SAND20XX-XXXXR
LDRD PROJECT NUMBER: 214921
LDRD PROJECT TITLE: Post-translational Modifications (PTMs) of Cas9 protein: A Potential Biomarker for Gene-editing
PROJECT TEAM MEMBERS: PI: Joshua Podlevsky, 08631; PM: James Carney, 08631; Team member: Dongmei Ye, 08635

ABSTRACT: (250 word limit)

The clustered regularly interspaced short palindromic repeats (CRISPR) arrays and the CRISPR associated (Cas) proteins comprise a prevalent prokaryotic and archaeal adaptive immune system. The CRISPR/Cas9 system has been coopted for and become the ubiquitous gene-editing system due to the simplicity of requiring minimally the CRISPR RNA components and Cas9 protein for specific DNA sequence alteration. CRISPR/Cas9 has been extensively used for gene-editing a wide range of species with human patient trails currently underway. However, unsanctioned genome editing is a national security and public health threat that can cause serious permanent illness and death as well as having the potential for very long-lasting effects over generations due to genetic inheritance of the gene-edit. While the Cas9 protein would appear as a highly specific indicator of exposure to gene-editing reagents, the bacterial origins of CRISPR/Cas9 creates a daunting problem for detection. Bacterial Cas9 would then generate false-positives for detecting gene-editing by conventional molecular biology techniques. Antibody-based assays for Cas9 would be unable to distinguish between Cas9 expressed in human cells for gene-editing and highly common unrelated Cas9 from bacterial infections. Post-translational modifications of proteins are highly cell specific and hold the potential for discerning the cellular origins of a Cas9 protein and the differentiating between bacterial and gene-editing CRISPR/Cas9. The work described herein is in progress towards the identification of Cas9 post-translational modifications from bacterial and human cell expressed Cas9.

INTRODUCTION: (> 500 words with no upper limit)

The clustered regularly interspaced short palindromic repeats (CRISPR) arrays and the CRISPR associated (Cas) proteins comprise a prokaryotic and archaeal adaptive immune system specialized for defense against invading foreign nucleic genetic elements [1-3]. These foreign nucleic genetic elements include bacteriophages and plasmids. CRISPR/Cas systems function as ribonucleoprotein (RNP) complexes composed minimally of a CRISPR RNA molecules that guides a nucleolytic Cas protein to the target sequence of the invading nucleic acid molecule(s). The CRISPR RNA harbors protein binding regions for binding Cas proteins as well as a short 20 nucleotide (nt) sequence complementary to the target sequence for nucleolytic cleavage[4]. The protein composition of CRISPR/Cas systems varies greatly from the single effector Cas9 protein found in *Streptococcus pyogenes* to large multimer complexes such as the CRISPR-associated complex for antiviral defense (Cascade) found in *Escherichia coli[5]*. The simplicity of the *S. pyogenes* CRISPR/Cas9 system that requires minimally the CRISPR RNAs and Cas9 protein for targeted DNA cleavage was coopted for targeted gene-editing[3, 6]. This system was further

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

streamlined with an engineered single guide RNA (sgRNA), resulting in a system that requires the single Cas9 protein and sgRNA for targeted DNA cleavage[7-10].

CRISPR/Cas gene-editing development has rapidly outpaced technologies for specific and accurate detection. Detection of CRISPR/Cas gene-editing reagents is complicated by its bacterial origins. Antibodies with high affinity and specificity for Cas9 cannot differentiate between bacterial Cas9 and gene-editing Cas9, recognizing the same protein with the same affinity and specificity. Biological samples are commonly contaminated with bacteria and CRISPR/Cas systems are prevalent across bacterial species. Further compounding this issue is that the most widely employed CRISPR/Cas systems are from *S. pyogenes* and *Staphylococcus aureus*, highly common human pathogens and infections[11, 12]. Thus, the use of antibody-based assays for identifying the presence of gene-editing CRISPR/Cas9 reagents would suffer a high incidence of false-positives for common bacterial infections or bacterial contamination.

DNA-based assays for detecting the activity of CRISPR/Cas9 are after-the-fact measurements and require extensive knowledge of the gene-editing attempt[13]. Sequencing or PCR analysis of the genome for modifications require successful gene-editing; exposure to gene-editing reagents without the DNA modification would be undetectable. Knowledge of the targeted locus or loci within the genome is necessary for PCR based assays or else requiring whole genome resequencing that is complicated by naturally occurring polymorphisms within the population[14]. Moreover, tissue specific targeting would further obfuscate detection of the gene-edits by requiring knowledge of the tissue(s) targeted and is incompatible for sampling highly delicate or sensitive regions such as the nervous or cardiac systems.

Post-translational modifications (PTMs) of the proteins are chemical modifications following protein synthesis and are ubiquitous across all domains of life, found in bacterial and human proteins[15]. These chemical modifications include the addition of simple chemical groups (phosphate or acetate), more complex molecules (carbohydrates or lipids), the covalent linkages of small proteins (ubiquitin and ubiquitin-like proteins), and amino acid side chain residue modifications (deamidation or eliminylation). PTMs affect protein complexity by altering function, regulation, localization, retention, and enzymatic activity of the proteins within a cell[16-19]. Different species, and even different cell types within a single organism will have proteins with distinct PTM profiles that are necessary for generating discrete cellular physiologies. It seems highly probable that Cas9 expressed in bacterial or human cells would have distinct PTMs that could then serve as biomarkers for the cellular origins of the Cas9 protein.

There is a need for a technology that can either specifically detect exclusively gene-editing CRISPR/Cas9 reagents or be able to differentiate the origins of the detected CRISPR/Cas9 systems. The ability to detect CRISPR/Cas9 reagents, and not the after-the-fact genomic alteration, affords greater therapeutic or prophylactic treatment options. This project was focused on developing a means of expressing, purifying, and analyzing Cas9 protein generated in the native bacterial and host human cells for PTM differences.

DETAILED DESCRIPTION OF EXPERIMENT/METHOD: (> 700 words with no upper limit)

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

The experimental procedure was divided into three major tasks: (1) expression of recombinant Cas9 with affinity tags for downstream purification, (2) affinity purification of the recombinant Cas9 protein, and (3) mass-spectrometry analysis and identification of PTMs. The fusion recombinant Cas9 protein construct for expression within *S. pyogenes* was designed with an N-terminal 6xHis tag, di-alanine spacer, 3xFlag tag, di-alanine threonine serine spacer, a TEV cleavage site, and a tri-alanine threonine serine spacer (N-HHHHHH*AA*DYKDHDGDYKDHDIDYKDDDDK*AATS*ENLYFQG*AAATS* -C) before the native full-length *S. pyogenes* Cas9 protein sequence. The N-terminal affinity tags and spacers were codon optimized for expression within *S. pyogenes* cells. The fusion recombinant Cas9 protein construct for expression within human HEK293 cells was identical with a few small modifications. Following an identical N-terminal 6xHis tag, di-alanine spacer, 3xFlag tag, di-alanine threonine serine spacer, a TEV cleavage site, and a tri-alanine threonine serine spacer was an SV40 nuclear localization signal (NLS) and short linker (N- PKKKRKV*GIHGVPAA* -C) before the full-length *S. pyogenes* Cas9 protein sequence with the first methionine removed and a C-terminal NLS (N- KRPAATKKAGQAKKKK -C). The N-terminal affinity tags, NLSs, and spacers were codon optimized for expression within mammalian cells.

Construction of the S. pyogenes Cas9 expression plasmid. The N-terminal 6xHis and maltose binding protein (MBP) tagged Cas9 plasmid pMJ806 (Addgene #39312)[4] was modified to replace the MBP tag with a 3xFlag tag by subcloning a codon optimized sequence for 3xFlag (Genscript). The pMJ806 plasmid is compatible for *Escherichia coli* protein expression. For *S. pyogenes* protein expression, the 6xHis-3xFlag-Cas9 DNA cassette was subcloned into pIB169 as shuttle vector for protein expression in *S. pyogenes*[20] to generate pIB169-His-Flag-Cas9. Subcloning was verified by restriction digestion fragment analysis.

Construction of the HEK293 expression plasmid. The N-terminal 3xFlag and C-terminal eGFP tagged Cas9 plasmid PX458 (Genscript) was N-terminus appended with a 6xHis tag by subcloning a codon optimized sequence for 6xHis and a stop codon inserted downstream of the Cas9 gene to prevent expression of the eGFP to generate pcDNA-His-Flag-Cas9. This Cas9 sequence had been codon optimized for expression within mammalian cells and was flanked by NLS sequences. Subcloning was verified by sequence analysis.

S. pyogenes transformation. Electrocompetent *S. pyogenes* M49 NZ131 (ATCC BAA-1633) cells were prepared as previously described with minor modifications[21]. Briefly, 10 ml of brain heart infusion (BHI) media (BD Bacto) supplemented with 20 mM glycine was inoculated with an isolated *S. pyogenes* colony from a BHI agar plate and incubated for 16 hours at 37°C. The overnight culture was expanded into 150 ml of BHI with 20 mM glycine and grown at 37°C to OD_{600} 0.4, incubated on ice for 30 min, pelleted centrifugation, washed 3 times with ice-cold sterile 10% glycerol (v/v) and resuspended in 1 ml of 10% glycerol (v/v) for 200 µl aliquots. One aliquot of the electrocompetent *S. pyogenes* cells was transformed with 2.5 µg of pIB169-His-Flag-Cas9 or the parental pIB169 in a 2 mm electroporation cuvette at 1.75 kV, 400 ohms, and 25 µF with a Gene Pulser Xcell (BioRad) then plated onto BHI agar plates with 25 µg/ml chloramphenicol. Isolated colonies were screened for chloramphenicol resistance and western blot performed for Cas9 protein expression.

S. pyogenes growth and Cas9 expression. A pIB169-His-Flag-Cas9 or pIB169 transformed *S. pyogenes* isolated colony was inoculated into 5 ml of BHI media with 25 µg/ml chloramphenicol

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

and incubated for 16 hours at 37°C. The overnight culture was expanded into 150 ml of BHI with 25 µg/ml chloramphenicol and grown until OD_{600} 0.7. Wet cell pellets were resuspended in 1.5 ml of His-binding buffer (50 mM Na PO_4 pH 7.7, 300 mM NaCl, 10 mM imidazole, 0.03% Triton X-100), sonicated 5 times at 70% amplitude for 5-1 sec pulses on ice, and clarified by centrifugation.

HEK293 transient transfection and Cas9 expression. HEK293 cells were grown in DMEM medium (Gibco) supplemented with 10% FBS, 1× Penicillin–Streptomycin mix (Lonza) and 5% CO2 at 37°C to 80–90% confluency. Cells in a 6-well plate were transfected with 0.6 µg of pcDNA-His-Flag-Cas9 and and 6 µl of FuGENE HD transfection reagent (Promega) following manufacturer's instruction. Cells were harvested 48 h post-transfection, homogenized in HEPES lysis buffer (20 mM HEPES-KOH, pH 7.9, 400 mM NaCl, 0.2 mM EGTA, 2 mM MgCl2, 10% glycerol, 5 mM β-mercaptoethanol), incubated on ice for 30 min, and the lysate clarified by centrifugation.

Ni2+ purification. The 6xHis-3xFlag-Cas9 protein was purified from lysate following the manufacturer's instruction with minor modification. Briefly, Three hundred microliters of the lysate was applied to 250 µl of His-affinity gel (Zymo), incubated for 5 min at 4°C with agitation, and washed twice with His-wash buffer (50 mM Na PO_4 pH 7.7, 300 mM NaCl, 50 mM imidazole, 0.03% Triton X-100). Protein was eluted with 150 µl of His-elution buffer (50 mM Na PO_4 pH 7.7, 300 mM NaCl, 250 mM imidazole).

Western blot. Samples were electrophoresed on a 4-12% Bis-Tris protein gels (Invitrogen) and semi-dry transferred to a polyvinylidene difluoride (PVDF, Millipore) membrane with a Mini Gel Tank and Blot Module (Life) following manufacturer's instruction. The PVDF membrane was blocked with 5% milk in 1x Tris-buffered Saline and Tween 20 [(TBST), 50 mM Tris pH 7.6, 150 mM NaCl, 0.05% Tween 20] prior to probing with anti-FLAG primary antibody (clone M2, Sigma-Aldrich) and horseradish peroxidase coupled goat anti-mouse IgG secondary antibody (ab6789, Abcam). The enhanced luminol-based chemiluminescence from Plus Western Blotting Substrate (Pierce) was imaged with a ChemiDoc XRS+ Imaging System (BioRad).

RESULTS: (> 700 words with no upper limit)

Ectopic expression of recombinant affinity tagged Cas9 protein in bacteria and human cells. For the comparative analysis of PTMs, Cas9 protein was expressed in *S. pyogenes* and HEK293 cells. The *S. pyogenes* CRISPR/Cas9 system is the most prolific gene-editor system with a myriad of derivatives such as base-editors and a deactivated transcriptional regulator. *S. pyogenes* is also a highly common human pathogen, found on the surface of human skin, and a common biological contaminant. To determine PTMs that are unique to human cells and not found from *S. pyogenes* expressed protein, an *S. pyogenes* cell line was constructed for affinity purification of Cas9 protein. This was accomplished by creating an expression plasmid that is replication competent in *S. pyogenes*, has selection for plasmid maintenance by chloramphenicol, and has affinity tags for Cas9 purification/enrichment. The final plasmid construct, pIB169-His-Flag-Cas9 (Fig. 1B), encodes for native Cas9 codon sequence N-terminally appended with a 6xHis and 3xFlag tag for downstream purification. The Cas9 recombinant protein was ectopically expressed under the *veg* promoter from *Bacillus subtilis* that is a strong promoter in *S. pyogenes[20]*.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

As the initial baseline for human PTMs, HEK293 cells were chosen. HEK293 cells are derived from human embryonic kidney cell culture transformed with a fragment of human adenovirus type 5 DNA[22] that created an immortalized cell line. For the analysis of Cas9 PTMs, transient transfection offers high levels of protein expression and does not suffer gene-silencing issues that arise with the time intensive genome integration. The short window of DNA insertion into the cell is too rapid for long term gene-silencing yet does require the design and construction of a plasmid with the gene of interest. The design of the recombinant affinity tagged Cas9 construct was as similar as possible to the *S. pyogenes* Cas9 construct to limit differences due to protein composition. The Cas9 protein was codon optimized for high level expression and to avoid truncated protein products from prematurely terminated protein translation. The N-terminus was appended with a 6xHis and 3xFlag tag for downstream purification. However, the Cas9 protein was flanked with NLS sequences to better represent Cas9 for gene-editing, where the Cas9 is necessary to enter the nucleus. The final plasmid construct of pcDNA-His-Flag-Cas9 (Fig. 1B) has Cas9 expressed from the strong synthetic CMV enhancer chicken β-actin (CAG) promoter for high protein yield.

Samples of the transformed *S. pyogenes* and transiently transfected HEK293 cells were analyzed by western blot. *S. pyogenes* cells transformed with either pIB169-His-Flag-Cas9 or the parental pIB169 grown under chloramphenicol selection for plasmid maintenance and Cas9 expression or as a negative control, respectively. Three separate isolated transformed colonies were grown for comparison of Cas9 expression. Three separate preparations of the pcDNA-His-Flag-Cas9 plasmid were transiently transfected into HEK293 cells. After 24 hours, the media was changed and continued to incubate with 5% CO2 at 37°C. At 48 hours post transfection, the cells were harvested. The Cas9 protein was appended with an N-terminal 3xFlag tag for imaging (Fig. 1). For western blot analysis of the clarified cell lysates, the blots were probed with anti-Flag antibodies (Fig. 2). Minor differences in the expression of Cas9 for *S. pyogenes* was observed for the separate isolates of pIB169-His-Flag-Cas9 transformed cells (Strains 1-3), while no signal was observed for cells transformed with the empty parental vector pIB169 (Fig. 2A, lanes 1-4). Greater differences were visible for the transient transfection of HEK293 cells with separate plasmid preparations. Strain 1 was a plasmid preparation from cells sent by the manufacturer, strain 2 was the plasmid preparation from the manufacturer, and strain 3 was a plasmid preparation from a transformation of TOP10 cells with the plasmid from strain 2. The TOP10 cell plasmid preparation yielded the greatest signal and there was no significant signal from untransfected cells (Fig. 2B, lanes 1-4). Of note, there were many lower molecular weight bands present in all lanes that expressed the Cas9 protein that would indicate partial degradation of the protein.

Purification of affinity tagged Cas9 protein from bacteria and human cells. The Cas9 proteins expressed were appended with an N-terminal 6xHis tag for purification (Fig. 1). For purification of the His-tagged Cas9 protein, the cell lysates from transformed *S. pyogenes* and transiently transfected HEK293 cells were run over Ni^{2+} purification columns and bound protein eluted with 250 mM imidazole. Despite similar Cas9 expression from the clarified cell lysates for transformed *S. pyogenes* cells (Fig. 2A, lanes 1-3), there was a dramatic difference in the signal from the Ni^{2+} purified samples. Strain 2 had a much stronger signal than strain 1, and strain 3 had a signal that was barely visible. As expected, the negative control did not generate any

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

signal (Fig. 2A, lanes 5-8). The lower molecular weight bands present in the clarified lysate samples were not visible within the Ni^{2+} purified samples.

An identical Ni^{2+} column purification was performed for the cell lysates from transiently transfected HEK293 cells and the untransfected negative control. The signal for the Ni^{2+} column purified samples followed as expected from the clarified cell lysates (Fig. 2B, lanes 1-3). Strain 3 had the highest signal from the cell lysates and similarly the highest yield for the purified samples, strain 1 had a somewhat lower signal, while strain 2 had no signal that was similar to the untransfected sample (Fig. 2B, lanes 5-8). Interestingly, there was a dominant doublet of bands from the HEK293 cell lysates that expressed Cas9, where there was only a single dominant band of the expected size (Fig. 2). Similar to the *S. pyogenes* samples, the lower molecular weight bands were not present in the Ni^{2+} purified samples.

DISCUSSION: (> 700 words with no upper limit)

CRISPR/Cas9 systems are a highly prolific system for gene-editing through nucleolytic cleavage of a target nucleic acid sequence. The simplicity of a single effector protein for the DNA cleavage has set this system apart from other more complicated CRISPRs, such as Cascade found in *E. coli*, that require numerous protein subunits. However, the rapid development of CRISPR/Cas9 for gene-editing has outpaced the development of detection technologies that are crucial for surveillance of accidental or intentional exposure to CRISPR/Cas9 reagents. The development of technologies for detecting specifically CRISPR/Cas9 for gene-editing has been complicated by the bacterial origins of CRISPR. CRISPR is a highly prevalent bacterial and archaeal adaptive immune system for their defense against invading foreign DNA from bacteriophages and plasmids. The implementation of CRISPR for gene-editing has primarily focused on CRISPR/Cas9 systems from human pathogens, *S. pyogenes* and *S. aureus*. *S. pyogenes* and *S. aureus* are highly common bacterial contaminants of human biological samples. This is due to *S. pyogenes* and *S. aureus* inhabiting the surface skin and respiratory tract of humans and being causative agents of several types of infections and diseases such as skin abscesses and pharyngitis (Strep throat). This increases the incidence of *S. pyogenes* and *S. aureus* human biological sample contamination. Thus, the most prevalent gene-editing systems share primary components with the most common human infections.

Attempts for the detection of gene-editor CRISPR/Cas9 reagents with antibodies is expected to generate a high rate of false-positives results. While Cas9 would seem an appealing target for detecting the presence of CRISPR/Cas9 with several commercially available antibodies already available; detection of Cas9 could signal exposure to either exposure to gene-editor reagents or common bacterial pathogen contamination. A detection platform that can be readily fooled by infections that are prevalent at rates as high as 30% would be unacceptable and useless from a surveillance perspective. The issue with Cas9 antibodies is further compounded by the proteins used for the immune response. The Cas9 protein is expressed within bacterial, albeit *E. coli* and not *S. aureus*. *S. pyogenes* and *S. aureus*. However, it would be expected that any antibodies with an epitope overlapping a PTM(s) would be in common with *S. aureus*. *S. pyogenes* and *S. aureus* and not human cell expressed Cas9 that is specific for gene-editors. This a population of antibodies generated would have selection for bacterial and not human expressed Cas9 protein.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

This project has focused on the detection of CRSIPR/Cas9 by investigating PTMs as possible biomarkers for the specific expression of Cas9 within human cells compared with common bacterial expressed Cas9. Towards this goal, the first step was the design of constructs for expressing an affinity tagged version of the Cas9 protein. This construct was design with an N-terminal 6xHis tag for affinity purification with Ni^{2+} resin to enrich for the Cas9 protein from clarified cell lysates. The N-terminus also harbors a 3xFlag tag for the visualization of the Cas9 protein by western blot analysis. The 3xFlag tag also affords another route for protein purification if the Ni^{2+} purification is insufficient for the mass spectroscopy analysis for the detection of PTMs. Downstream of the 6xHis and 3xFlag tags was a TEV cleavage recognition sequence. This would allow for the site-specific endopeptidase TEV to cleave the Cas9 fusion protein for removal of the affinity tags. This was added for the ability to specifically elute the Cas9 fusion protein for protein cleave. The M2 clone of the anti-Flag antibody has been shown to have very high affinity for the 3xFlag tag that can result in low yield from competition-based elution with 3xFlag peptides. TEV cleavage can overcome this issue with site-specific cleavage and elution of the Cas9 protein that is free of the affinity tags.

Plasmids pIB169-His-Flag-Cas9 and pcDNA-His-Flag-Cas9 were generated based off these design specifications (Fig. 1). The two fusion proteins are high similar with identical affinity tags located at the N-terminus of the fusion protein. However, the Cas9 fusion protein generated from pcDNA-His-Flag-Cas9 has the Cas9 flanked at the N- and C-termini with NLS, which better emulates Cas9 employed for gene-editing application where the Cas9 need to localize to the nucleus for DNA editing activities. The pIB169-His-Flag-Cas9 was specifically designed for protein expression within *S. pyogenes* and Cas9 sequence is identical to the native protein sequence to better emulate the native protein. While the pcDNA-His-Flag-Cas9 is a common plasmid for protein expression within mammalian cells. Both plasmid systems overexpress the Cas9 fusion protein so that the initial cell lysates begin enriched with the Cas9 protein for PTM analysis (Fig. 2).

ANTICIPATED OUTCOMES AND IMPACTS: (> 700 words with no upper limit)

The current next steps in the project are developing a relationship with a collaborator for the mass spectrometry analysis of the *S. pyogenes* and HEK293 cells expressed Cas9 fusion proteins. The collaboration would entail the use of the *S. pyogenes* and HEK293 cells expressed Cas9 fusion proteins either purified from the clarified lysate or the clarified lysate itself for the mass spectrometry analysis. This would allow for the collaborator to have access to new and useful tools for studying the Cas9 protein when expressed for the original bacterial host or in a recombinant state within a human cell host system. In exchange for access to these new constructs, tools, and materials, would be the mass spectrometry analysis and the identification of minimally a single PTM that is found within human cell expressed Cas9 proteins and not found within the bacterial expressed Cas9 proteins.

This collaboration could answer the lingering question of whether PTMs can function as biomarkers for determining the cellular origins of the Cas9 protein, which in turn would then reveal if the protein was expressed within human cells—a unnatural state—would imply the Cas9 was a gene-editor reagent. Prediction software indicates that there would be a large multitude of PTMs with a range of confidences and numerous sites where there is the possibility

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

of multiple types of PTMs are a single amino acid location within the Cas9 protein primary sequence. Cas9 is a larger bacterial protein at 1,368 amino acids in length and it seems highly unlikely there would either be no PTMs found across the entire length of the protein within either the native bacterial cells or the recombinant unnatural expression within human cells. Moreover, it is unlikely that the native bacterial and and human cells would have the same PTMs across the length of the protein due to the massive evolutionary distances between prokaryotic bacteria and eukaryotic human cells; especially as even closely related species and even different cell types from a single organism have been shown to have different PTM profiles for examined proteins.

Initial follow on work with the identification of a candidate PTM would be replicates necessary to verify the authenticity of the PTM. This would include biological and technical replicates for determining the PTM was not an artifact of growth conditions or from the mass spectrometry analysis. With the technical replicates, the next steps would be to determine whether the human PTM is tissue or cell type specific, or if the PTM could function as a universal biomarker for human cell expression of Cas9 protein within mammalian cells. This would involve expanding the study from exclusively HEK293 cells to additional cells types that would be representative of other tissue types for similar ectopic Cas9 fusion protein expression, purification, and mass spectrometry analysis.

With the identification of a single PTM—ideally a PTM that is found within human expressed Cas9 proteins, yet lacking in bacterial expressed Cas9 proteins—the next steps would be a proposal for how this knowledge could be used as part of the detection platform for determining the cellular origins of the Cas9 protein and whether the Cas9 was from CRISPR/Cas9 for gene-editing or simply the result of bacterial contamination and unrelated to gene-editing. Initial design ideas center around generating antibodies that would target an epitope that harbors the site and specific PTM. This could be commercially accomplished by an outside vendor performing contracted synthesis of the peptide with the PTM with the subsequent inoculation of animals, suck as rabbits, for generating antibodies. However, this has the possibility that either a small portion to no antibodies would recognize the PTM and the bulk of the antibodies would be unable to differentiate between the presence or absence of the PTM within this region of the Cas9 protein. Moreover, peptide antibodies pose the risk of lacking key structural information of the full-length protein and would therefore require denaturation of the Cas9 protein prior to analysis. While more complicated and higher resource intensive, with the identification of a human cell specific PTM a mass spectrometry-based detection system could be developed. Depending on the sensitivity, this could be an initial first-generation detection platform.

The knowledge of a PTM that could serve as biomarker for human cell expression of Cas9 would itself be beneficial. Study of the PTM, or any subsequent PTMs, could reveal location for protein regulation, enhancing Cas9 protein activity within human cells for improved CRISPR/Cas9 activity or fidelity, as well as the potential for deactivating Cas9 when no longer desired within the cell. Cas9 is a bacterial protein that evolved within a massively dissimilar environment from the eukaryotic system. It would be highly expected that Cas9 is not optimized for function within human cells and that PTMs could be altering Cas9 function, specificity, or overall activity within human cells.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

CONCLUSION: (500 word limit)

This project has successfully designed a Cas9 fusion protein cassette system for the expression of the Cas9 fusion protein in the native *S. pyogenes* bacterial cells and in HEK293 cells as a representative cell line for human cell expression. *S. pyogenes* was chosen as the representative bacterial species, being the native species from which the Cas9 sequence had originated. Moreover, the *S. pyogenes* CRISPR/Cas9 system has become the gold standard for gene-editors and is the most ubiquitous CRISPR/Cas9 or general gene-editor system in current use. The Cas9 fusion protein cassettes were matched in the primary amino acid sequence—with minor modifications for the inclusion of the flanking NLS sequences—for comparative analysis of the proteins without bias for sequence differences. Additionally, there was included N-terminal affinity tags and a TEV cleavage site for several options of protein purification including endopeptidase protein cleavage and elution. The 6xHis tags functioned well for protein purification, while the 3xFlag tag excelled for visualizing the Cas9 protein by western blot detection.

The major impact of this project has been the development of useful tools for expressing Cas9 proteins within different species, specifically the native bacteria representing common biological sample contamination and HEK293 cells that represented CRISPR/Cas9 for gene-editing applications. These materials can function as proof of concept for the later mass spectrometry analysis and for the attraction of future collaborators. Moreover, this work can be applied for other Cas9 protein analysis projects and gene-editor detection projects.

There have been several lessons learned for the firs-year PI of the project. Primarily has been understanding the rapid pace of Exploratory Express projects. With that, the PI has gained valuable experience with the proper pacing of a project as well as designing the scope of the project to better accommodate fast paced research. The most import lesson learned has been to reduce the complexity of the experimental question. This project involved numerous tasks and subtasks and relied on external vendors for the generation of initial materials critical for downstream analysis. Additionally, a critical piece of equipment for the analysis was not available. Alternative plans should have been in place for the conditions that there would be delays from vendors, equipment, or other unforeseen events.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

REFERENCES

1. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science (New York, N.Y.), 2007. **315**: p. 1709-12.
2. Karvelis, T., et al., *crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus* RNA Biology, 2013.
3. Barrangou, R., *RNA events. Cas9 targeting and the CRISPR revolution.* Science (New York, N.Y.), 2014. **344**: p. 707-8.
4. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science (New York, N.Y.), 2012. **337**: p. 816-21.
5. Jackson, R.N., et al., *Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli.* Science (New York, N.Y.), 2014. **345**: p. 1473-9.
6. Pennisi, E., *The CRISPR craze.* Science (New York, N.Y.), 2013. **341**: p. 833-6.
7. Doudna, J.A. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9.* Science (New York, N.Y.), 2014. **346**: p. 1258096.
8. Heigwer, F., et al., *CRISPR library designer (CLD): software for multispecies design of single guide RNA libraries.* Genome biology, 2016. **17**: p. 55.
9. Rahman, M.K. and M.S. Rahman, *CRISPRpred: A flexible and efficient tool for sgRNAs on-target activity prediction in CRISPR/Cas9 systems.* PloS one, 2017. **12**: p. e0181943.
10. Brazelton, V.A., et al., *A quick guide to CRISPR sgRNA design tools.* GM crops & food, 2015. **6**: p. 266-76.
11. Cattoir, V., *Mechanisms of Antibiotic Resistance.* Streptococcus pyogenes : Basic Biology to Clinical Manifestations, 2016. **30**: p. Unit 9D.3.
12. Koonin, E.V., K.S. Makarova, and F. Zhang, *Diversity, classification and evolution of CRISPR-Cas systems.* Current opinion in microbiology, 2017. **37**: p. 67-78.
13. Sentmanat, M.F., et al., *A Survey of Validation Strategies for CRISPR-Cas9 Editing.* Scientific reports, 2018. **8**: p. 888.
14. Feuk, L., A.R. Carson, and S.W. Scherer, *Structural variation in the human genome.* Nature reviews. Genetics, 2006. **7**: p. 85-97.
15. Silva, A.M.N., et al., *Post-translational modifications and mass spectrometry detection.* Free radical biology & medicine, 2013. **65**: p. 925-941.
16. Chicooree, N., R.D. Unwin, and J.R. Griffiths, *The application of targeted mass spectrometry-based strategies to the detection and localization of post-translational modifications.* Mass spectrometry reviews, 2015. **34**: p. 595-626.
17. Tokmakov, A.A., et al., *Multiple post-translational modifications affect heterologous protein synthesis.* The Journal of biological chemistry, 2012. **287**: p. 27106-16.
18. Ribet, D. and P. Cossart, *Post-translational modifications in host cells during bacterial infection.* FEBS letters, 2010. **584**: p. 2748-58.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

19. Audagnotto, M. and M. Dal Peraro, *Protein post-translational modifications: In silico prediction tools and molecular modeling.* Computational and structural biotechnology journal, 2017. **15**: p. 307-319.

20. Biswas, I., J.K. Jha, and N. Fromm, *Shuttle expression plasmids for genetic studies in Streptococcus mutans.* Microbiology (Reading, England), 2008. **154**: p. 2275-82.

21. Le Breton, Y. and K.S. McIver, *Genetic manipulation of Streptococcus pyogenes (the Group A Streptococcus, GAS).* Current protocols in microbiology, 2013. **30**: p. Unit 9D.3.

22. Kavsan, V.M., A.V. Iershov, and O.V. Balynska, *Immortalized cells and one oncogene in malignant transformation: old insights on new explanation.* BMC cell biology, 2011. **12**: p. 23.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.