopic detection of the WSLP/pEGFP-N3 complexes cytotoxicity

The microscopic images (100× magnification) of the cells treated with WSLP/pEGFP-N3 complexes after 24 hours of incubation show that the majority of the cells are destroyed by the excess of the WSLP. Cytotoxicity was reasonably low even at 10/1 N/P ratio for A549 cells, nevertheless it depended remarkably on the type of the applied cell.

A

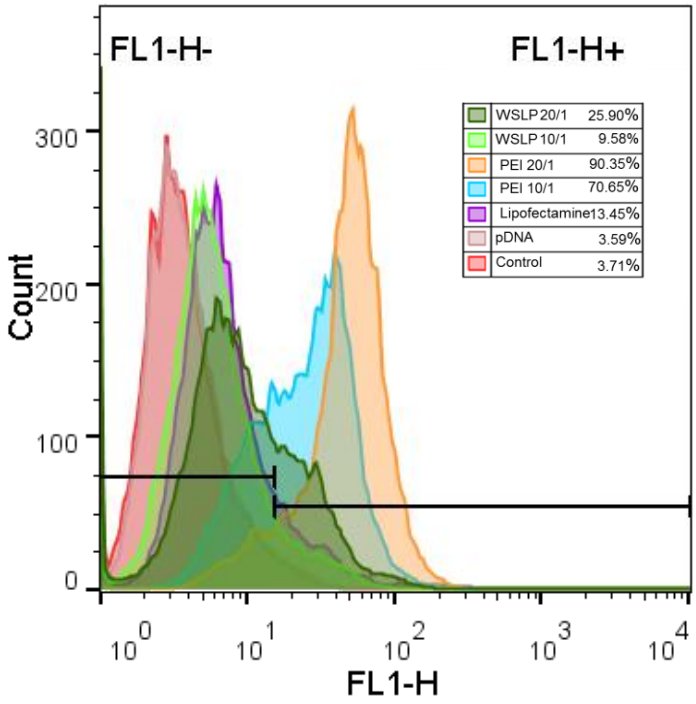


B

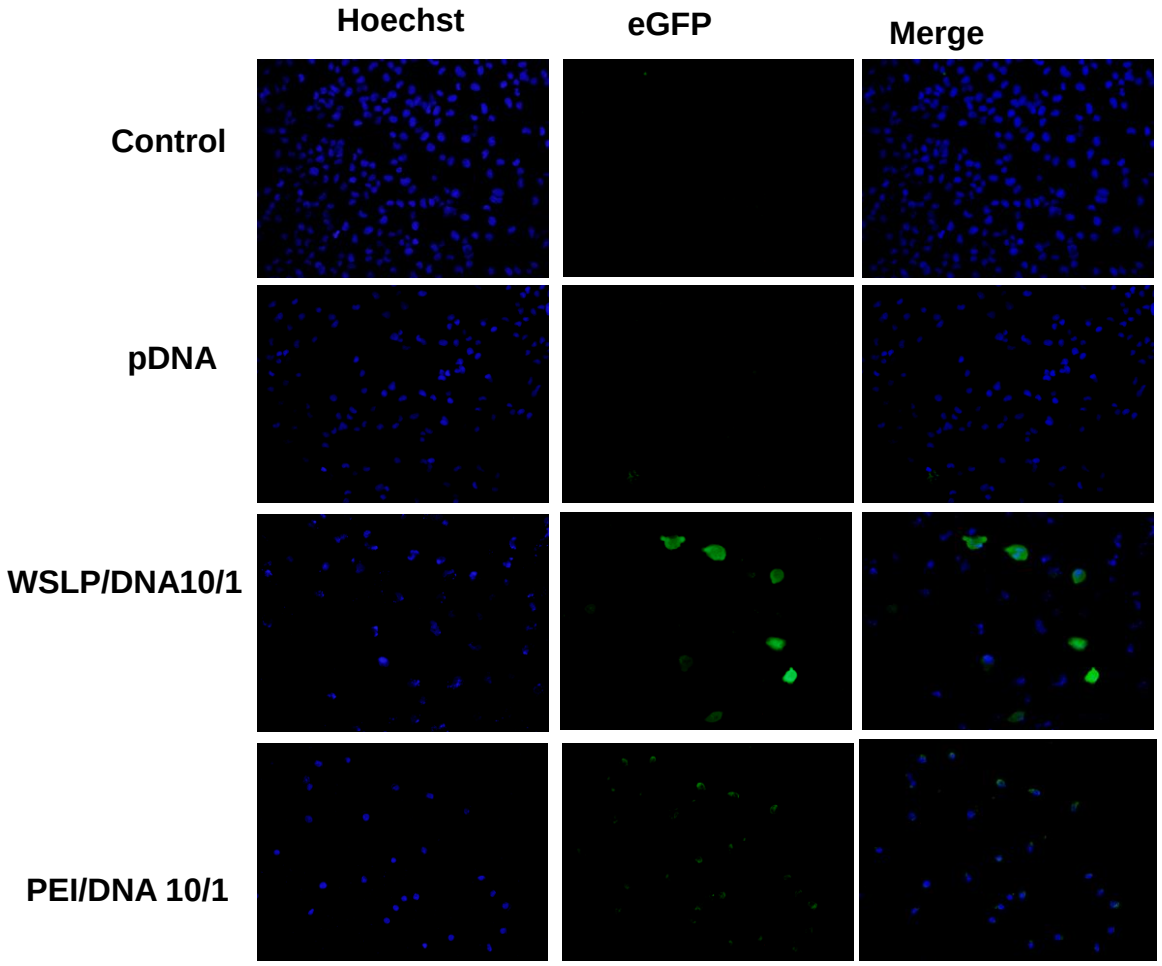


A.12. In vitro transfection

A549 cells were transfected with pX458 complexes with WSLP, PEI and Lipofectamine at various N/P ratios. The transfection efficiency of the starting HMW PEI, as well as the commercial standard transfection reagent Lipofectamine^{TM} 3000 (Thermo Fisher Scientific) was also checked by Flow cytometric analysis for comparison. The results demonstrated high transfection efficiency of highly cytotoxic PEI. The efficiency of the WSLP/pX458-E17-GGG at N/P 10/1 ratio was similar to Lipofectamine.



A.13. Confocal microscopic images of A549 cells transfected by WSLP/pEGFP-N3 complexes at 10/1 N/P ratio. PEI was used as a control transfection agent. For each triplicate image, from left to right: cell nuclei stained by Hoechst; eGFP signals; merged image. The scale bar represents 200 μm.



A.14. Assessment of WSLP/pX458-E17-GGG complexes activity by in vitro gene knockout assay

CodonCode Aligner sequences of EGFR demonstrating pX458-E17-GGG activity after seven days transfection by WSLP. Top sequence represents control sequence without gene knock-out. Highlighted region shows the region complementary to CRISPR single guide RNA. Second and third sequencing show slight increase in background noise, implying low CRISPR efficiency resulted from WSLP pX458-E17-GGG 10/1 and 20/1 respectively. b, c, TIDE data shows the percentage probability of insertion or deletion mutations on test sequence which correlates to mutation efficiency in the case of WSLP/pX458-E17-GGG 10/1 and 20/1 respectively.