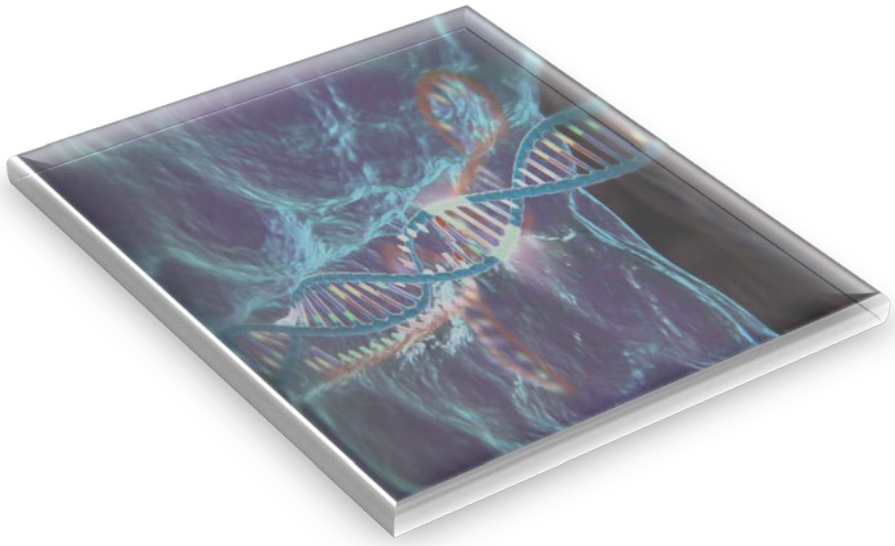
Master thesis,
Large scale expression and purification of a high fidelityCas9 nuclease

Graduate Programme in Protein Biotechnology

University of Crete – Biology Department Foundation for Research and Technology
Institute of Molecular Biology & Biotechnology
Ioanna-Maria Kefala, Heraklion 2017
Scientific supervisor: Dimitris Koutsioulis



Contents:

Acknowledgments……..3

Summary……4

Περίληψη……5

INTRODUCTION……6

1.1 Genome modification era……6

1.2 Genome editing, or engineering or hijacking……6

1.3 Repair pathways of the cell in presence of DSB……7

1.4 Nucleases: ZFNs, TALENs, RGENs……7,8

1.5 Zinc finger nucleases (ZFNs)……. ……8, 9

1.6 Transcription activator-like effectors nucleases (TALENs)……9, 10

1.7 CRISPR system biological role: from bacterial immunity to genome editing……10

1.7.1 CRISPR/Cas RNA-guided nucleases' natural work ……10, 11

1.7.2 Subtypes of CRISPR associated nucleases……11, 12, 13

1.7.3 CRISPR/Cas9 system exploitation for genome editing ……13, 14

1.8 Cas9 structure and mechanism……14, 15

1.9 The Off-target Problem and Cas9 variants engineering……16, 17, 18

1.10 High fidelity Cas9 nucleases……18, 19

1.11 The high fidelity variants of Cas9 mechanism of function……19, 20

1.12 CRISPR-Cas9 system: delivery strategies……20, 21

1.13 CRISPR-Cas9 system: Various delivery methods……22, 23

1.14 CRISPRCas9 system Vs other programmable nucleases……23, 24

1.15 Advantages of CRISPRCas9……24, 25

CRISPR-Cas9 system and applications……25, 26, 27

2. SCIENTIFIC SIGNIFICANCE – AIMS ……28

3. MATERIALS AND METHODS ……29

3.1 Over expression of SpCas9 Cluster 1 (HypaCas9) variant in E. coli cells……29,30

3.2 Cell disruption ……31

3.3 Chromatography with Ni-NTA agarose matrix ……31, 32, 33, 34

3.4 Western blotting protocol……..34

3.5 Purification of HypaCas9-TEV digest……34

3.5.1 Ni-NTA Chromatography for HypaCas9-TEV digest…….34, 35

3.5.2 Gel filtration chromatography for HypaCas9-TEV digest……..35

3.6 Quality Control Assays……..35, 36

3.7 sgRNA synthesis and purification……36,37, 38

4. RESULTS……38

4.1 Cas9 expression plasmids……38, 39

4.2 HypaCas9 expression and purification strategy……39

4.3.1 TEV protease purification scheme 1……40, 41

4.3.2 TEV protease purification scheme 2……41

4.4.1 HypaCas9 purification scheme 1……41, 42

4.4.2 HypaCas9 purification scheme 2……42

4.5 Digest of HypaCas9 by HisTEV protease scheme 1……43, 44

4.6 Digest of HypaCas9 by HisTEV protease scheme 2……44, 45

4.7.1 Purification and Quality assessment of Scheme 1 digestion reaction……45, 46, 47

4.7.2 Purification and Quality assessment of Scheme 2 digestion reaction……47, 48

Discussion……49

References……50, 51, 52, 53, 54

Acknowledgments

I would like to express my appreciation in my supervisor Dimitris Koutsioulis, who provided me with constant, everyday guidance during this research. Furthermore, I would like to thank Professors Michael Kokkinidis and Vasilis Bouriotis, as well as, Dr. Michael aivaliotis for participating in my tripartite committee. Of course, I have to extend my thanks also to the people who helped me with everyday problems during this thesis. First and foremost, I owe a big thank you to Mrs. Chrisoula and Mrs. Chara, as well as Mr. Dimitris Dialektakis, for their advice and suggestions, and for being always willing to help me every time I asked. Additionally, to Natasha Tomatsidou from Prof. Bouriotis lab and Dina Kotsifaki, from Prof. Kokkinidis lab a big thank you for your help and advice on a part of my experiments, which was conducted in your lab. Special appreciation I own to Giannis Vlatakis who helped me with my experiments during the last period of my master thesis providing me with better results and to my friend and fellow student Ioanna Makri for her everyday support and cooperation. I would also like to express my gratitude to Mrs. Peph Karpelh, Sofia Markakiou, Panagiotis Kopsiaftis and Maria Tsontaki I had the pleasure to work with every day in Minotech.

Above all, my thanks and love goes to my family, who, supported my choices, were always there for me and for giving me the greatest gift anyone could give another person, they believed in me.

Summary

CRISPRCas9 technology has been the most enlightening tool for genome editing compared to until now existing systems (ZFNs, TALENs). It's been originally recognized as a bacterial immunity system. Its function is based in the cooperation of a nuclease (Cas9) that introduces double stranded breaks and a gRNA molecule. The avoidance of protein engineering and the easily manageable gRNA molecule gave to scientists the ability of targeting any possible place in the genome and made this particular system the most efficient, simple and accurate way of editing. The last decade this breakthrough came out of the blue to "rock the boat" in scientific world and find a huge spectrum of applications.

The extensive use of the system has attracted the interest for more scientific research in purpose of making the system even more efficient and overcome its weaknesses. Scientists manage to manipulate the RNA molecule for multiplex targeting and most importantly manufactured new variants of Cas9 nuclease with dynamic properties. Cas9 variants such as high fidelity nucleases have been the latest development of this technology with the most promising of all the newest model: HypaCas9.

The current master thesis is focusing on expression and purification of HypaCas9. The aim of this project is to use HypaCas9 in a ribonucleoprotein complex for genome editing in the animal house facility of IMBB-FORTH. In this master thesis is presented the procedure followed for large scale expression and the development of an effective purification protocol. The final protein should be pure enough to pass all the quality control assays and the final *in vitro* testing that proves its functionality.

Περίληψη

Η τεχνολογία της CRISPRCas9 αποτελεί πλέον το πιο διαφωτιστικό εργαλείο για την επεξεργασία του γονιδιώματος σε σύγκριση με τα υπάρχοντα συστήματα (ZFNs, TALENs). Αρχικά αναγνωρίστηκε ως σύστημα βακτηριακής ανοσίας. Η λειτουργία του βασίζεται στη συνεργασία μιας νουκλεάσης (Cas9) η οποία εισάγει θραύσεις στη διπλή έλικα του DNA και σε ένα μόριο gRNA. Η αποφυγή χρήσης πρωτεϊνικής μηχανικής και το εύκολα διαχειρίσιμο μόριο gRNA έδωσαν στους επιστήμονες τη δυνατότητα να στοχεύσουν οποιαδήποτε πιθανή θέση στο γονιδίωμα και να καταστήσουν αυτό το σύστημα τον πιο αποτελεσματικό, απλό και ακριβή τρόπο παρέμβασης στη γενετική πληροφορία. Την τελευταία δεκαετία αυτή η ανακάλυψη προέκυψε ξαφνικά για να “ταράξει τα νερά” του επιστημονικού χώρου και να χρησιμοποιηθεί σε ένα τεράστιο φάσμα εφαρμογών.

Η εκτεταμένη χρήση του συστήματος έχει προσελκύσει το ενδιαφέρον για περισσότερη επιστημονική έρευνα προκειμένου να καταστεί το σύστημα ακόμη πιο αποτελεσματικό και να ξεπεραστούν τυχόν αδυναμίες του. Οι επιστήμονες κατορθώνουν να αναπλάθουν το μόριο RNA ώστε να στοχεύει ταυτόχρονα σε πολλαπλά σημεία του γονιδιώματος καθώς και να κατασκευάζουν νέες παραλλαγές των Cas9 νουκλεασών με ξεχωριστές ιδιότητες. Παραλλαγές της Cas9, όπως οι νουκλεάσες υψηλής πιστότητας, ήταν η τελευταία εξέλιξη αυτής της τεχνολογίας με την πιο ελπιδοφόρα από όλες το νεότερο μοντέλο: HypaCas9.

Η παρούσα διπλωματική εργασία επικεντρώνεται στην έκφραση και τον καθαρισμό της HypaCas9. Σκοπός αυτού είναι η χρήση της πρωτεΐνης σε ριβονουκλεοπρωτεΐνικο σύμπλοκο για πραγματοποίηση πειραμάτων που αφορούν την επεξεργασία γονιδιώματος στο ζωοτροφείο του IMBB-FORTH (ITE). Σε αυτή τη μεταπτυχιακή εργασία παρουσιάζεται η διαδικασία που ακολουθείται για την μεγάλης κλίμακας έκφραση και την ανάπτυξη ενός αποτελεσματικού πρωτοκόλλου καθαρισμού. Η τελική πρωτεΐνη πρέπει να είναι αρκετά καθαρή ώστε να περάσει όλες τις δοκιμές ελέγχου ποιότητας και την τελική *in vitro* δοκιμή που αποδεικνύει τη λειτουργικότητά της.

1. INTRODUCTION

1.1 Genome modification era

The modification of a particular DNA sequence (genotype) of an organism followed by the observation of the impact of this change on the organism (phenotype) is called reverse genetics and its significance in modern biology lies in its relative simplicity. Initially the modification of a sequence was achieved by methods as site-directed mutagenesis. Later scientists focused on recombination based methods that utilize the natural ability of cells to exchange DNA between its own genetic information and an exogenous DNA.

There is an expanding toolbox for genome construction and manipulation such as integrases, recombinase-mediated cassette exchange, group II introns and insertional elements, oligo-mediated allelic replacement, chemical DNA cutters, transposons. The procedure that draw scientist's attention most is homologous recombination using double-stranded DNA cassettes that enables programmable target replacement, which can be stimulated via site-specific cleavage using zinc-finger, CRISPR, or TAL nucleases *(Genome-scale engineering for systems and synthetic biology, Kevin M Esvelt, 2013)*

Although there are resemblances among the mechanism of function of those programmable nucleases, the targeting efficiency and specificity of those tools vary.

1.2 Genome editing, or engineering or hijacking

The first targeted genomic changes were introduced in yeast and in mice in the 1970s and 1980s (Genome Editing: Past, Present, and Future, Dana Carroll, 2017). This way of genome modification depended on the process of homologous recombination, which was remarkably precise but time consuming and not applicable in human cells mainly due to the high rate of **off-targeting**. Surprisingly due to a new wave of technology, termed "**gene editing**" that came out of the blue, researchers are allowed to manipulate cells by provoking specific genome modifications. The basis of this innovation lies in targetable **DNA cleavage reagents** and **cellular DNA repair** pathways. **Synthetic nucleases** are able to insert targeted double stranded breaks in chromosomal DNA that generate mutations via **non-homologous** end joining and provide the means for gene editing via **homologous** recombination. (*Genome Engineering with Targetable Nucleases, Dana Carroll et al., 2014*.) A variety of genetic alterations can be precisely and efficiently introduced into mammalian cells, ranging from knock in of single nucleotide variants to insertion of genes and deletion of chromosomal regions.
(*Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9, Rajat M. Gupta and Kiran Musunuru, 2014).*

1.3 Repair pathways of the cell in presence of DSB

Engineered nucleases work as molecular scissors that generate double strand breaks **(DSBs**) at a target area of choice, and stimulate the cells implant DNA repair pathways to end up in desirable **modifications**. The most frequent way of DSBs restitution is by non-homologous end joining (**NHEJ)**, which is an error prone process and can end up either on **insertions or deletions** *(Figure1).* More specifically, NHEJ repairs DSBs by **ligating** the two broken DNA ends using specific protein factors in absence of homologous DNA template (*Mechanism of DNA double-strand break repair by non-homologous end joining, Hefferin ML1, Tomkinson AE., 2005)*. NHEJ can give rise to small insertions or deletions (**indels**) mutating the target site. Especially indels can provoke frameshift mutations which disrupt genetic information and result in gene knockout. Alternatively, if the DSB incidence occurs in the presence of homologous donor DNA sequence it will be restored via homology-directed repair (HDR) *(Figure1)*. In this manner precise insertion or substitution edits can be introduced at the target site. Generally, plasmid DNA and viral vectors can be exploited to alter genomic sites using HR. Although, co-delivery of programmable nucleases and targeting vectors can surprisingly increase the specificity of mutagenesis.

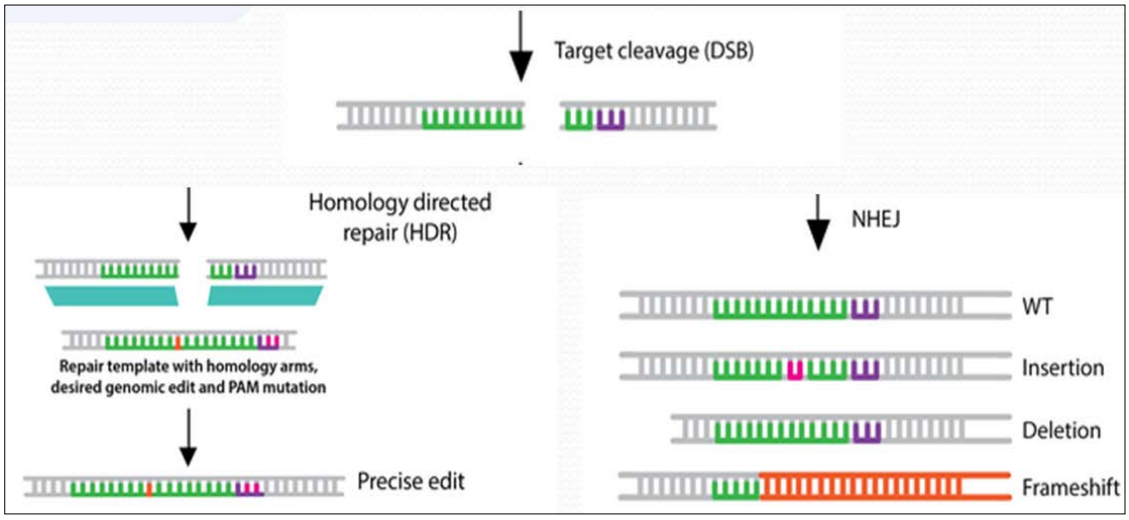


Figure 1. , *When the DSB incidence occurs in the presence of homologous donor DNA sequence it will be restored via homology-directed repair (HDR) and a precise edit will be introduced. Alternatively, restoration can be succeeded via non-homologous end joining (**NHEJ)**, which is an error prone process and can end up either on **insertions, deletions or frameshift mutations.***

1.4 Nucleases: ZFNs, TALENs, RGENs

The **Programmable nucleases** technology is evolving in exponential rates. Until recently**, meganucleases** and zinc-finger nucleases (**ZFNs**) were the only option available. Around 2011, transcription activator-like effector nucleases (**TALENs**) came to the surface as tools for genome editing. Last but not least, on January 2013, a new

class of genome editing nucleases — termed RNA-guided engineered nucleases (**RGENs**) has been reported. RGENs derived from the bacterial clustered regularly interspaced short palindromic repeat (CRISPR)–Cas (CRISPR-associated) system. The functionality of the system is ruled by small guide RNAs instead of DNA binding proteins.

Although those three types of nucleases share similar mechanism of action, they differ in composition, targetable sites, specificity, success rates and size among other characteristics. The huge value of these enzymes in research, medicine and biotechnology arises from their ability to induce site-specific DNA cleavage in the genome, the repair of which allows high-precision genome editing. Until now, programmable nucleases have enabled targeted genetic modifications in cultured cells, as well as in whole animals and plants.
(*https://www.horizondiscovery.com/resources/webinars/an-introduction-to-crispr-genome-editing)*

1.5 Zinc finger nucleases (ZFNs)

ZFNs were the first efficient way of genome hijacking. To start with, the Cys2-His2 (C2H2) domain is the most abundant type of DNA-binding motif in eukaryotes. Multiple cysteine and histidine residues are contained in the C2H2 ZF domain, which are the most common ligands for the zinc ion in proteins since they use zinc coordination that offers great stability (*Genome engineering at the dawn of the golden age. Segal DJ, Meckler JF, 2013).* One of the biggest challenges for researchers was to create ZF proteins (ZFPs) that recognize any possible DNA sequence. This ability sets ZFPs as new promising tools for biomedical research applications such as gene regulation, genome engineering and diagnostics. It is known that amino acids contained in ZF have affinity towards specific nucleotides and each finger is able to recognize 3 to 4 nucleotides of DNA. ZFs arranged into a specific array recognize a specific sequence of nucleotides on the DNA. Indeed scientists managed to create modules of six ZFs that can potentially recognize 18 bp of DNA, which would be sufficient enough to recognize a unique DNA sequence in the human genome. This has been achieved via a combination of rational design and selection *(Engineering polydactyl zinc-finger transcription factors. Beerli RR, Barbas CF 3rd, 2002*). Since a way for specific recognition of the target sequence was established, scientists came to face one more challenge: how could they create a cleavage in this area? What they did was to link ZFs to a nuclease domain, so that ZFNs not only to recognize but also to cleave DNA at the desired location. The cleavage domain fused to ZFs came from type II restriction enzyme FokI. The introduction of a DSB at the target site results after dimerization of FokI domains. Two ZFs are fused with two FokI cleavage domains to assemble functional ZFNs. Two ZF-FokI monomers bind independently in an inverted tail-to-tail orientation and with a 5–7 bp spacer sequence recognized by the cleavage domain between the binding sites. (Figure 2) *(ZFN, TALEN, and CRISPR/Cas-based methods for genome*

engineering, Gaj T, 2013. Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9, LaFountaine, 2015. Engineering and Application of Zinc Finger Proteins and TALEs for Biomedical Research, Moon-Soo Kim, 2017)

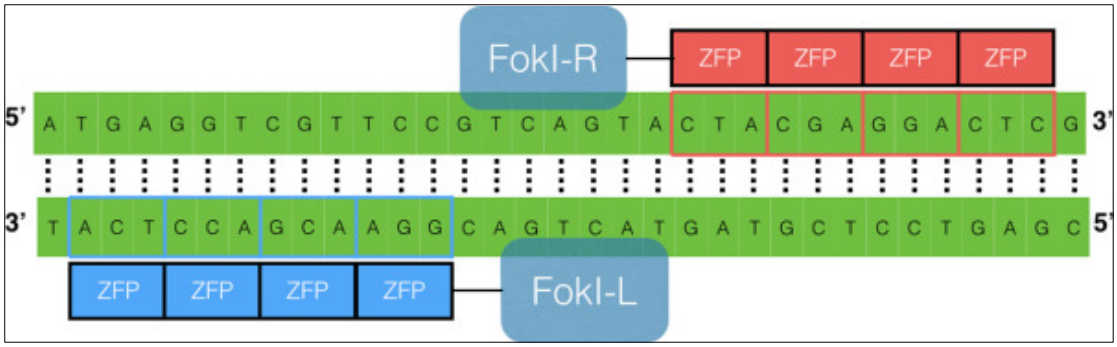


Figure 2. *The introduction of a DSB at the target site results after dimerization of FokI domains. (Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9, LaFountaine, 2015)*

It is worth mentioning that zinc finger nucleases are the most mature way of genome editing with a great variety of **applications** in therapeutics ranging from inherited disorders such as cystic fibrosis and SCID, to acquired infections as HIV or HBV. It was the first genome editing tool utilized in experiments with **animal and human** cell lines with admirable success. Specifically, the efficacy of ZNFs have been demonstrated in **vivo** in correcting genetic disorders such as haemophilia B (*In vivo genome editing restores haemostasis in a mouse model of haemophilia, Li H, 2011)* or Duchenne Muscular Dystrophy *(Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study, 2009).* Furthermore, among the gene editing technologies available the ZFNs were the first utilized in **human clinical trials** in diseases such as HIV.

1.6 Transcription activator-like effectors nucleases (TALENs)

TALENs are one of several types of programmable, engineered nucleases that bind and cleave specific DNA sequences triggering the cellular machinery to repair the cleaved DNA area by introducing indels. Transcription activator-like effectors (TALEs) are proteins from the *Xanthomonas* pathogen that activate transcription of specific plant host proteins. (*AL effectors: finding plant genes for disease and defense. Bogdanove AJ, Schornack S, Lahaye T, 2010)* Similarly to ZFNs, TALENs result from the combination of a TALE DNA - binding region and a FokI restriction endonuclease domain at their carboxyl termini.*(A One-Step System for Convenient and Flexible Assembly of Transcription Activator-Like Effector Nucleases (TALENs).Zhao J, 2016).* In contrast to ZFNs where each finger module recognizes three target DNA nucleotides, TALE proteins contain a highly conserved, central domain, usually consisting of 33–35 amino acid TALE repeats for which each protein monomer is capable of recognizing single base pairs of the target DNA (*TALENs: a widely applicable technology for targeted genome editing. Joung JK, Sander JD Nat*

Rev Mol Cell Biol. 2013). After the dimerization of two protein domains a precise double stranded break is introduced (Figure 3).The key of specificity of TALE recognition for target DNA sequences is dictated by two hypervariable residues called repeat variable di-residues (RVDs) found at amino acid positions 12 and 13 of each monomer (*A simple cipher governs DNA recognition by TAL effectors., Moscou MJ, Bogdanove AJ, 2009)*. Four different RVD modules, Asn-Asn, Asn-Ile, His-Asp and Asn-Gly recognize guanine, adenine, cytosine and thymine, respectively.

The single-nucleotide precision give rise to superior editing efficiency with minimal off-target and cytotoxicity effects when compared to ZFNs thereby making TALENs good candidates for sequence-specific genome modification (*TALEN-Mediated Knockout of CCR5 Confers Protection Against Infection of Human Immunodeficiency Virus.Shi B, Li J, Shi X, 2017).* Although TALENs are cost effective when compared to ZFNs, they are difficult to generate. The bulkiness of both ZFNs and TALENs makes it more difficult to deliver these reagents to several targeted cell types (*Novel lentiviral vectors with mutated reverse transcriptase for mRNA delivery of TALE nucleases. Mock U, Riecken K., 2014).*

Similar to ZFNs, TALENs have great potential as far as therapeutics are concerned as they have been a powerful tool in cellular and animal models for a variety of genetic diseases (SCID, DMD, A1ATD, PV, RDEB, sickle cell anemia, hemophilia A) and acquired disorders. *(TALEN gene editing takes aim on HIV, Ronald Benjamin, 2016), (The therapeutic landscape of HIV-1 via genome editing, Alexander Kwarteng, 2017).*

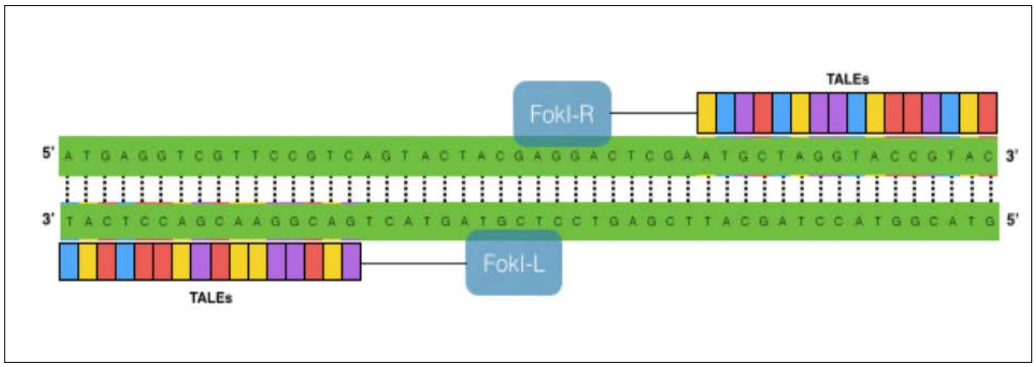


Figure 3. *After the dimerization of two protein domains a precise double stranded break is introduced* (*Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9, LaFountaine, 2015)*

1.7 CRISPR system biological role: from bacterial immunity to genome editing

1.7.1 CRISPR/Cas RNA-guided nucleases' natural work

The most popular report to targetable nucleases derives from a prokaryotic adaptive immune system mediated by clustered regularly interspaced short palindromic repeats (CRISPR). Invading DNA from viruses or plasmids is captured into small fragments (protospacers) inside the genome in copies of a repeat sequence (around

20 bps). Since the virus reenters the system the loci are transcribed, and transcripts are then processed to generate small RNAs (crRNA – CRISPR RNA), which are used to guide effector endonucleases that target invading DNA based on sequence complementarity (Figure 4). After that the cleavage of incoming genome is imminent.

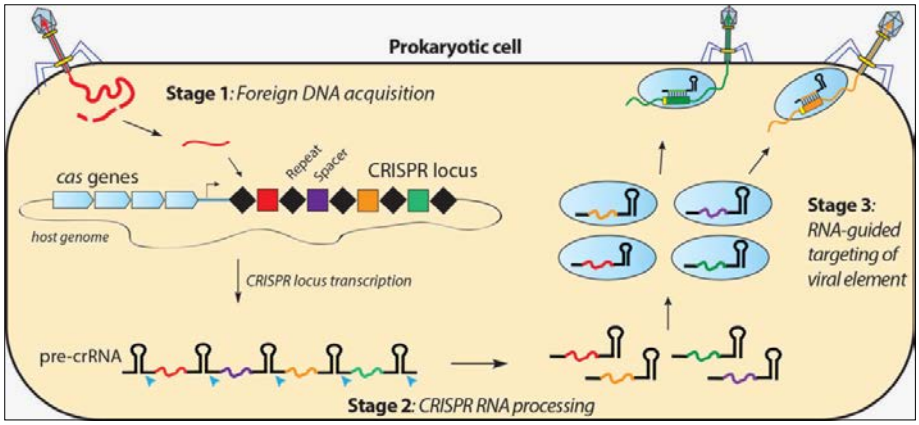


Figure 4. *The CRISPR locus is transcribed to generate mature CRISPR RNAs, each encoding a unique spacer sequence. Each crRNA associates with Cas effector proteins that use crRNAs as guides to silence foreign genetic elements that match the crRNA sequence.*
http://doudnalab.org/research_areas/crispr-systems/

1.7.2 Subtypes of CRISPR associated nucleases

The high diversion of CRISPR–Cas systems made their division into subtypes necessary. .

The Type I-E system in *E. coli* is responsible for RNA-guided silencing of foreign DNA. The silencing is achieved after the recruitment of the Cascade complex which is composed of Cse1, Cse2, Cas7, Cas5e, and Cas6e subunits and one crRNA, forming a structure that binds and unwinds dsDNA to form an R-loop (Figure 5.)

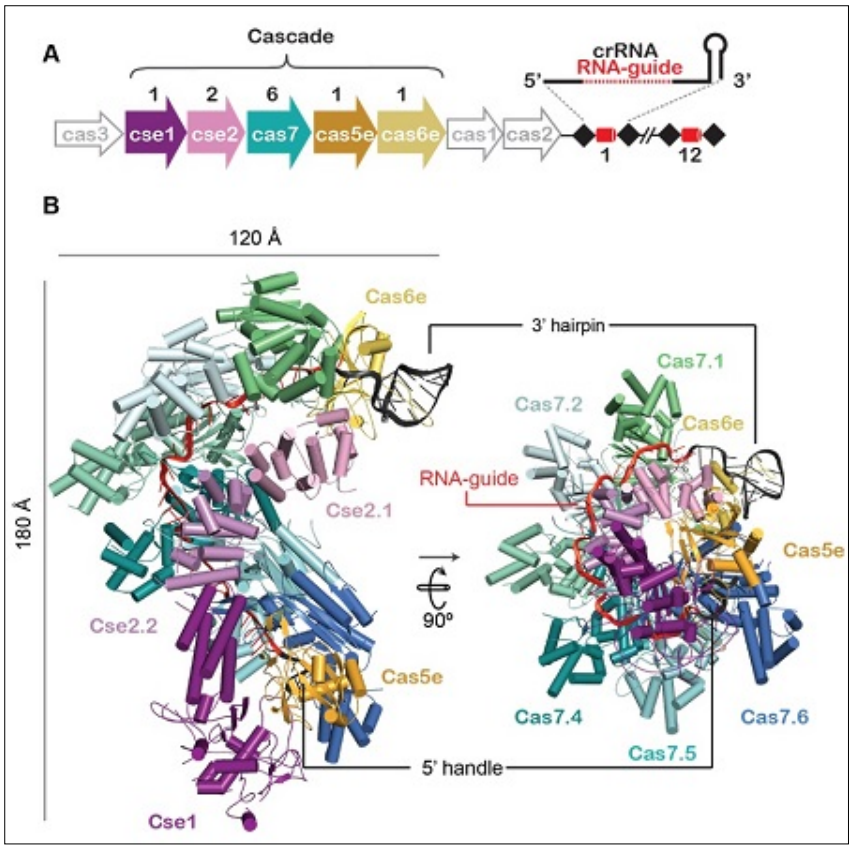


Figure 5. *A) Type IE CRISPR operon in E. coli. The Cas genes cse1, cse2, cas7, cas5e, and cas6e are expressed and combine with processed crRNA to produce Cascade. B) The Cascade crystal structure before and after guide RNA binding.*
(https://www-ssrl.slac.stanford.edu/content/science/highlight/2015-01-30/crispr-rna-guided-surveillance-escherichia-coli).

TypeIICRISPR-Cas9 is the most routinely used CRISPR gene-editing system. The DNA cleaving enzyme included in **type II CRISPR** system is the Cas9 protein which is an RNA-guided nuclease. In order to generate **double-strand breaks** in invasive DNA during an adaptive bacterial immune response Cas9 is recruited by the **guide RNA**. The gRNA is composed of the transactivating crRNA (**tracrRNA**) and **crRNA**. The crRNA contains a 20-nt protospacer element and an additional sequence that is complementary to the tracrRNA. The tracrRNA hybridizes to the crRNA and binds the Cas9 protein, forming the CRISPR-Cas9/gRNA complex to create cleavage at target sites in the genome. (Figure 6)The dual-tracrRNA: crRNA is normally engineered as a single-strand sgRNA. CRISPR RNA-guided surveillance complexes target foreign DNA for degradation through RNA–DNA base-pairing and recognition of a unique sequence adjacent to the target DNA called the protospacer adjacent motif (**PAM**). In other words, bacteria collect “protospacers” from foreign DNA sequences, incorporate them into their genomes, and use them to express short guide RNAs, which can then be used by a CRISPR-Cas system to destroy any DNA sequences matching the protospacers. The Cas9 protein is practically an endonuclease containing two nuclease domains, **RuvC and HNH**. The RuvC domain cleaves non complementary DNA strands, while the HNH domain cleaves complementary DNA

strands. After DSBs are introduced, the repair pathways of the cell (NHEJ, HDR) are activated as mentioned above.

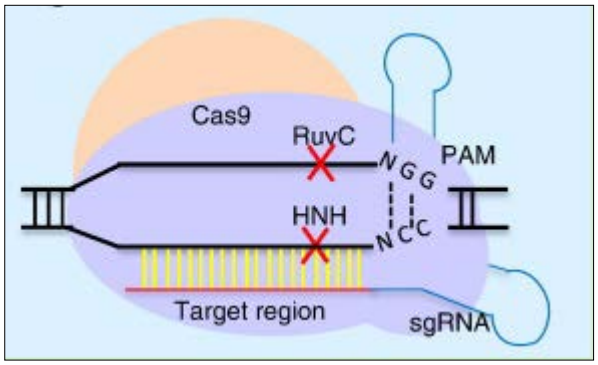


Figure 6. *Cas9 is recruited by guide RNA to the area of DNA with the specific PAM sequence in order to create cleavage. Cleavage is created after the cooperation of Puvc and HNH domains.*

In contrast to Type I and Type II CRISPR–Cas complexes that target double-stranded DNA, **Type III** complexes are able of targeting single-stranded RNA and DNA. Near-atomic resolution cryo–electron microscopy reconstructions of native Type III Cmr (CRISPR RAMP module) complexes in the absence and presence of target RNA reveal a helical protein arrangement that positions the crRNA for substrate binding. Thumblike β hairpins intercalate between segments of duplexed crRNA: target RNA to facilitate cleavage of the target at 6-nucleotide intervals. The Cmr complex is architecturally similar to the Type I CRISPR-Cascade complex, suggesting divergent evolution of these immune systems from a common ancestor. (Figure 7)

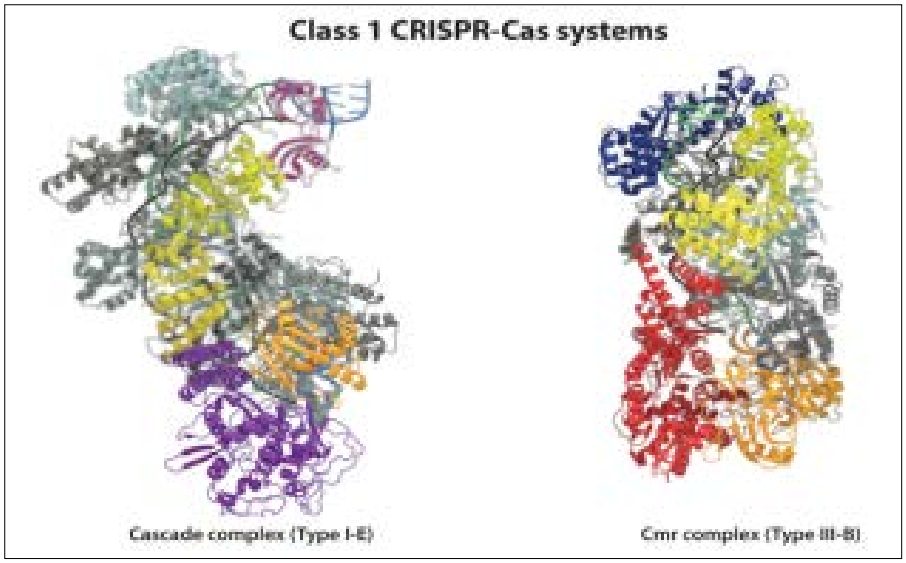


Figure 7. *Cascade complex of type I-E and Cmr complex of type III-B*

1.7.3 CRISPR/Cas9 system exploitation for genome editing

Scientists studied the molecular mechanisms by which the CRISPR pathway operates and eventually managed to hijack this system in order to exploit it to delete or insert genes in desired areas of genome. Among the three types of CRISPR/Cas system, the type II system requires only the **Cas9** nuclease to degrade DNA that matches a

single guide RNA (sgRNA). The ability to **program Cas9 for DNA cleavage** at specific sites defined by **guide RNAs** has led to its adoption as a versatile platform for genome engineering and gene regulation. Specifically, scientists managed to build chimerical crRNA-tracrRNA hybrids that direct Cas9 cleavage within the mammalian cellular genome to stimulate NHEJ or HDR-mediated genome editing. The design of the gRNA molecule is simple and the hybridization happens by base pairing. This permits the insertion of mutation in any possible place among the human genome. **Multiple guide RNAs can also be used to target several genes at once**.

1.8 Cas9 structure and mechanism

The key step of efficient perturbation in an organism's genome is specific targeting of the desired sequence of DNA. As mentioned above, after the complex of the Cas9 protein and guide RNA formation the identification of the target happens with high selectivity.
As it is depicted in figure 8, Cas9 protein is constituted from six domains, **REC I, REC II, Bridge Helix, PAM, HNH and RuvC,** the proper interaction of which is responsible for locating and cleaving target DNA. (*Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease, Carolin Anders, Ole Niewoehner, Alessia Duerst, Martin Jinek, 2014*).
Rec I domain, the largest one, interacts with the guide RNA. The vaguest of all is the **REC II** domain the role of which is not yet determined. There is an arginine-rich bridge helix that is crucial for activating the cleavage event upon binding of target DNA (*Crystal structure of Cas9 in complex with guide RNA and target DNA, Nishimasu H et al, 2014)*.
The **PAM**-Interacting domain confers PAM specificity and is therefore responsible for initiating binding to target DNA *(DNA interrogation by the CRISPR RNA-guided endonuclease Cas9, Sternberg SH, 2014*). As mentioned above, **HNH and RuvC** domains are nuclease domains that catalyze the break of DNA helix. They are highly homologous to HNH and RuvC domains found in other proteins (Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease, Carolin Anders, Ole Niewoehner, Alessia Duerst, Martin Jinek, 2014).

In absence of guide RNA Cas9 retains its inactive form. In engineered CRISPR systems, guide RNA is comprised of a single strand of RNA that forms **a T-shape** (figure 8)comprised of one tetraloop and two or three stem loops (*Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease, Carolin Anders, Ole Niewoehner, Alessia Duerst, Martin Jinek, 2014*). The guide RNA is engineered to have a 5' end that is complimentary to the target DNA sequence.

After the binding of the artificial guide RNA, Cas9 protein undergoes a conformational change that converts the inactive protein into its active form. (*Structures of Cas9 endonucleases reveal RNA-mediated conformational activation, Jinek M, 2014*). This particular mechanism is not completely understood, but the general hypothesis is that steric interactions or weak binding between protein side chains and RNA bases may be responsible (figure 8).

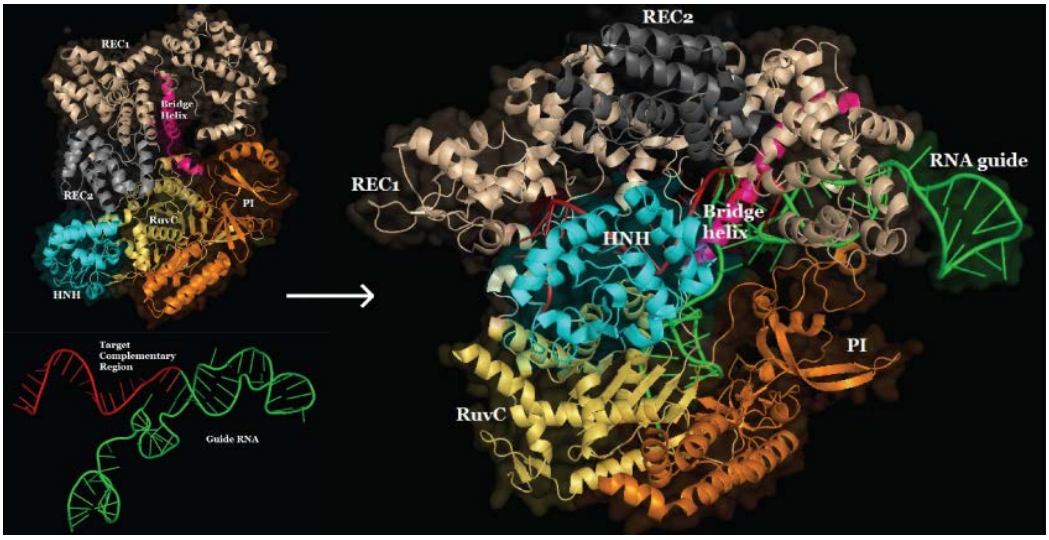


Figure 8. *After the binding of the guide RNA, Cas9 protein undergoes a conformational change that converts the inactive protein into its active form. In the picture above are illustrated the domains that compose the Cas9 protein as well as the T-shape of gRNA.*

After its activation, Cas9 scans for target DNA by binding with sequences that match its protospacer adjacent motif (**PAM**) sequence (*DNA interrogation by the CRISPR RNA-guided endonuclease Cas9, Sternberg SH, 2014*). A PAM sequence is parted by two or three bases and located within one nucleotide downstream of the region complementary to the guide RNA. PAMs have been identified in all CRISPR systems, and the specific nucleotides that define PAMs are indicative of the particular category of CRISPR system (*Short motif sequences determine the targets of the prokaryotic CRISPR defense system. Mojica FJ, 2009*). For instance, PAM in *Streptococcus pyogenes* is 5'-NGG-3'. *(A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Jinek M, 2012*). As soon as the Cas9 protein locates a potential target with the specific PAM, the protein will pair the bases upstream the PAM with the complementary region on the guide RNA. (Figure9) (*DNA interrogation by the CRISPR RNA-guided endonuclease Cas9, Sternberg SH, 2014*). If the complementary and the target region pair properly, the RuvC and HNH nuclease domains will cut the target DNA after the third nucleotide base upstream of the PAM. ***(****Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease, Carolin Anders, 2014)*

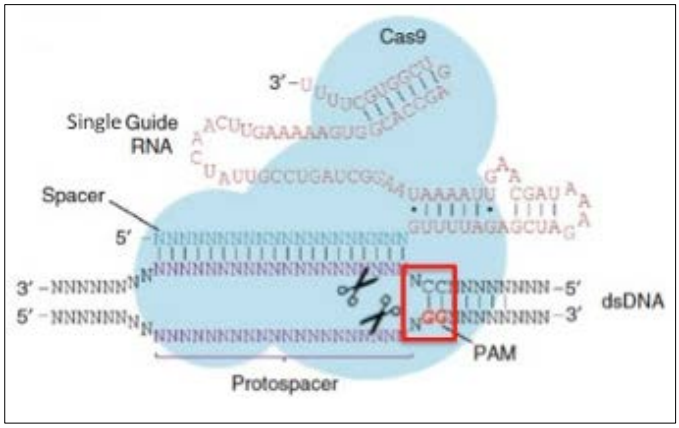


Figure 9. *As soon as the Cas9 protein locates a potential target with the specific PAM, the protein will pair the bases upstream the PAM with the complementary region on the guide RNA.*

1.9 The Off-target Problem and Cas9 variants engineering

It's been widely reported that CRISPR/Cas9 as every other system of genome editing it's not foolproof. There are additional sites throughout the genome where partial homology exists with the gRNA that are mistakenly edited. These sites are called off-targets. In their trial to eliminate the "**off-target effects**" scientists have adopted solutions such as the use of **truncating gRNA sequences** for targeting that are less tolerant to mismatches, or the **engineering of new Cas9 variants (Figure 10).**

Mutation of both, the HNH and RuvC catalytic domains of the Cas9 results to a DNA binding protein, called "**deadCas9 " (dCas9**) (Figure 10), which can be deployed as a platform to recruit various effector proteins to sites of interest *(CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes., Gilbert LA, 2013*) CRISPR - mediated modular RNA -guided regulation of transcription in eukaryotes. (RNA -guided gene activation by CRISPR -Cas9 -based transcription factors. Perez -Pinera, P. et al. 2013).

Another way researchers have attempted to minimize off-target effects is with the use of "**paired nickases**". (Figure 10). The wild type Cas protein needs two conserved nuclease domains, HNH and RuvC, to cleave DNA by nicking the sgRNA-complementary and non-complementary strands, respectively. A "nickase" mutant (**Cas9n)** results from alanine substitution at key catalytic residues of these domains. Specifically, **D10A** inactivates RuvC *(A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Jinek M, 2012)*, while **N863A** has been found to inactivate HNH *(Crystal structure of Cas9 in complex with guide RNA and target DNA. Nishimasu H, 2014).*

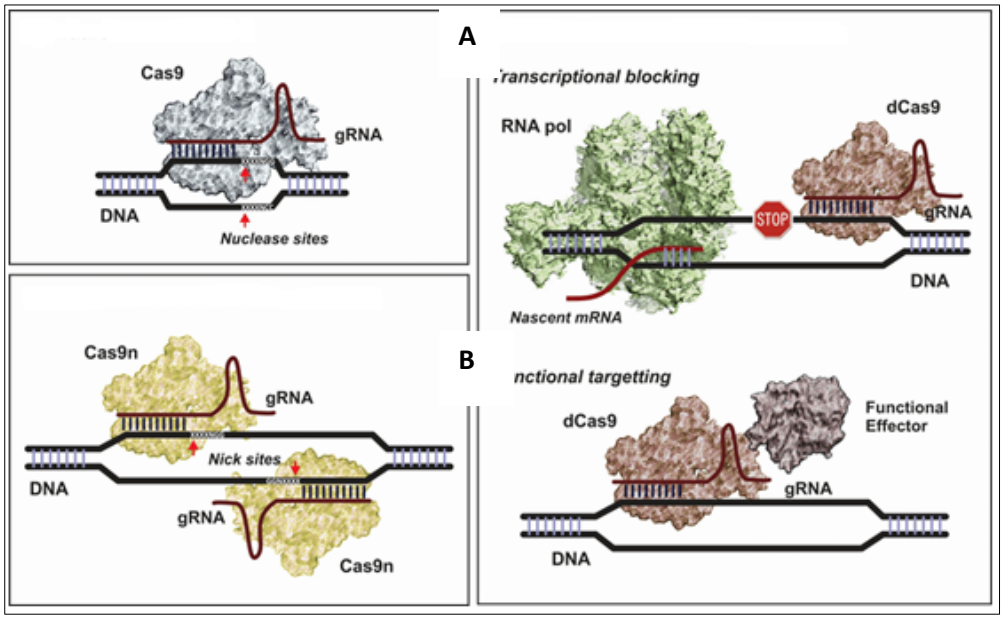


Figure 10. *Here is depicted a comparison between A) the CRISPR-Cas system, B) the CRISPR with nickase Cas9n mutant and C) the CRISPR with inactive dCas9 mutant. By altering the Cas protein the specificity of CRISPR system can be sensibly improved.*

Two nicking enzymes, appropriately spaced and oriented at the same locus, effectively generate DSBs as in the wild-type case *(RNA-guided human genome engineering via Cas9. Mali P, 2013, Genome engineering using the CRISPR-Cas9 system. Ran FA, 2013*). This can be achieved by Cas9n and two sgRNAs complementary to the adjacent area on opposite strands of the target site. While this induces DSBs in the target DNA, it is expected to create only single nicks in off-target locations and, therefore, result in minimal off-target mutations.

About two years ago, another type of Cas protein, **Cas12** formally known as **Cpf1**, was reported (Table 1). The interesting part in Cas12 is that requires a **small crRNA** while it's specificity is determined by a **longer spacer** (at least 22 nt for maximum efficiency) (Zetsche, B. et al., Cpf1 is a single RNA -guided endonuclease of a class 2 CRISPR -Cas system., 2015) . Cas12a **recognizes T -rich PAMs** instead of G –rich ability that increases the number of potential target sites. A recent structural study discovered a second nuclease domain, the **Nuc** domain, which together with the RuvC domain is responsible for target site cleavage (Yamano, T. et al., Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA., 2016). Another useful characteristic of this individual is that compared to Cas9 which generates blunt ends after cleavage, Cpf1 leaves sticky 5' overhangs that may be used for directional cloning. This property of Cpf1 has been exploited to perform highly specific DNA assembly in vitro. (Genome Engineering Using the CRISPR-Cas9 System. Ran, FA. et al. 2013).

A very recent and highly useful addition to the CRISPR toolbox is **Cas13**, also known as C2c2. This type acts exclusively on RNA via its two HEPN (higher eukaryotes and prokaryotes nucleotide -binding) domains *(Abudayyeh, O.O. et al., C2c2 is a single -component programmable RNA -guided **RNA - targeting** CRISPR effector, 2016*). Mutation in RuvC or in HEPN domains can convert both Cas12 and Cas13 into Cas dead nucleases. In this way, a programmable DNA or RNA binding protein is obtained.

Originally, the CRISPR/Cas9 system has proven to be a simple tool for site -specific mutagenesis. Also, Cas9 orthologues from Streptococcus pyogenes (**SpCas9**) and Staphylococcus aureus (**SaCas9**) (Table 1) provide interestingly high mutagenesis frequencies, thus providing an alternative and more efficient choice for genome -editing applications. (*Transforming plant biology and breeding with CRISPR/Cas9, Cas12 and Cas13, Schindele P et al, 2018)*

While spCas9 is the most commonly used CRISPR nuclease, saCas9 has recently drawn much attention due to its size which is approximately 1 kb smaller than spCas9, allowing more effectively packaging. Also, saCas9 opens up new possibilities for genome editing due to its different targeting capabilities compared to spCas9. spCas9 recognizes a PAM sequence of 5'-NGG-3', while saCas9 recognizes 5'-NNGRRT-3'. A greater variety in PAM sequences available for use means an increased number of loci are available for genome editing.
(https://www.abmgood.com/marketing/knowledge_base/CRISPR_Cas9_Introduction_Part7.php)

	spCas9	saCas9	Cpf1
Gene Length	~4.1 kb	~3.3 kb	~3.8 kb
Cleavage Type	Blunt ends	Blunt ends	5' overhangs
Repeated Cleavage	No	No	Yes
Cleavage Site	Within recognition sequence	Within recognition sequence	Downstream of recognition sequence
PAM Sequence	5'-NGG-3'	5'-NNGRRT-3'	5'-TTN-3' or 5'-TTTN-3'
RNA Required	tracrRNA + crRNA	tracrRNA + crRNA	crRNA

Table 1: *Comparison of spCas9, saCas9, and Cpf1 nucleases*

An engineered high -fidelity SpCas9 variant (SpCas9-HF1) has demonstrated strongly reduced off-target activity, while on target activity remained comparable to wild type *(Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases, Tsai SQ, 2016).*

1.10 High fidelity Cas9 nucleases

The effectiveness of Cas9 is based on two events: The first is about interactions between the non-target strand and a positively charged groove formed by HNH and RuvC domains. The second one is about interactions between the target strand and the gRNA. Scientists reasoned that, if they somehow **decreased the positive charge** in the **HNH/RuvC groove** (Figure 11a), they could weaken interactions between the groove and the negatively charged DNA. This would decrease strand separation and combined with the weak separation afforded by gRNA binding at non-target sites would theoretically decrease off-target cutting. In order to decrease its electropositivity, Slaymaker et al. made a set of **alanine substitutions** using rational mutagenesis to develop two high fidelity Cas9 variants: The enhanced specificity **eSpCas9 and** High fidelity **SpCas9-HF1** variants (Figure 11b)**.** These variants have proven to exhibit improved on-target specificity. The most recent advance over this field is the **hyper accurate** Cas9 variant **(HypaCas9),** which was developed by researchers in the Doudna Lab at Berkeley and displays comparable on-target activity and decreased off-target activity compared to eSpCas9 and SpCas9-HF1, probably due to alanine substitutions (SpCas9 N692A/M694A/Q695A/H698A) achieved after targeted mutagenesis. All those variants are tools that promise precision in mammalian cells editing and simultaneously elimination of off target effects in applied or therapeutic settings. *(http://blog.addgene.org/enhancing-crispr-targeting-specificity-with-espcas9-and-spcas9-hf1).*

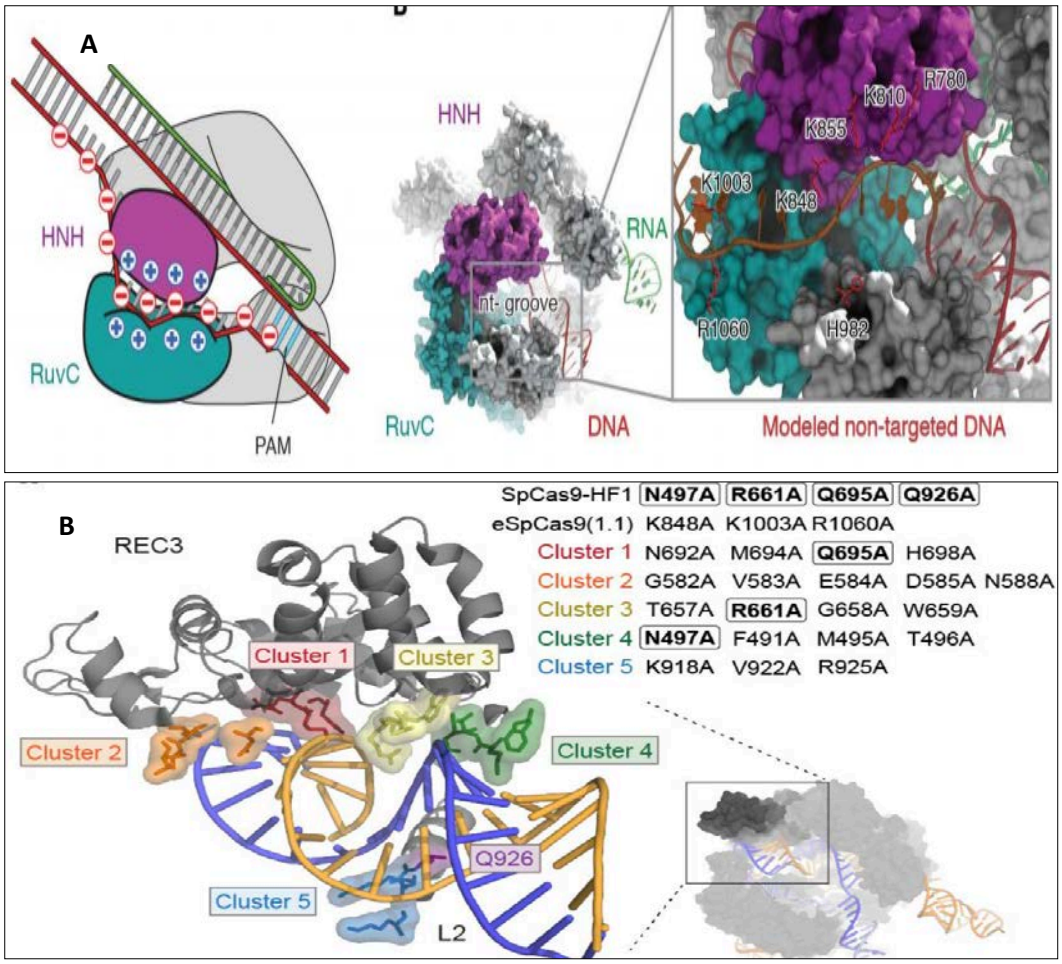


Figure 11. *A) Neutralization of positively-charged residues within the non-target strand groove (nt-groove) existing between HNH and RuvC, could weaken non-target strand binding and encourage re-hybridization between the target and non-target DNA strands, by requiring more stringent Watson-Crick base pairing between the RNA guide and the target DNA strand (Rationally engineered Cas9 nucleases with improved specificity, Ian M. Slaymaker, 2015). B) Alanine substitutions researchers did using rational mutagenesis to develop: the enhanced specificity eSpCas9 and High fidelity SpCas9-HF1 variants (Enhanced proofreading governs CRISPR–Cas9 targeting accuracy, Jennifer A. Doudna et al, 2017)*

1.11 The high fidelity variants of Cas9 mechanism of function

It's generally accepted as fact that **HNH** nuclease domain of Cas9 undergoes conformational rearrangement upon target binding, which activates the **RuvC** nuclease for cleavage. In case of binding to mismatched targets, HNH domain it is trapped to an inactive conformational shape. What happens in the **high fidelity variants** of Cas9 is that they reduce off-target cleavage by **raising the threshold** for **HNH activation**. Since HNH domain does not directly contact the DNA, another Cas9 domain govern it's mobility after it tracks complementarity. It is about **REC3**, a non-catalytic domain, which acts as an **allosteric effector** that recognizes the RNA-DNA heterodublex and cooperates with **REC2 to allow HNH domain activation (**Figure 12).

Exploiting this observation, scientists designed high fidelity variants that include mutation of residues within REC3 involved in RNA-DNA recognition (Figure 11). Those mutations prevent transitions by REC2 domain, which more stringently make HNH domain to stall in an inactive conformational stage in presence of mismatches. Consequently, the variants mentioned above with emphasis in HypaCas9 demonstrate **high genome-wide specificity** without compromising on-target activity in human cells. *(Enhanced proofreading governs CRISPR–Cas9 targeting accuracy, Jennifer A. Doudna et al, 2017)*

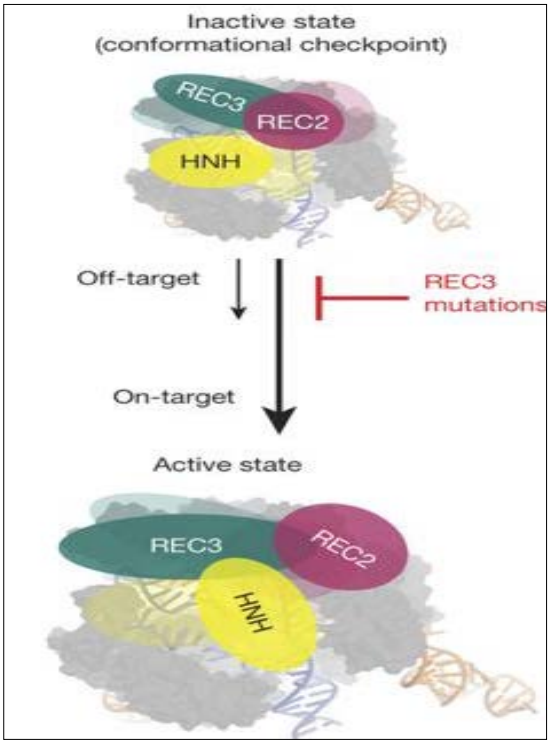


Figure 12. *REC3 recognizes the RNA-DNA heterodublex and cooperates with REC2 to allow HNH domain activation. Mutations in REC3 subunit raise the threshold for HNH activation.*

1.12 CRISPR-Cas9 system: delivery strategies

There are **three** different **strategies** to **deliver** successfully CRISPR-Cas9 system into cells and achieve precise gene editing. **Cas9** can be transferred as **DNA, RNA or protein** form. (Figure 13) The use of a **single plasmid** encoding for the cas9 gene and the sgRNA may offer the most **stable** expression of the Cas9 protein, however the gene undergoes transcription and translation. This leads to a **delay** in the onset of gene **editing**. (F.A.Ran, et al., Genome engineering usingthe CRISPR-Cas9 system, Nat. Protoc. 8 (2013)**.** In mammalian cells, DNA generally must be delivered directly into the nucleus as well. *(Engineering the Delivery System for CRISPR-Based Genome Editing Zachary Glass)*. The second strategy is to deliver the mixture of the **Cas9 mRNA and sgRNA.** The mRNA will be translated to protein in cells and the Cas9/sgRNA complex will be formed. The advantages of administering mRNAs is that the duration of gene-editing and the off-

targets are limited in competence to plasmid-based CRISPR-Cas9 systems. Apart from subsiding off-target effects, the use of the mRNA results in low cytotoxicity in primary cells and cell lines. The drawback of the above strategy lies to the poor stability of mRNA. (Generation of gene-modified cynomolgus monkey via Cas9/RNA mediated gene targeting in one-cell embryos, Y. Niu, et al., 2014) since mRNAs are susceptible to enzymatic degradation. Another drawback is that therapeutic effect is delayed due to the process of translation. mRNA, also needs to be delivered into the nucleus. (*Engineering the Delivery System for CRISPR-Based Genome Editing Zachary Glass, 2018).*

The newest and most promising is the third strategy which includes the direct delivery of Cas9/sgRNA complex into cells. The only demand is a purified Cas9 protein, positively charged that can efficiently form a complex with sgRNA. This is called **Cas9/sgRNA ribonucleoprotein complex (RNP**). (J.A. Zuris, et al., Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo, 2015) Direct delivery of RNPs has numerous advantages, including rapid action, high gene editing efficiency, no requirement of codon optimization and promoter selection and reduced off-target effects, toxicity and immune responses. (S. Kim, D. Kim, S.W. Cho, J. Kim, J.S. Kim, Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins, 2014), (Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes., Staahl BT, 2017). This strategy suffers from the additional challenge of delivery across two membranes the cellular and nuclear. (*Engineering the Delivery System for CRISPR-Based Genome Editing Zachary Glass, 2018)*

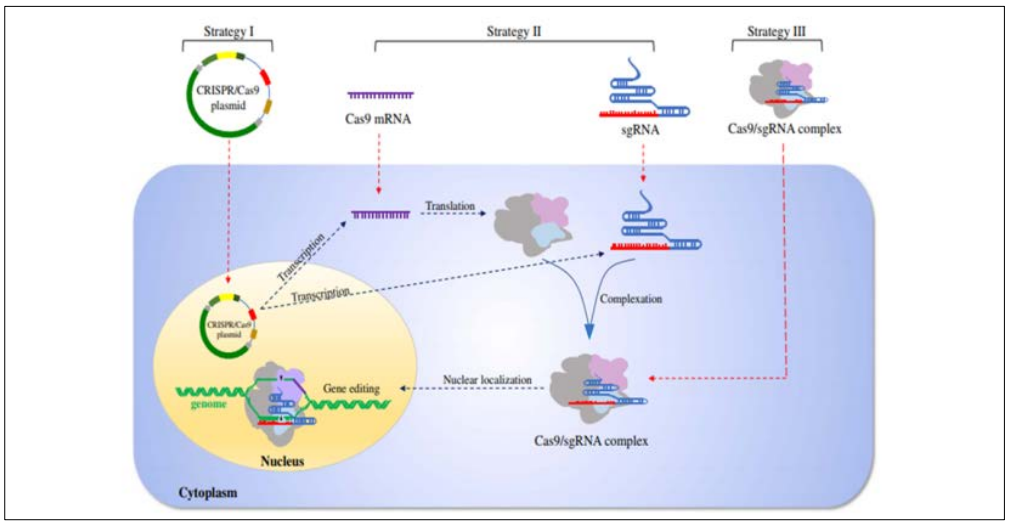


Figure 13. *There are three different strategies to deliver successfully CRISPR-Cas9 system into cells and achieve precise gene editing. Cas9 can be transferred as DNA, RNA or protein form*, *(Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications, Chang Liu et.al, 2017)*

1.13 CRISPR-Cas9 system: Various delivery methods

As it was already discussed Cas9 can be delivered as DNA, RNA, or as a functional RNP. (Figure 13) Regardless of cargo format, one more challenge remains: the proper

way of **delivering the cargo across the cell and nucleus membrane** avoiding degradation by the host's protective mechanisms. There are viral ways available as Adeno-associated virus (**AAV**) or Intergrase deficient Lentivirus vectors (**IDLVs**) and non-viral such as gold nanoparticle (**AuNP**), cell-penetrating peptide (**CPP**), lipid nanoparticle (**LNP**) and nuclear localization sequence (**NLS**). (Figure 14)

To start with, the use of **viral** vectors is widespread in the laboratory environment, but is not appropriate for clinical settings. After all there is a limit in the maximum capacity of the cargo a virus can transfer. The **non-viral** ways are divided into **physical** methods that disrupt cellular barriers and **chemical** modifications that improve the cargo's delivery. **Electroporation** is a technique that physically exerts a strong electric field onto a cell's membrane making it temporarily permeable in macromolecules as DNA, RNA, and even RNPs. **Hydrodynamic injection** can deliver large macromolecules by injecting a liquid solution intravenously at extremely high volume and pressure Although, those approaches are both not suitable for therapeutic use in patients. Microinjection, on the other hand, has been successfully applied on human embryos. This technique demands micron-scale needles to simply pierce the cell membrane and directly inject the cargo. The need for manual injection is the downsize of this method.

Chemical delivery approaches utilize the formulation of complementary molecules to assist Cas9 to bypass cellular barriers as well as protect it from degradation pathways.

Lipids are naturally occurring amphiphilic organic molecules. The manufacturing of synthetic **cationic lipids**, is an effective delivery vehicle especially in the case of Cas9 RNPs. Synthetic lipids also contain a bioreducible disulfide bond within the hydrophobic tail chain, which can be cleaved intracellularly to facilitate payload delivery and reduce lipid toxicity.

Encapsulating the Cas9 protein, DNA, or RNA within polymeric carriers is another popular approach due to the wide array of polymers available for use in biological transfection. Last but not least, proteins and nucleic acids can be chemically modified to achieve improved delivery. One common approach for this is the use of cell-penetrating peptides (**CPPs**) that can cross the cell membrane easily, or nuclear localization sequences (**NLSs**) whose natural work is to tag proteins that are synthesized in the cytoplasm for transport into the nucleus.

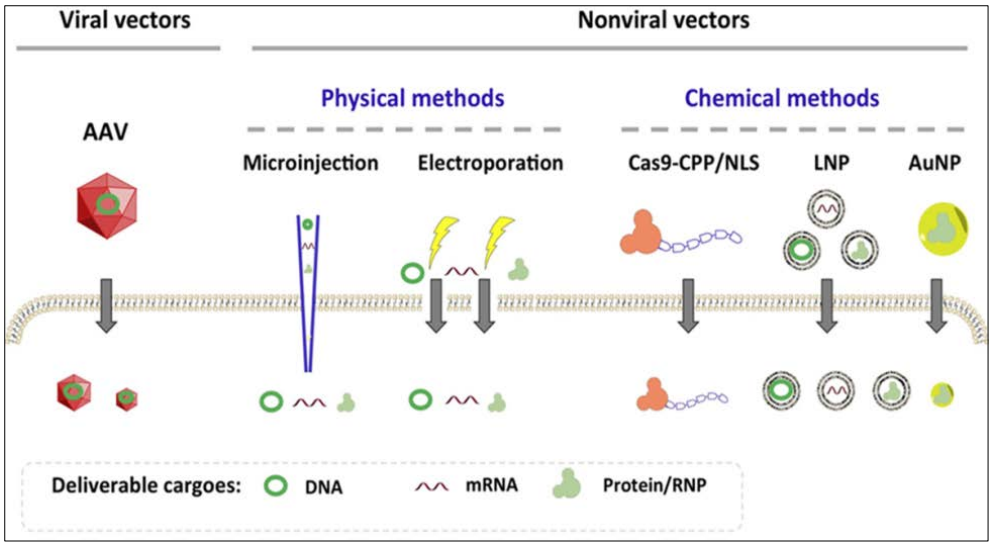


Figure 14. *There are viral ways available for delivering as Adeno-associated virus (**AAV**) or Intergrase deficient Lentivirus vectors (**IDLVs**) and non-viral such as gold nanoparticle (**AuNP**), cell-penetrating peptide (**CPP**), lipid nanoparticle (**LNP**) and nuclear localization sequence (**NLS**). (Engineering the Delivery System for CRISPR-Based Genome Editing Zachary Glass, 2018)*

Proper selection of a delivery system and the type of CRISPR-Cas9 is significant for reducing the off-target effects especially in case of clinical applications. In general, non-viral delivery system has lower off-target effects than that of viral delivery system. Various strategies have been developed to modify Cas9 protein so that it can form a stable COMPLEX with sgRNA and non-viral vectors. Pioneer studies demonstrated that chemical modification of Cas9 protein does not attenuate its nuclease activity. *(Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications, Chang Liu et.al, 2017), (Engineering the Delivery System for CRISPR-Based Genome Editing Zachary Glass, 2018)*

1.14 CRISPRCas9 system Vs other programmable nucleases

There are specific prerequisites on choosing the programmable nuclease needed for the conduction of an experiment. According to the purpose of research or the organism involved scientists have to take into consideration a variety of aspects such as success rates, specificity, off-target effects, cytotoxicity, delivery ways and of course the unique characteristics of each nuclease. In table 2 a comparison between the available nucleases is summarized. The first conclusion that can be derived is that the most challenging nuclease to program is ZFN mainly due to the poor targeting density and their often failure to combine high activity and low cytotoxicity. TALENs seem to be more enlightening in genome editing field as they are able to target almost any given DNA sequence while they combine low off-targeting and high success rate. Although, their weakness to cleave methylated DNA make them seem inferior to RGENs that tend to be the most promising way of genome editing as they are easy to handle and able to overcome all their flaws.

	ZFNs	TALENs	RGENs
DNA targeting specificity determinant	Zinc-finger proteins	Transcription activator-like Effectors	crRNA or sgRNA
Nuclease	FokI	FokI	Cas9
Success rate	Low	High	High
Average mutation rate	Low or variable	High	High
Specificity-determining length of target site	18–36bp	30–40bp	22bp (total length 23bp)
Restriction in target site	G-rich	Start with T and end with A (owing to the heterodimer structure)	End with an NGG or NAG (lower activity) sequence (that is, PAM)
Design density	One per ~100bp	At least one per base pair	One per 8bp (NGG PAM) or 4bp (NGG and NAG PAM)
Off-target effects	High	Low	Variable
Cytotoxicity	Variable to high	Low	Low
Size	~1kb×2	~3kb×2	4.2kb (Cas9 from Streptococcus pyogenes)+0.1kb (sgRNA)

Table 2. *A comparison between the available nucleases is summarized in the table above. Although they all result in DNA cleavage each one has special features. (A guide to genome engineering with programmable nucleases, Hyongbum Kim, Jin-Soo Kim, 2014)*

1.15 Advantages of CRISPRCas9

One of the most thrilling advantage of CRISPRCas9 system is that of the **target design simplicity**. In any other editing system specificity relies on protein-DNA recognition. On the contrary, this system demands only ribonucleotide complex formation, fact that eliminates the **need for protein engineering** while **gRNAs can be designed easily**. CRISPR-Cas9 can be easily adapted to target any genomic sequence by just changing the 20-bp protospacer sequence of the guide RNA, which can be accomplished by **cloning** this particular nucleotide sequence in a vector encoding either for crRNA or sgRNA. Lately an even easier way for new RGENs manufacturing came to surface that do not require cloning. Specifically, crRNA or sgRNA can be prepared by simply **annealing** two complementary oligonucleotides followed by

in **vitro transcription**. (*A guide to genome engineering with programmable nucleases. Hyongbum Kim, Jin-Soo Kim, 2014)*

Moreover, **ease of use** for CRISPR-Cas9 is a significant advantage over ZFNs and TALENs, especially in generating a large set of vectors to target numerous sites or even genome-wide libraries.

Another crucial advantage of CRISPR-Cas9 is the ability to **multiplex** that is to use multiple guide RNAs in parallel to target multiple sites simultaneously in the same cell by retaining the same Cas9 protein. To achieve the same outcome ZFN or TALEN should be in pairs while RGENs function as monomers.

Site selection which is another important issue **is limited** to 23-bp sequences on either strand that end in an NGG motif (the PAM for S. pyogenes Cas9), which occurs on average once every 8 bps. The limitation of the necessary existence of a guanine at the 5' end is circumvented by making guide RNAs with one or two additional G bases at their 5' ends.

Though RGENs present **better output** over ZFNs, TALENs in features as success rate, specificity, and cytotoxicity, there are fields that need **further optimization** for the refinement of the system.

The first that needs optimization is the quest for the proper **delivery system**. The **size** of the Cas9 protein it's been always a limp point of the CRISPR system. The cDNA encoding S. pyogenes Cas9 is approximately 4.2 kb in size, making it somewhat larger than a TALEN monomer and much larger than a ZFN monomer

The eternal speculation over every programmable nuclease is that of the **off target effects and specificity.** Undoupdetly, RGENs **are tolerant to single/multiple mismatches.** To overcome this, scientists, as mentioned above, proceeded to the manufacturing of Cas9 variants (high fidelity nucleases, nickases, Cas12, Cas13 variants) that increase the specificity and simultaneously decline the off-target effect. Another solution is the use of gRNAs reported with no off-target effects or reducing the length of the protospacer portion of the gRNA, which makes it less tolerant of mismatches.

1.16 CRISPR-Cas9 system and applications

CRISPR-Cas9 system has a wide variety of applications and no one can deny the contribution of CRISPR in **life sciences, therapeutics, biotechnology and molecular biology** (Table 3). As far as life sciences are concerned CRISPRCas9 system has contributed to: the understanding of gene function and physiological roles of genetic variations (SNPs), as well as to rewire biological pathways/mechanisms. Since the first time that came into light, CRISPR/Cas9 system has demonstrated strong therapeutic potential. There are several reports for exploitation of CRISPR in potential treatment for genetic diseases and cancer. CRISPR–Cas was indicated as a promising tool for drug discovery or development. A lot of recent researches have used CRISPR system in cellular models, screening, animal models, drug delivery, gene

drives for control. The application of CRISPRCas9 system seems to be substantial in agriculture, crop engineering, dairy industry, and biofuels.

Finally, yet importantly, the CRISPR technology gave rise to a lot of molecular biology tools for transcriptional repression or activation, epigenetic modifications, multiplex genome engineering, screening and visualizing genomic loci. More specifically, CRISPR was exploited for making precise modifications Using Homology Directed Repair (HDR), point mutations by engineering CRISPR base editors that fuse Cas9 nickase or dCas9 to a cytidine deaminase (base editors are targeted to a specific locus by a gRNA, and they can convert cytidine to uridine), or by developing RNA base editor that converts adenosine to inosine in RNA fusing an adenosine deaminase to catalytically dead Cas13b. Also, Cas enzymes fused to epigenetic modifiers end up in heritable epigenetic modifications without inducing double-strand breaks.

Another CRISPR ability is multiplex genome engineering (through expressing several gRNAs from the same plasmid) that demands use of dual nickases to generate a knockout or gene edit, Cas9 to generate large genomic deletions by removing the sequence between two gRNA target sites, modifying multiple genes at once. CRISPR, due to easy design and synthesis of gRNA as well as the ability to target almost any genomic locus CRISPR, has already been used extensively to screen for novel genes that regulate known phenotypes, including resistance to chemotherapy drugs, resistance to toxins, cell viability, and tumor metastasis. (The method involves the use of pooled lentiviral CRISPR libraries).

As mentioned above the deprivation of catalytic ability via mutation in HNH, RuvC, results in a Cas9 protein (dCas9) with binding abilities. dCas9 can be tagged with transcriptional repressors or activators, and targeting these dCas9 fusion proteins to the promoter region results in robust transcriptional repression or activation. dCas9 can also be fused to a fluorescent marker like GFP, turning up into a customizable DNA labeler compatible with fluorescence microscopy in living cells. Alternatively, gRNAs can be fused to protein-interacting RNA aptamers, which recruit specific RNA-binding proteins (RBPs) tagged with fluorescent proteins to visualize targeted genomic loci. Identifying molecules associated with a genomic region of interest in vivo is essential to understanding locus function.

Using CRISPR, researchers have expanded chromatin immunoprecipitation (ChIP) to allow purification of any genomic sequence specified by a particular gRNA. dCas9 is used to purify genomic DNA bound by the gRNA. An epitope tag(s) can be fused to dCas9 or gRNA for efficient purification. To conclude, there is no doubt that CRISPR is a largely evolving system that allows working on multiple types of cells that are more related to the biological questions asked.

Life sciences	Therapeutics biotechnology	Biotechnology	Molecular biology https://www.addgene.org/crispr/guide/#base-edit
Understanding of gene function	**Genetic diseases**: Sickle cell anemia (*Xie et. al., 2014*, Duchenne muscular dystrophy (DMD) *(Li et al., 2014, Long et al., 2014*), Cystic fibrosis (*Schwank et al., 2013*), Cardiovascular disease *(Carroll, 2016)*	**Crop engineering** (*Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes, Zhen Liang, 2017)*	**Precise Modifications (HDR), point mutations (CRISPR base editors or RNA base editors), heritable epigenetic modifications**
Understanding physiological roles of genetic variations (SNPs)	**Cancer types:** Human papillomavirus (HPV) and cervical cancer (Hu et al., 2014, Zheng Hu, et al, 2014), Osteosarcoma (Feng et al., 2014)	**Dairy industry** (*Song X, et al, 2017)*	**Genome-Wide Screens** (screen for novel genes that regulate known phenotypes, including resistance to chemotherapy drugs, resistance to toxins, cell viability, and tumor metastasis.
Rewire biological pathways/mechanisms	**Drug discovery or development**: generation of cellular models, large-scale functional screening, rapid generation of animal models, for studying human diseases, drug delivery (*Christof Fellmann et. Al.,2017*)	**Engineering industrial microorganisms for biofuels** *(Becker J, et al., 2015), (Lee SK1, 2008*	**Multiplex Genome Engineering** (**dual nickases, large genomic deletions by removing the sequence between two gRNA target sites, modification of multiple genes at once**
	Synthetic gene drives for control (*Douglas W. Drury, 2017)*		**Transcriptional repression or activation**
			Visualization of Genomic Loci (using the fluorescent CRISPR system)
			Purification of Genomic Regions (Using dCas9 or guide RNA)

Table 3. *Contribution of CRISPR-Cas9 system in life sciences, therapeutics, biotechnology and molecular biology.*

2. SCIENTIFIC SIGNIFICANCE – AIMS

The aim of the current thesis is the large scale **production and purification** of a Cas9 variant, **HypaCas9** which as it was previously mentioned, is a high fidelity nuclease. The ultimate purpose of Cas9 production is to be utilized for in house *in vivo* experiments (IMBB-FORTH). In order to accomplish that, it is necessary to have a **pure** and **active** form of the target protein. The initial steps that will be followed are, a **small** scale expression of the protein, and protein purification through a variety of chromatography tests. After the standardization of the most appropriate **purification** scheme, it is necessary to proceed with the **quality control** assays that will verify not only the **quality** and **purity** of the protein but also whether the protein is **active** and ready to use for *in vivo* experiments. Since we achieve these goals a **large scale** production, via fermentation, will follow.

The **delivery** strategy of the Cas9 system that will be used is the most recently advanced that includes transport of the protein in a complex with the guide RNA (**Ribonucleoprotein complex**). This delivery strategy also demands an absolutely clear protein from DNAses that will destabilize or mislead the RNP complex or RNAses that will fragment the guide RNA deactivating the Cas9 functionality.

3. MATERIALS AND METHODS

3.1 Over expression of SpCas9 Cluster 1 (HypaCas9) variant in *E. coli* cells.

BL21 Star (DE3) pLysS *E.coli* **competent** cells were prepared according to the following protocol:

Materials: BL21 Star (DE3) pLysS cells, LB medium (10 g/liter tryptone, 5 g/liter yeast extract, 5 g/liter NaCl), LB agar (LB medium containing 15 g/liter agar), TFB1 (100 mM KCl, 10 mM CaCl2, 50 mM MnCl2, 30 mM CH3COOK, 15% glycerol, sterile-filter), TFB2 (10 mM KCl, 10 mM MOPS, 75 mM CaCl2, 15% glycerol, sterile-filter), Chloramphenicol stock solution (34μg/ml)

A trace of BL21 Star (DE3) pLysS cells was removed from the vial with an inoculating loop, and streaked out on LB agar containing 34 μg/ml Chloramphenicol. Incubation followed at 37°C overnight. A single colony was picked and inoculated on 10 ml of LB-Chloramphenicol (34 μg/ml). Incubation followed overnight at 37°C. 1 ml overnight culture was added to 100 ml prewarmed LB medium containing 34 μg/ml chloramphenicol in a 250 ml flask, and shaked at 37°C until an OD600 of 0.5 reached. After 5 min cooling on ice the culture was transferred to a sterile, round-bottom centrifuge tube. Cells were collected by centrifugation at low speed (5 min, 4000 x g, 4°C). Supernatant was discarded while cells were kept always on ice. Resuspending of cells followed gently in cold (4°C) TFB1 buffer (20 ml for a 100 ml culture) for approximately 20 minutes. Cells were collected by centrifugation (5 min, 4000 x g, 4°C). Supernatant was discarded carefully. Resuspending of the cells followed in 4 ml ice-cold TFB2 buffer. Aliquots of 60μl were prepared in sterile microcentrifuge tubes and froze in a dry-ice–ethanol mix. Competent cells were stored at –80°C.

BL21 Star (DE3) pLysS *E.coli* **competent** cells were **transformed** with the appropriate **pJSCO11** plasmid construct, provided by **Addgene (Figure 15)**, (*Enhanced proofreading governs CRISPR–Cas9 targeting accuracy, Janice S. Chen, 2017)* carrying the following **mutations**: C80S, C574S, S355C, S867C, N692A, M694A, Q695A and H698A. The BL21 Star (DE3) pLysS efficiency was $9{,}12e^{7}$ transformants /μg DNA. According to the transformation protocol an aliquot of the competent cells was transferred (60 μl or less) on ice for 5 -10 min. About 100ng of plasmid DNA was transferred to 60 μl of competent cells and incubated on ice for 20 -30 min. Heat shock was followed to a 42°C water bath or heating block for 90 sec. 900 μl SOC medium was added to the cells and incubated for 45-60 min at 37°C. Shaking increases transformation efficiency. The mixture was centrifuged at 3,500 rpm for 5 min and 900 μl from supernatant were discarded. Cell pellet got resuspended in the remaining 100μl. Plating with the appropriate antibiotics (100 μg/ml **Ampicillin** and 34 μg/ml

Chloramphenicol) was performed on LB-agar plates. The plates were incubated at 37°C overnight.

After the heat- shock transformation protocol was performed a fresh colony was chosen for the **inoculation** of an LB **pre-culture**, in the presence of the above-mentioned concentrations of the two antibiotics and incubated overnight (16 hours) at 37 °C under vigorous shaking. The pre-culture was used to inoculate 1 liter of LB in a ration of 1:50, in the presence of Ampicillin and Chloramphenicol. The inoculated **cultures** were grown at 30 °C until the OD_{600} reached 0.5-0.6 and, subsequently **induced** with 0.5 mM **IPTG** (IsoPropyl- beta-D-thioGalactopyranoside). Cultivation continued for another 16 hours at 16 ° C under shaking. The cells were collected by **centrifugation** of the cultures at 5000 g, 4°C for 15 minutes and stored at -20 °C until use. The same conditions were followed for large scale expression in 30 liter fermentor (Bioengineering AG). A 30 liter fermentation culture gave 97 gr cell paste.

The same conditions were followed for the transformation and expression of a T7 inducible plasmid (Koukaki Marina, IMBB-FORTH) that express a His-Tev protease in the same competent cells. Large scale expression in 30 liter fermentor resulted to 40gr cell paste.

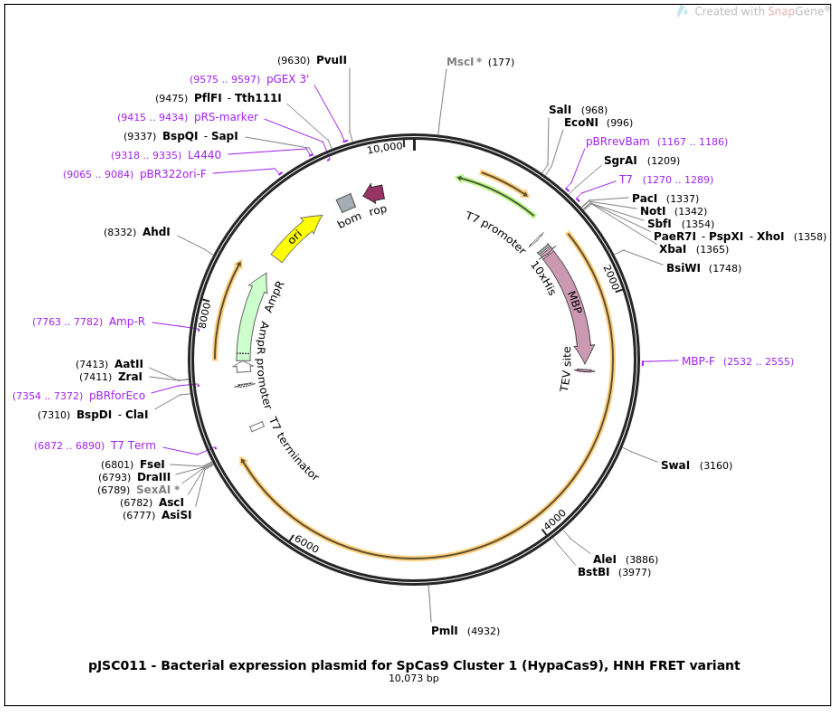


Figure 15. *pJSC11-Bacterial expression plasmid for SpCas9 cluster 1 (HypaCas9) provided by Addgene.* https://www.addgene.org/101229/

3.2 Cell disruption

Frozen **cell paste** expressing the Cas9 variant or the TEV protease was **thawed** on ice and resuspended in lysis buffer in a ration 1:3 at 4° C. The cells were disrupted by **sonication** (12x30 seconds sonication steps, with resting intervals on a dry ice-ethanol bath) and the **lysate** was centrifuged at 15000 g, 4 °C for 20 minutes. Alternatively, when the cell paste was more than 10 gr, **French press** was used using the lysis buffer in a ration 1:1. The collected lysate was also centrifuged in same conditions. Subsequently, the cytoplasmic supernatant was collected and diluted in a ration 1: 3 with calibration buffer.

Lysis buffer for **TEV** protease: 20 mM Tris-HCl pH: 8, 200 mM NaCl, 1 mM DTT (Dithiothreitol) and 1 mM PMSF (PhentlMethylSulfonylFluoride)

Lysis buffer for **HypaCas9**: 50 mM Tris-HCl pH: 8, 500 mM NaCl, 5 mM imidazole, 10% glycerol and 1 mM PMSF

Calibration buffer/ TEV protease: 50 mM Tris-HCl pH: 8, 1 M NaCl, 5 mM imidazole, 20% glycerol

Calibration buffer/HypaCas: 50 mM Tris-HCl pH: 8, 1 M NaCl, 5 mM imidazole, 10% glycerol,

3.3 Chromatography with Ni-NTA agarose matrix

The purification schemes followed for TEV protease and HypaCas9 were based on the QIAexpress system (Figure 16):

3.3.1. Scheme 1 His TEV

In case of **His-TEV protease**, a chromatography column containing Ni-NTA Agarose beads (GE Healthcare) was packed and equilibrated with buffer A (50 mM Tris-HCl pH: 8, 1 M NaCl, 20% glycerol, 5mM Imizazole). For 5gr TEV protease, 1 ml (1ml=1 column volume, CV) beads were calibrated. The calibration volume was 5 CV. Next, the 3 times diluted supernatant was loaded in the column using a peristaltic pump (flow rate: 1,5 ml/min). After sample loading, the column was washed using buffer A (flow rate: 2 ml/min) (5CV), buffer B (50 mM Tris-HCl pH: 8, 50 mM NaCl, 20% glycerol, 5 mM Imizazole) (5CV) and buffer C (50 mM Tris-HCl pH: 8, 50 mM NaCl, 20% glycerol, 30 mM Imizazole) (10CV). The bonded proteins were eluted in presence of Buffer D ((50 mM Tris-HCl pH: 8, 50 mM NaCl, 20% glycerol, 300 mM Imizazole), and collected in 2 fractions of 10 ml each (flow rate: 3 ml/min). Finally, the fractions were analyzed by SDS-PAGE in order to assess their relative concentration in His-TEV protease.

3.3.2. Scheme 2 His-TEV.

The procedure followed was exactly the same described above but when the elution buffer D (10 CV) was added, Nikel beads were stirred smoothly by a glass pippete (5 ml) allowing them to mix well before collecting. 17 fractions were collected (D1-D10), 1 ml each and D11-D17 1,5 ml each. For 2 gr pellet, 2 ml beads were used.

3.3.3. Scheme 1 HypaCas9

In case of **His tagged HypaCas9**, a chromatography column containing Ni-NTA agarose beads (GE Healthcare) was packed and equilibrated with buffer A (50 mM Tris-HCl pH: 8, 1 M NaCl, 10% glycerol, 5mM Imizazole). For 20 gr cell paste, 5 ml (5ml=1CV) of agarose beads were calibrated. The calibration volume was 10 CV. Next, the 3 times diluted supernatant was loaded in the column using a peristaltic pump (flow rate: 1, 5 ml/min). After sample loading, the column was washed using buffer A (flow rate: 2 ml/min) (10CV) and buffer B (50 mM Tris-HCl pH: 8, 50 mM NaCl, 10% glycerol, 5Mm Imizazole) (10CV). The bonded proteins were eluted by a linear gradient of imidazole in Buffer B, starting from 5 mM and reaching 300 mM and collected in 18 fractions of 7.5 ml each (flow rate: 3 ml/min). Finally, the fractions were analyzed by SDS-PAGE in order to assess their relative concentration in HypaCas9, and choose which ones will be further purified.

3.3.4. Scheme 2 HypaCas9

The main process was the same as described above with 20 gr pellet re-suspended in 30 ml equilibration buffer plus PMSF (1 mM final concentration). For cells lysis French Press (3x times) and 3 sonication cycles (3 X30’’), in order to degrade genomic DNA, was used. The lysate was centrifuged at 4°C, 15000 g for 30 min.The soluble lysate was diluted 2 X before loaded to the Ni-NTA column. 2,5 ml beads were loaded onto a chromatography column and washed with 4 CV ddH20 and equilibrated with 10 CV Eq-buffer. The Lysate was loaded onto the column and allowed to pass through by gravity flow. The column was washed with 10 CV of Eq-buffer and 10 CV Wash I buffer. The protein was eluted with 4 CV Elution buffer I (collected in a single fraction), 4CV Elution buffer II (collected in 5 fractions- first four fractions 1 ml each and fraction 5 contained last 5 ml) and 10 CV elution buffer III (collected in 5 fractions of 5 ml each). (Table 4) The purification procedure was analyzed by 12 % SDS PAGE.

Eq-buffer	Wash I	Elution I	Elution II	Elution III
50 mM Tris-HCl pH 8 500mM NaCl 5 mM Imidazole 10% glycerol	50 mM Tris-HCL pH 8 100 mM NaCl 5 mM Imidazole 10 % glycerol	50 mM Tris-HCl pH 8 100 Nacl 20 mM Imidazole	50 mM Tris-HCl pH 8 100 Nacl 50 mM Imidazole	50 mM Tris-HCl pH 8 100 Nacl 300 mM Imidazole

		10% glycerol	10% glycerol	10% glycerol

Table 4. *Buffers involved in the alternative purification for HypaCas9.*

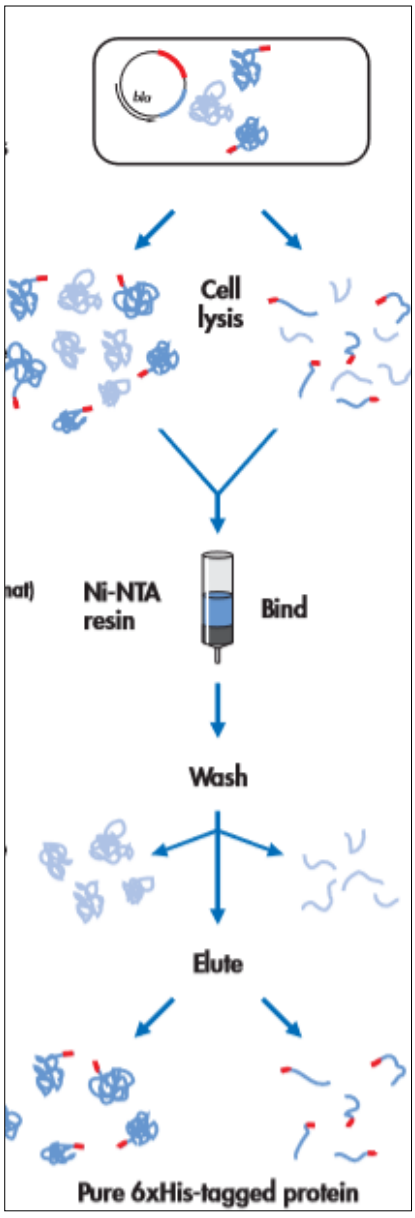


Figure 16. *Purification of His-tagged proteins using the QIAexpress system.*

The two resulting proteins HypaCas9 and His-TEV protease derived from the first purification protocol proposed, were embedded in dialysis (50mM NaCl, 50 mM Tris PH=8, 50% glycerol), overnight at cold room so as to minimize the concentration of imidazole and avoid protein precipitation. HypaCas9 and TEV samples were measured in **nanodrop and incubated** in a ration 4: 0, 8 mg, at 37° C for 1 hour or in 8° C for a day, gently shaking. A site specific cleavage on His-HypaCas9 is expected by

His-TEV protease in order to remove the His-MBP fusion tag from HypaCas9. The cleavage can be easily visualized by an SDS-PAGE 12,5%. To further elucidate the cleavage event, a **western blotting (3.4)** experiment took place.

As far as the D3-D17 samples of His-TEV, derived from the alternative purification scheme, they were pooled and along with E9 from Hypa fractions (also alternative scheme), were subjected to dialysis against 50 mM Tris –HCL PH=8, 100 mM NaCl, 10 % glycerol and 5 mM b-mercaptoethanol, o/n at 4° C. After dialysis Hypa and TEV fractions' concentration was estimated by **nanodrop and incubated** in a ration 4: 0,8 mg, o/n at 4° C, gently shaking. The next day the results were analyzed in SDS-PAGE 9 %.

3.4 Western blotting protocol

From two identical SDS gels that included the TEV protease, the HypaCas9 and the digest of the two, the one was incubated in blue Coomasie and the other one was placed properly in a transfer tank with 4 wattman papers and a nitrocellulose membrane. The tank that included the gel was filled with transfer buffer (10x electrophoresis buffer, 20% methanol, H2O) and remain for 1hour at 138Volt in cold room. Following transfer the blot was incubated in blocking solution (1x TTBS, 5% Dry milk powder), 1h at R.T. gently shaking. One wash in 1xTTBS (20mM Tris pH: 7.5, 150mM NaCl, 0.3% Tween-20), for 15 min and 2 washes for 5min each, followed at R.T. gently shaking. Next step was the overnight incubation, in primary mouse anti-His Ab solution (1x TTBS, 1% Dry milk Powder) (1:200), at cold room, gently shaking. Washes in 1xTTBS followed before the 1 hour incubation in secondary mouse Ab (1:25.000), at R.T., gently shaking. Washes in 1xTTBS, as well as, one wash in 1xTBS for 5min followed, before the 1 min incubation of blot in the ECL reagents mix. ChemiDoc Imaging System was used for the visualization of the result.

3.5 Purification of HypaCas9-TEV digest

3.5.1 Ni-NTA Chromatography for HypaCas9-TEV digest (scheme 1)

A part of the digest derived from the scheme 1 procedure, was loaded in a Ni-NTA chromatography column in order to isolate the untagged HypaCas9 from His-TEV and remaining contaminants. The protocol followed was the same as the purification protocol of His-TEV protease. The same wash and elution buffer were used and the ration of Ni-NTA agarose beads to the digest sample was 1:10. HypaCas9 is expected in the flow through as it will not have the 10xhistidine tag that interacts with the Ni-NTA resin. The undigested HypaCas9 and the His-TEV is expected in the buffer D due to the high concentration in imidazole that antagonizes the binding of tagged proteins to Ni-NTA resin due to its chemical structure that resembles the histidine.(Figure 17)

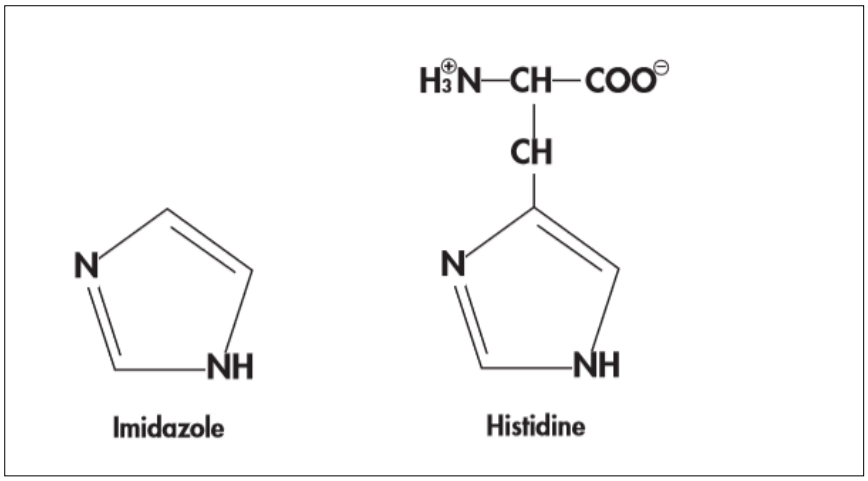


Figure 17. *Chemical structures of histidine and imidazole*

As far as the part of digest derived also from the first procedure, that was not purified with Ni-NTA chromatography is concerned, ~1 ml was placed in dialysis buffer A (50mM Tris PH=8, 1M NaCl, 20% glycerol, 20mM imidazole), ~1ml in dialysis buffer B (50mM Tris PH=8, 1M NaCl, 20% glycerol, 20mM imidazole, 1% Triton-X), ~1ml in dialysis buffer C (50mM Tris PH=8, 1M NaCl, 20% glycerol, 20mM imidazole, 1MUrea), overnight at cold room. The samples A, B, C were incubated in falcons with Ni agarose beads for 1 hour and cetrifugated at 1,200rpm for 2 min. After flow through was collected, centrifugation was repeated after 2 wash steps with dialysis buffer A and 1 elution wash (50mM Tris PH=8, 50mM NaCl, 20% glycerol, 300mM imidazole) for each sample. Among the three samples, flow C (the one treated with Urea) was chosen for the following quality control assays. The concentration of our protein in flow through of sample C was determined by nanodrop measurements.

3.5.2 Gel filtration chromatography for HypaCas9-TEV digest (scheme 2)

For the digest derived from the purification schemes 2 an S 200 Sephacryl HR column was used (360ml, 1CV). The first step included wash with 30 ml 0,2 N NAOH, and equilibration with 2 CV Gel filtration buffer (20 mM Tris pH 7.4, 250 mM NaCl, 10 % glycerol and 5 mM b-mercaptoethanol). 20 ml of Hypa –tev cleavage reaction were concentrated to 2ml by Amicon 10 kD cutoff.

After loading to gel filtration, 36 fraction were collected of 10 ml each. After incubation of 100 µl of each fraction with 900 µl Bradford working solution the fractions that evidently presented protein load were analysed in an SDS PAGE gel 12 %.

3.6 Quality Control Assays

The next step of the process is to check the purity condition of the protein via some quality control assays:

Endonuclease Activity (Nicking): protein samples are tested in a reaction containing a supercoiled DNA substrate (pBR322). The first step was the linearization of pBR322

by PVUII enzyme after 1,5 hour at 37 ◌̊C , visualization of the digest via agaroze gel 2% (1hr/ 100Volt), purification (PCR purification kit by NEB) and nanodrop measurement. After that the DNA was incubated with the protein sample for 2 hours at 37 ◌̊ C and after a heat inactivation at 65 ◌̊C for 5 min, the percent converted to the nicked form was determined by 2% agarose gel electrophoresis at 100Volt for 1 hour. For the reaction a 10x Cas9 buffer (200mM HEPES PH=6,5, 1M NaCL, 50mM MgCL2, 1mM EDTA) was used as well as 1 μg pBR322 with 1 picomole of protein.

Protein Purity (SDS-PAGE): The physical purity is assessed by comparing contaminating protein bands in a concentrated sample to the protein of interest band in a sample of known dilution. The purity is determined by 12,5% SDS-PAGE at 120 Volt for 2 hours.

RNase Assay (Extended Digestion): A 10 μl reaction in Cas9 Nuclease Reaction Buffer containing 40 ng of RNA extracted from Hella cells and 1 picomole of protein was incubated at 37°C for 4 hours. The percent of substrate RNA that remains intact was determined by 1% agarose gel electrophoresis (100Volt/1hr).

Functional assay (Cas9 nuclease, targeted digestion): A 30 μl reaction in 1X Cas9 Nuclease Reaction Buffer containing 1 nM PvuII linearized pBR322 DNA (one targeted site CGCTTGTTTCGGCGTGGGTA), 40 nM sgRNA and 20 nM Cas9 Nuclease NLS, S. pyogenes were incubated for 1 hour at 37°C. The digestion percentage in was determined by 2% agarose gel electrophoresis (100Volt/1 hour).

3.7 sgRNA synthesis and purification

Last but not least, we needed a sgRNA to lead our protein to a target DNA in order to accomplish the functional assay. The sgRNA synthesis was according to EnGen (NEB Kit). The first step is to select target sequence and then to design and order target-specific oligo(s). The synthesis kit provides the ability to synthesize dsDNA and sgRNA in a single reaction, (Figure18).

Primer for sgRNA synthesis:

pBR322 PvuII digested will be targeted at site CGCTTGTTTCGGCGTGGGTA

According to NEB biocalculators sgRNA designer
http://nebiocalculator.neb.com/#!/sgrna

The sequence of the oligo to order was:

TTCTAATACGACTCACTATAGCGCTTGTTTCGGCGTGGGTAGTTTTAGAGCTAGA (55nts)

Transcribed sgRNA sequence (5'→3'):

GCGCUUGUUUCGGCGUGGGUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGC
UAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (101nts)

REAGENT AMOUNT: Nuclease-free water 3 µl, EnGen 2X sgRNA Reaction Mix, S. pyogenes 10 µl, Target-specific DNA Oligo (1 µM) 5 µl, EnGen sgRNA Enzyme Mix 2 µl (Total volume 20 µl) →Incubate at 37°C for 30 minutes.

DNase treatment: adjustment of reaction volume to 50 µl by adding 30 µl of nuclease-free water. Addition of 2 µl of DNase I (RNase-free, provided), incubation at 37°C for 15 minutes.

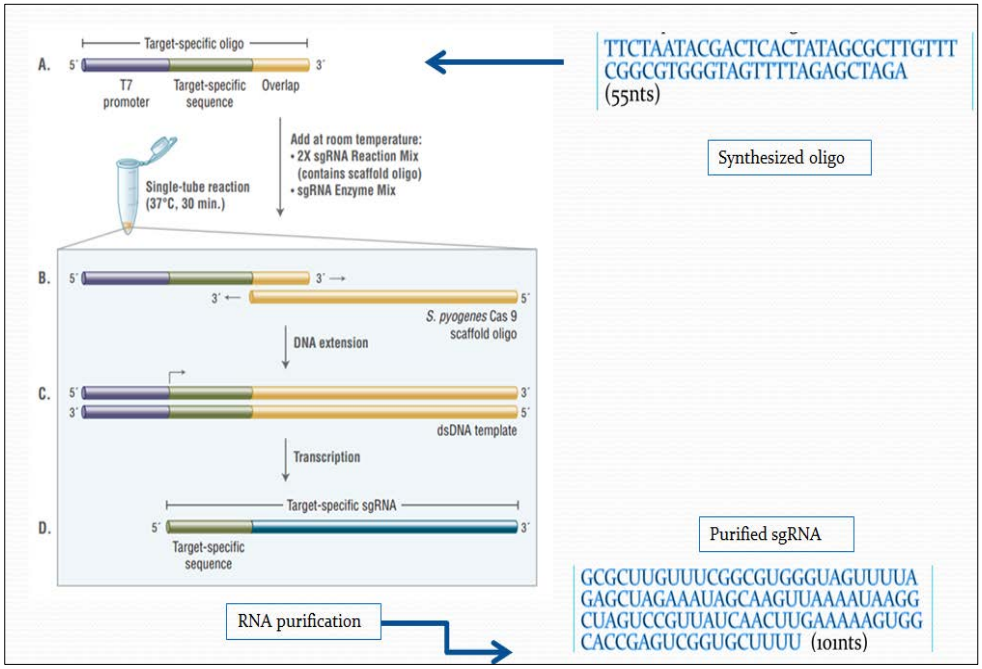


Figure 18. *After designing and ordering target specific oligos, sgRNA synthesis happened according to sgRNA synthesis kit provided by EnGen.*

Purification of Synthesized RNA (E2040)

Protocol: Phenol-chloroform Extraction and Ethanol Precipitation

For removal of proteins and most of the free nucleotides, phenol: chloroform extraction and ethanol precipitation of RNA transcripts is the preferred method:

The reaction volume was adjusted to 180 µl by adding 160 µl nuclease-free water. Next step was the addition of 20 µl of 3 M sodium acetate, pH 5.2, (thoroughly mixed). Extract with an equal volume of 1:1 phenol/chloroform mixture, followed by two extractions with chloroform. Collection of the aqueous phase and transfer to a new tube was necessary. Precipitation of the RNA was achieved by adding 2 volumes of ethanol. After a 30 min incubation at -20°C the pellet was collected by centrifugation. Supernatant was removed and the pellet was rinsed with 500 µl of

cold 70% ethanol. RNA was resuspended in 50 µl of 0.1 mM EDTA and stored at -20°C or below. All centrifugation steps were conducted at 20.000g/ at 4°C.

4. RESULTS

4.1 Cas9 expression plasmids

Three bacterial expression plasmids were provided by Addgene (Table 5):

- pJSCO11 that codes for SpCas9 Cluster 1 (HypaCas9)
- 4xNLS-pMJ915v2 that codes for modified SpyCas9 in *E.coli*
- pET28a-Cas9-His that codes for a Ca9 protein

	Full protein	**Protein ready to use (vs wt Cas9)**	**Applications**	**Advantages**
SpyCas9	**Mutation:** 4x NLS(N terminal on backbone), 2xNLS(C terminal on backbone) **Tags / Fusion Proteins:** His/ MBP/ TEV site (N terminal on backbone)	**Mutation:** 4x NLS(N terminal on backbone), 2xNLS(C terminal on backbone)	**Post-mitotic neurons in mouse brain**	Multiple NLS enabled a tenfold increase in the efficiency of neuronal editing in **vivo**.
Cas9His	**Mutation**: NLS-Cas9-NLS **Tag / Fusion Protein**: 6xHis (C terminal on backbone)	**Mutation:** NLS-Cas9-NLS	**Bread wheat cells**	Reduced off-target due to CRISPRCas9 RNPs (elimination of off-target mutations in mutant plants-no transgene integration)
SpCas9 (HypaCas9)	**Mutation**: C80S, C574S, S355C, S867C, N692A,	**Mutation**: C80S, C574S, S355C, S867C, N692A,	**Human cells**	High fidelity Reduced off-

	M694A, Q695A and H698A **Tags / Fusion Proteins:** 10x His/ MBP/ TEV site (N terminal on backbone)	M694A, Q695A and H698A		target cleavage in human cells

Table 5. *The three bacterial expression plasmids provided by Addgene and their special features*

Each one of the bacterial plasmids above were expressed in *E.coli* strains (*Rosetta blue, Star PlysS*). *Star PlysS* were chosen for the following experiments as they gave better expression levels. We focused on the expression of the bacterial plasmid pJSCO11 because it codes for the HypaCas9, the Cas9 variant with the highest fidelity mentioned so far. The other two bacterial plasmids will be part of future experiments.

4.2 HypaCas9 expression and purification strategy

HypaCas9 has in its N-terminal a 10x **His-MBP fusion tag**, for the easier isolation of the protein during purification via **Ni-NTA chromatography**, and a **TEV site**, for the removal of the fusion tag after the purification process. Due to this fact it was necessary to proceed with the production and purification of a **TEV protease** that will create a site specific cleavage in the 203kD purified His tagged HypaCas9 thus decreasing its molecular weight to 163kD and facilitating an easier transportation via the cell membrane. After the identification of the proper *E.coli* strain, the expression of HypaCas9 and TEV protease followed. First of all, a small scale expression took place for the optimization of the expression and purification protocol. After that we proceeded to large scale production of HypaCas9 and TEV protease (30 lt fermentor). The key steps followed for the expression and purification of both TEV protease and HypaCas9 were the same and are depicted in Figure 19.

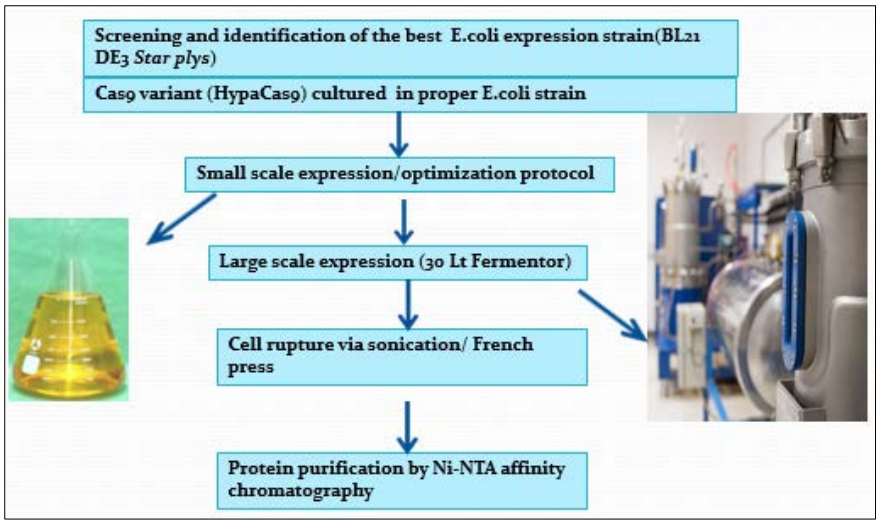


Figure 19. *Experimental procedure followed for the expression and purification of both TEV protease and HypaCas9.*

4.3.1 TEV protease purification scheme 1

The Ni-NTA purification procedure followed for His-TEV was mentioned previously in materials and methods of the current thesis (section 3.3.1). The purification results are depicted in the figure below (Figure 20). The purification gave way to dialysis overnight at cold room in a buffer (50mMNaCl, 50mMTris PH=8, 20% glycerol) free of Imidazole in order to reduce its concentration in the samples of interest and avoid protein precipitation. Samples 11,13,14,15 were united in one sample (1) and placed in dialysis membrane. The sample 16 was diluted 6 times and placed in dialysis membrane as sample 2. Sample 17 was also diluted 3 times and placed in dialysis membrane as sample 3. Samples 16, 17 were judged necessary to be diluted as it was noticed by previous experiments that His-TEV is precipitated in high concentrations (over 1mg/ml). Samples 18, 19 that exhibited the lowest percentage of protein contaminants were united in one sample (4) and also placed in dialysis membrane.

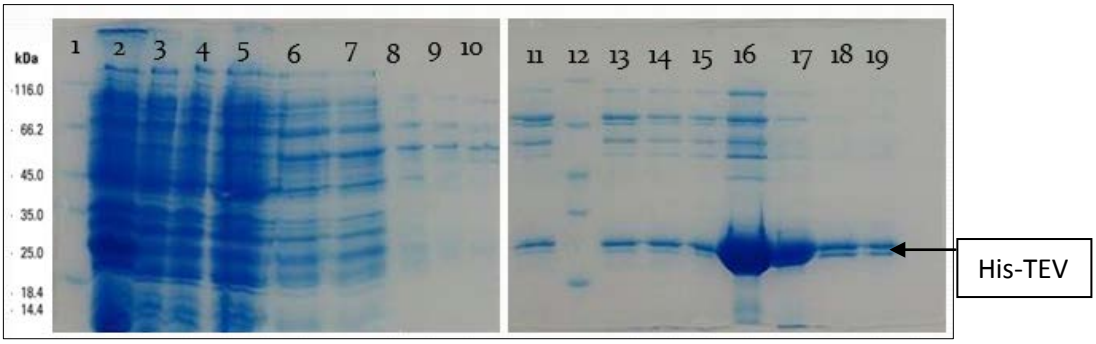


Figure 20. *Purification of His-Tev protease analyzed with SDS-PAGE and visualized with Coomasie staining. 1: protein marker, 2: load sample, 3, 4: flow through samples, 5, 6, 7: buffer A wash samples,*

8,9,10: buffer B wash samples, 12: protein marker, 11,13,14,15: buffer C elution samples, 16,17,18,19: buffer D elution samples.

Nanodrop measurements followed (Table 6) in order to calculate the exact concentration of each sample.

	mg/ml	A280	160/280
Sample 1	0,16	0,178	0,95
Sample 2	0,30	0,344	0,82
Sample 3	0,25	0,287	0,92
Sample 4	0,20	0,232	1,23

Table 6. *Nanodrop measurements for His-TEV, (extinction coefficient: 32, molecular weight: 28kD).*

4.3.2 TEV protease purification scheme 2

The Ni-NTA purification procedure followed for His-TEV was mentioned previously in materials and methods section of the current thesis (section 3.3.2). The results are depicted in the figure below (Figure 21). Elutions at 300mM imidazole were all combined and dialyzed against buffer 50 mM Tris –HCL PH=8, 100 mM NaCl, 10 % glycerol.

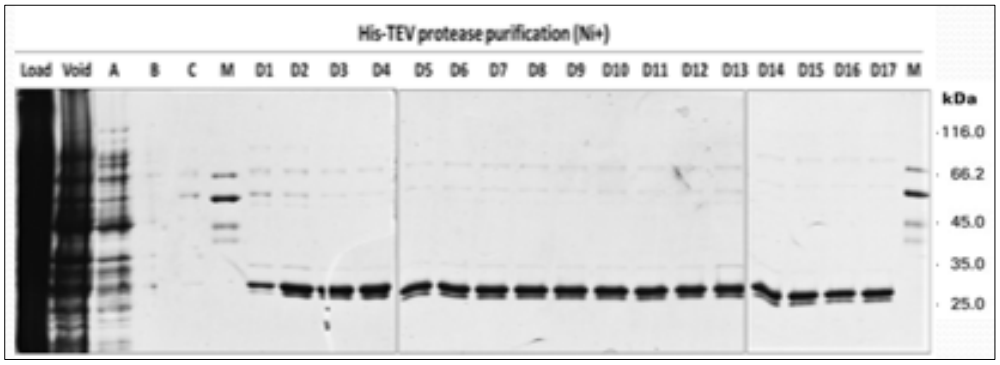


Figure 21. *His-TEV purification: load, flow through, wash sample A, wash sample B, wash sample C, protein marker, D1,D2,D3,D4,D5,D6,D7,D8,D9,D10, D11,D12,D13,D14,D15,D16,D17: elution sample with 300mM Imid., protein marker. His-TEV is visible in D1-D17 elution samples in same concentration and clarity.*

4.4.1 HypaCas9 purification scheme 1

The Ni-NTA purification procedure followed for HypaCas9 was mentioned previously in materials and methods section of the current thesis (section 3.3.3). The purification gave way to dialysis overnight at cold room in a buffer (100mMNaCl, 50mMTris PH=8, 10% glycerol) free of Imidazole in order to reduce its concentration

in the samples of interest and avoid protein precipitation. Samples 15, 16 were united in one sample (A) and placed in dialysis membrane. Samples 11, 13,14,17,18 were also united in one sample (B) and placed in dialysis membrane.

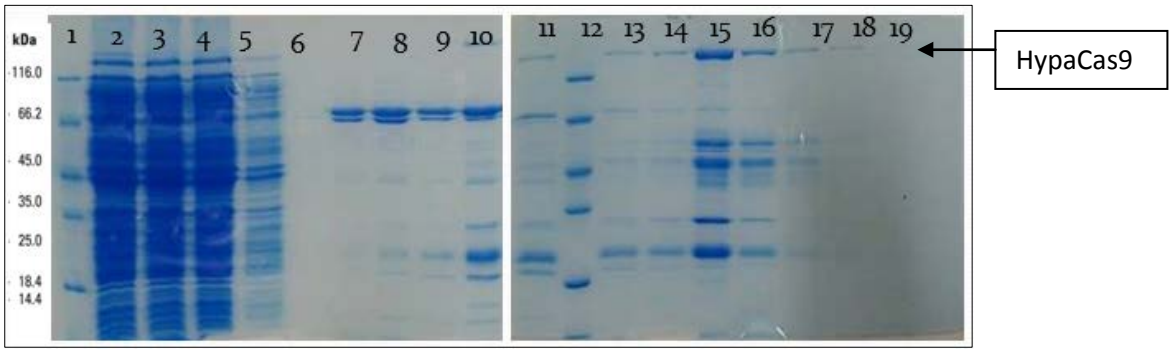


Figure 22. *Purification of HypaCas9 analyzed with SDS-PAGE and visualized with Coomasie staining. 1: protein marker, 2: load sample, 3, 4: flow through samples, 5, 6: wash samples, 7,8: elution 1 samples (20mM imidazole), 9,10: elution 2 samples (40mM imid.), 12: protein marker, 11,13: elution 3 samples (50mM imid.), 14,15,16: elution 4 samples (100mM imid.), 17,18,19: elution 5 samples (150mM imid.), 20: elution 6 sample (200mM imid.).*

Nanodrop measurements followed (Table 7) in order to calculate the exact concentration of each sample.

	mg/ml	A280	260/208
Sample A	0,56	0,522	0,56
Sample B	0,20	0,183	0,65

Table 7. *Nanodrop measurements for HypaCas9, (extinction coefficient: 188, molecular weight: 203kD)*

4.4.2 HypaCas9 purification scheme 2

The Ni-NTA purification procedure followed for HypaCas9 was mentioned previously in materials and methods section of the current thesis (section 3.3.4). E9 fraction was dialyzed against buffer 50 mM Tris –HCL PH=8, 100 mM NaCl, 10 % glycerol.

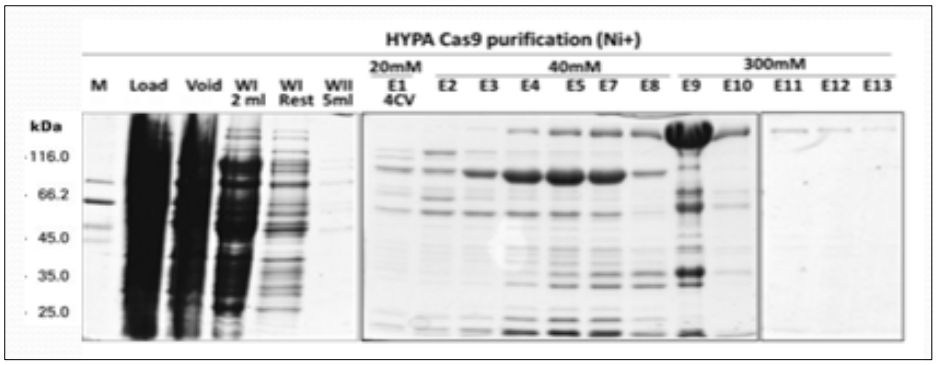


Figure 23. *HypaCas9 purification: protein marker, load, flow through, first 2 ml wash 1, rest of wash1, wash 2, E1: elution sample with 20mM Imid., E2,E3,E4,E5,E6,E7,E8: elution samples with 40mM imid., E9,E10,E11,E12,E13: elution samples with 300mM imid. HypaCas9 is visible in E3-E13 elution samples with a pick in elution sample E9.*

4.5 Digest of HypaCas9 by HisTEV protease scheme 1

HypaCas9 (from Scheme 1) was incubated with His-TEV protease (from Sheme 1 in various combinations of concentration rations. The best ratio was found to be 4mg HypaCas9:0,8mg His-TEV. Two incubation conditions were tested. The first one was overnight at cold room (10^{o}C) facilities and the second was for 1 hour at 37^{o}C. No difference at the cleavage efficiency was noticed so the second condition was chosen as the faster one. SDS-PAGE that depicts the cleavage event of HypaCas9 sample A and B after incubation with all the TEV samples available is presented below (Figure 24). Samples of small volumes (~40µl) were chosen in order to test more conditions.

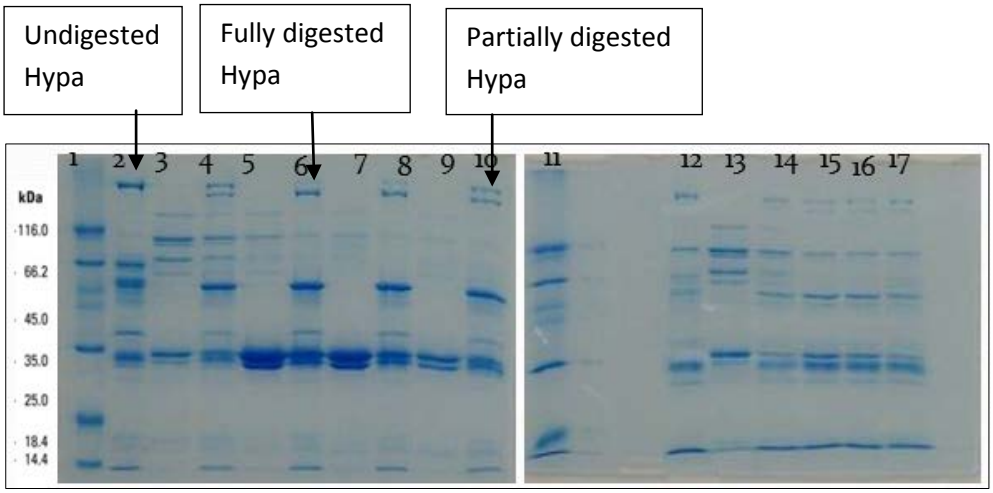


Figure 24. *SDS-PAGE that depicts the cleavage event of HypaCas9 sample A and B after incubation with all TEV samples available. 1: protein marker, 2:HypaA, 3: TEV1, 4:digest A1, 5: TEV2, 6:digestA2, 7: TEV3, 8:digest A3, 9:TEV4, 10:digest A4, 11:protein marker, 12:Hypa B, 13:TEV1, 14:digest B1, 15:digest B2, 16:digest B3, 17:digest B4.*

Among those tests, digest A2 of HypaCas9 (A) and TEV 2, was the best as it had the highest percentage of digested HypaCas9. The next step was to incubate larger

volume (~10ml) of HypaCas9 and TEV protease (overnight at 10°C) for the preparation of HypaCas9 for further purification. A2 digest was analyzed through SDS-PAGE (Figure 25) that depicts the cleavage efficiency and through a western blot that verifies the absence of his-tag in the digested Cas9 protein and its presence in the His-HypaCas9 and His-TEV and His-MBP product after TEV digest (Figure 26).

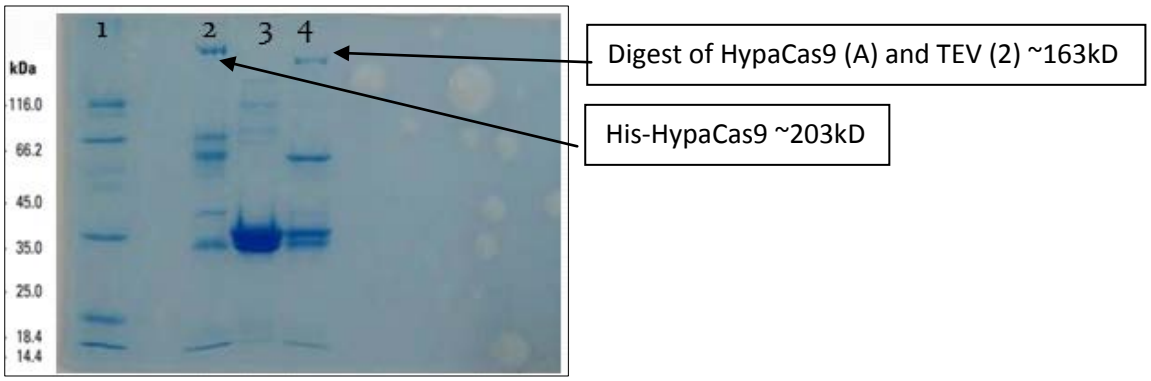


Figure 25. *1: protein marker, 2: HypaCas9 (A), 3: His-TEV (2), 4: Digest of HypaCas9 (A) and TEV protease (2). We can visualize His-HypaCas9 in ~203Kd, His-TEV in ~28Kd, Digest A2 in ~163Kd.*

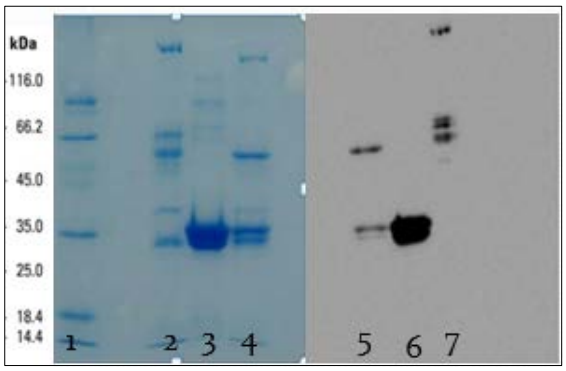


Figure 26. *1: protein marker, 2: HypaCas9 (A), 3: His-TEV (2), 4: Digest of HypaCas9 (A) and TEV protease (2), the result of western blot for 5: digest (TEV – HypaCas9), 6: His-TEV, 7: His-HypaCas9. The digested HypaCas9 depicted in sample 4 at 163kd is absent in the respective place in the western blot. This fact is indicative of the cleavage event. In contrast, we can visualize the presence of His tag in sample 5 and 6 (at 28kd where there is His-TEV), in sample 5, 7(at 40kd where the His-MBP fusion tag exists), and in sample 7(at 200kd where His-HypaCas9 exists).*

4.6 Digest of HypaCas9 by HisTEV protease scheme 2

HypaCas9 -TEV Protease Reaction included 16ml (3,2 mg) (from Scheme 2) TEV incubated with 6,8 ml(14,28 mg) HYPA Cas9(4:0,8) (from Scheme 2) o/n at 40 C, gently shaking. The results that were analyzed in SDS-PAGE 9 % gave a crystal clear visualization of the successful cleavage. (Figure 27).

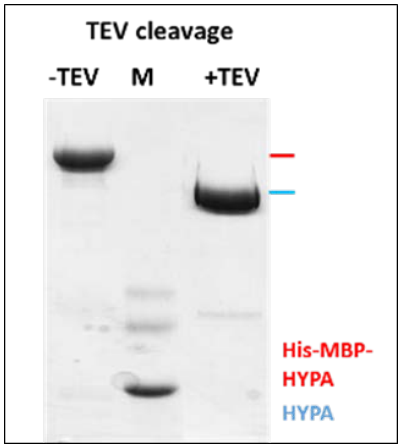


Figure 27. *HypaCas9 sample without TEV presence, protein marker, HypaCas9 with TEV. It is clear that in presence of TEV protease a site specific cleavage is introduced and the His-MBP tag is removed from HypaCas9.*

4.7.1 Purification and Quality assessment of Scheme 1 digestion reaction

Further purification of the digest (TEV – HypaCas9) was necessary through Ni-NTA or gel filtration chromatography (Figure 28) (section 3.5.1).

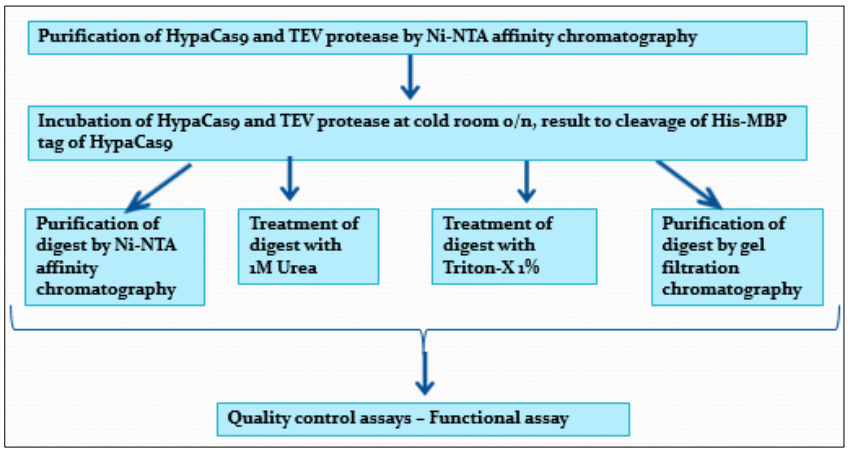


Figure 28. *Synopsis of the purification procedure conducted for the eventual isolation of HypaCas9*

Only a part of the digest was used for purification with Ni-NTA chromatography. The results showed in an SDS-PAGE were negative, as the imminent presence of our protein in flow through was never visualized (Figure 29). Nanodrop measurements for the flow through were also negative.

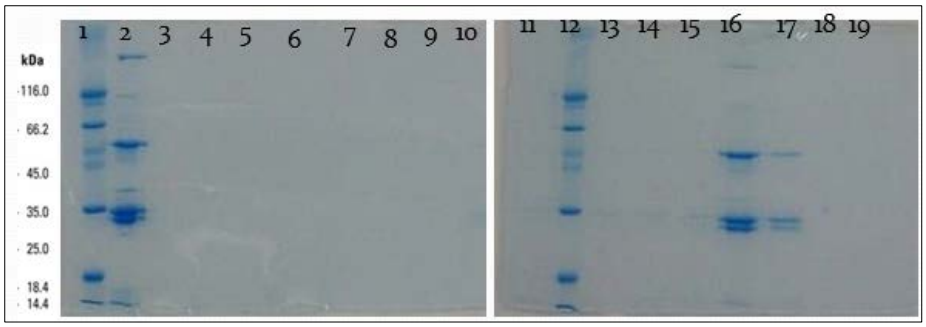


Figure 29. *1: protein marker, 2: load, 3: flow through, 4,5,6: wash A samples, 7,8,9: wash B samples, 10,11, 13,14,15: wash C samples, 12: protein marker, 16,17: elution D samples. HypaCas9 is not visible in flow through or anywhere else. In samples 16, 17 TEV is eluted.*

What we did next was to put a small volume ~1 ml of the digest in dialysis buffer A (50mM Tris PH=8, 1M NaCl, 20% glycerol, 20mM imidazole), ~1ml in dialysis buffer B (50mM Tris PH=8, 1M NaCl, 20% glycerol, 20mM imidazole, 1% Triton-X), ~1ml in dialysis buffer C (50mM Tris PH=8, 1M NaCl, 20% glycerol, 20mM imidazole, 1MUrea), overnight in cold room. The samples A, B, C were incubated in falcons with Ni agarose beads for 1 hour and centrifuged at 1,200rpm for 2 min. After flow through was collected, centrifugation repeated after 2 wash steps with dialysis buffer A and 1 elution wash (50mM Tris PH=8, 50mM NaCl, 20% glycerol, 300mM imidazole) for each sample. In flow through A, B, C HypaCas9 was visualized in high clarity but low concentration (Figure 30). Among the three samples, flow C (the one treated with Urea) was chosen as the best for the following quality control assays. The concentration of our protein in flow C sample was determined by nanodrop measurements (0,07mg/ml).

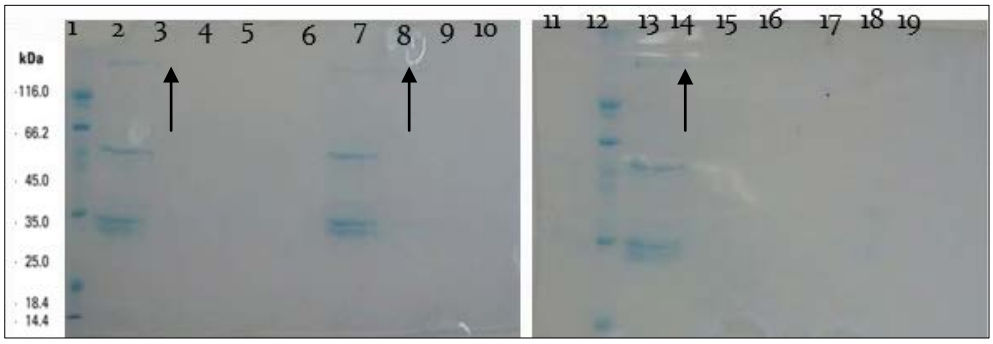


Figure 30. *1: protein marker, 2: load A, 3: flow A, 4, 5: wash A, 6: elution A, 7: load B, 8: flow B, 9, 10: wash B, 11: elution B, 12: protein marker, 13: load C, 14: flow C, 15, 16: wash C, 17: elution C. Elution A, B, C are clearly depicted in samples 2, 6, 13 respectively. Flow through A, B, C (samples 3, 8, 14) show our protein as trace (Due to the image's failure to not ideally present the protein, errors were placed to exhibit its position).*

Quality controls: Nuclease and RNAse activity assay (Figure 31) revealed absence of endonuclease or RNAse activity.

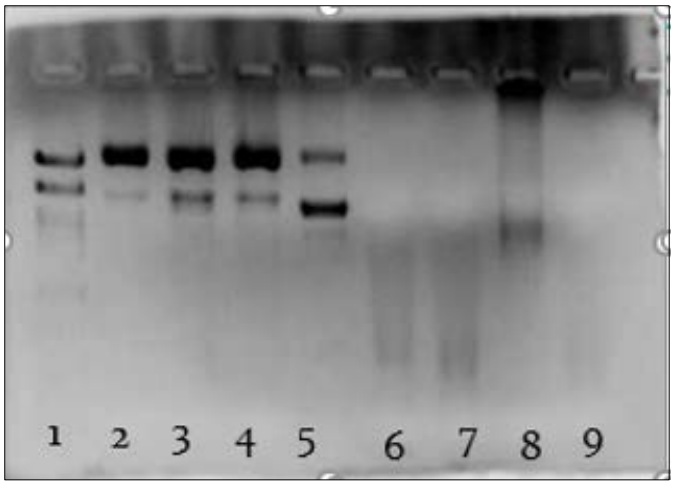


Figure 31. *1: λ/styI marker, 2: flow B – PBR, 3: flow C – PBR, 4: flow B –PBR – RNAse inhibitor, 5: Control (PBR –no HypaCas9), 6: flow B – RNA, 7: flow C – RNA, 8: flow B –RNA – RNAse inhibitor, 9: Control (RNA –no HypaCas9).*

As far as the sgRNA synthesis is concerned, we used the kit provided by EnGen multiple times, yet the result was negative. After the control reaction included in kit gave also negative result, we assumed that the kit was not functional and that we need to figure out alternative ways for sgRNA synthesis. Functional test was not applied due to the low concentration of protein, the presence of RNAses and the failure of sgRNA synthesis.

4.7.2 Purification and Quality assessment of Scheme 2 digestion reaction

The resulting digest(figure of digest 2) had the best clarity and quantity of any other digest until now, so we proceeded in gel filtration chromatography that seperates protein according to their molecular weight (section 3.5.2). The finest fractions 17 and 18 were pooled along with 16 and dialyzed against (20 mM Tris-pH 7,4, 250 mM NaCl, 50% glycerol and 5 mM b-mercaptoethanol) for 24 hours at 4 C. After that we proceeded with sample concentration with 50 kD cut off of 5 ml of fraction 17/18 to 500 μl, in order to reach protein concentration similar to that of NEB spCas9.

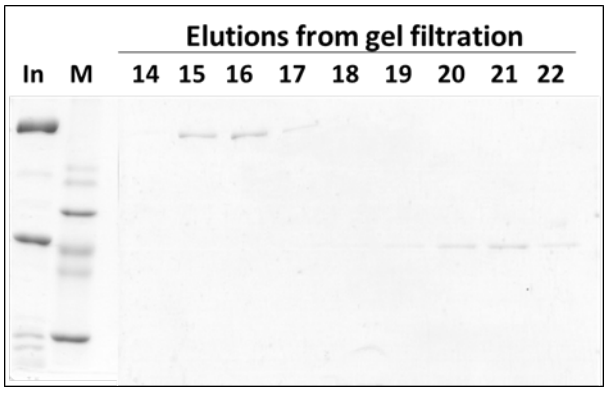


Figure 32. *The digest was further purificated by gel filtration chromatography. Initial sample, protein marker, elutions 14,15,16,17,18,19,20,21,22. The target protein was visible in elution samples 14-18*.

Since the NEB gRNA synthesis failed we proceeded to a functional assay with a sgRNA designed by Dr.Talianidis lab (IMBB-FORTH_against HNF4 gene which is cloned into PCMx vector). sgRNA guides HypaCas9 to target site of cloned HNF4 gene and linearize PCMx vector. (Figure 33) Only when HypaCas9 and the specific sgRNA are both present the reaction leads to the specific product. When components are either missing or when Hypa was incubated with an unspecific sgRNA (sgRNA targets pBR322 vector sequence) the substrate remained intact. The functional assay result was visualized in an agaroze gel.(Figure 33)

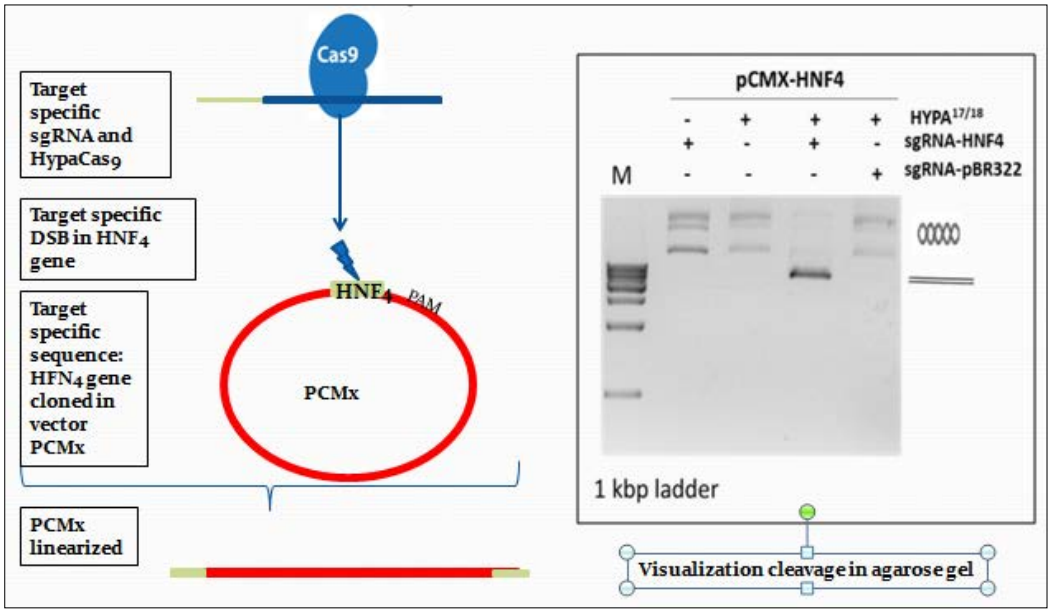


Figure 33. *sgRNA guides HypaCas9 to target site of cloned HNF4 gene and linearize PCMx vector. λ/styI marker, sgRNA–HFN4, Hypa 17/18, sgRNA–HFN4 and Hypa 17/18, sgRNA–PBR322 and Hypa 17/18(negative control).*

5. DISCUSSION

HypaCas9, the newest artificial Cas9 variant, has engrossed our interest as its high fidelity action can minimize the off targeting thus broading the horizons of genome editing technology. The aim of this master thesis was the expression and purification of this particular Cas9 variant. This goal was achieved after many small-scale expression trials, which allowed us to figure out a proper expression and purification scheme. The large-scale expression ended up to a satisfying concentration and purity level of the target protein. In parallel with HypaCas9 the TEV protease's expression and purification also resulted to an effective scheme. Ni-NTA chromatography was applied with success in both cases. As it was imminent, the upcoming incubation of the two proteins resulted in full removement of His-MBP tag from HypaCas9, resulting to a HypaCas9 protein sample fulfilling all necessary quality controls.

At this point, it should be declared that the best purification scheme for the digested HypaCas9 has proven to be the attempt that included gel filtration chromatography. This scheme not only gave the higher concentration of HypaCas9, but also the best clarity for the target protein (HypaCas9). Previous treatment assays with Urea, Triton-x 1% or Ni-NTA chromatography have not concluded to a combination of protein clarity and high concentration.

It should be emphasized that HypaCas9 is going to be used in a ribonucleoprotein complex for genome editing in the animal house facility of IMBB-FORTH, fact that demands a flawless condition. The so far applied quality control assays showed absence of DNAses and RNAses and an active enzyme judging by the *in vitro* assay that was used.

To conclude, HypaCas9 expression and purification is of great value if we consider the wide spectrum of applications included. The progress made during this current master thesis will help to set a state-of the art genome editing tool in IMBB-FORTH.

6. REFERENCES

Advanced biotechnology: metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products. Becker J, 2015),

A guide to genome engineering with programmable nucleases, Hyongbum Kim, Jin-Soo Kim, 2014

AL effectors: finding plant genes for disease and defense. Bogdanove AJ, Schornack S, Lahaye T, 2010

A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9, Kelli J. Carroll, 2016

A One-Step System for Convenient and Flexible Assembly of Transcription Activator-Like Effector Nucleases (TALENs).Zhao J, 2016

A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Jinek M, 2012

A simple cipher governs DNA recognition by TAL effectors. Moscou MJ, Bogdanove AJ, 2009

Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo, J.A. Zuris, et al., 2015

C2c2 is a single -component programmable RNA -guided RNA - targeting CRISPR effector, Abudayyeh, O.O. et al., 2016

Cornerstones of CRISPR–Cas in drug discovery and therapy, Christof Fellmann, 2017

Cpf1 is a single RNA -guided endonuclease of a class 2 CRISPR -CAS system. Zetsche, B. et al., 2015

CRISPR-Cas9D10A Nickase-Assisted Genome Editing in Lactobacillus casei. Song X, 2017

CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations, Douglas W. Drury, 2017

CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Gilbert LA, 2013

Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Yamano, T. et al., 2016

Crystal structure of Cas9 in complex with guide RNA and target DNA, Nishimasu H et al, 2014

Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases, Tsai SQ, 2016

Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9, LaFountaine, 2015

Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications, Chang Liu et.al, 2017

Disruption of HPV16-E7 by CRISPR/Cas System Induces Apoptosis and Growth Inhibition in HPV16 Positive Human Cervical Cancer Cells, Zheng Hu, 2014

DNA interrogation by the CRISPR RNA-guided endonuclease Cas9, Sternberg SH, 2014

Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes, Zhen Liang, 2017

Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. Staahl BT, 2017

Engineering and Application of Zinc Finger Proteins and TALEs for Biomedical Research, Moon-Soo Kim, 2017

Engineering polydactyl zinc-finger transcription factors. Beerli RR, Barbas CF 3rd, 2002

Engineering the Delivery System for CRISPR-Based Genome Editing Zachary Glass, 2018

Enhanced proofreading governs CRISPR–Cas9 targeting accuracy, Jennifer A. Doudna et al, 2017

Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9, Rajat M. Gupta and Kiran Musunuru, 2014

Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients, Schwank G, 2013

Generation of gene-modified cynomolgus monkey via Cas9/RNA mediated gene targeting in one-cell embryos, Y. Niu, et al., 2014

Genome Editing: Past, Present, and Future, Dana Carroll, 2017

Genome engineering at the dawn of the golden age. Segal DJ, Meckler JF, 2013

Genome Engineering with Targetable Nucleases, Dana Carroll et al., 2014.

Genome Engineering Using the CRISPR-Cas9 System. Ran, FA. Et al. 2013

Genome-scale engineering for systems and synthetic biology, Kevin M Esvelt, 2013

Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins, S. Kim, D. Kim, S.W. Cho, J. Kim, J.S. Kim, 2014

http://blog.addgene.org/enhancing-crispr-targeting-specificity-with-espcas9-and-spcas9-hf1

http://doudnalab.org/research_areas/crispr-systems/

http://nebiocalculator.neb.com/#!/sgrna

https://international.neb.com/

https://international.neb.com/protocols/2016/05/11/engen-sqrna-synthesis-kit-s-pyogenes-protocol-e3322

https://www.abmgood.com/marketing/knowledge_base/CRISPR_Cas9_Introduction_Part7.php

https://www.addgene.org/

https://www.addgene.org/89610/

http://www.chem.uky.edu/courses/che554/6_AffinityChro/QIAexpressionist.pdf

https://www.horizondiscovery.com/resources/webinars/an-introduction-to-crispr-genome-editing

https://www-ssrl.slac.stanford.edu/content/science/highlight/2015-01-30/crispr-rna-guided-surveillance-escherichia-coli

In vivo genome editing restores haemostasis in a mouse model of haemophilia, Li H, 2011

Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study, 2009

Mechanism of DNA double-strand break repair by non-homologous end joining. Hefferin ML1, Tomkinson AE., 2005

Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. Lee SK1, 2008

Novel lentiviral vectors with mutated reverse transcriptase for mRNA delivery of TALE nucleases. Mock U, Riecken K., 2014

Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy, Chengzu Long, 2015

Precise Correction of the Dystrophin Gene in Duchenne Muscular Dystrophy Patient Induced Pluripotent Stem Cells by TALEN and CRISPR-Cas9, Hongmei LisaLi, 2015

Rationally engineered Cas9 nucleases with improved specificity, Ian M. Slaymaker, 2015

RNA -guided gene activation by CRISPR -Cas9 -based transcription factors., Perez - Pinera, P. et al., 2013

RNA-guided human genome engineering via Cas9. Mali P, 2013, Genome engineering Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease, Carolin Anders, Ole Niewoehner, Alessia Duerst, Martin Jinek, 2014

Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyback, 2014, Fei Xie

Short motif sequences determine the targets of the prokaryotic CRISPR defense system. Mojica FJ, 2009

Structures of Cas9 endonucleases reveal RNA-mediated conformational activation, Jinek M, 2014

TALENs: a widely applicable technology for targeted genome editing. Joung JK, Sander JD Nat Rev Mol Cell Biol. 2013

TALEN gene editing takes aim on HIV, Ronald Benjamin, 2016

TALEN-Mediated Knockout of CCR5 Confers Protection against Infection of Human Immunodeficiency Virus.Shi B, Li J, Shi X, 2017

Targeting CDK11 in osteosarcoma cells using the CRISPR-Cas9 system, Feng Y, 2015

Transforming plant biology and breeding with CRISPR/Cas9, Cas12 and Cas13. Schindele P et al, 2018

The therapeutic landscape of HIV-1 via genome editing, Alexander Kwarteng, 2017

ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering, Gaj T, 2013

“ In memory of my grandmother”