Developing a expression and purification strategy for Cas9 proteins

Jia Le, Lim CID: 00865029

Department of Biology, Imperial College London,
South Kensington Campus, London, U.K.

Interned at: Protein Expression and Purification Core Facility,
European Molecular Biology Laboratory,
Heidelberg, Germany

Supervised by

Dr. Kim Remans & Ms Ines Racke

Last updated: June 3, 2016

Contents

Abstract

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) protein system is naturally found in prokaryotes as an adaptive immune response against pathogens. Cas9 proteins interact with double-stranded RNAs (gRNA) to target and cleave DNA complementary to the protospacer element found on the gRNA. By changing the protospacer element sequence, Cas9 proteins can be easily manipulated for genome engineering, making it increasingly popular in many fields, such as the study of evolutionary biology or disease modelling. In this research project, a method for expression and purification of Cas9 proteins was tried and tested for its efficiency and ability to produce functional Cas9 proteins. Cas9 proteins were easily purified in high concentrations and were proven effective in in vitro assays. Nonetheless, they were not functional in in vivo assays, indicating that this purification strategy will need to be further modified.

1 Introduction.

1.1 Genome engineering.

Genome engineering is a top-down method of synthetic biology, in which the effects of gene modifications are studied. The main aim of genome engineering is to introduce site-specific modifications and to subsequently remove any exogenous DNA or proteins. Much research nowadays is based on genome engineering, which can correct disease-causing genes, silence viral sequences, uncover gene functions and more (Gupta and Musunuru, 2014; Kaminski et al., 2016; Ran et al., 2013b).

Zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are popular approaches of genome engineering. These proteins, when coupled to nuclease domains of restriction enzymes, were able to directly bind DNA and create a double-stranded break (DSB) at specific sites. However, new ZFNs and TALENs have to be designed to target novel DNA sequences as they directly bind to DNA. Design, expression and validation of such proteins are hence difficult and problematic (Doudna and Charpentier, 2014).

Another conventional method involves the use of RNA interference (RNAi). Small RNAs are designed to be complementary to targeted RNA. When introduced into cells, RNAi binds to mRNAs and results in double-stranded RNA that will be degraded by eukaryotic cells. This method reduces the expression of proteins of interest but has been shown to have unpredictable off-target effects and is unable to completely inhibit gene expression. Moreover, RNAi mainly targets mRNA and its inhibition of gene expression is only temporary (Gilles and Averof, 2014). Researchers were hence looking for easier and better methods of genome engineering when they came across the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system.

1.2 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR).

CRISPR is originally an adaptive immune response used by prokaryotes (Sorek et al., 2013). Each CRISPR locus comprises of short repeat sequences interrupted by unique spacer sequences derived from viral elements. CRISPR-associated (Cas) proteins bind to CRISPR RNAs (crRNAs) transcribed from CRISPR loci, creating ribonucleoproteins (RNP) that target and cleave viral sequences complementary to crRNA, hence preventing viral infections (Sternberg and Doudna, 2015).

There are three main types of CRISPR-Cas systems. Type I and III CRISPR-Cas systems consist of large protein complexes. These systems are hence not feasible for experimental use as multiple genes or

proteins have to be simultaneously introduced into cells for the CRISPR-Cas system to be functional. The Type II system, on the other hand, consists of only one protein, the Cas9 protein. This system was therefore chosen for genome engineering (Doudna and Charpentier, 2014; Jinek et al., 2012).

1.2.1 Type II CRISPR-Cas system.

The DNA endonuclease, Cas9, contains two nuclease domains homologous to HNH and RuvC endonucleases. In this system, the crRNA forms a double-stranded structure with the trans-activating crRNA (tracrRNA). This double-stranded RNA or guide RNA (gRNA) binds to Cas9 proteins and directs it to the targeted DNA sequence, which is base-paired with the protospacer element of the crRNA. Thereafter, Cas9 proteins create a DSB three base pairs upstream of the Protospacer adjacent motif (PAM). PAM is used by prokaryotes to differentiate between exogenous and intracellular DNA and must be found in targeted DNA sequences for DSBs to occur. Single chimeric RNAs were also recently designed to form a similar double-stranded gRNA structure (Figure 1). This streamlines the process of gRNA purification or its expression in cells (Jinek et al., 2012).

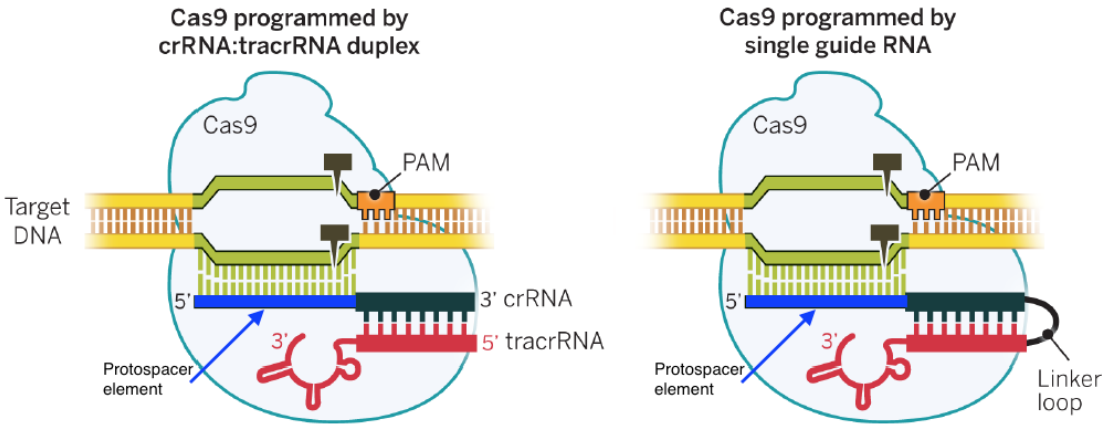

Figure 1: **Structure of Cas9.** crRNA binds to tracrRNA to create a double-stranded RNA that guides Cas9 proteins to the cleavage site. Cleavage sites are complementary to the protospacer element sequence (blue bar) found on the crRNA. HNH and RuvC-like domains each cut a single strand of DNA at a site three base pairs upstream of the Protospacer adjacent motif (PAM), creating a double-stranded break. Figure modified from Doudna and Charpentier (2014).

By changing the 20bp protospacer element of the crRNA (Figure 1), researchers can easily manipulate the Cas9 to cleave their genes of interest. This can be performed easily by subcloning the nucleotide sequence into the gRNA backbone plasmid. Moreover, with the introduction of multiple gRNAs, multiple genes can be modified simultaneously with the same protein (Jinek et al., 2012). Hence, in contrast to ZFNs and TALENs, Cas9 can be easily designed for genome engineering.

DSBs caused by endonucleases induce homology-directed repair or inaccurate repair by non-homologous

end joining (NHEJ). These mechanisms are commonly found in cells and introduce indels at the site of DSB, hence silencing the gene of interest. Unlike the temporary repression of gene expressions by RNAi, such modifications switch-off targeted genes permanently. Creating DSBs has also been the rate-limiting step in genome modifications. With the introduction of the CRISPR-Cas9 system, this limitation has therefore been largely mitigated (Chen et al., 2015; Ramalingam et al., 2013).

1.3 Cas9 Variants.

Mutant forms of the wild type *Streptococcus pyogenes* Cas9 (SpyCas9) were subsequently created. Cas9 D10A nickase and Cas9 H840 nickase were created with a non-functional HNH or RuvC endonuclease respectively. Cas9 nickases create a DSB when both of them cause nearby single-stranded breaks, hence reducing off-target DSBs caused by single Cas9 nucleases (Ran et al., 2013a).

Another method of reducing unwanted DSBs is to attach Fok1 endonucleases to dead Cas9 (dCas9) proteins. When two dCas9 proteins bind to the targeted gene, Fok1 dimerises and creates a DSB. dCas9 proteins have non-functional endonuclease activity but similarly bind to their targets (Guilinger et al., 2014). Hence, transcriptional repressors and activators have been coupled to dCas9 to conduct genome-scale screens for many types of targets, such as essential genes, tumor suppressor genes and drug resistance genes. (Gilbert et al., 2014; Shalem et al., 2015).

1.4 Project aim.

Wild type SpyCas9 nucleases and Cas9 nickases are currently available commercially (Thermo Fisher Scientific, New England BioLabs). However, Cas9 variants, such as recombinant dCas9 proteins, are not readily available.

Moreover, due to the steep price of Cas9 proteins, most researchers transfect their cells with plasmids containing gRNA and Cas9 sequences. gRNA and Cas9 proteins will thereafter be expressed in cells. However, this requires an incubation time of ~72 hours (Kleinstiver et al., 2016) and it is unknown if this may result in unwanted effects such as the integration of introduced DNA into the genome.

As such, the Protein Expression and Purification Core Facility at the European Molecular Biology Laboratory(EMBL), Heidelberg hope to be able to provide a cheaper, faster and more flexible alternative for the other researchers of EMBL that use Cas9 proteins. In this research, SpyCas9 with differing NLS are expressed, purified and tested for their functionality. By developing a purification strategy, functional Cas9 proteins can be quickly produced for future research.

2 Materials and Methods.

2.1 Expression and purification of Cas9.

The first fusion Cas9 used in this study carries at the N-terminus a His_6 tag and a nuclear localisation sequence(NLS), and will thereafter be known as NLS-Cas9. A second fusion Cas9 protein carries a N-terminal 6xHis-maltose binding protein tag (His_6-MBP tag), followed by a tobacco etch virus (TEV) protease cleavage site. It also has a C-terminal 2xSV40 NLS, and will thereafter be called Cas9-2NLS. Two different fusion proteins were used to investigate if the length of NLS affects the protein's translocation into the nucleus.

NLS-Cas9 and Cas9-2NLS were expressed from pET-NLS-Cas9-6xHis (Zuris et al., 2014) and pMJ915 (Sambrook and W Russell, 2001) respectively. Both plasmids were obtained from Addgene. His_6 tags of both proteins were used as affinity tags for protein purification. The N-terminal His_6-MBP tag of Cas9-2NLS was eventually cleaved off by TEV protease.

Plasmids were transformed into BL21(DE3) CodonPlus-RIL cells (Stratagene) using standard heat shock protocols with 45 seconds at 42°C (Sambrook and W Russell, 2001). Selection of transformed cells was performed on agar plates containing 100 µg/ml ampicillin and 33 µg/ml chloramphenicol.

Purification of NLS-Cas9 and Cas9-2NLS were carried out as previously described in Jinek et al. (2012) and Zuris et al. (2014) respectively with some modifications. Bacterial cells were grown in Luria-Bertani (LB) medium instead of a 2xTY medium. In lieu of an overnight expression, a short expression of 2NLS-Cas9 was carried out, induced by addition of 0.5mM of β-D-thiogalactopyranoside(IPTG) at a optical density of A_{600nm} 0.6. Thereafter, cells were harvested after being incubated for 3 hours at 28°C. An overnight expression was carried out for NLS-Cas9 as in Zuris et al. (2014). Instead of using a homogeniser, all cells were lysed by sonication with 4 cycles of 30 seconds of pulsing at 60 µm amplitude with 30 second pauses. Lysis buffer consisted of 20 mM Tris, 500 mM NaCl and 20mM Imidazole at pH 8.0.

Protein samples were purified by a combination of affinity and ion exchange chromatographic steps. In immobilised-metal affinity chromatography (IMAC), His_6 tags bind to the column, allowing Cas9 to be separated from other proteins or from the cleaved His_6-MBP tag of Cas9-2NLS after digestion by TEV protease. Running buffer for IMAC consisted of 20 mM Tris, 500 mM NaCl, 20mM Imidazole at pH 8.0 while proteins were eluted in 20mM Tris, 250mM NaCl, 10% glycerol and 300mM Imidazole at pH 8.0. For ion-exchange chromatography (IEX), columns were washed in 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), 100 mM KCl, 1mM TCEP and 10%

glycerol at pH 7.5 and proteins were eluted in 20 mM HEPES, 1M KCl, 1mM TCEP and 10% glycerol at pH 7.5.

Lastly, protein concentration was measured using Bio-Rad Bradford Protein Assay according to the manufacturer's protocol before being flash-frozen in liquid nitrogen. Both NLS-Cas9 and Cas9-2NLS were divided into 100-250ml aliquots and stored in 20 mM HEPES, 300mM KCl, 1mM TCEP and 10% glycerol at pH 7.5 at -80°C.

1 µl of cell lysate, 1 µl of supernatant after centrifugation and 20 µl of protein samples were denatured with loading buffer (0.6% SDS, 15% glycerol, 7.5% β-mercaptoethanol, 0.0015% bromphenol blue and 0.075 M Tris HCl, pH 6.8) and incubated at 75°C for 10 mins before being ran in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Staining of SDS-PAGEs with Coomassie Brillant Blue was done according to the standard protocol (Lawrence and Besir, 2009).

2.2 Structural analysis.

Thermofluor assays were carried out using CFX Connect^{TM} Real-Time PCR Detection System (Bio-Rad) to analyse the stability of purified Cas9 proteins. Protein samples were stained with SYPRO^{TM} Orange protein gel stain and analysed using the protocol and software from the manufacturer.

One-dimensional nuclear magnetic resonance spectroscopy (1D-NMR) was performed to compare between A5 and A6 NLS-Cas9. Sample buffer was changed into 20mM Na_xPo_4, 100mM NaCl and 0.1mM TCEP at pH 6.9 using a 5mL HiTrap desalting column(GE Healthcare) and concentrated to ~80µM using a Amicon ultracentrifugal filter (Millipore, 100-kDa molecular weight cut-off). Proteins were analysed as in Anglister et al. (1993) and the molecular weight of both proteins are calculated by: Molecular Weight (kDa) $\approx 2\tau_c$(ns), where $\tau_c \approx \frac{1}{5T_2(s)}$ at 20°C and $T_2(s) = \frac{2(\Delta_A - \Delta_B)}{\ln(I_A/I_B)}$
τ_c, the rotational correlation time, was adjusted to account for the temperature difference using the equation in Daragan and Mayo (1997). I represents the intensity at time points A and B while T_2 stands for transverse relaxation time (Anglister et al., 1993).

2.3 In vitro assays.

2.3.1 PCR amplification of target region.

A 553bp region of the cdc20 loci, consisting of the target site, was PCR amplified using the following primers: forward 5'-TTGGAACTAGATTTGCCTGGGG-3' and reverse 5'-CGGTTAAGACAGTTGG

GTTTTGAA-3'. The PCR reaction was performed using Phusion Hot Start II PCR kit (Thermo Fischer Scientific) with 200ng of mouse genomic DNA and the thermocycler at the following setting: one cycle of 98°C for 1.5 min, 30 cycles of 98°C for 15 s, 64°C for 20 s, 72°C for 45 s, and one cycle of 72°C for 5 min. PCR products were loaded onto a 2% agarose gel containing SYBR Safe (Life Technologies) in 1xTBE buffer and purified using the QIAquick Gel Extraction Kit (Qiagen). DNA concentration was measured using NanoDrop (Thermo Fisher Scientific).

2.3.2 Cas9 RNP assembly and assay analysis.

Alt-R^{TM} CRISPR tracrRNA and crRNA were ordered from IDT with the following protospacer element in the crRNA: 5'-CUCGAACACGAACUGCGCCAGUU-3'. Formation of the crRNA:tracrRNA complex (gRNA) is performed according to the user manual from IDT.

RNP complexes are assembled right before the experiment with equimolar amounts of gRNA and Cas9 in Cas9 buffer (20mM HEPES, 150mM KCL, 1mM DTT, 5% glycerol at pH7.5). The components were mixed and left at room temperature for 10 mins. Thereafter, a 20 µl reaction, containing 625nM RNP complex with 104nM PCR product, was incubated in Cas9 buffer with 10mM $MgCl_2$ at 37°C for one hour. The reaction was quenched by addition of 6xPurple Gel Loading Dye (NEB) and loaded immediately onto a 2% agarose gel containing SYBR Safe (Life Technologies) in 1xTBE buffer. Commercially produced SpyCas9 from IDT was used as a positive control and the negative control contained no gRNA nor Cas9 proteins.

The experiment was repeated using Cas9 Nuclease Reaction Buffer (NEB) to demonstrate if Cas9 proteins are active in the buffer for subsequent in vivo assays. Positive control in this experiment comprised of the A6 NLS-Cas9 in Cas9 buffer, shown in the previous experiment to be active.

Percentages of cleavage were calculated using the following equation: $\frac{b+c}{a+b+c}$, where a is the band intensity of uncleaved PCR products while b and c reflect those of the cleavage bands. GelQuant.NET software, provided by biochemlabsolutions.com, was used in the analysis of bands on agarose gels.

2.4 In vivo assays.

HeLa Kyoto cells were cultured in high glucose Dulbecco's modified Eagle's medium (DMEM; Life Technologies) supplemented with 10% (v/v) fetal bovine serum (FBS), 100 units/ml penicillin, 0.1 mg/ml streptomycin, 2 mM Glutamine and 1 mM (v/v) Sodium pyruvate at 37°C and 5% CO_2. Cells were passaged the day before to maintain <70% confluency.

2.4.1 Cas9 RNP assembly and lipofection.

Alt-R^{TM} CRISPR tracrRNA (IDT) contained the following protospacer element: 5'-ACCAGACGGACACTTACTGA-3', which targeted a region in the bub1 gene. Formation of the gRNA and RNP complexes were performed as in the in vitro assays (subsubsection 2.3.2). 1:1 and 1:2 ratios of gRNA to Cas9 proteins in 10X Cas9 Nuclease Reaction Buffer (NEB) were topped up to 12.5 µl with Opti-MEM® Medium (Optimem; Thermo Fisher Scientific) in RNP assembly.

Simultaneously, the transfection complex was formed using a mixture of 1.2 µl of Lipofectamine® RNAiMAX Reagent (Thermo Fisher Scientific) and 11.3 µl of Optimem and incubated for 5 mins at room temperature. Thereafter, RNP complexes were added to the transfection complex and incubated at room temperature for 20 mins. Each reaction was subsequently mixed with 10000 cells in a well of a 96-well plate. The experiment was first conducted with 1:1 and 1:2 ratios of 10nM of gRNA to 10nM and 20nM of Cas9 proteins respectively. Afterwards, a 1:1 ratio of gRNA:Cas9-2NLS was used at final RNP concentrations of 10, 60 and 180nM in a subsequent in vivo assay.

Cell medium was changed after 4hrs to Optimem with 10% Fetal Bovine Serium (FBS; Gibco) and incubated at 37°C with 5% CO_2 for 48 hours to allow NHEJ to take place. Thereafter, the cell medium was removed and 30 µl of lysis buffer (0.45% Tween, 0.45% TX100, 2.5mM $MgCl_2$, 50mM KCL, 10mM Tris-HCL pH8.3, 100 µg/ml proteinase K) was mixed with cells. The mixture was left on a rocker for 5 min at room temperature. The extract was then placed in a thermocycler and incubated at 56°C for 1 hour, followed by 95°C for 10 mins and placed on hold at 4°C.

2.4.2 PCR amplification of target region.

The target region was amplified with the following primers: forward 5'-AGAAGAAACGCGACTCTCAG-3' and reverse 5'-GGATAATTCGTTAGGCCCTC-3' with the HotStar HiFidelity Polymerase Kit (Qiagen) and 7 µl of the extract as template. The thermocycler settings are as follows: one cycle of 95°C for 5 mins, 40 cycles of 95°C for 30s, 60°C for 45s, reduced at 0.5°C/cycle and maintained at 50°C for the last 20 cycles, 72°C for 2 mins and one cycle of 72°C for 2 mins.

2.4.3 T7 endonuclease I assay

Mismatches caused by the hybridisation of wild-type and mutant DNA strands are cleaved by the T7 endonuclease 1 (T7E1), resulting in cleavage bands. The T7E1 assay was performed as in Lin et al. (2014) with PCR products. The reaction was quenched with 6xPurple Gel Loading Dye (NEB) and

resolved on 2% agarose gel and analysed as in the in vitro assays (subsubsection 2.3.2). The percentage cleavage was calculated using the following equation: $(1 - (1 - (\frac{b+c}{a+b+c})^{\frac{1}{2}}))$ x 100, where b and c represent the intensity of cleavage bands and a represents the total intensity of uncut PCR products and Cas9 bound PCR products (Ran et al., 2013a).

3 Results.

3.1 Purification.

Both NLS-Cas9 and Cas9-2NLS were soluble and easily separated from the cell debris, as seen by comparing the cell lysis and supernatant samples on the SDS-PAGE (Figure 2, Figure 4).

However, NLS-Cas9 sample eluted as two peaks in IEX (Figure 3). NLS-Cas9 proteins from the first and second peaks were hence stored separately as A5 NLS-Cas9 and A6 NLS-Cas9 respectively.

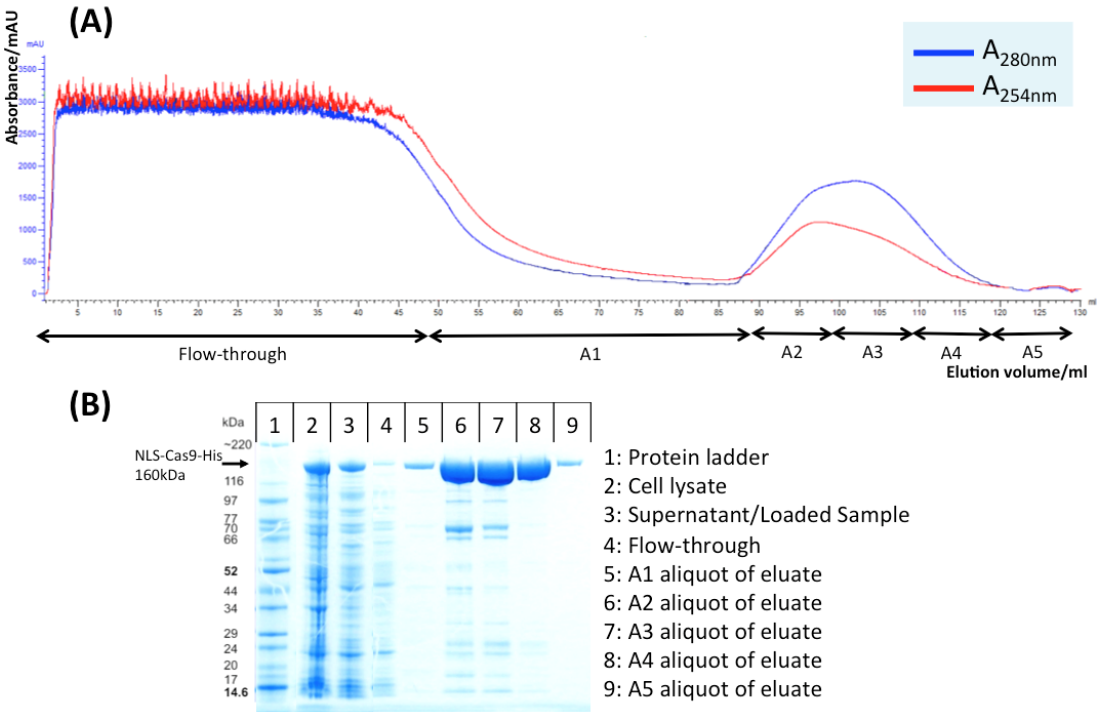

Figure 2: **Purification of NLS-Cas9 by immobilised-metal affinity chromatography (IMAC). (A)** NLS-Cas9 was expressed in *E. coli* as a fusion protein containing a N-terminal His_6 tag and purified first by IMAC. NLS-Cas9 eluted in a single peak, as shown by the absorbances at 280 and 260nm on the chromatogram of the IMAC presented here. IMAC was performed using a 5ml Ni-NTA column (GE Healthcare). **(B)** SDS-PAGE analysis of IMAC fractions. Analysis was performed on a 4-12% gradient polyacrylamide gel stained with Coomassie Brillant Blue.

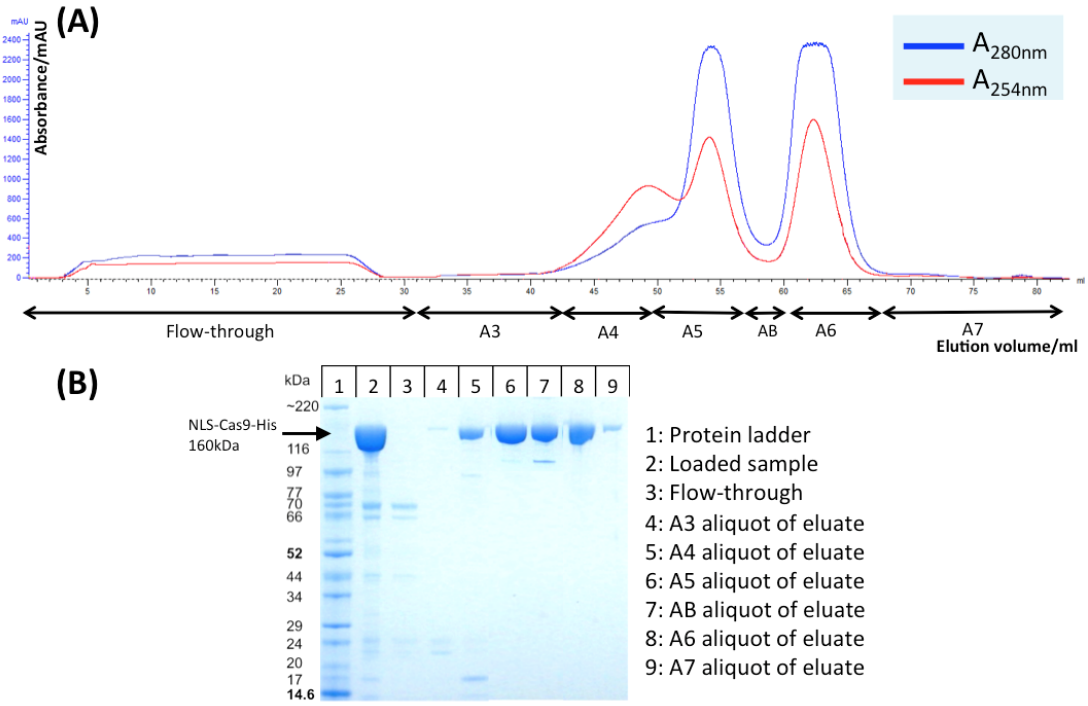

Figure 3: **Purification of NLS-Cas9 by ion-exchange chromatography (IEX). (A)** Aliquots A2-4 of IMAC in Figure 2 were further purified by IEX on a 5mL HiTrapSP column (GE Healthcare). NLS-Cas9 eluted in two peaks, with little contaminants as indicated by the ratio of absorbances at 280 and 260nm on the chromatogram of the IEX shown here. **(B)** SDS-PAGE analysis of IEX fractions, carried out as described in Figure 2. A5 and A6 aliquots of eluate were stored separately and used in subsequent experiments as A5 NLS-Cas9 and A6 NLS-Cas9 respectively.

The cleaving of the His_6-MBP tag of Cas9-2NLS by TEV protease was successful and the tag was removed by its binding to the Ni-NTA column. Cas9-2NLS was eluted in the flowthrough in the IMAC as confirmed by the SDS-PAGE (Figure 5). Cas9-2NLS eluted in a single peak in the subsequent IEX (Figure 6).

All purified proteins from IEX contain little contaminants, as deduced from the ratio of the absorbances at 280 and 260 nm on the chromatogram, which agrees with the SDS-PAGE analysis (Figure 3, Figure 6).

High concentrations of proteins were obtained from the purification. Concentrations of A5 NLS-Cas9 and A6 NLS-Cas9 were 4.04mg/mL and 9.03mg/mL respectively. Cas9-2NLS samples contained 0.8mg/mL of proteins.

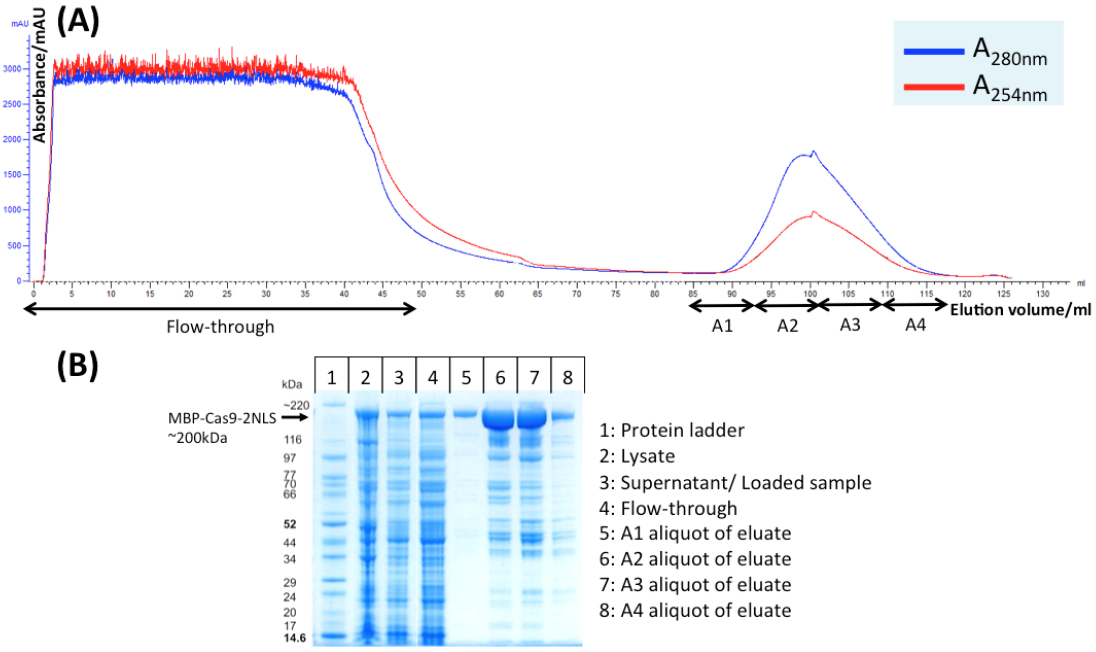

Figure 4: **Purification of Cas9-2NLS by immobilised-metal affinity chromatography (IMAC). (A)** Cas9-2NLS was expressed in *E. coli* as a fusion protein containing a N-terminal His_6-MBP tag and purified first by IMAC. Cas9-2NLS eluted in a single peak on the chromatogram of the IMAC presented here. IMAC was performed using a 5ml Ni-NTA column (GE Healthcare). **(B)** SDS-PAGE analysis of IEX fractions, performed as described in Figure 2.

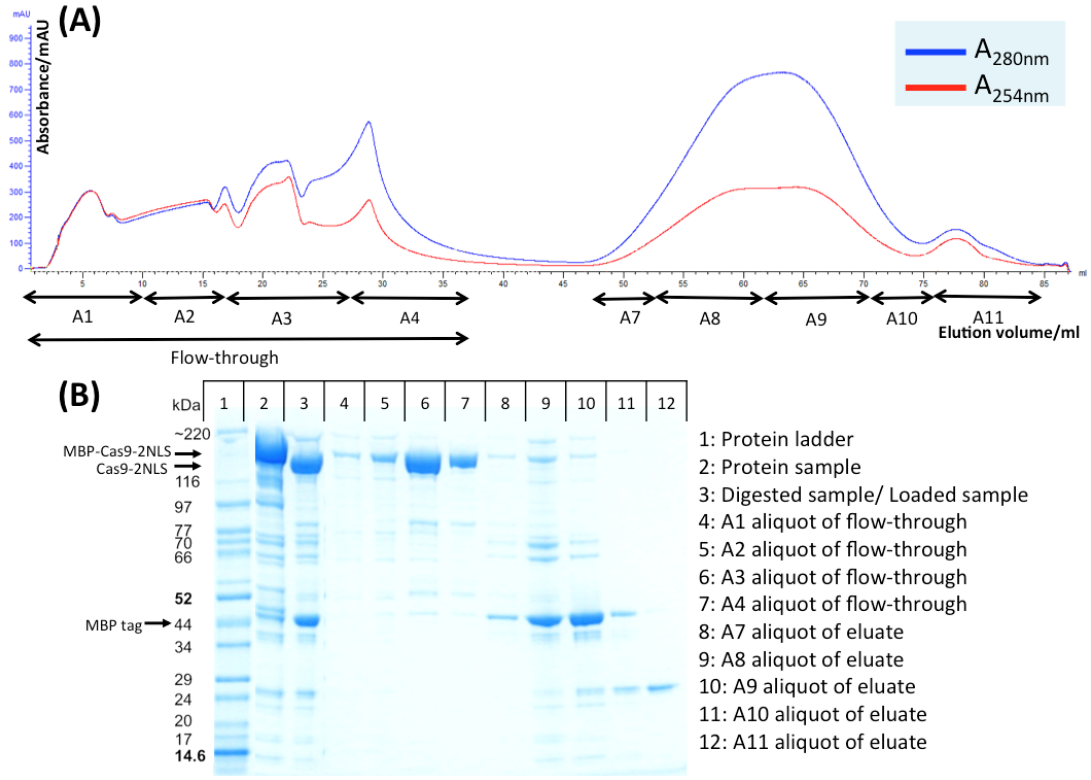

Figure 5: **Removal of N-terminal His_6-MBP tag by IMAC. (A)** After digesting A2 and A3 aliquots in Figure 4, the His_6-MBP tag was successfully removed by its binding to the 5ml Ni-NTA column (GE Heathcare). Cas9-2NLS was found in the flow-through while the last peak in the chromatogram depicted here reflects the tag found in the eluate. **(B)** SDS-PAGE analysis of IMAC fractions, carried out as described in Figure 2.

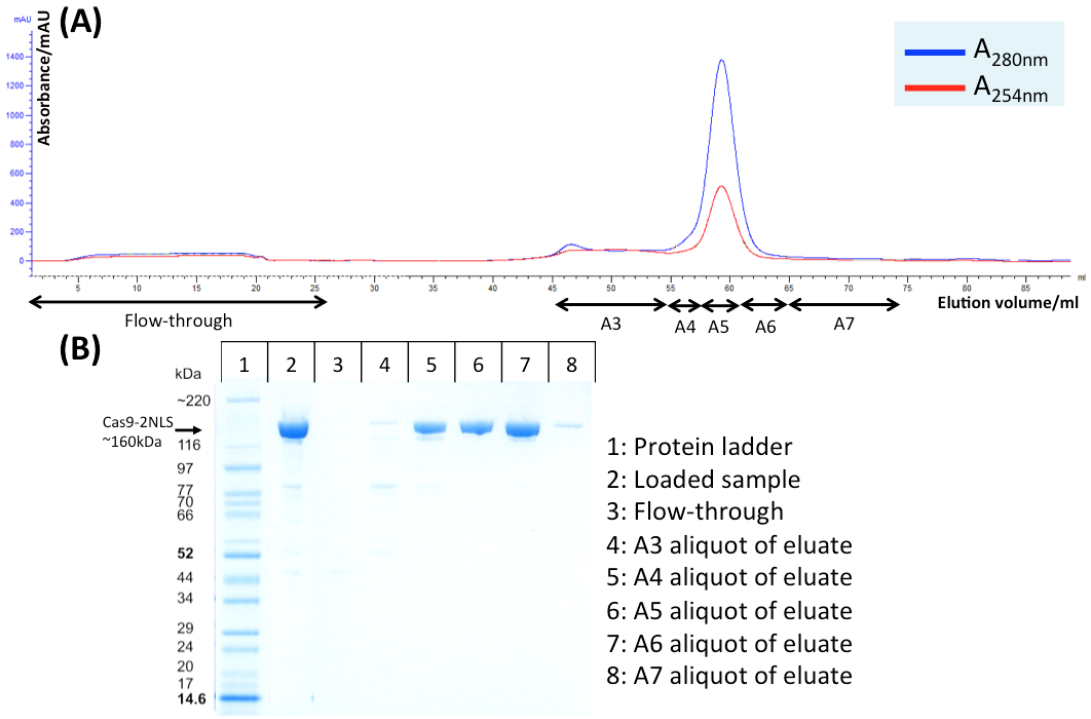

Figure 6: **Purification of Cas9-2NLS by ion-exchange chromatography (IEX). (A)** Aliquots A3 and A4 of IMAC in Figure 5 were further purified by IEX on a 5mL HiTrapSP column (GE Healthcare). Cas9-2NLS was eluted in a single monomeric peak, with little contaminants as indicated by the ratio of absorbances at 280 and 260nm on the chromatogram depicted here. **(B)** SDS-PAGE analysis of fractions of IEX, performed as described in Figure 2. A4-6 aliquots of eluate were pooled together and used in subsequent experiments.

3.2 Structural analysis.

Using thermofluor assays, denaturation temperature of both A5 NLS-Cas9 and A6 NLS-Cas9 were determined to be ∼44°C. This suggests little protein folding differences between the two NLS-Cas9.

Diagram generated by the 1D-NMR shows that A5 NLS-Cas9 and A6 NLS-Cas9 are similar (Figure 7). $\frac{I_A}{I_B} = 0.38$ for both proteins. As a delay of 1ms was used, $\Delta_A = 1ms$ and $\Delta_B = 0$. After correcting for temperature and using the equations shown in methods, the molecular weight of both proteins were calculated to be ∼ 165kDa, which agrees with the protein size. This indicates that A5 NLS-Cas9 and A6 NLS-Cas9 are similar monomers.

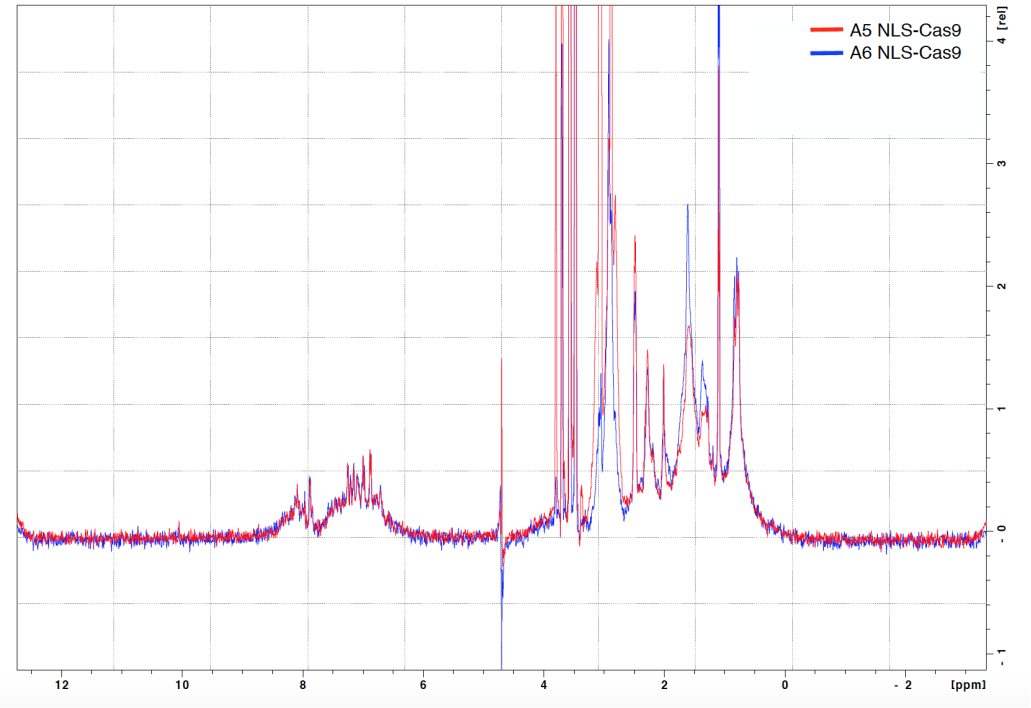

Figure 7: **1D-NMR shows that A5 NLS-Cas9 and A6 NLS-Cas9 are similar.** Graph plots frequency against intensity of signals obtained in the 1D-NMR experiment performed as in Anglister et al. (1993). The diagram shows that the aromatic and amphipathic groups of A5 NLS-Cas9 and A6 NLS-Cas9 are similar as the signals overlap.

3.3 Functional analysis.

All Cas9 proteins were functional in in vitro assays. Using the equation listed in methods, percentage cleavage of PCR products in Cas9 buffer are as follows: 84.0% for IDT Cas9, 84.0% for A5 NLS-Cas9, 86.3% for A6 NLS-Cas9 and 76.0% for Cas9-2NLS. In NEB buffer, A5 NLS-Cas9, A6 NLS-Cas9 and Cas9-2NLS cleaved 84.6%, 77.9% and 54.3% of PCR products respectively (Figure 8).

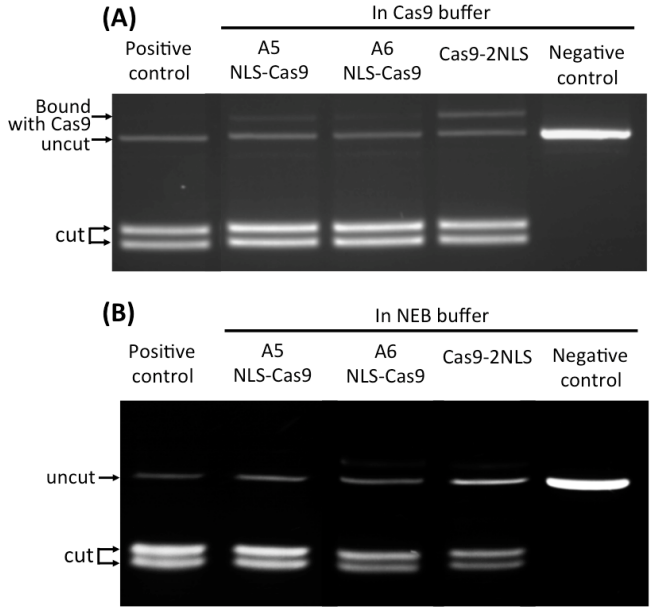

Figure 8: **In vitro assay shows that self-produced Cas9 proteins are functional. (A)** 1:1 ratio of gRNA:Cas9 RNP complexes used to cleave PCR products in Cas9 buffer. Cleavage products were resolved on a 2% agarose gel containing SYBR Safe (Life Technologies) in 1xTBE buffer. **(B)** 1:1 ratio of gRNA:Cas9 RNP complexes used to cleave PCR products in Cas9 Nuclease Reaction Buffer (NEB).

However, in both in vivo assays, self-produced Cas9 proteins were not active, despite the NEB Cas9 in the positive control being functional (Figure 9, Figure 10).

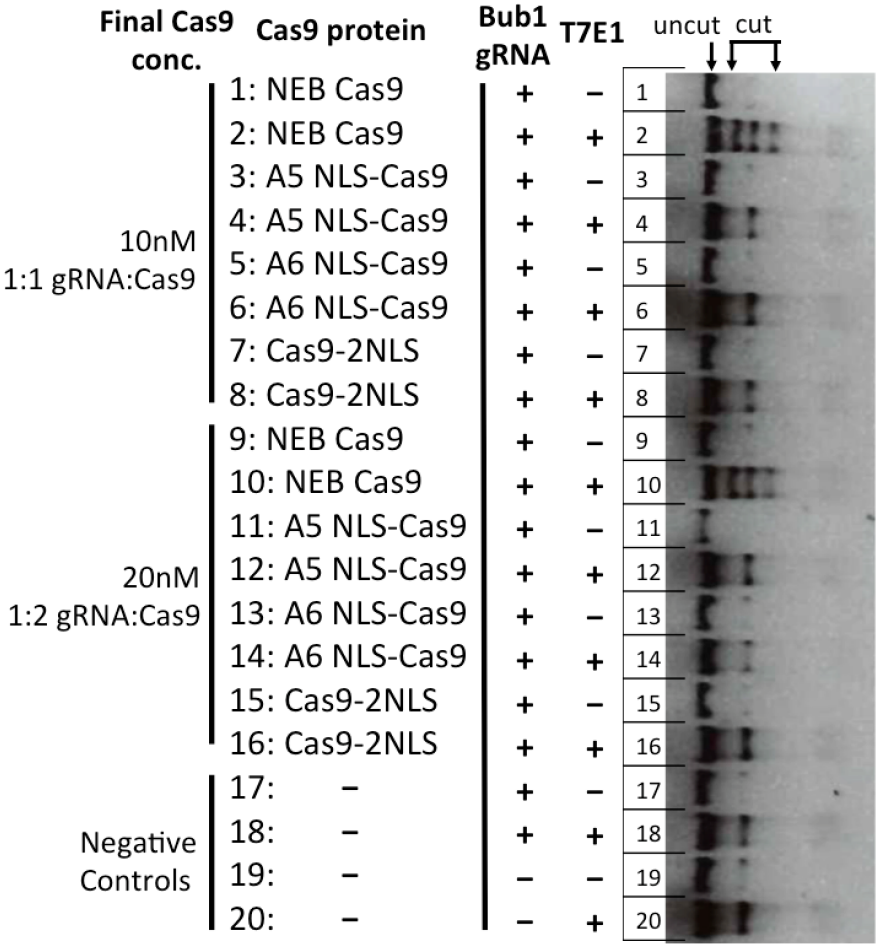

Figure 9: **Self-produced Cas9 proteins are not functional in cells.** 1:1 and 1:2 ratio of gRNA:Cas9 RNP complexes used to infect HeLa Kyoto cells. T7 endonuclease 1 assay used to detect non-homologous end joining mutations. Cleavage products were resolved on a 2% agarose gel containing SYBR Safe (Life Technologies) in 1xTBE buffer.

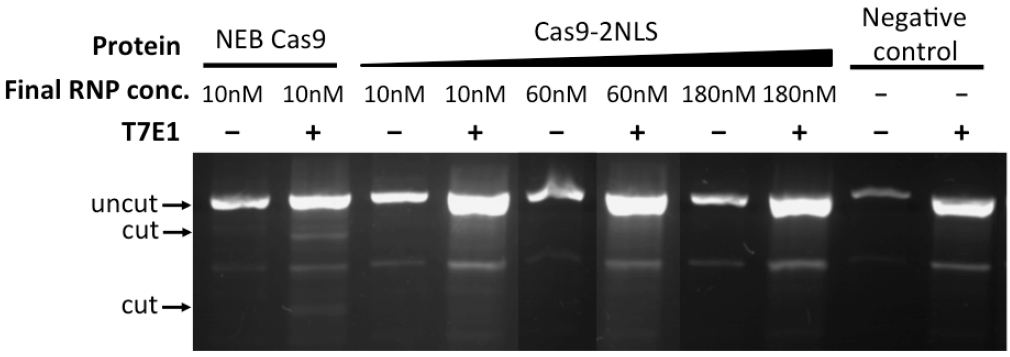

Figure 10: **Self-produced Cas9 proteins are not functional in cells.** 1:1 ratio of gRNA:Cas9 RNP complexes used to infect HeLa Kyoto cells. Analysis carried out as in Figure 9.

4 Discussion.

4.1 Analysis of self-produced Cas9 proteins

Four weeks of expression and purification yielded ~270mg of NLS-Cas9 and ~8mg of Cas9-2NLS. These amounts are sufficient for at least 500 reactions, assuming that each reaction requires a minimum of 100 pmol of Cas9 proteins (Lin et al., 2014). Purified proteins also contain little contaminants (Figure 3, Figure 6). Moreover, these proteins are functional in in vitro assays (Figure 8), with percentage of PCR products cleaved comparable to commercially produced Cas9 and to Cas9 proteins used in other research (Jinek et al., 2012).

Structural and functional assays by NMR, thermofluor and in vitro assays (Figure 7, Figure 8) show that A5 and A6 NLS-Cas9 are similar proteins with similar molecular weight. This shows that they are both correctly folded and are both monomers of Cas9. However, it remains unknown why they were eluted in two separate peaks (Figure 3).

Cas9-2NLS was not eluted in one smooth peak in its second IMAC (Figure 5) as it was mistakenly dialysed against the running buffer for IEX instead of IMAC. The difference between the running buffer of the IEX and the running buffer of the IMAC affected the absorbances recorded in the chromatogram, hence resulting in a jagged curve. This could have affected the yield of Cas9-2NLS but should not have affected the proteins themselves.

Nonetheless, all self-produced Cas9 proteins did not work in in vivo assays (Figure 9, Figure 10). This was baffling as the same buffer and concentration of Cas9 in the in vitro assays were used for the in vivo assays. Moreover, a similar approach has been used in the literature (Zuris et al., 2014). The positive control, which uses the commercially produced Cas9 protein, worked in both assays, indicating that the problem did not lie with the protocol used. Perhaps, Cas9 proteins were degraded from both the 5' and the 3' end during purification, resulting in the loss of the NLS and its inability to translocate into the nucleus.

A mass spectrometry can be conducted to show if the NLS is indeed present. This purification strategy can also be further modified by the addition of protease inhibitors to the protein sample during purification, in order to minimise protein degradation. Moreover, Cas9 proteins can also be tagged by fusion to green fluorescent protein (GFP) to show its movement and entry into the cell, to confirm that the problem lies with the translocation of the protein and not the transfection of cells (Zuris et al., 2014). Unfortunately, due to time restrictions, these assays were not carried out.

4.2 Protein transfection.

As the in vivo assays were unsuccessful, it was not possible to compare the efficiency of Cas9 cell delivery methods: protein transfection and plasmid transfection. Nonetheless, current literature suggests that percentage cleavage by Cas9 proteins transfected by both methods are similar. However, plasmid transfection requires cells to be incubated for up to ∼72 hours as cells have to express Cas9 proteins and its gRNAs before the genome can be edited. In comparison, by directly delivering proteins into cells, cells can be harvested after ∼48 hours (Kleinstiver et al., 2016; Zuris et al., 2014). By shortening the incubation time, the length of the experiment is reduced. Moreover, this decreases the chance of unexpected cell death and risk of contamination.

Furthermore, protein transfection does not involve the introduction of exogenous DNA as the RNP complex is the only substrate introduced into the cell. Cas9 RNP complexes have a half-life of a few hours in cells. Hence, after creating a DSB at specific sites, the complex will be broken down by ribonucleases and proteases. This makes it ideal for human genome editing as it poses no risk of the integration of exogenous DNA into the human genome (Hendel et al., 2015; Ramalingam et al., 2013).

Problems with low efficiency of intracellular delivery of proteins have now been resolved with the use of cationic lipid-mediated delivery (Zuris et al., 2014) and cell-penetrating peptide mediated delivery (Ramakrishna et al., 2014), making it more feasible to transfect cells with Cas9 proteins. With the increased flexibility and lower cost introduced by this purification strategy, more researchers will be able to enjoy such benefits of protein transfection in genome engineering.

4.3 Limitations of SpyCas9 proteins.

Nevertheless, many areas of Cas9-mediated gene-editing still require much improvement.

Off-targeting of Cas9 nucleases remains a problem in gene editing. SpyCas9 variants with amino acid substitutions, better gRNA designs and the use of recombinant dCas9 proteins have significantly reduced off-target effects, but more accurate tests, such as genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq), will be needed to test and reaffirm their off-targeting effects before human trials can begin. As off-targeting effects seem to be dependent on the gRNA sequences, this may limit the number of genes that can be corrected in human cells (Fu et al., 2013; O'Geen et al., 2015; Ramalingam et al., 2013; Stemmer et al., 2015).

Moreover, in *Streptococcus pyogenes*, the PAM sequence is NGG and this has to be present in the

targeted site for it to be recognised by Cas9. This limits the number of sites that Cas9 can target. However, range of Cas9 targets can be increased by using Cas9 proteins of other species, which have different PAMs (Hou et al., 2013), or by genetic engineering (Kleinstiver et al., 2015).

Furthermore, Cas9 proteins have a molecular weight of 160kDa, and are larger than ZFN and TALEN monomers. Since the discovery of SpyCas9, researchers have found smaller Cas9 proteins from other bacteria, such as *Neisseria meningitidis* (*Nme*). The *Nme* Cas9 is 287 amino acids smaller than that from *Streptococcus pyogenes* and contains a longer PAM sequence. Depending on the gene of interest, the *Nme* CRISPR-Cas9 system can have similar on-target cleavage efficiencies but lower off-target effects, making it more ideal for precise genome engineering. Nonetheless, the *Nme* CRISPR-Cas9 system is not as well-studied as the CRISPR-Cas9 system from *Streptococcus pyogenes*, and will require further trials and modifications to improve its targeting efficiencies (Lee et al., 2016).

Nonetheless, the purification strategy developed for SpyCas9 proteins may not necessarily produce functional Cas9 proteins from other bacteria and new strategies will have to be developed if other types of Cas9 proteins are found to be more efficient in genome engineering.

5 Conclusion.

The CRISPR-Cas9 system is easy to use, manipulate and control to achieve specific genetic modifications (Kaur et al., 2015; Sternberg and Doudna, 2015). Thus, it is unsurprising that this system is becoming increasingly popular in research, being used in the study of evolutionary biology, curing of diseases and genome-scale studies of genes (Gilbert et al., 2014; Kaminski et al., 2016; Pál et al., 2014). With growing interests shown in this system, the gene-editing capabilities of CRISPR-Cas9 will hopefully be further improved, thereby boosting the benefits of genome engineering.

Acknowledgements

I would like to thank Dr. Kim Remans and Ms Ines Racke at the Protein Expression and Purification Core Facility for taking their time to teach and guide me through this internship. I would also like to thank Ms Birgit Koch and Ms Bianca Nijmeijer for teaching and guiding me through the in vitro and in vivo assays and Dr. Bernd Simon for conducting the 1D-NMR experiment. Lastly, I would like to thank EMBL for providing the equipment and materials needed for this research.

References

Anglister, J., Grzesiek, S., Ren, H., Klee, C. B. and Bax, A. (1993). Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS. Journal of Biomolecular NMR *3*, 121–126.

Chen, F., Pruett-Miller, S. M. and Davis, G. D. (2015). Chromosomal Mutagenesis, vol. 1239,. Springer New York.

Daragan, V. A. and Mayo, K. H. (1997). Motional model analyses of protein and peptide dynamics using 13C and 15N NMR relaxation. Progress in Nuclear Magnetic Resonance Spectroscopy *31*, 63–105.

Doudna, J. A. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096–1258096.

Fu, Y., Foden, J. a., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K. and Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology *31*, 822–6.

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M. and Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Gilles, A. F. and Averof, M. (2014). Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution. EvoDevo *5*, 43.

Guilinger, J. P., Thompson, D. B. and Liu, D. R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nature biotechnology *32*, 577–582.

Gupta, R. M. and Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9.

Hendel, A., Bak, R. O., Clark, J. T., Kennedy, A. B., Ryan, D. E., Roy, S., Steinfeld, I., Lunstad, B. D., Kaiser, R. J., Wilkens, A. B., Bacchetta, R., Tsalenko, A., Dellinger, D., Bruhn, L. and Porteus, M. H. (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nature biotechnology *33*, 985–989.

Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L.-F., Sontheimer, E. J. and Thomson, J. a. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proceedings of the National Academy of Sciences of the United States of America *110*, 15644–9.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Kaminski, R., Chen, Y., Fischer, T., Tedaldi, E., Napoli, A., Zhang, Y., Karn, J., Hu, W. and Khalili, K. (2016). Elimination of HIV-1 Genomes from Human T-lymphoid Cells by CRISPR/Cas9 Gene Editing. Scientific reports *6*, 22555.

Kaur, K., Tandon, H., Gupta, A. K. and Kumar, M. (2015). CrisprGE: A central hub of CRISPR/Cas-based genome editing. Database *2015*, 1–8.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z. and Keith Joung, J. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P., Li, Z., Peterson, R. T., Yeh, J. R., Aryee, M. J. and Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

Lawrence, A.-M. and Besir, H. U. S. (2009). Staining of proteins in gels with Coomassie G-250 without organic solvent and acetic acid. Journal of visualized experiments : JoVE *None*, 2–4.

Lee, C. M., Cradick, T. J. and Bao, G. (2016). The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. Molecular therapy : the journal of the American Society of Gene Therapy *24*, 645–54.

Lin, S., Staahl, B., Alla, R. K. and Doudna, J. a. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. eLife *3*, 1–13.

O'Geen, H., Yu, A. S. and Segal, D. J. (2015). How specific is CRISPR/Cas9 really? Current Opinion in Chemical Biology *29*, 72–78.

Pál, C., Papp, B. and Pósfai, G. (2014). The dawn of evolutionary genome engineering. Nature reviews. Genetics *15*, 504–12.

Ramakrishna, S., Kwaku Dad, A.-B., Beloor, J., Gopalappa, R., Lee, S.-K. and Kim, H. (2014). Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome research *24*, 1020–7.

Ramalingam, S., Annaluru, N. and Chandrasegaran, S. (2013). A CRISPR way to engineer the human genome. Genome biology *14*, 107.

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380–9.

Ran, F. A., Hsu, P. D. P. P. D., Wright, J., Agarwala, V., Scott, D. a. and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281–308.

Sambrook, J. and W Russell, D. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY *None*, 999.

Shalem, O., Sanjana, N. E. and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nature reviews. Genetics *16*, 299–311.

Sorek, R., Lawrence, C. M. and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annual review of biochemistry *82*, 237–66.

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. and Mateo, J. L. (2015). CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. PLoS ONE *10*, 1–11.

Sternberg, S. H. and Doudna, J. A. (2015). Expanding the Biologist's Toolkit with CRISPR-Cas9. Molecular Cell *58*, 568–574.

Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., Maeder, M. L., Joung, J. K., Chen, Z.-Y. and Liu, D. R. (2014). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nature biotechnology *33*, 73–80.