IDENTIFICATION OF THE MAJOR SUMOYLATION SITE ON CRISPR CAS9 PROTEIN

by

Duygu Yeşildağ

B.S., Molecular Biology and Genetics, Istanbul Technical University, 2016

Submitted to the Institute for Graduate Studies in Science and Engineering in partial fulfillment of the requirements for the degree of Master of Science

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

2019

ACKNOWLEDGEMENTS

This project was supported by the funds of European Molecular Biology Organization (EMBO).

ABSTRACT

IDENTIFICATION OF THE MAJOR SUMOYLATION SITE ON CRISPR CAS9 PROTEIN

CRISPR/Cas9 system is a type of adaptive immune system, which operates in prokaryotic organisms as an antiviral defense. Ever since its adaptation to eukaryotic systems as a gene editing tool, CRISPR has been one of the most popular genetic engineering technologies. Its diverse applications in molecular biology and medicine such as gene therapy, drug design, epigenome regulation, etc. have made this technology worth exploring further and further. In order to set up a more specific and efficient gene targeting system, it is essential to learn how the CRISPR pathway is regulated, in particular the Cas9 enzyme (CRISPR-associated protein 9), which is the key player of this system.

In the lights of strongly supported data provided by a current lab member, it is revealed that Cas9 is covalently modified by SUMO-1 and SUMO-2 proteins. This was the first ever demonstration for a post-translational modification to which Cas9 is subjected. In the scope of this study, we intended to identify the major sumoylation site of Cas9 protein and generate a Cas9 mutant that is defective in sumoylation. For this purpose, a series of site-directed mutagenesis experiments were conducted. As a result, 10 single mutants for the 10 sumoylation motifs found on Cas9 protein were generated, as well as an 11^{th} mutant in which all 10 sumoylation consensus motifs were destroyed altogether.

ÖZET

CRISPR CAS9 PROTEİNİNİN ANA SUMOLASYON BÖLGESİNİN TANIMLANMASI

CRISPR / Cas9 sistemi, prokaryotik organizmalarda antiviral savunma olarak çalışan bir tür adaptif bağışıklık sistemidir. Ökaryotik sistemlere gen düzenleme aracı olarak adapte edildiğinden beri, CRISPR en popüler genetik mühendisliği teknolojilerinden biri olmuştur. Moleküler biyoloji ve tıp alanlarında yapılan gen tedavisi, ilaç tasarımı, epigenom düzenlemesi vb. çeşitli uygulamalar, bu teknolojiyi daha da keşfedilmeye değer kılmıştır. Daha spesifik ve etkili bir gen hedefleme sistemi kurmak için, CRISPR yolağının, özellikle bu sistemin asıl oyuncusu olan Cas9 enziminin (CRISPR ile ilişkili protein 9) nasıl düzenlendiğini öğrenmek elzemdir.

Mevcut laboratuvar üyesi tarafından sağlanan ve güçlü şekilde desteklenen verilerin ışığında, Cas9'ın SUMO-1 ve SUMO-2 proteinleri tarafından kovalent olarak modifiye edildiği ortaya çıkmıştır. Bu, Cas9'ın uğradığı gösterilen ilk post-translasyonal modifikasyondur. Bu çalışma kapsamında, Cas9 proteininin ana sumolasyon bölgesini tanımlamayı ve sumolasyonu engellenmiş bir Cas9 mutantı üretmeyi amaçladık. Bu amaçla, bir dizi yönlendirilmiş mutajenez deneyi yapılmıştır. Sonuç olarak, Cas9 proteini üzerinde bulunan 10 sumolasyon motifi için 10 adet tekli mutant ve bu 10 sumolasyon motifin 10 unun da yok edildiği 11. bir mutant üretildi.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

bp	Base pair
g	Gram
kb	Kilobase
L	Liter
ml	Milliliter
mm	Millimeter
M	Molar
ng	Nanogram
nm	Nanometer
rpm	Revolutions per minute
v	Volume
w	Weight
µg	Microgram
µm	Micrometer
µl	Microliter
°C	Degree Celcius

LIST OF ACRONYMS/ABBREVIATIONS

Ago-2	Argonaute-2
ATL	Adult T-cell Leukemia
Cas9	CRISPR-associated protein 9
Cpf1	CRISPR from Prevotella and Francisella 1
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
dCas9	dead Cas9
DMD	Duchenne Muscular Dystrophy
DNA	Deoxyribonucleic Acid
DSB	Double Strand Break
FISH	Fluorescently-Labelled In Situ Hybridization
FRET	Fluorescence Resonance Energy Transfer
HDAC1	Histone Deacetylase 1
HDR	Homology-Directed Repair
HIV	Human Immunodeficiency Virus
HP1	Heterochromatin Protein 1
HPV	Human Papilloma Virus
HTLV-1	Human T Cell Leukemia Virus Type 1
mRNA	messenger RNA
nCas9	Cas9 nickase
NHEJ	Nonhomologous End Joining
NUC	Nuclease
PAM	Protospacer Adjacent Motif
PCR	Polymerase Chain Reaction

PI	PAM-interacting
PTM	Post-Translational Modification
REC	Recognition
RNA	Ribonucleic Acid
SDM	Site-Directed Mutagenesis
SENP	Sentrin-Specific Protease
sgRNA	single guide RNA
SIM	SUMO-Interacting Motif
STUbLs	SUMO-Targeted Ubiquitin Ligases
SUMO	Small Ubiquitin-Like Modifier
T1S TRP120	Type 1 Secretion Tandem Repeat Protein 120
TALEN	Transcription Activator-Like Effector Nuclease
tracrRNA	trans-activating crRNAs
WED	Wedge
ZFN	Zinc Finger Nuclease

1. INTRODUCTION

The ability to modify desired genes has always been an exciting matter in the field of genetics. Ever since DNA has been identified as the genetic material, scientists have dreamed of ways to alter genes in order to study gene function, disease modeling, genomic imaging, and various other areas. For the alteration of a desired gene, the gene has to be successfully targeted and cleaved by a nuclease, and then the cleaved DNA needs to be repaired.

For the past two decades, various gene editing mechanisms have been introduced. Among several gene manipulation technologies, the most popular ones have utilized zinc finger nucleases (ZFN) or Transcription activator-like effector nucleases (TALEN) (Figure 1.1). Zinc finger nucleases are comprised of zinc finger domains and a restriction enzyme called Fok1 which needs to form a dimer in order to become active and induce a double strand break; when zinc finger domains at both sides bind their target strands, Fok1 enzymes fused with these domains can form a dimer and introduce a double-strand break (DSB) on target genome (Kim *et al.,*1996). A similar system is applied with TALEN where instead of zinc finger domains, DNA-binding repeat motifs are fused with Fok1 enzyme (Miller *et al.,* 2011). Both of these methods are based on a protein's ability to interact with a specific DNA sequence which amplifies the work load on designing new motifs that can bind desired DNA sequences.

Thanks to the advent of CRISPR Cas9 system, gene targeting has become more practical and relatively more specific due to the fact that it relies on nucleotide base pairing between target DNA and a secondary structure of RNA called single guide RNA (sgRNA). Upon introduction of the CRISPR Cas9 system as a gene engineering technology, its practice has been disseminated among researchers all around the globe.

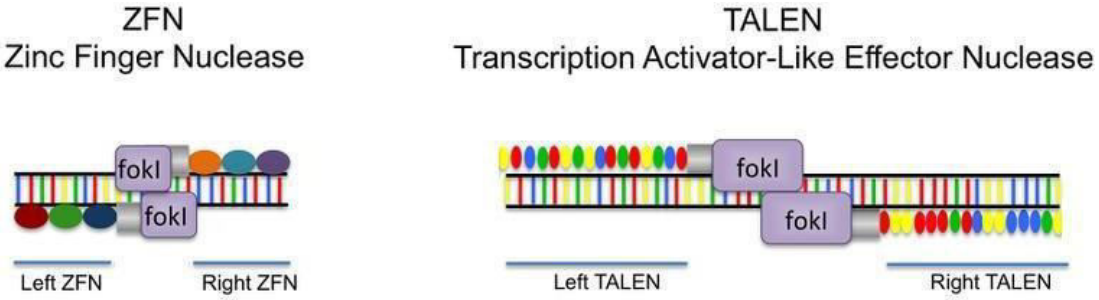


Figure 1.1 Zinc finger nucleases (ZFN) and Transcription activator-like effector nucleases (TALEN). They both utilize the ability of certain protein domains to interact with DNA to target specific regions; ZFN possess zinc finger domains and TALEN possess TALE-DNA binding domains fused with Fok1 nuclease which introduce double strand break when DNA binding domains at both sides bind their target regions (adapted from Moore *et al.*, 2012).

1.1. CRISPR System

CRISPR is an adaptive immunity mechanism which has evolved in many types of bacteria and archaea in order to ward off viruses through short circuiting the integration of viral genome into their own genome. The main purpose of this system is to recognize viral genome pieces, transcribe them into RNA and use these to guide an endonuclease (or a group of endonucleases) to destroy viral genome pieces that have infiltrated the host DNA (Bhaya, Davison, & Barrangou, 2011). CRISPR is divided into two sub-systems: Class I and Class II CRISPR systems. Class I CRISPR system requires a group of endonucleases to carry out foreign DNA cleavage, whereas Class II CRISPR system can manage this with a single endonuclease (Makarova *et al.,* 2015). In Class II Type II system, Cas9 is the major player which cleaves viral DNA off, and in Class II Type V system, Cpf1 accomplishes viral DNA cleavage (Murovec, Pirc, & Yang, 2017).

CRISPR system is comprised of three steps; acquisition, processing, and interference (Figure 1.2). In the acquisition step, viral gene fragments called protospacers (comprised of spacer and repeat sequences) are recognized by a set of Cas proteins and consecutively get cut off in order to be packed in the CRISPR locus of the bacterial genome, forming a CRISPR array at this region (Hille & Charpentier, 2016). In the processing step, spacer fragments are transcribed into pre-crRNA (pre-crispr RNA) fragments, and then they get processed into crRNAs by Cas proteins. In the last step called the interference step, crRNA fragments are joined together with different types of endonucleases (depending on the particular CRISPR system) which initiate the cleavage of viral DNA in sequence-dependent manner.

On an important note, Type II system possesses RNA structures called tracrRNAs (trans-activating crRNAs) which form a complex with crRNAs (as a part of crRNA processing) and this enhances the efficiency of crRNA interaction with Cas9 endonuclease (Westra, Buckling, & Fineran, 2014). Another critical detail is the presence of PAM (protospacer adjacent motif) sequences in the vicinity of the invading genome which determines where Cas9 will come and bind, and if there is a match between crRNA and viral gene fragment, Cas9 will be able to initiate the cleavage of viral DNA (Lone *et al.,* 2018). The fact that host genome doesn't have PAM sequences makes sure that Cas9 won't bind and cleave host DNA.

1.2. CRISPR Cas9 System

The ability of a nuclease to bind and cleave a target gene in a sequence-specific manner through RNA guidance is an excellent system which is worth exploiting, which scientists eventually did. Using an endonuclease which can be guided to a specific gene locus by a secondary structure of RNA, eventually named as single guide RNA (sgRNA), has opened new avenues in genomic targeting. By creating constructs containing a 20-nt sgRNA, a PAM sequence, and a Cas9 endonuclease, any gene locus can be targeted for the means of gene editing, gene regulation, gene imaging, or many other purposes.

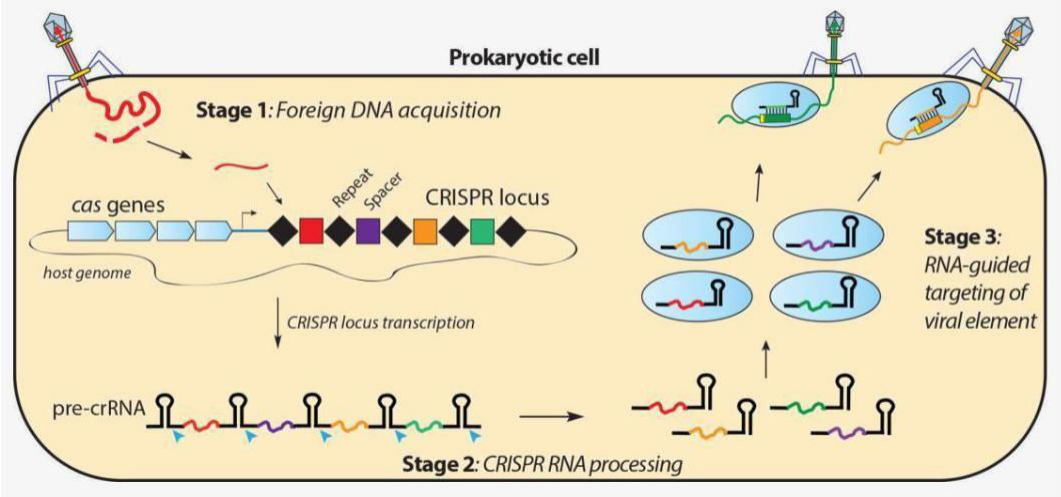


Figure 1.2 Stages of CRISPR system. CRISPR is an adaptive immunity mechanism comprised of three steps: acquisition, crisprRNA processing, and RNA-guided interference (adapted from Doudna, n.d.).

After a double strand break is introduced by Cas9 endonuclease, there are two mechanisms this break can be repaired in the cell; nonhomologous end joining (NHEJ) or homology-directed repair (HDR) (Figure 1.3) (Mali *et al.*, 2013). In nonhomologous end joining, while putting the broken DNA fragments back together, mutations in the form of either insertions or deletions are bound to happen; this case is well exploited in gene editing where a knockout version of a desired gene can easily be created by introducing a frameshift mutation.

For the case of homology-directed repair system, the addition of a donor DNA makes relatively more specific sequence alterations such as single-base mutations possible. HDR is more frequently resorted when a gene needs to be corrected since homologous recombination can make the correction if the donor DNA possessing the corrected version is supplied.

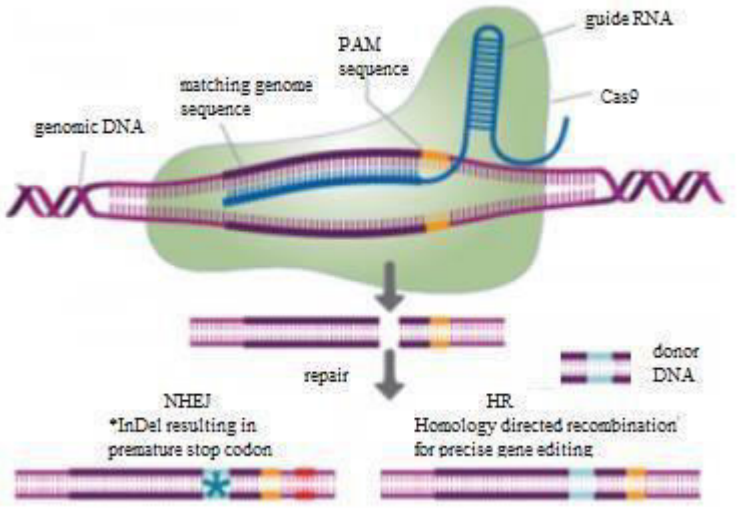


Figure 1.3 CRISPR mechanism. Cas9 will induce a double strand when there is a sequence match between genomic DNA and guide RNA nearby a PAM sequence; this break can be repaired by either nonhomologous end joining which introduces indels, or homology directed recombination which enables more precise gene alteration when a donor DNA is provided (adapted from CRISPR Genome Editing, n.d.).

1.3. Protein Structure of Cas9

There are various types of Cas9 proteins in numerous different microorganisms. The most popular one among others belongs to *Streptococcus pyogenes,* which is followed by *Staphylococcus aureus, Streptococcus thermophiles, Neisseria meningitides,* etc. All theseCas9 proteins have their specific PAM sequences that they are able to recognize, giving a wider choice for gene targeting. For instance, SpCas9 can recognize NGG (N being any) and SaCas9 can recognize NNGRRT (R being a purine base) sequences (Hu *et al.,* 2018). Crystal structures of SpCas9 in unbound, sgRNA-bound, and sgRNA-DNA-bound forms have shed light on specific locations where interactions between Cas9 and sgRNA, Cas9 and PAM sequence, and Cas9 and sgRNA-target DNA complexes occur.

SpCas9 is comprised of 2 lobes; Recognition (REC) and Nuclease (NUC) lobes (Jiang & Doudna, 2015). According to crystal structure of spCas9, single guide RNA seems to sit right between these two lobes (Figure 1.4). Recognition lobe is essential for the recognition of sgRNA-target DNA complex. Nuclease lobe possesses 4 different domains; PAM-interacting (PI), Wedge (WED), HNH nuclease, and RuvC nuclease domains. PI domain is where Cas9 fully interacts with the PAM sequence around the target DNA in a base-specific manner. Wedge domain is where Cas9 interacts with sgRNA and also the PAM sequence. HNH nuclease domain introduces a break at target strand where sgRNA recognizes, and RuvC nuclease domain introduces a break at the opposite strand; overall, the two domains result in a double strand break 3 nt downstream of the PAM sequence (Chen, Choi, & Bailey, 2014).

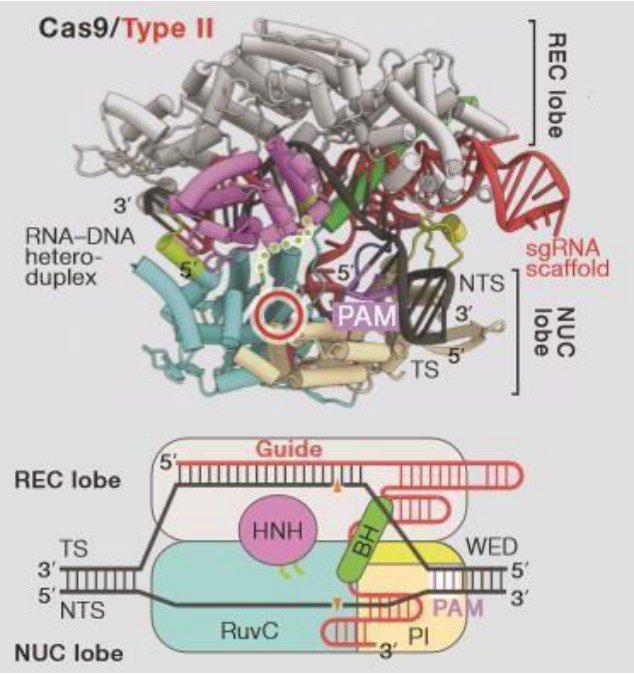


Figure 1.4 Crystal structure of spCas9 in complex with DNA-gRNA heteroduplex. Cas9 is comprised of recognition and nuclease lobes which are bound by a bridge helix (green). Its nuclease lobe possesses HNH nuclease, RuvC nuclease, Wedge, & PAM-interacting domains (adapted from Makarova, Zhang, & Koonin, 2017).

1.4. Exploiting CRISPR Cas9 System for Gene Editing

Since its discovery as a gene engineering tool in 2012, CRISPR Cas9 technology is one of the most studied topics in numerous research areas. By constructing a fused version of crRNAs and tracrRNAs (forming sgRNAs) which is complementary to a desired gene locus, guidance of Cas9 enzyme to any DNA region can be simulated in mammalian systems. The idea to use a bacterial adaptive immunity mechanism as a gene targeting tool has revolutionized science as we know it. Previously used gene targeting methods like TALEN and ZFN technologies have been replaced with a more practical and highly specific CRISPR technology. Most popular applications for CRISPR are gene editing, gene activation/repression, and genomic imaging (Figure 1.5).

One of the most popular applications of this technology is gene editing which can include generating knock-in/knock-out versions of desired genes, and/or making particular corrections on them. CRISPR has become an essential tool to create models for well-studied diseases. For this purpose, previously identified mutations known to cause certain disease phenotypes are introduced in corresponding genes to study the downstream effects of these genes and hopefully get some insight on molecular and cellular basis of disease mechanisms. By introducing multiple single guide RNAs, it is also possible to make double strand breaks at multiple locations and mimic certain disease phenotypes originating from particular chromosomal rearrangements.

If we can introduce mutations to mimic certain disease phenotypes, we can also reverse them by correcting faulty genes. For this approach, scientists are actively using CRISPR to correct mutations in animal germ and somatic cells, and also in induced pluripotent stem cells. For the purposes of gene therapy, mutant genes have been successfully corrected for diseases such as β-thalassemia, Duchenne muscular dystrophy (DMD), cystic fibrosis, etc (Wang, La Russa, & Qi, 2016). Another way of therapy has been made possible by CRISPR for diseases caused by HIV (human immunodeficiency virus), Hepatitis B virus or HPV (human papilloma

virus); by utilizing CRISPR Cas9, viral genome can be removed from host DNA (Hu *et al*., 2014; Liu, Hao, Chen, Guo, & Chen, 2015; Zhen *et al*., 2014).

CRISPR is a popular application in the area of reverse genetics, where the function(s) of a desired gene is studied by making a gene knock-out. When Cas9-mediated DNA cleavage is coupled with nonhomologous end joining, a frameshift mutation results in complete loss of function in the gene of interest. Moreover, by creating a sgRNA library, functions of thousands of genes can be studied with great ease. Compared to RNA interference technology which used to be popular for genome-wide functional analysis studies, CRISPR is more effective since it leads to complete function loss as opposed to RNAi which usually causes partial loss in gene function.

1.5. Exploiting CRISPR Cas9 System for Gene Activation and Gene Repression

The ability of CRISPR to target DNA regions with great specificity has opened new avenues for not just gene editing, but many other well-studied fields such as gene regulation and epigenome editing. For these purposes, a Cas9 variant lacking nuclease activity has been constructed; it is called deactivated or dead Cas9 (dCas9) (Qi *et al*., 2013). This Cas9 variant is utilized for its ability to reach specific target DNA regions through the guidance of sgRNAs. By fusing dCas9 to certain gene regulator proteins such as activators or repressors, regulation of desired genes is enabled with high specificity.

In order to induce gene activation, dCas9 is fused to a transcriptional activator domain such as VP64; with this technique, any desired gene can be activated (Perez-Pinera *et al*., 2013). It is also possible to downregulate a gene by fusing dCas9 to a repressor such as KRAB domain of Kox1 (zinc finger protein 10) and this construct is directed to the target region by sequence-specific guide RNAs (Thakore *et al.*, 2015). Epigenetic regulation can also be achieved through dCas9 coupling with epigenetic modifiers such as aceytlases and

demethylases. It has been shown that when dCas9 is coupled to p300, the targeted DNA region can be acetylated at H3K27 (Hilton *et al*., 2015).

1.6. Exploiting CRISPR Cas9 System for DNA Imaging

For the purposes of proper intervention on the genome, its three-dimensional structure has to be elucidated. Before the advent of CRISPR technology, DNA imaging used to be performed utilizing either DNA-DNA or protein-DNA interactions. To give brief examples to some early imaging techniques, fluorescently-labelled in situ hybridization (FISH) method uses fluorescently tagged DNA probes to bind desired gene loci in a sequence-specific manner (Langer-Safer, Levine, & Ward, 1982). Another technique utilizes transcription activator-like effectors (TALEs) or Zinc-fingers (ZFs) fused with fluorescent proteins for genomic imaging purposes (Ma, Reyes-Gutierrez, & Pederson, 2013). The rise of CRISPR technology has made the utilization of both DNA-DNA and protein-DNA interactions for genomic imaging possible. With the coupling of dCas9 with a fluorescent protein, specifically designed guide RNAs can direct dCas9-fluoroscent protein complex to target DNA; in this way, multiple genomic regions can be imaged by simultaneous introduction of multiple sgRNAs and dCas9s fused with different fluorescent proteins (Chen *et al*., 2013).

1.7. Challenges in Exploiting CRISPR Cas9 System

Although CRISPR technology is by far the most practical and efficient way of gene targeting, unfortunately it has its own flaws which make the jobs of scientists challenging. First of all, there is a chance of off-target binding which can lead to gene alterations at unwanted regions. In order to decrease this chance, sgRNA design is the key step; sgRNA constructs should be unique to the target region and sensitive to any kind of mismatches. Another way to heighten target specificity is using two sets of Cas9 mutants called Cas9 nickases (nCas9) guided by two different guide RNAs. In this application, one of the Cas9 proteins is mutated at the active site of its HNH domain, and the other one is mutated at the

active site of its RuvC domain, so they can each cleave a single strand (Ran *et al*., 2013). This technique slightly resembles TALEN and ZFN technologies, but in this case, two sgRNAs direct two Cas9 nickases to target region, making it far more specific compared to not just TALENs or ZFNs, but also conventional CRISPR Cas9 system.

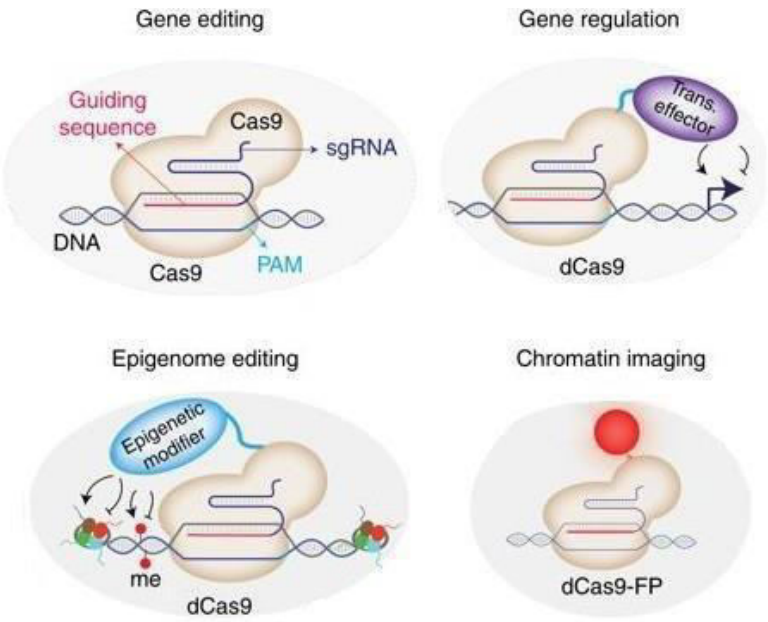


Figure 1.5 Applications of CRISPR. CRISPR Cas9 system can be utilized as a genome targeting tool for not just gene editing purposes, but also various other applications such as gene activation/repression, epigenetic modification, or genomic imaging studies (adapted from Adli, 2018).

There is another problem regarding the targeting of desired genes. Since not all genes may possess an NGG sequence nearby, it is difficult to adapt CRISPR to all desired genes. In order to increase the options of successful targeting, Cas9 proteins from various other organisms such as *Staphylococcus aureus, Streptococcus thermophiles, Neisseriameningitides,* and *Actinomyces naeslundii,* all of which recognizing distinct PAM sequences can be used, making the span of CRISPR broader for gene targeting (Esvelt *et al.*, 2013; Ran *et al*., 2015).

The best way to cope with the challenges in CRISPR is to get more insight on how this intricate system can be regulated, so that we can achieve gene targeting with higher specificity and greater efficiency. This regulation can be either at transcriptional level or post-translational level. For the purposes of transcriptional regulation, there are chemical- or light-induced systems available which give scientists a chance to control Cas9 expression (Richter *et al*., 2016). As for post-translational regulation, a study showed that when intein is fused with Cas9, it inhibits Cas9 nuclease activity inside the cell and when intein gets spliced through chemical induction, fully active Cas9 can work freely (Truong *etal*., 2015). This can be a valuable method to modulate the initiation of Cas9 cleavage activity. Moreover, as a novel approach to regulate Cas9 activity, in 2017, Tu *et al*. showed that when Cas9 is fused with ubiquitin protein, it can be directed to proteosomal degradation path. Another group achieved a similar result when they fused Cas9 to a protein called geminin, which is a known target of ubiquitin-proteosome machinery (Gutschner, Haemmerle, Genovese, Draetta, & Chin, 2016). Apart from these advances, there isn't much input on how CRISPR system can be regulated through post-translational modifications. In fact, there is no study which showed any post-translational modification on Cas9 protein in mammalian systems. To be able to show a post-translational modification on the major player of CRISPR technology would be a serious advance for comprehending CRISPR Cas9 system more deeply and getting a chance to create more specific and efficient Cas9 variants and highly contribute to CRISPR technology.

1.8. Post-Translational Modifications

Throughout the evolutionary process, from only a limited set of genes, a vast diversity of proteins with high level of complexity and numerous functions has been able to originate thanks to certain mechanisms that have also evolved along with everything else. In order to extend the variety of the transcriptome (and thus the proteome), there is alternative splicing, which takes place after the gene is transcribed and as a result, different combinations of exons from a single gene lead to different mRNA products. Different promoter usage and mRNA editing have also been known to serve the same purpose: to increase proteome diversity. These

mechanisms have enabled the generation of around 100,000 transcriptome from 21,000 protein-coding genes in humans.

In order to further increase the diversity of the proteome, there are post-translational modifications (PTMs). These processes take place after the mRNA is translated into its corresponding protein product. The modification of a protein usually occurs by the covalent attachment of either a small protein or a chemical group. Besides this, proteins can also get trimmed by special enzymes, usually to achieve their active, functional forms inside the cell. The major PTMs which have been studied extensively are phosphorylation, methylation, acetylation, glycosylation, ubiquitination, neddylation, and sumoylation. The most important role of PTMs in cells is to regulate protein structure, enzymatic activity, stability, subcellular localization, and interactions with other proteins. The fact that many of these modifications are reversible favors the dynamic feature of the proteome, making it easier to create a fast response inside the cell.

1.9. Sumoylation

Covalent attachment of a small protein called SUMO to a target protein is known as the sumoylation event. SUMO, which stands for Small Ubiquitin-like Modifier, is a relatively small protein with a size of 13 kDa. As the name suggests, SUMO shares a structural identity with the ubiquitin protein. Yet, it happens to diverge from ubiquitin in terms of sequence identity and function. SUMO protein has 4 isoforms; SUMO1, SUMO2, SUMO3, and SUMO4. Because SUMO2 and SUMO3 share 97% identity, they are referred to as SUMO2/3. On the other hand, SUMO1 protein shares 50% identity with SUMO2/3; it has a long peptide extension at its N-terminus, which possibly contributes to its diverse effects on target proteins (Figure 1.6) (Citro, 2013). There are various ways sumoylation can have an impact on a target protein. When covalently attached, SUMO proteins can alter a protein's structure, function, stability, solubility, cellular sub-localization, and interactions with partner proteins. For instance, when a protein gets sumoylated, it gets a chance of interacting with many other proteins if these proteins contain specific motifs called SUMO-interacting motifs (SIMs) (Han,

Feng, Gu, Li, & Chen, 2018). On an important note, SUMO proteins are only expressed in eukaryotic organisms, making sumoylation a eukaryotic modification. Smt3 is a SUMO-1 paralog found in yeast and its conjugation is so far implicated in DNA metabolism, transcription and nuclear organization (Newman *et al*., 2017).

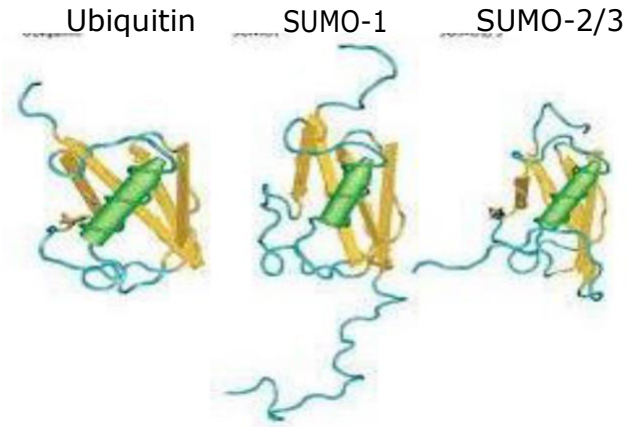


Figure 1.6 Protein structure of SUMO proteins. Small ubiquitin-like modifier (SUMO) proteins share structural identity with ubiquitin protein. SUMO2 and SUMO3 share 97% identity, whereas SUMO1 shares only 50% identity with SUMO2 and SUMO3 (adapted from Martin, Wilkinson, Nishimune, & Henley, 2007).

A significant feature of a protein targeted for sumoylation is the possession of at least one consensus sumoylation motif on its surface. This Ψ-K-x-D/E motif consists of a hydrophobic residue (Ψ), a lysine residue (K), any residue (x), and finally an acidic residue (D/E) in that order (Geiss-Friedlander & Melchior, 2007). After the motif is recognized by the SUMO-conjugating enzyme Ubc9, SUMO protein is covalently attached to the lysine residue on this motif.

There are three ways substrates can get sumoylated; they can be monosumoylated, multisumoylated (sumoylation on multiple motifs), or polysumoylated, in which case a poly-SUMO chain is conjugated to the target protein. In fact, SUMO2/3 proteins possess a

consensus sumoylation motif of their own, therefore only they have the ability to form poly-SUMO chains on target proteins (Ahner, Gong, & Frizzell, 2016).

Another similarity of sumoylation with ubiquitination is the mechanism of this modification. Just like the ubiquitin pathway, the SUMO pathway also comprises of E1, E2, and E3 enzymes that carry out the covalent attachment on the substrate (Figure 1.7). Before a target protein gets sumoylated, the precursor SUMO protein needs to be cleaved by special proteases called sentrin-specific proteases (SENPs), thus revealing a di-glycine motif at the C-terminus (Flotho & Melchior, 2013). Heterodimeric SUMO-activating E1 enzyme Aos1/Uba2, then activates SUMO through adenylation in an ATP-dependent manner and forms a thioester link with SUMO. This thioester link is transferred to SUMO-conjugating E2 enzyme, Ubc9 which mediates the SUMO conjugation of the target protein on the lysine residue of its sumoylation motif. SUMO-conjugated proteins can also get de-sumoylated by SENPs, making this a reversible modification (Di Bacco *et al*., 2006).

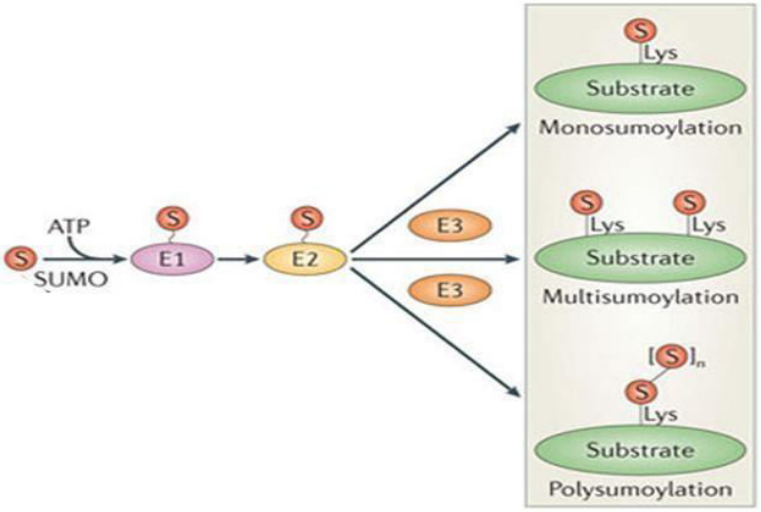


Figure 1.7 SUMO mechanism. Sumoylation is a reversible modification carried out by E1, E2, and E3 enzymes. As a result of this modification, substrates can get monosumoylated, multisumoylated, or polysumoylated (adapted from Hickey, Wilson, & Hochstrasser, 2012).

One major difference between SUMO and ubiquitin pathways is that in the ubiquitination machinery, substrate specificity is well-established through several substrate-specific E2 enzymes and hundreds of different E3 enzymes, whereas SUMO pathway comprises of a single universal E2 ligase that directly interacts with the sumoylation motif on the substrate and most of the time carries out the covalent attachment of SUMO to the substrate by itself, and a few E3 ligases which are not always required for conjugation (unlike the ubiquitin E3 ligases). SUMO E3 ligases only facilitate the reaction by positioning the SUMO-E2 ligase complex and the substrate by interacting with both of them, and in some cases, the presence of a SUMO E3 ligase favors modification by a certain SUMO isoform over other SUMOs (Pichler, Gast, Seeler, Dejean, & Melchior, 2002).

1.10. SUMO As A Major Regulator

SUMO conjugation has fundamental regulatory roles in numerous cellular processes such as nuclear import, DNA repair, transcription control, and cell cycle. In the past 2 decades, numerous proteins taking part in important cellular mechanisms are found out to be sumoylated. One example to this is the major player of RNA interference, Argonaute-2 protein. In 2014, Sahin *et al*. showed SUMO modification on Ago-2 protein at K402 position, and that this modification actually decreased the stability of this protein. In 2015, Josa-Prado *et al*. also showed that Ago-2 is sumoylated, plus this modification enhanced the RNA interference activity of Ago-2. After all, showing a novel post-translational modification on a protein carrying out a fundamental cellular mechanism is a big step on elucidating the intricate regulation system of a well-studied protein, thus a well-studied cell mechanism.

p53, another well-studied protein, is a tumor suppressor which is known to become active when DNA damage presents itself. When activated, p53 protein has an impact on cell cycle through either the induction of growth arrest or apoptosis in order to short circuit the genomic instability problem (Jin & Levine, 2001). In 1999, Rodriguez *et al*. showed that SUMO1 modification regulated p53 in a way that it enhanced p53-dependent transcription activity in cells. In 2000, another study from Muller *et al*. showed that p53 sumoylation by

SUMO1 impaired its impact on apoptosis induction. In conclusion, these and many other studies imply that SUMO machinery is a prominent regulator of p53 protein.

Another known target of SUMO proteins is histone proteins which are important elements in gene regulation. It has been shown that H4 gets sumoylated and this modification leads to gene silencing through the recruitment of histone deacetylase HDAC1 and heterochromatin protein 1 (HP1), both of which are known to cause gene repression (Shiio & Eisenman, 2003).

Apart from popular proteins and the cellular machinery they operate, there are many disease phenotypes which are heavily influenced by the downstream effects of SUMO regulation. In numerous cases, the disease mechanisms are able to successfully turn this in their favor. A study has shown that *Listeria monocytogenes* disrupted SUMO machinery to replicate and propagate inside the host by inducing the degradation of SUMO E2 ligase, Ubc9 through a toxin named LLO, thereby impairing global sumoylation in the cell (Koronakis & Hume, 2010). In another study, it has been shown that, in Ehrlichia chaffeensis, an intracellular bacterium known to cause tick-borne zoonosis, an effector protein called T1S TRP120 (type 1 secretion tandem repeat protein 120) got sumoylated after the infection of the host, and this modification enhanced the survival of the organism through the interaction of TRP120 with various host proteins possessing SUMO-interacting motifs (SIMs) (Dunphy *etal.*, 2014). It is the first study that showed a bacterial protein getting sumoylated.

Bacterial proteins are not the only non-eukaryotic proteins that can be conjugated with SUMO inside the host. In 2006, Nasr *et al*. for the first time showed the sumoylation of Tax, a viral oncoprotein which is the major player in ATL (adult T-cell leukemia), and that this modification lead to enhanced NFκB activation, important step for T-cell transformation. Tax isn't the only viral protein that gets sumoylated inside its host. BZLF1 and rta proteins of Epstein-Barr virus are also some of the targets of SUMO machinery; in fact, BZLF1 sumoylation interferes with transition to lytic phase and rta sumoylation induces viral replication (Murata *et al*., 2010; Chang *et al*., 2008). Another example is L2 protein, which

belongs to human papilloma virus (HPV) and it also undergoes SUMO modification which enhances viral assembly inside the host (Marusic *et al*., 2010).

Based on all these information presented so far, it is safe to say that first of all, sumoylation is an important type of post-translational modification which takes part in diverse cellular mechanisms and regulates these mechanisms in eukaryotic systems; and second of all, non-eukaryotic (bacterial or viral) proteins have been shown to undergo sumoylation inside the host cells in multiple studies. In the light of these facts, our lab proposed that when it is expressed in mammalian systems, Cas9 protein (of *Streptococcus pyogenes*) could be modified by SUMO proteins.

In silico studies performed by current lab member, Yusuf Tunahan Abacı have revealed that spCas9 has 10 consensus sumoylation motifs all of which on the surface of the protein. After finding this out, he performed immunoprecipitation experiments to show whether Cas9 is modified by SUMO proteins. As a result of these experiments, he showed for the first time that Cas9 is in fact conjugated with both SUMO1 and SUMO2 proteins (Figure 1.8). This is the first time any post-translational modification is shown on Cas9 protein and it is an important step towards elucidating a novel regulatory mechanism on Cas9 protein.

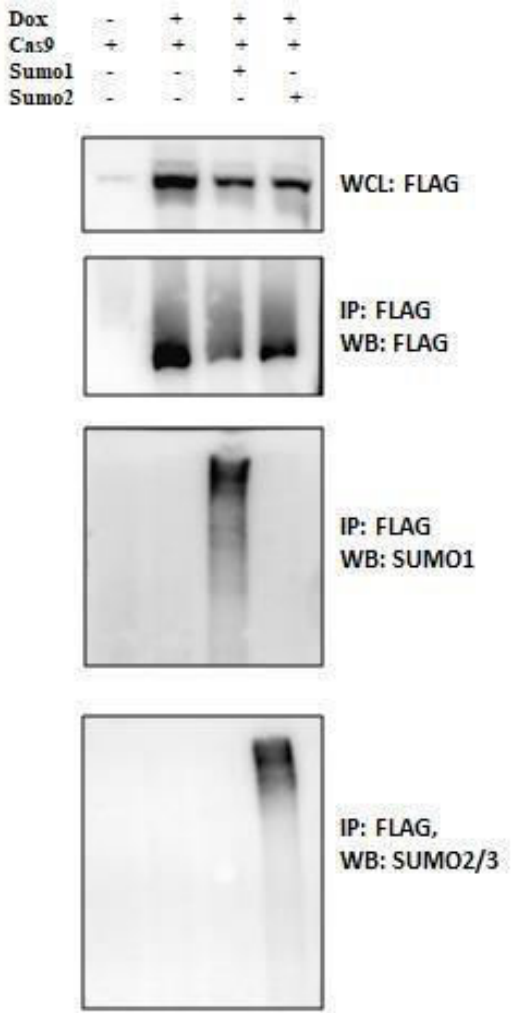


Figure 1.8 CRISPR Cas9 protein is sumoylated. Immunoprecipitation experiments persistently showed that when FLAG-Cas9 and SUMO1/SUMO2 are overexpressed in HEK293T cell line, Cas9 is conjugated with SUMO1 and SUMO2. FLAG-Cas9 vector is doxycycline-inducible (adapted from Abacı, 2017, unpublished).

2. AIM OF THE STUDY

In order to discover whether sumoylation has any regulatory role on CRISPR Cas9 protein, first objective was to create a sumoylation-defective Cas9 mutant construct. For that purpose, for each sumoylation site, single lysine-to-arginine mutations on pCW-Cas9 plasmid were made. In addition to those, a Cas9 mutant which possesses lysine-to-arginine mutations on all 10 sumoylation sites was generated.

Our next objective was to find out which sumoylation site is the major contributor on Cas9 sumoylation. For that purpose, His pulldown experiments were performed by overexpressing Cas9 wild type and the sumoylation-defective mutants along with His-tagged SUMO1/SUMO2 in order to show the potential abolishment of sumoylation for the mutant versions of Cas9.

Final objective was to show the interaction between Cas9 protein and the universal SUMO E2 ligase, Ubc9 by co-immunoprecipitation experiment.

3. MATERIALS AND METHODS

3.1. Equipment

The equipment used in the experiments of this project are listed in Table 3.1.

Table 3.1 Equipment.

Equipment	**Supplier**
Agarose Gel Electrophoresis System	EASY-CAST, ThermoFisher, USA
Autoclaves	Midas 55, Prior Clave, UK ASB260T, Astell, UK
Carbon dioxide Tank	Genç Karbon, Turkey
Cell Culture Dishes	TPP, Switzerland
Cell Culture Incubator	WTC, Binder, Germany
Cell Scraper	TPP, Switzerland
Centrifuges	Ultracentrifuge J2MC, Beckman, USA VWR CT15RE, Japan Allegra X-22, Beckman USA
Centrifuge Tubes	CAPP, Denmark
Cold room	Birikim Elektrik Soğutma, Turkey
Cryovial Tubes (2ml)	CAPP, Denmark

Table 3.1 Equipment (cont.).

Equipment	**Supplier**
Documentation System	Gel Doc XR System, Bio-Doc, ITALY Stella, Raytest, Germany G-BOX Chemi XX6, Syngene, UK
Deepfreezers	(-20) Ugur, UFR 370 SD, Turkey (-80) ULT deep freezer, Thermo, UK (-150) Sanyo MDF-1156, UK
Dish Washer	Mielabor G7783, Miele, Germany
Gel Electrophoresis Equipment	Mini-Protean III Cell, Bio-Rad, USA
Heat blocks	Block heater analog, VWR, USA
Ice Machine	Scotsman Inc. AF20, Italy
Inverted Microscope	Z1 Axio Observer, Zeiss, USA
Bottles	VWR, USA
Laminal Flow Cabinet	Class II B, Tezsan, Turkey
Magnetic Stirrer	VMS-C7,VWR, USA
Microfuge Tubes	CAPP, Denmark
Micropipettes	Finnpipette, Thermo, USA
Micropipette Tips	Axygen, USA

Table 3.1: Equipment (cont.).

Equipment	**Supplier**
Microscopes	Inverted Microscope, Nikon, Eclipse TS100, Netherlands Fluoroscence Microscope, Observer.Z1, Zeiss, Germany
Microwave Oven	Arçelik, Turkey
Multiwell Plates	TPP, Switzerland
Nitrocellulose Blotting Membrane (0.2 µm)	Amersham Protran, GE Healthcare Life Sciences
Oven	Gallenkamp 300, UK
PCR Tubes (0.2ml)	Axygen, USA
Petri Dishes	Fırat Plastik, Turkey
pH Meter	Hanna Instruments, USA
Pipettor	S1 pipet filler, Thermo Fisher, USA
Pippette Tips (Bulk)	CAPP, Denmark
Pippette Tips (Filtered)	BioPointe, USA
Power Supplies	EC XL 300, Thermo Fisher, USA Bio-Rad, USA

Table 3.1 Equipment (cont.).

Equipment	**Supplier**
Softwares	Quantity One, Bio-Rad, ITALY ImageJ, Image Analysis Software, NIH, USA XStella 1.0, Stella, GERMANY FlowJo, USA Syngene-Genetools, UK
Syringes	Set Medikal, Turkey
Syringe Filter Units	EMD Millipore, USA
Test Tubes	CAPP, Denmark
Thermal Cycler	Bio-Rad, USA
Vortex	VWR, USA
Water Purification	WA-TECH UP, Germany
Water Purification System	UTES, TURKEY
Watmann Filter Paper-Extra Thick	Thermo Scientific, USA

3.2. Cell Culture

HEK293 cells were kindly provided by Prof. Nurhan Özlü from Koç University.

3.3. Plasmids and Primers

Plasmids used in this project are listed in Table 3.2.

Table 3.2 Plasmids.

Construct	**Origin**	**Backbone**
pCW-FLAG-Cas9	Provided by Tolga Emre, Boğaziçi University	pCW
GFP	Provided by Nesrin Özören, Boğaziçi University	pEGFP C1
His-SUMO1	Provided by Hugues de The, College de France	Unknown
His-SUMO2	Provided by Hugues de The, College de France	Unknown
GFP-Ubc9	Provided by Hugues de The, College de France	Unknown
FLAG-Ago-2	Addgene #72207	pCMV5

Table 3.3 Sequences of primers.

Oligo ID	**Sequence (5'-3')**	**Application**
p.K91R-For.	GAGATGGCCAGGGTGGACGAC	SDM
p.K91R-Rev.	GTCGTCCACCCTGGCCATCTC	SDM
p.K564R-For.	GAAGCAGCTGAGAGAGGACTAC	SDM
p.K564R-Rev.	GTAGTCCTCTCTCAGCTGCTTC	SDM
p.K684R-For.	CTGGATTTCCTGAGGTCCGACGGC	SDM
p.K684R-Rev.	GCCGTCGGACCTCAGGAAATCCAG	SDM
p.K704R-For.	GCCTGACCTTTAGAGAGGACATC	SDM
p.K704R-Rev.	GATGTCCTCTCTAAAGGTCAGGC	SDM

Table 3.3 Sequences of primers (cont.)

Oligo ID	**Sequence (5'-3')**	**Application**
p.K847R-For.	GAGCTTTCTGAGGGACGACTCC	SDM
p.K847R-Rev.	GGAGTCGTCCCTCAGAAAGCTC	SDM
p.K1003R-For.	CAAAAAGTACCCTAGGCTGGAAAGCGAG	SDM
p.K1003R-Rev.	CTCGCTTTCCAGCCTAGGGTACTTTTTG	SDM
p.K1024R-For.	GGAAGATGATCGCCAGGAGCGAGCAGGAAAT C	SDM
p.K1024R-Rev.	GATTTCCTGCTCGCTCCTGGCGATCATCTTCC	SDM
p.K1047R-For.	GAACTTTTTCAGGACCGAGATTAC	SDM
p.K1047R-Rev.	GTAATCTCGGTCCTGAAAAAGTTC	SDM
p.K1148R-For.	GTGGTGGCCAGAGTGGAAAAG	SDM
p.K1148R-Rev.	CTTTTCCACTCTGGCCACCAC	SDM
p.K1191R-For.	CAAAGAAGTGAGAAAGGACCTG	SDM
p.K1191R-Rev.	CAGGTCCTTTCTCACTTCTTTG	SDM

3.4. General Kits, Enzymes and Chemicals

The following tables indicate the kits, enzymes and chemicals used in this study.

Table 3.4 Kits and enzymes.

Name	**Supplier**
DpnI restriction endonuclease	NEB, USA
ECL-Femto	Thermo, USA
ECL-Pico	Thermo, USA
Q5 High-Fidelity DNA Polymerase	NEB, USA
ZymoPURE^{TM} MaxiPrep Kits	ZymoResearch, USA
ZymoPURE^{TM} MidiPrep Kits	ZymoResearch, USA
ZymoPURE^{TM} Mini Prep Kits	ZymoResearch, USA

Table 3.5 Chemicals.

Chemical	**Supplier**
Acetic Acid	Sigma-Aldrich, USA
Acrylamide	Bio-Rad, USA
Agar	Sigma-Aldrich, USA
Agarose	Peqlab, USA
Ammonium Persulfate (APS)	AppliChem, Germany

Table 3.5 Chemicals (cont.).

Chemical	**Supplier**
Ampicillin	Sigma-Aldrich, USA
B-Mercaptoethanol	Merck, Germany
Bovine Serum Albumin (BSA)	Capricorn Scientific, Germany
Bromophenol Blue	Sigma-Aldrich, USA
Calcium chloride dehydrate	Sigma-Aldrich, USA
Complete Mini Protease Inhibitor Cocktail	Roche, Switzerland
Cycloheximide	Sigma-Aldrich, USA
DAPI	Sigma-Aldrich, USA
Disodium hydrogen phosphate	Merck, Germany
DMSO	Sigma-Aldrich, USA
DNA Ladder (1 kb)	NEB, USA
Dulbecco's Modified Eagle Medium	Gibco, Fisher Scientific, USA
EDTA	Wisent Bioproducts, Canada
Ethanol	Merck, Germany
Ethidium Bromide	Sigma-Aldrich, USA
Fetal Bovine Serum (FBS)	Gibco, Fisher Scientific, USA

Table 3.5 Chemicals (cont.).

Chemical	**Supplier**
Glycerol	MP Biomedicals, USA
Glycine	NeoFROXX, Germany
Guanidine Hydrochloride	NeoFROXX, Germany
HEPES Buffered Saline Solution (HBS)	Lonza, Switzerland
Hydrochloric Acid	Sigma-Aldrich, USA
Imidazole	Santa Cruz' USA
Kanamycin	Gold Biotechnology, USA
LB Broth	Caisson Laboratories, USA
Methanol	Merck, Germany
N-ethylmaleimide (NEM)	Sigma-Aldrich, USA
Nickel NTA Beads	QIAGEN, Germany
PageRuler Prestained Protein Ladder	Thermo, USA
Penicillin/Streptomycin (100X)	Lonza, Switzerland
Potassium chloride	Sigma-Aldrich, USA
Potassium dihydrogen phosphate	Merck, Germany
Sodium Chloride (NaCl)	Merck, Germany

Table 3.5 Chemicals (cont.).

Chemical	**Supplier**
Sodium Dodecyl Sulfate (SDS)	Merck, Germany
Sodium Hydroxide	Merck, Germany
TEMED	Sigma-Aldrich, USA
Tris-Base	Sigma-Aldrich, USA
Triton-X-100	VWR, USA
Trypsin 0.05% EDTA	Thermo Fischer, USA
Tween-20	Merck, Germany

3.5. Solutions and Buffers

The following table shows the solutions and buffers used in this project.

Table 3.6 Solutions and buffers.

Solution/Buffer	**Content**
Co-IP Buffer	150 mM NaCl 50 mM Tris-HCl pH 7.4 1 mM EDTA 1% Triton-X-100 Protease Inhibitor Cocktail
4X Laemmli Buffer	200mM TrisHCl pH 6.8 8% (w/v) SDS 40% (w/v) 100% Glycerol 4% (w/v) B-mercaptoethanol 50 mM EDTA 0.08% (w/v) Bromophenol Blue

Table 3.6 Solutions and buffers (cont.).

8% Resolving Gel	375 mM TrisHCl pH 8.8 0.1% (w/v) SDS Acrylamide:Bisacrylamide (8% w/v) 0.05% (w/v) APS 0.005% (w/v) TEMED
4% Stacking Gel	0.125 mM TrisHCl pH 6.8 0.1% (w/v) SDS Acrylamide:Bisacrylamide (4% w/v) 0.05% (w/v) APS 0.0075% (w/v) TEMED
10X SDS Running Buffer	1% (w/v) SDS 3% (w/v) Tris Base 14.4% (w/v) Glycine
10X Transfer Buffer	3% (w/v) Tris Base 14.4% (w/v) Glycine
Buffer A	5M Guanidine Hydrochloride 10mM Imidazole 0.1M $NaH_2PO_4*2H_2O/Na_2HPO_4*2H_2O$
Buffer TI	20mM Imidazole 25mM Tris-HCl pH 6.8
Buffer A/TI	Buffer A-Buffer TI ratio: 1:3

Table 3.6 Solutions and buffers (cont.).

Solution/Buffer	**Content**
1X Phosphate Buffered Saline (PBS)	137mM NaCl 2.7mM KCl 10mM Na_2HPO_4 1.8mM KH_2PO_4
1X Phosphate Buffered Saline with Tween-20 (PBS-T)	137mM NaCl 2.7mM KCl 10mM Na_2HPO_4 1.8mM KH_2PO_4 0.1% (w/v) Tween-20
Western Blot Blocking Solution	5% (w/v) skim milk in PBS-T
1^{st} Antibody Solution	5% (w/v) skim milk in PBS-T
2^{nd} Antibody Solution	5% (w/v) skim milk in PBS-T

3.6. Antibodies

The following table shows the list of antibodies used in this project.

Table 3.7 Antibodies.

Antibody	**Supplier**	**Source**	**Dilution**
Actin	Cell Signaling Technologies, USA	Rabbit	1:1000
Cas9	Cell Signaling Technologies, USA	Mouse	1:1000
FLAG	Sigma-Aldrich, USA	Mouse	1:1000
His	Santa Cruz Biotechnology, USA	Mouse	1:1000
Mouse IgG, HRP	Cell Signaling Technologies, USA	Goat	1:10000
Rabbit IgG, HRP	Cell Signaling Technologies, USA	Goat	1:10000
SUMO-1	Cell Signaling Technologies, USA	Rabbit	1:1000
SUMO-2	Abcam, United Kingdom	Rabbit	1:1000
Tubulin	Santa Cruz Biotechnology, USA	Mouse	1:1000

3.7. Cell Culture Techniques

3.7.1. Cell Maintenance

HEK293T cells were grown in complete DMEM with 10% FBS and incubated at 37°C in 5% CO_2. Cell passaging was done every 2 days.

3.7.2. Cell Passaging

Medium was discarded, and cells were washed with 1x PBS. Since HEK293T cells are adherent, they require trypsin and 3 min incubation at 37°C while on trypsin. Then, trypsin was deactivated by complete DMEM (cDMEM). Cells were split in either 1:5 or 1:6 ratio depending on their confluency, fresh cDMEM was given, and cells were maintained at 37°C in 5% CO_2.

3.7.3. Cell Seeding

For His pulldown experiments, 2.5 to 3 million cells were seeded on 10cm plates 36 hours prior to immunoprecipitation. For Cas9 expression studies, $5*10^5$ cells were seeded on 6-well plates 60 hours prior to whole cell lysis.

3.7.4. Transfection

HEK293T cells were transfected with calcium phosphate transfection method. DNA-dH_2O mixes were prepared in 2.0ml Eppendorf tubes. 2M $CaCl_2$ was added to DNA-dH_2O mix drop by drop. Then, 2x HEPES-buffered saline (HBS) was added drop by drop. The final mixture was mixed thoroughly by harsh pipetting with aeration and incubated for 10min at room temperature to let the DNA-$Ca_3(PO_4)_2$ complex to form. Then, the mixture was given to the cells slowly and the plates were swirled gently to spread the mix evenly. For His pulldown

experiments, transfection was conducted 24 hours prior to the experiment. For Cas9 expression studies, it was done 48 hours prior to whole cell lysis.

3.7.5. Doxycycline Treatment

Since Cas9 plasmid has a pCW promoter that is doxycycline-inducible, cells require doxycycline treatment in order to express Cas9 protein. The amount of doxycycline needed for more than sufficient Cas9 expression is 2mg/ml. For His pulldown experiments, dox treatment was done 8 hours after transfection, 16 hours prior to the experiment. For Cas9 expression studies, it was done 24 hours after transfection and 24 hours prior to whole cell lysis.

3.7.6. Whole Cell Lysis

Medium was discarded and HEK293T cells were washed with 1X PBS. PBS was discarded and cells were lysed with 2x Laemmli buffer and the plate was swirled slowly for a couple of minutes until the Laemmli buffer became viscous. The samples were collected in 1.5ml Eppendorf tubes and boiled at 95°C for 10min and stored at -20°C for further use.

3.8. Molecular Biological Techniques

3.8.1. His Pulldown Assay

Before collecting with 1x PBS, HEK293T cells were washed with 20mM NEM in PBS, then they were collected with 1.5ml 1x PBS and spinned down @3000g for 3mins. The pellets were resuspended in 500µl 1X PBS and 20µl was taken for whole cell lysate (WCL). 50µl 2x Laemmli buffer was added to WCL and they were boiled @95^0 C for 10mins. The rest of the cell suspension was spinned down @3000g for 3mins. After removing the supernatant, the pellet was lysed with 1ml Buffer A by pipetting and vortexing. The samples were sonicated and then spinned down @13000g for 30mins. 900µl of each supernatant was collected and mixed with 50µl Nickel NTA beads that were equilibrated with Buffer A, and the mixture was left for 3-hour incubation at room temperature.

After the incubation, the supernatant was discarded, and the beads were washed with Buffer A twice, then washed with Buffer A/TI twice, and finally washed with Buffer TI once. After getting rid of all liquid from the beads, the samples were eluted with 100µl 2x Laemmli buffer and boiled @95^0 C for 10mins.

3.8.2. Western Blot

Protein samples collected with 2x Laemmli Buffer were loaded in homemade polyacrylamide gel and run at 70V for 20 mins, then 100V for 2-3 hours. The gel was transferred to a nitrocellulose membrane at 100V for 3 hours. The membrane was blocked for

1 hour in 5% skim milk in PBS-T, and then it was blotted with primary antibody overnight at 4°C.

The next day, the membrane was washed with 1x PBS-T three times for 5 mins, then incubated with secondary antibody (specific to the primary antibody used) for 1 hour at room temperature. The membrane was washed with 1x PBS-T three times for 10 mins, and then visualized (with required HRP substrates) using GeneSys program.

3.8.3. Site-Directed Mutagenesis

Primers required for site-directed mutagenesis (SDM) were designed by using Primer3 Input Tool and Baker Laboratory Tools from the website of Washington University. For SDM experiments, two different protocols, one from Stratagene and one from NEB, were used. For both protocols, NEB's Q5 DNA Polymerase was used. The compositions of reaction mixtures are shown in Tables 3.8 and 3.9. PCR cycles and temperatures for both protocols are shown in Tables 3.10 and 3.11.

After PCR is complete, template was digested by DpnI. For 50ul reaction mix, 6ul of Cutsmart and 1ul of DpnI was added. The digestion is complete after 3 hours incubation @37^0 C.

Table 3.8 Stratagene's reaction mixture.

Reagent	**Volume (µl)**
5x Q5 reaction buffer	10
DMSO	3
DNA template (50-100 ng)	X
dNTP	1
Forward primer (100-150 ng)	X
Reverse primer (100-150 ng)	X
Q5 DNA Polymerase	1
ddH_2O	X

Table 3.9 NEB's reaction mixture.

Reagent	**Volume (µl)**
5x Q5 reaction buffer	10
DMSO	3
DNA template (300-500 ng)	X
dNTP	1
Forward primer (100-150 ng)	X
Reverse primer (100-150 ng)	X
Q5 DNA Polymerase	1
ddH_2O	X
Total	50

Table 3.10 Stratagene's PCR cycle protocol.

Temperature	**Time**	
95 ^{o}C	3 min	
95 ^{o}C	30 sec	x15
55-72 ^{o}C	30 sec	
72 ^{o}C	6 min	
72 ^{o}C	5 min	
4 ^{o}C	∞	

Table 3.11 NEB's PCR cycle protocol.

Temperature	Time	
98 °C	30 sec	
98 °C	10 sec	x25
55-72 °C	30 sec	
72 °C	6 min	
72 °C	2 min	
4 °C	∞	

3.8.4. Agarose Gel Electrophoresis

1% agarose gel (w/v) was prepared with 1X TAE buffer and 30 ng/ml ethidium bromide. By making the final concentration of loading dye 1x, DNA samples were prepared and loaded in the gel along with DNA markers (100bp or 1kb). The gel was run at 90 V for 30 min, then it was visualized under the UV transilluminator.

3.8.5. Co-immunoprecipitation

HEK293 cells were grown in 10cm plates. 40ul agarose beads were incubated at room temperature with 2ul FLAG antibody for 1 hour. 48 hours after transfection, cells were lysed with 1ml co-IP buffer and collected. Samples were centrifuged at 13,000rpm for 30mins. Supernatant is incubated with antibody-bead mix at 4°C overnight.

The next day, beads were washed with 1x PBS three times. 60ul 2x Laemmli buffer is added and samples were boiled at 95°C for 10mins.

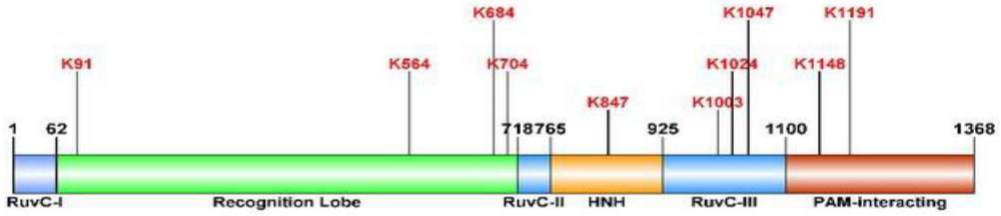
4. RESULTS

In silico studies had previously shown 10 sumoylation consensus motifs (ΨKxD/E) on the surface of SpCas9 protein; 4 of which in the Recognition lobe and 6 in the Nuclease lobe. In the Nuclease lobe, 1 motif is in the HNH domain, 3 in the RuvC domain, and 2 in the PAM-interacting domain.

a SpCas9 Wild Type Amino Acid Sequence (1368 AA)

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSN
EMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV
KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY
SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK*

b



c

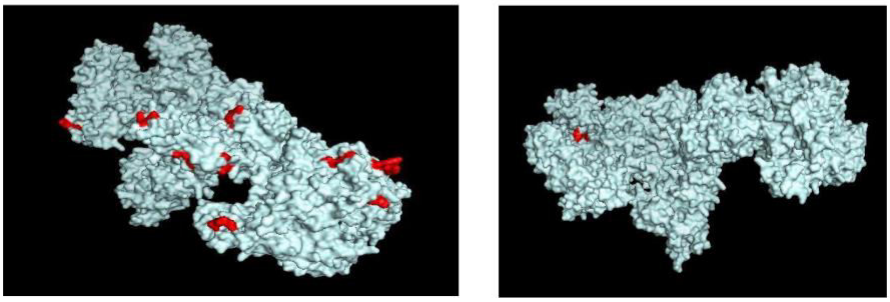


Figure 4.1 SpCas9 has 10 sumoylation consensus motifs. a. Full amino acid sequence of spCas9 with 10 sumoylations motifs shown in red. b. SpCas9 protein domains with 10 sumoylation motifs shown in red. c. Three-dimensional structure of spCas9 with 10 sumoylation motifs shown in red. The left image shows 9 of these motifs while the right image shows the remaining 1 motif.

4.1. Generation of Sumoylation-defective Cas9 Mutants

In order to find out which sumoylation site or sites contribute to Cas9 sumoylation, for each sumoylation motif, the lysine (K) residue was mutated to arginine (R) by performing site-directed mutagenesis. As a result of these experiments, 10 mutant constructs, each possessing a single K-to-R mutation at each sumoylation site, were generated. After amplifying these mutant constructs with PCR, the mutations were verified by sequencing.

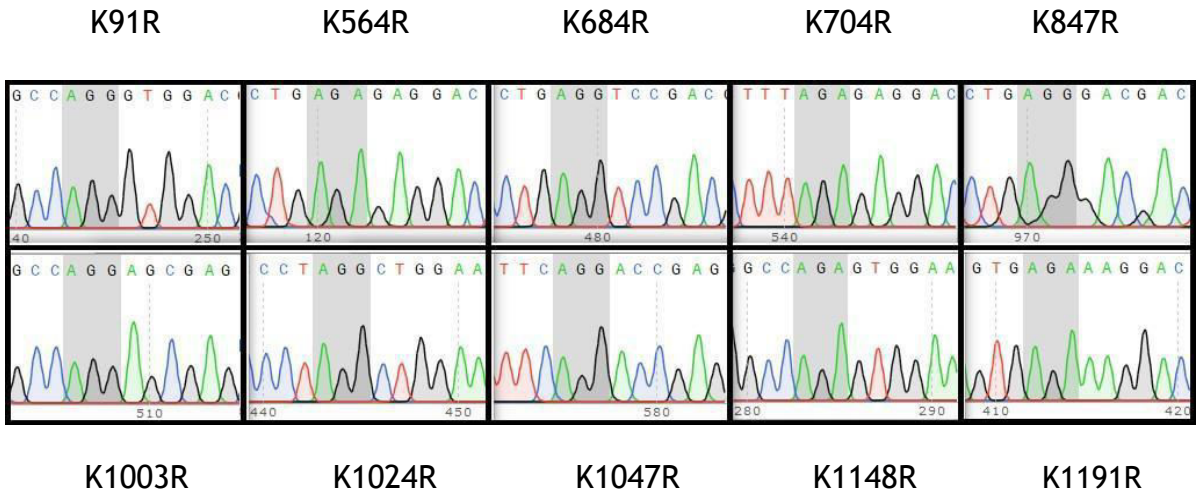


Figure 4.2 Validation of Cas9 mutants. Sequences of 10 Cas9 mutants where lysine residues were mutated into arginine at each sumoylation site.

In addition to these mutant constructs, again by site-directed mutagenesis, 10KR Cas9 mutant, which possesses 10 K-to-R mutations at 10 sumoylation motifs was created. After the PCR amplification step, each mutation on the construct was verified one at a time by sequencing. The lysine residues that were mutated are shown in Figure 4.3 and Table 4.1.

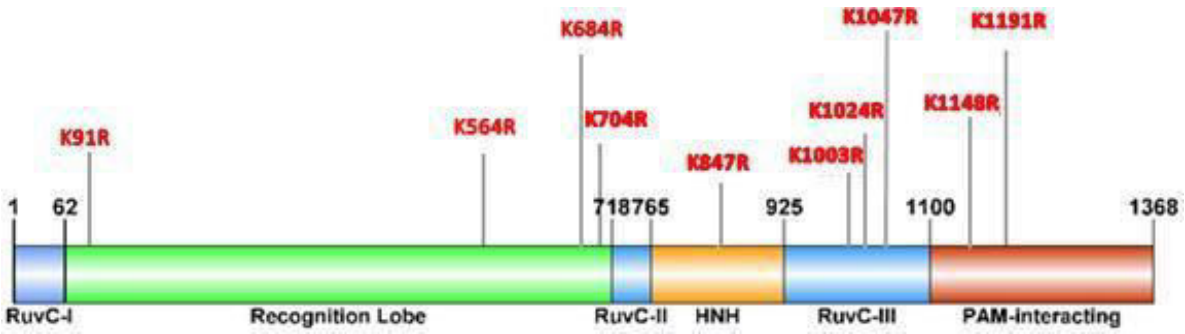


Figure 4.3 Illustration of 10KR mutant of Cas9. 10KR mutant was created by mutating the lysine residues to arginine in all 10 sumoylation motifs on Cas9 protein. Those 10 sumoylation motifs are depicted in red.

Table 4.1 List of the lysine residues mutated on 10 sumoylation sites of Cas9.

Site Name	Amino Acid Change	Site Name	Amino Acid Change
Site 1	K-91-R	Site 6	K-1003-R
Site 2	K-564-R	Site 7	K-1024-R
Site 3	K-684-R	Site 8	K-1047-R
Site 4	K-704-R	Site 9	K-1148-R
Site 5	K-847-R	Site 10	K-1191-R

After successfully generating 11 Cas9 mutants, each construct was transfected to HEK293T cells to check their expression levels. Figure 4.4 shows the expression of all 11 Cas9 mutants along with wild type Cas9.

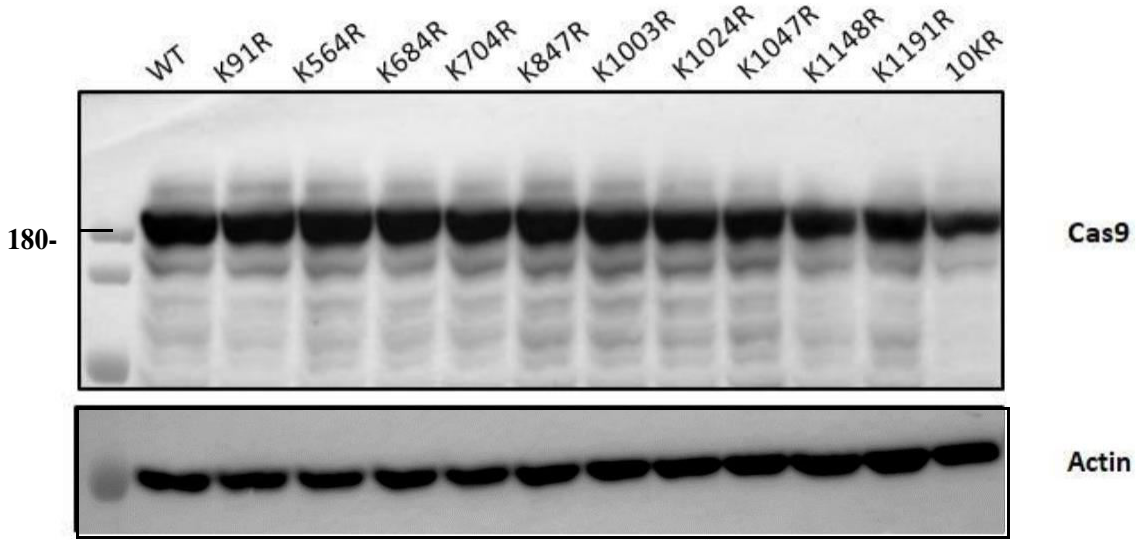


Figure 4.4 Cas9 mutants are properly expressed in HEK293T cells. Western blot experiment depicts the expression levels of WT FLAG-Cas9 and 11 FLAG-Cas9 mutants with their loading control. FLAG and Actin antibodies were used.

4.2. SUMO-1 Conjugation Levels of Wild Type Cas9 & 10KR Cas9 Mutant

Before determining which site or sites of Cas9 are the major sites of sumoylation, we wanted to show the complete abolishment of sumoylation in 10KR Cas9 mutant by performing His pull-down assay. In this assay, we transfected His-tagged SUMO-1 along with FLAG-tagged Cas9, then all proteins conjugated with His-tagged SUMO-1 were pulled down by Nickel NTA beads. In order to detect SUMO-1-conjugated Cas9, we performed Western blot using FLAG antibody. With this method, we were able to pull down His-SUMO-1-conjugated Cas9 proteins. Figure 4.5 depicts two FLAG blots and

one SUMO-1 blot which correspond to Cas9 expression in whole cell lysate, His-SUMO-1-conjugated Cas9 and all His-SUMO-1-conjugated proteins respectively.

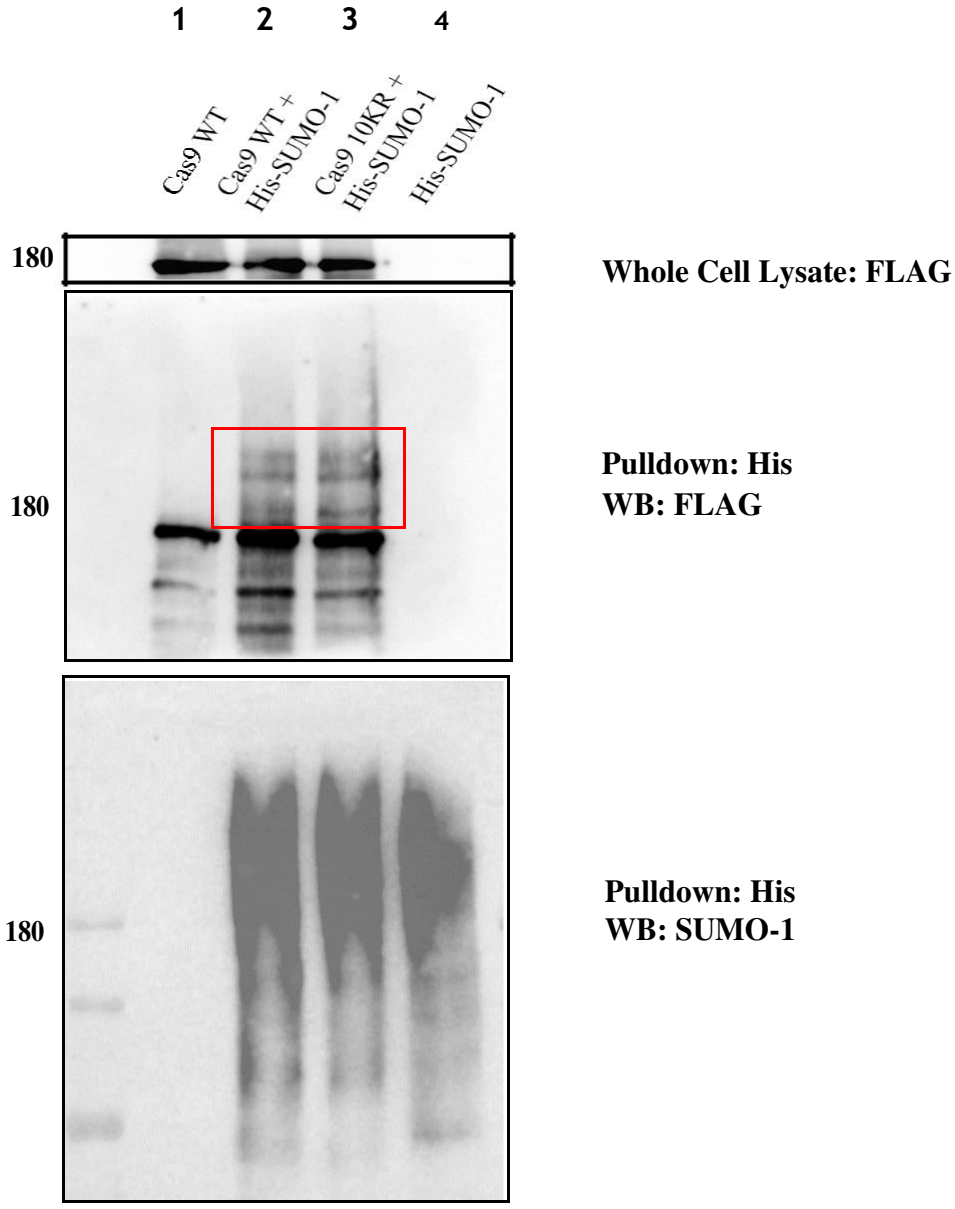


Figure 4.5 His pulldown assay for the detection of SUMO-1-conjugated Cas9. Top figure is Cas9 expression in whole cell lysate. Middle figure shows Cas9 proteins conjugated with His-SUMO-1. 3 bands framed by the red box in Lanes 2 and 3 depict His-SUMO-1-conjugated Cas9. Single band below 180 kDa in Lanes 1, 2, and 3 corresponds to unmodified Cas9 bound to beads non-specifically. Bottom figure represents the pulldown control.

In Figure 4.5, in FLAG blot, we observed 3 separate bands (shown inside the red box) in Lanes 2 and 3 which correspond to sumoylated Cas9 proteins. Lanes 2 and 3 show similar levels of SUMO-1-conjugated Cas9. We do not observe these 3 bands in Lanes 1 and 4, which correspond to lysates collected from only Cas9- and only His-SUMO-1- transfected cells; this indicates that those 3 bands are in fact Cas9 proteins conjugated to His-tagged SUMO-1. In FLAG blot in the middle panel, the smear of bands below the non-specifically bound Cas9 band is the degradation products of Cas9 proteins which are absent in only His-SUMO-1-transfected sample in Lane 4.

4.3. SUMO-2 Conjugation Levels of Wild Type Cas9 & 10KR Cas9 Mutant

In this assay, instead of His-SUMO-1, we transfected His-tagged SUMO-2 along with FLAG-tagged Cas9, then all proteins conjugated with His-tagged SUMO-2 were pulled down by Nickel NTA beads. In order to detect SUMO-2-conjugated Cas9, we performed Western blot using FLAG antibody. With this method, we were able to pull down His-SUMO-2-conjugated Cas9 proteins. Figure 4.6 depicts two FLAG blots and one SUMO-2 blot which correspond to Cas9 expression in whole cell lysate, His-SUMO-2-conjugated Cas9 and all His-SUMO-2-conjugated proteins respectively.

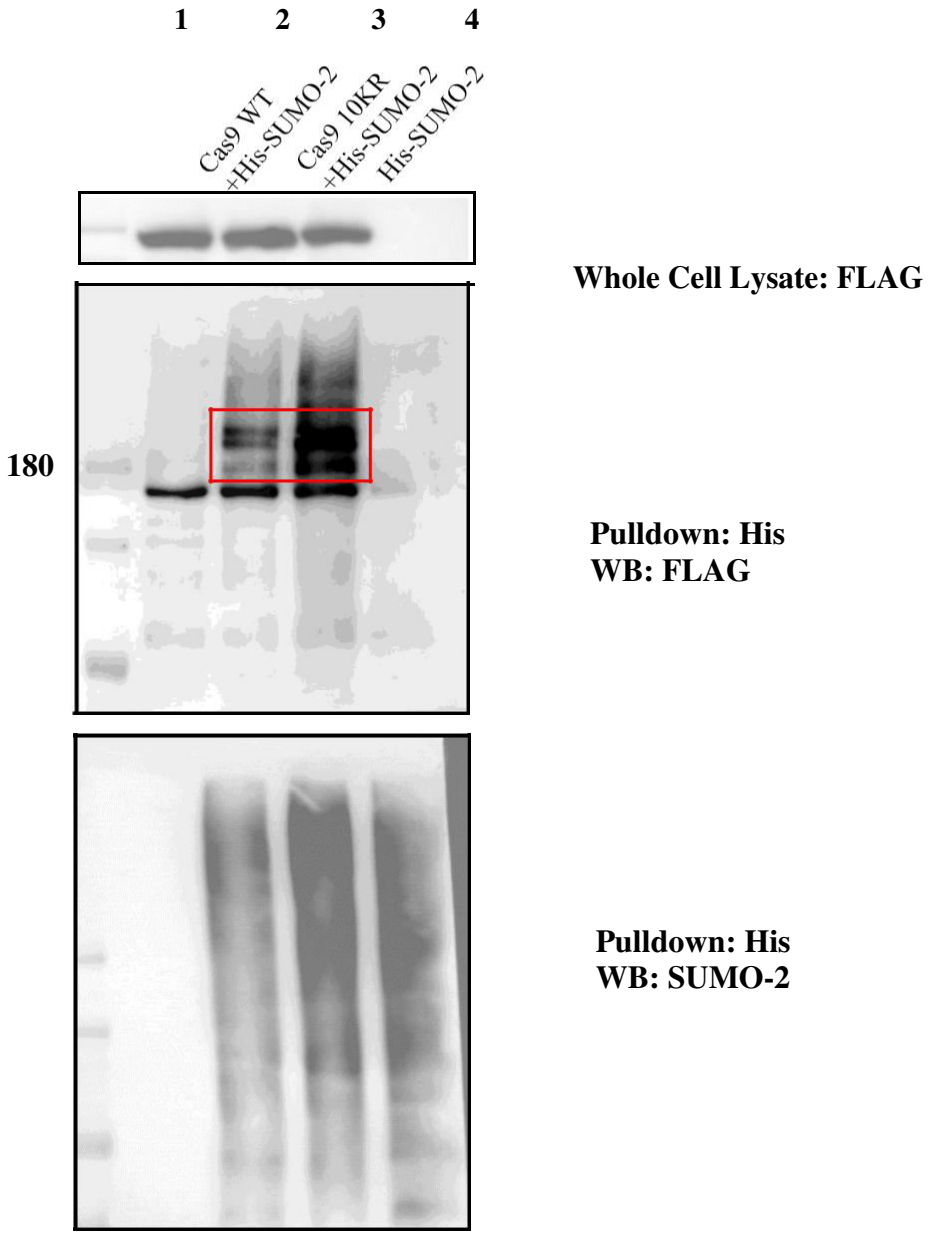


Figure 4.6 His pulldown assay for the detection of SUMO-2-conjugated Cas9. Top figure depicts Cas9 expression in whole cell lysate. Middle figure shows Cas9 proteins conjugated with His-SUMO-2. 3 bands framed by the red box in Lanes 2 and 3 depict His-SUMO-2-conjugated Cas9. Single band below 180 kDa in Lanes 1, 2, and 3 corresponds to unmodified Cas9 proteins bound to beads non-specifically. Bottom figure represents the pulldown control.

In Figure 4.6, in the FLAG blot in the middle panel, we observed 3 separate bands (shown inside the red box) in Lanes 2 and 3 which correspond to sumoylated Cas9 proteins. Plus, we observed a smear above those 3 bands. Lane 3 shows increased levels of SUMO-2-conjugation in Cas9 10KR mutant when compared to Lane 2 which depicts sumoylated wild type Cas9. However, in the SUMO-2 blot which represents the pulldown control, Lane 3 has more protein than Lane 2; this means that wild type Cas9 and 10KR mutant have similar levels of SUMO-2-conjugation. We do not observe these 3 bands in Lanes 1 and 4, which correspond to protein lysates collected from only wild type Cas9- and only His-SUMO-2-transfected cells; this indicates that those 3 bands are in fact Cas9 proteins conjugated to His-tagged SUMO-2.

4.4. Depiction of Interaction of Cas9 and Ubc9, the Universal SUMO E2 Ligase

For the purpose of showing the interaction between Cas9 and Ubc9, co-immunoprecipitation experiment was performed. As a positive control, Argonaute-2 (Ago-2) protein, a well-known interaction partner of Ubc9 was used. Pulldown was done by an antibody against FLAG tag which both Cas9 and Ago-2 proteins have. As a negative control, only Ubc9-transfected sample was used.

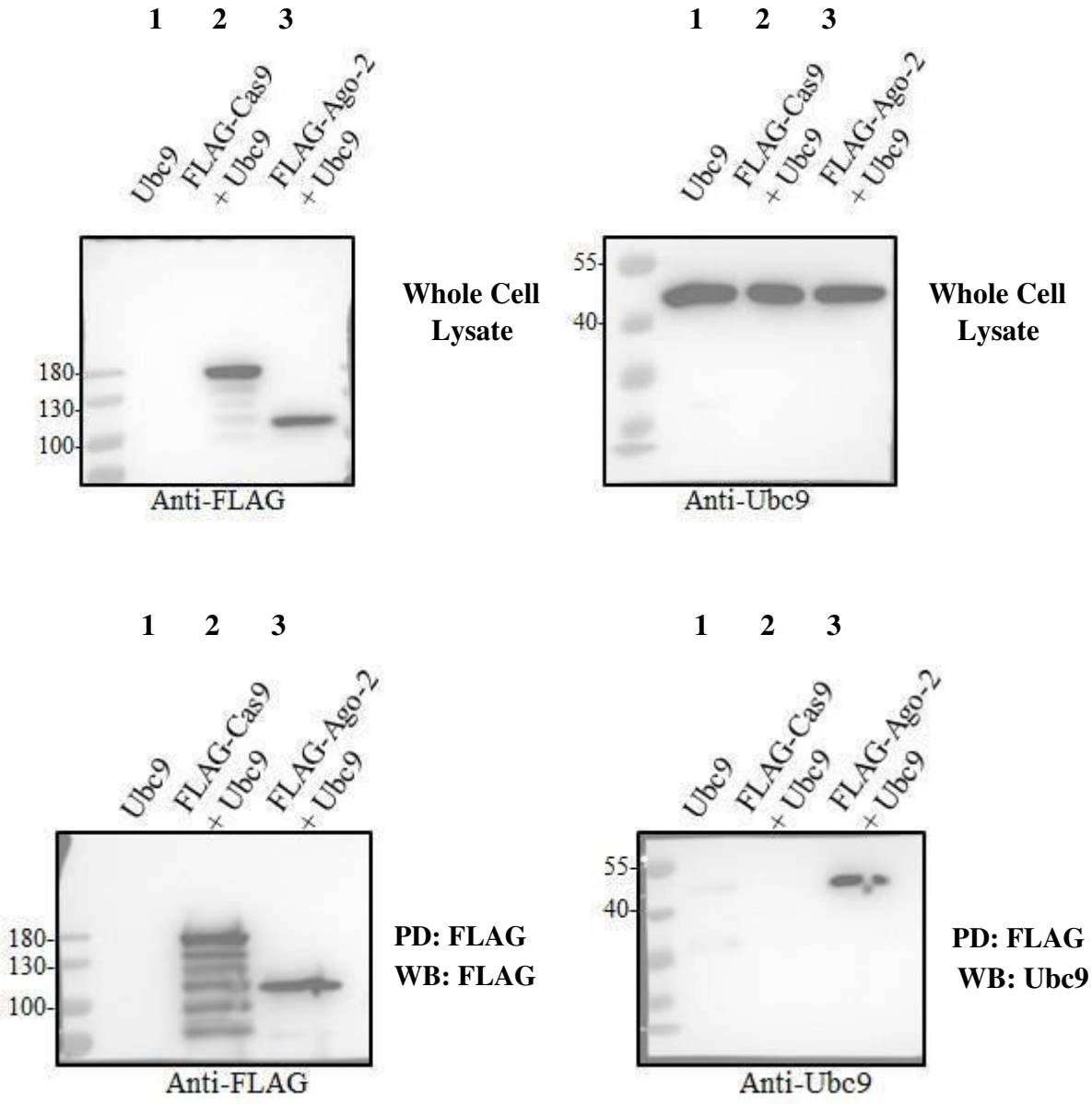


Figure 4.7 The interaction between Cas9 and Ubc9 cannot be observed, whereas Ago-2 and Ubc9 interaction is visible.

In Figure 4.7, top 2 panels show the expression levels of Cas9, Ago-2, and Ubc9 in whole cell lysate. Bottom left panel is the FLAG blot of the pulldown. In Lane 2, the bands below the 180 kDa band are degradation products of Cas9. Bottom right panel is the Ubc9 blot of the pulldown. Lane 3 clearly depicts the captured interaction of Ubc9 with Ago-2, however, Lane 2 has no band which means Cas9 and Ubc9 interaction could not be captured here.

5. DISCUSSION

5.1. Implications of Sumoylation on Cas9 Regulation

CRISPR Cas9 technology is a powerful gene editing tool that enables gene alterations in very specific manner, making this a revolutionary method which can be adapted to multiple research areas such as drug discovery, disease modeling, epigenome engineering, and genome imaging, hence making it even more urgent to study its major player, CRISPR Cas9 protein. Although many research groups are well aware of the significance of CRISPR system and its regulation, and that thousands of groups work vehemently on improving this system for the purposes of enhancing gene targeting efficiency, altering Cas9 stability, or generating novel molecular biology techniques by utilizing Cas9, there is still not nearly enough knowledge that can shed light on how Cas9 can be properly regulated. More specifically, there is not a single post-translational modification shown on Cas9 protein in eukaryotic systems which is one of the major types of protein regulation in eukaryotic cells.

Ever since current lab member Yusuf has shown that Cas9 protein is conjugated to SUMO1 and SUMO2, it has been a great enigma to discover how SUMO proteins can regulate Cas9 after being conjugated to it. Sumoylation is known to be involved in various key cellular mechanisms like cell cycle, DNA repair, stress responses, transcription regulation, subcellular localization, chromatin organization, and many others through regulating the proteins involved in these processes by covalently attaching to them. Even though Cas9 is a bacterial protein and sumoylation is a eukaryotic post-translational modification, Cas9 can still be possibly regulated by SUMO machinery once it is inside a eukaryotic cell. Ever since its adaptation as a gene editing tool, Cas9 protein has been introduced in various types of eukaryotic cells for gene targeting purposes. We propose that after being introduced in a eukaryotic cell, Cas9 protein becomes a target of SUMO machinery and it can be regulated by sumoylation through

covalent attachment. SUMO conjugation could affect Cas9 in many aspects such as DNA binding, sgRNA binding, nuclease activity, and protein stability.

Sumoylation can affect the DNA binding property of Cas9, because it has been known to alter the DNA binding affinity of many transcription factors such as p53, Sp1, c-Myc, and also histone proteins which are also chromatin-associated proteins. When conjugated to its target, SUMO can modify the activity of a chromatin-associated protein by altering its DNA binding affinity. In 2014, Gong *et al*. showed that when Sp1 protein is conjugated to SUMO2, its binding affinity for its target genes diminished. In 2015, another group showed that disruption of c-Myc sumoylation lead to increased c-Myc activity as a transcriptional regulator (González-Prieto *et al*., 2015). There are many other examples to SUMO machinery altering transcriptional activity through affecting DNA binding affinity. The same type of alteration can be observed for Cas9 protein, too.

It is possible that sumoylation can affect the single guide RNA binding of Cas9 protein. There are many RNA binding proteins that undergo sumoylation. One major example is Argonaute-2 (Ago-2), the main player of RNA interference. Ago-2 is a target of SUMO machinery, and it has been shown that Argonaute-2 sumoylation lead to altered gene silencing activity (Josa-Prado *etal*., 2015). With that in mind, it is safe to say that sumoylation can have an impact on the interaction of Cas9 with guide RNA.

Another property sumoylation can modify is the nuclease activity of Cas9 protein. There is a chance sumoylation can alter the cleavage activity of Cas9 on target DNA. By creating a sumoylation-defective Cas9 mutant, we might possess a Cas9 variant in our hands which could be either hyperactive or hypoactive.

One final possibility is the effect of sumoylation on the stability of Cas9 protein. There are many studies depicting how sumoylation can alter a target protein's half-life in eukaryotic cells. One mechanism involves the proteosomal degradation of target proteins through SUMO-dependent ubiquitylation mediated by SUMO-targeted ubiquitin ligases (STUbLs) such as

RNF4. One significant example to this would be the degradation of a viral oncoprotein named Tax. HTLV-1 (human T cell leukemia virus type 1) originated Tax protein which is the major cause of adult T cell leukemia is found out to be destroyed by SUMO/RNF4-dependent proteosomal degradation (Dassouki *et al*., 2015).

5.2. Generation of Sumoylation-defective Cas9 Variants

In order to tell if sumoylation has any regulatory role on Cas9 protein, it is absolutely essential to generate a Cas9 variant that is sumoylation-defective. In this way, it will be possible to make a comparison between wild type Cas9 and sumoylation-defective Cas9 variant regarding various features of this protein such as DNA-binding, nuclease activity, protein stability, etc. For that purpose, site-directed mutagenesis experiments were performed to create various Cas9 mutants which are all candidates for a sumoylation-defective mutant. Normally, we expect 1 sumoylation site on the target protein to be preferred over other sites by SUMO E2 ligase Ubc9, and when the lysine residue on that site is mutated, we expect to ablate sumoylation altogether. For that purpose, we generated 10 single mutants, one for each sumoylation site which possessed a single K-to-R mutation on their designated sumoylation motif (Figure 4.2 & Table 4.1). In addition to those single mutants, we generated another Cas9 construct possessing all 10 K-to-R mutations on the sumoylation motifs and called it Cas9 10KR mutant (Figure 4.3).

5.3. Representation of Cas9 Sumoylation with His Pulldown Experiments

After creating all possible candidates of a sumoylation-defective Cas9 variant, the next step was to confirm the abrogation of Cas9 sumoylation. For that reason, His pulldown experiments were performed to show the sumoylated levels of Cas9 WT and Cas9 10KR mutant. After co-transfecting FLAG-Cas9 WT/ FLAG-Cas9 10KR and His-SUMO-1/ His-SUMO-2 to HEK293T cells, His-SUMO-1- and His-SUMO-2-tagged proteins were enriched by Nickel pulldown. Then, by blotting against the FLAG tag, we were able to show Cas9 proteins that were conjugated to SUMO-1 and SUMO-2 (Figures 4.5 & 4.6).

In Figure 4.5, in the FLAG blot in middle panel, 3 separate Cas9 bands that correspond to SUMO-1-conjugated Cas9 were observed. Our certainty on this information is due to the fact that these 3 bands are absent in Lanes 1 and 4, which correspond to only FLAG-Cas9 - and only His-SUMO-1 –transfected samples respectively. The reason there are 3 bands but not a single band can be explained by the feature of SUMO-2 peptides to form poly-SUMO chains through the sumoylation motif they possess, and SUMO-1 can terminate this chain by conjugating to it. The presence of 3 bands hints that there are 3 different versions of sumoylated Cas9 depending on the number of SUMO peptides conjugated to it.

One interesting observation was the levels of SUMO-1-conjugated Cas9 WT and Cas9 10KR. When we looked at Figure 4.5, no decrease in SUMO-1 conjugation of Cas9 10KR was observed compared to wild type Cas9. Even though, all 10 sumoylation sites were successfully mutated on this Cas9 variant, we still observed same levels of sumoylation. It seems like Cas9 sumoylation was not even remotely affected by the disruption of 10 consensus sumoylation sites all of which residing on the protein surface.

The same phenomenon was seen with SUMO-2-conjugated Cas9 levels in Figure 4.6. Again, we observed 3 separate bands representing different versions of sumoylated Cas9. The difference in band sizes resulted from the difference in the number of SUMO peptides on the poly-SUMO chain conjugated to Cas9 protein. Furthermore, once again, Cas9 10KR mutant seemed to be sumoylated as much as the wild type Cas9 (Figure 4.6). Even though 10KR mutant seemed to be sumoylated more than the wild type, it was only because there was more protein in the pulldown, which can be seen in the SUMO-2 blot in Figure 4.6.

Both of the results from Figures 4.5 and 4.6 hint at the existence of at least one sumoylation site on Cas9 protein other than the ones predicted before.

5.4. Investigation of Alternative Sumoylation Sites on Cas9 Protein

After multiple trials of His pulldown experiments showed that Cas9 10KR mutant still got sumoylated as much as the wild type Cas9 and there was no visible decrease in sumoylation levels, we accepted the fact that there is another site or possibly multiple sites that are preferred for SUMO conjugation over the canonical sumoylation sites mentioned before.

Josa-Prado and his colleagues' paper (published in 2015) on Argonaute-2 (Ago-2) sumoylation had an interesting result regarding the generation of a sumoylation-defective mutant. They seemed to have a problem with ablating Ago-2 sumoylation. In fact, they had detected Lysine 402 as the major sumoylation site, yet they still observed small amount of sumoylation on Ago-2. What they later proposed was that mutating the lysine residue might not always be sufficient to prevent SUMO conjugation; the hydrophobic residue and the acidic residue on the consensus sumoylation motif (ΨKxD/E) might be just as important for sumoylation. Their theory suggested that the hydrophobic residue at the first position and the acidic residue on the fourth position were essential for Ubc9, the universal SUMO E2 ligase, to recognize the sumoylation motif, and even if the lysine was mutated, once Ubc9 recognized the hydrophobic and acidic residues, it could attach the SUMO peptide on any lysine residue within its reach. In order to prove this idea, they created an AKAA mutant of the major sumoylation site, Lysine 402, and they finally succeeded to abolish Ago-2 sumoylation, concluding that the hydrophobic and acidic residues at the major sumoylation site were indeed important.

In the lights of the findings of Josa-Prado and his colleagues, we decided to look for the lysine residues in the vicinity of each consensus sumoylation site of Cas9 protein. PyMOL program was used to look into the protein structure of Cas9. As a result, we discovered various candidate lysine residues around 4 consensus sumoylation sites on Cas9 which could be targeted for sumoylation (Figure 5.1).

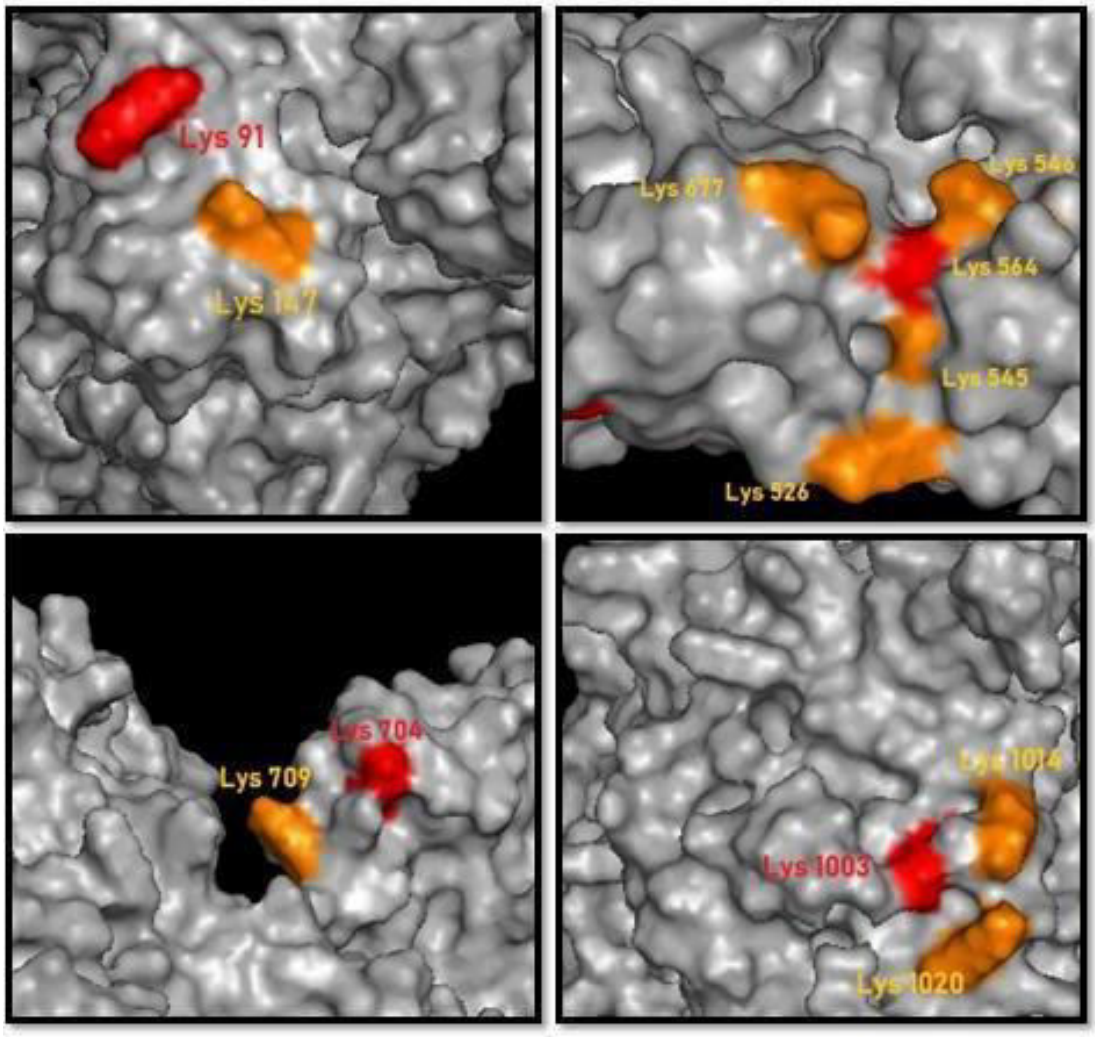


Figure 5.1 Illustration of neighboring lysine residues in the vicinity of the sumoylation sites. The residues shown in red are the lysines in the sumoylation motifs, whereas the ones in yellow are the neighboring lysine residues around the sumoylation motifs.

By looking at Figure 5.1, we counted a total of 8 neighboring lysine residues, and every one of them could be a candidate sumoylation site of Cas9 protein.

As an alternative approach to this, literature research was done for possible non-canonical sumoylation sites shown in other studies. As a result of this research, we came across another sumoylation motif other than the canonical ΨKxD/E, though not as popular. It is comprised of a hydrophobic residue at first position, a lysine at second, any residue on third, and finally a serine residue at fourth position (ψKxS). This motif counts as a conditional site for sumoylation. If the serine residue at fourth position gets phosphorylated, it will become negatively charged, thus making this motif a sumoylation site. When we scanned through the amino acid sequence of Cas9, we found 3 motifs that fit this profile (Figure 5.2 & Table 5.1).

Cas9 AA Sequence

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS
NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD
LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD
KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT
AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKA
GQAKKKK*

Figure 5.2 Detection of 3 non-canonical sumoylation sites depicted in pink. The red motifs are the canonical sumoylation sites which were predicted earlier.

Table 5.1 The list of newly found non-canonical sumoylation sites on Cas9.

Site Name	Amino Acid Sequence	Location of Lysine
Site 1	PKHS	509
Site 2	PKYS	1199
Site 3	LKGS	1245

5.5. Depiction of Interaction of Cas9 and Ubc9, the Universal SUMO E2 Ligase

After current lab member Yusuf showed Cas9 sumoylation, being able to show the interaction between Cas9 and Ubc9, the one and only protein responsible for sumoylation, was an important step. For that reason, co-immunoprecipitation was performed to capture the interaction between Cas9 and Ubc9. In order to make sure the protocol was working to capture interaction partners of Ubc9, Argonaute-2, a protein known to be sumoylated, thus a known interaction partner of Ubc9 was used. In order to eliminate the possibility of non-specific binding of Ubc9 to agarose beads, we also put only Ubc9-transfected sample as negative control.

Figure 4.7 shows that Ago-2 and Ubc9 interaction is captured quite well, whereas Cas9 and Ubc9 interaction could not be captured. Since Cas9 sumoylation was observed repeatedly by the immunoprecipitation experiments done by Yusuf, and this modification was also observed by 2 other methods (His pulldown assay and proximity ligation assay), it is certain that Cas9 and Ubc9 are interacting. After showing that the co-IP protocol we used was successful to capture a known interaction partner of Ubc9, we conclude that co-IP experiments are not sensitive enough to capture Cas9-Ubc9 interaction. As a solution to this problem, we are planning to perform proximity ligation assay which can show the interaction partners that are up to 16nm apart.

5.6. Future Studies

As the next step, we are planning to send Cas9 protein to mass spectrometry where we can detect the exact lysines SUMO peptide or peptides are conjugated to. After getting the mass spectrometry results, we are going to resume site-directed mutagenesis experiments to mutate these lysines and finally generate a sumoylation-defective Cas9 mutant. If the lysine residue or residues detected by mass spectrometry are in fact the neighboring lysines in close proximity of the sumoylation motifs we have shown in Figure 5.1, it will support the hypothesis of Josa-Prado and his colleagues on the significance of the hydrophobic residue at

first position and the acidic residue at fourth position for Ubc9 to recognize a sumoylation site, and subsequently to sumoylate any lysine residue in the vicinity of that site. Another scenario would be the detection of lysines in the non-canonical (conditional) sumoylation sites we have shown in Figure 5.2 & Table 5.1. This result would implicate that Cas9 could also be modified through phosphorylation prior to sumoylation.

After successfully generating a sumoylation-defective mutant of Cas9, the next step is to start functional experiments regarding the impact of sumoylation on various features of Cas9 protein. We will start with checking the regulatory effect of sumoylation on the nuclease activity of Cas9. For that purpose, a GFP reporter system will be set up, a sgRNA targeting GFP will be introduced to HEK293T cells along Cas9 and then the percentage of DNA cleavage by wild type Cas9 and sumoylation-defective Cas9 will be assessed.

In order to check the regulatory effect of sumoylation on DNA binding feature of Cas9, we will perform ChIP-seq experiments with wild type and sumoylation-defective Cas9. With this method, we can assess on-target and off-target binding of wild type Cas9 vs. sumoylation-defective Cas9.

To see if sumoylation has a regulating role for Cas9 and sgRNA interaction, we can perform fluorescence resonance energy transfer (FRET) experiments. We would assign donor and acceptor molecules (chromophores) to Cas9 protein (wild type/ sumoylation-defective mutant) and guide RNA and monitor the interactions between these two.

REFERENCES

Adli, M. (2018). "The CRISPR tool kit for genome editing and beyond." *Nature Communications*. https://doi.org/10.1038/s41467-018-04252-2.

Ahner, A., Gong, X., & Frizzell, R. A. (2016). Divergent signaling via SUMO modification: potential for CFTR modulation. *American journal of physiology. Cell physiology*, *310*(3), C175–C180. doi:10.1152/ajpcell.00124.2015

Alkhnbashi, O.S., Costa, F., Shah, S.A., Garrett, R.A., Saunders, S. J., & Backofen, R. (2014). CRISPRstrand: Predicting repeat orientations to determine the crRNA-encoding strand at CRISPR loci. *Bioinformatics. 30*(17):I489-I496. doi:10.1093/bioinformatics/btu459

Chang, L.K., Liu, S.T., Kuo, C.W., Wang, W.H., Chuang, J.Y., Bianchi, E., & Hong, Y.R. (2008). Enhancement of transactivation activity of rta of Epstein-Barr virus by RanBPM. *J Mol Biol*. *379*(2):231–42

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., … Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*; *155*(7), 1479–1491. doi:10.1016/j.cell.2013.12.001

Citro S, & Chiocca S. (2013). Sumo paralogs: redundancy and divergencies. *Front Biosci*; 544–553

Dassouki, Z., Sahin, U., Hajj, H. E., Jollivet, F., Kfoury, Y., Lallemand-Breitenbach, V., Hermine, O., de The, H., & Bazarbachi, A. (2014). ATL response to arsenic/interferon therapy is triggered by SUMO/PML/RNF4-dependent Tax degradation. *Blood.125*(3):474-482.

Di Bacco, A., Ouyang, J., Lee, H. Y., Catic, A., Ploegh, H., & Gill, G. (2006). The SUMO-specific protease SENP5 is required for cell division. *Molecular and cellular biology*; *26*(12), 4489–4498. doi:10.1128/MCB.02301-05

Dunphy, P. S., Luo, T., & Mcbride, J. W. (2014). Ehrlichia chaffeensis exploits host sumoylation pathways to mediate effector-host interactions and promote intracellular survival. *Infection and Immunity. 82*(10):4154-4168. doi:10.1128/iai.01984-14

Eifler, K., & Vertegaal, A. C. (2015). SUMOylation-mediated regulation of cell cycle progression and cancer. *Trends in Biochemical Sciences. 40*(12):779-793. doi:10.1016/j.tibs.2015.09.006

Enserink, J. M. (2017). Regulation of cellular processes by SUMO: Understudied topics. *SUMO Regulation of Cellular Processes Advances in Experimental Medicine and Biology. 963:* 89-97. doi:10.1007/978-3-319-50044-7_5

Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J., & Church, G. M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods*, *10*(11), 1116–1121. doi:10.1038/nmeth.2681

Flotho, A., & Melchior, F. (2013). Sumoylation: a regulatory protein modification in health and disease. *Ann Rev Biochem*; 82:357–85. 10.1146/annurev-biochem-061909-093311

Gareau, J. R., & Lima, C. D. (2010). The SUMO pathway: Emerging mechanisms that shape specificity, conjugation and recognition. *Nature Reviews Molecular Cell Biology.11*(12): 861- 871. doi:10.1038/nrm3011

Geiss-Friedlander, R., & Melchior, F. (2007). Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol*.; 8(12):947–956. doi: 10.1038/nrm2293

Gong, L., Ji, W.K., Hu, X.H., Hu, W.F., Tang, X.C., Huang, Z.X., Li, L., Liu, M., Xiang, S.H., & Wu, E. (2014). Sumoylation differentially regulates Sp1 to control cell differentiation. *Proc Natl Acad Sci USA* ; 111:5574-5579; https://doi.org/10.1073/pnas.1315034111

González-Prieto, R., Cuijpers, S. A., Kumar, R., Hendriks, I. A., & Vertegaal, A. C. (2015). c-Myc is targeted to the proteasome for degradation in a SUMOylation-dependent manner, regulated by PIAS1, SENP7 and RNF4. *Cell cycle (Georgetown, Tex.)*, *14*(12), 1859-72.

Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F., & Chin, L. (2016). Post-translational regulation of Cas9 during G1 enhances homology-directed repair. *Cell Rep.*; 14:1555–1566

Han, Z. J., Feng, Y. H., Gu, B. H., Li, Y. M., & Chen, H. (2018). The post-translational modification, SUMOylation, and cancer (Review). *International Journal of Oncology*, *52*(4), 1081–1094. doi:10.3892/ijo.2018.4280

Hickey, C. M., Wilson, N. R., & Hochstrasser, M. (2012). Function and regulation of SUMO proteases. *Nature reviews. Molecular cell biology*, *13*(12), 755–766. doi:10.1038/nrm3478

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology*, *33*(5), 510–517. doi:10.1038/nbt.3199

Hu, W., Kaminski, R., Yang, F., Zhang, Y., Cosentino, L., Li, F., … Khalili, K. (2014). RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection.

Proceedings of the National Academy of Sciences of the United States of America, *111*(31),11461–11466. doi:10.1073/pnas.1405186111

Ji, H., Jiang, Z., Lu, P., Ma, L., Li, C., Pan, H., … Zhu, H. (2016). Specific Reactivation of Latent HIV-1 by dCas9-SunTag-VP64-mediated Guide RNA Targeting the HIV-1 Promoter. *Molecular therapy : the journal of the American Society of Gene Therapy*, *24*(3), 508–521. doi:10.1038/mt.2016.7

Jiang, F., & Doudna, J. A. (2015). The structural biology of CRISPR-Cas systems. *Currentopinion in structural biology*; *30*, 100–111. doi:10.1016/j.sbi.2015.02.002

Jin, S. & Levine, A. J. (2001). The p53 functional circuit. *J. Cell Sci 114*:4139-4120

Kerscher, O., Felberbaum, R., & Hochstrasser, M. (2006). Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol*.; 22:159–80

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA*.; 93:1156–1160. doi: 10.1073/pnas.93.3.1156

Koronakis, V., & Hume, P. (2010). Faculty of 1000 evaluation for Listeria monocytogenes impairs SUMOylation for efficient infection. *F1000 - Post-publication Peer Review of the Biomedical Literature*. 464(7292):1192-1195. doi:10.3410/f.3013969.2693069

Langer-Safer, P. R., Levine, M., & Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy ofSciences of the United States of America*, *79*(14), 4381–4385

Lin, J., Ohshima, T., & Shimotohno, K. (2004). Association of Ubc9, an E2 ligase for SUMO

conjugation, with p53 is regulated by phosphorylation of p53. *FEBS Letters, 573*(1-3):15-18. doi:10.1016/j.febslet.2004.07.059

Liu, Y., Zhao, M., Gong, M., Xu, Y., Xie, C., Deng, H., et al. (2018). Inhibition of hepatitis B virus replication via HBV DNA cleavage by Cas9 from Staphylococcus aureus. *Antiviral Res.* 152:58–67. 10.1016/j.antiviral.2018.02.011

Ma, H., Reyes-Gutierrez, P., & Pederson, T. (2013). Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(52), 21048–21053. doi:10.1073/pnas.1319097110

Makarova, K. S., Zhang, F., & Koonin, E. V. (2017). "SnapShot: Class 1 CRISPR-Cas Systems." *Cell*. https://doi.org/10.1016/j.cell.2017.02.018

Martin, S., Wilkinson, K. A., Nishimune, A., & Henley, J. M. (2007). Emerging extranuclear roles of protein SUMOylation in neuronal function and dysfunction. *Nature reviews. Neuroscience*, *8*(12), 948–959. doi:10.1038/nrn2276

Marusic, M.B., Mencin, N., Licen, M., Banks, L., Grm, H.S. (2010). Modification of human papillomavirus minor capsid protein L2 by sumoylation. *J Virol*. *84*:11585–11589

Miller, J.C., Tan, S., Qiao G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. (2011). A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 29:143–148

Moore, F. E., Reyon, D., Sander, J. D., Martinez, S. A., Blackburn, J. S., Khayter, C., Ramirez, C. L., Joung, J. K., … Langenau, D. M. (2012). Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). *PloS one*, *7*(5), e37877

Murata, T., Hotta, N., Toyama, S., Nakayama, S., Chiba, S., Isomura, H., Ohshima, T., Kanda, T., & Tsurumi, T. (2010). Transcriptional repression by sumoylation of Epstein-Barr virus BZLF1 protein correlates with association of histone deacetylase. *The Journal ofbiological chemistry*. *285*(31), 23925-35

Müller, S., Berger, M., Lehembre, F., Seeler, J., Haupt, Y., & Dejean, A. (2000). C-Jun and p53 activity is modulated by SUMO-1 modification. *Journal of Biological Chemistry.275*(18):13321-13329. doi:10.1074/jbc.275.18.13321

Newman, H. A., Meluh, P. B., Lu, J., et al. (2017). A high throughput mutagenic analysis of yeast sumo structure and function. *PLoS Genet*.; 13(2): e1006612

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., … Gersbach, C. A. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods*, *10*(10), 973–976. doi:10.1038/nmeth.2600

Pichler, A., Gast, A., Seeler, J. S., Dejean, A., & Melchior, F. (2002). The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell.*; 108(1):109–120. doi: 10.1016/S0092-8674(01)00633-X

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183. doi:10.1016/j.cell.2013.02.022

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., … Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing

specificity. *Cell*, *154*(6), 1380–1389. doi:10.1016/j.cell.2013.08.021

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., … Zhang, F. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, *520*(7546), 186– 191. doi:10.1038/nature14299

Richter, F., Fonfara, I., Bouazza, B., Schumacher, C. H., Bratovič, M., Charpentier, E., & Möglich, A. (2016). Engineering of temperature- and light-switchable Cas9 variants. *Nucleic acids research*, *44*(20), 10003–10014. doi:10.1093/nar/gkw930

Rodriguez, M. S., Desterro, J. M., Lain, S., Midgley, C. A., Lane, D. P., & Hay, R. T. (1999). SUMO-1 modification activates the transcriptional response of p53. *The EMBO Journal.18*(22):6455-6461. doi:10.1093/emboj/18.22.6455

Sahin, U., Lapaquette, P., Andrieux, A., Faure, G., & Dejean, A. (2014). Sumoylation of human Argonaute 2 at lysine-402 regulates its stability. *PLoS ONE. 9*(7). e102957. doi:10.1371/journal.pone.0102957

Santiago, A., Li, D., Zhao, L.Y., Godsey, A., & Liao, D. (2013). P53 sumoylation promotes its nuclear export by facilitating its release from the nuclear export receptor CRM1. *Molecular Biology of the Cell. 24*(17):2739-2752. doi:10.1091/mbc.e12-10-0771

Shiio, Y. & Eisenman, R.N. (2003). histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci. 100*:13225–30. doi:10.1073/pnas.1735528100

Srikanth, C.V., & Verma, S. (2017). Sumoylation as an integral mechanism in bacterial infection and disease progression. *SUMO Regulation of Cellular Processes Advances inExperimental Medicine and Biology. 963:*389-408. doi:10.1007/978-3-319-50044-7_22

Tatham, M. H., Kim, S., Jaffray, E., Song, J., Chen, Y., & Hay, R. T. (2004). Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection. *Nature Structural & Molecular Biology, 12*(1):67-74. doi:10.1038/nsmb878

Thakore, P. I., D'ippolito, A. M., Song, L., Safi, A., Shivakumar, N. K., Kabadi, A. M., et al. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* 12, 1143–1149. doi: 10.1038/nmeth.3630

Truong, D. J., Kühner, K., Kühn, R., Werfel, S., Engelhardt, S., Wurst, W., & Ortiz, O. (2015). Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic acidsresearch*, *43*(13), 6450–6458. doi:10.1093/nar/gkv601

Tu, Z., Yang, W., Yan, S., Yin, A., Gao, J., Liu, X., … Li, X. J. (2017). Promoting Cas9 degradation reduces mosaic mutations in non-human primate embryos. *Scientific reports*, *7*, 42081. doi:10.1038/srep42081

Vialter, A., Vincent, A., Demidem, A., Morvan, D., Stepien, G., Venezia, N.D., & Rio, P.G. (2011). Cell cycle-dependent conjugation of endogenous BRCA1 protein with SUMO-2/3. *Biochimica Et Biophysica Acta (BBA) - General Subjects. 1810*(4):432-438. doi:10.1016/j.bbagen

Wang, H., La Russa, M., & Qi, L. S. (2016). CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem*; 85:227–64

Zhen, S., Hua, L., Takahashi, Y., Narita, S., Liu, Y. H., & Li, Y. (2014). In vitro and in vivo growth suppression of human papillomavirus 16-positive cervical cancer cells by CRISPR/Cas9. *Biochem Biophys Res Commun*.; 450:1422–1426. doi: 10.1016/j.bbrc.2014.07.014