sgRNA design and *in vitro* nucleolytic analysis of the Cas9-RNP complex for transgene-free genome editing of the *eIF4E1* gene from *Capsicum annuum* L.

Josefanny Tham[1], Alfred Patisenah[1], Tommy Octavianus Soetrisno Tjia[2], Santiago Signorelli[3], Intan Taufik[1], Karlia Meitha[1]

[1]School of Life Sciences and Technology, Institut Teknologi Bandung, West Java 40132, Indonesia
[2]School of Life Science and Technology, Tokyo Institute of Technology, Japan 152-8550
[3]Departamento de Biología Vegetal, Universidad de la República, Uruguay 11200
*Corresponding author: karliameitha@itb.ac.id

ABSTRACT Chili (*Capsicum annuum* L.) is a highly valued vegetable, renowned for its unique taste and aroma. However, chili production faces challenges in meeting the high demand due to infections caused by pathogens such as ChiVMV (potyvirus). Previous studies have suggested that chili *eIF4E1* plays a crucial role in potyvirus gene transcription. Therefore, this study explores the potential of CRISPR-Cas9-based genome editing to enhance chili resistance by introducing premature stop codons or truncated proteins. Two sgRNAs were designed, targeting the first and second intron of the *eIF4E1* gene. The production of Cas9 protein was assessed with varying IPTG concentrations in *Escherichia coli* BL21(DE3), carrying 4xNLS-pMJ915v2-sfGFP plasmid with a TEV protease cut-site at the N terminal. The findings indicate that the optimal IPTG concentration is 500 μM. Purification using an IMAC column confirmed the presence of Cas9 in the initial 2 mL of the eluted fractions, although there were numerous background proteins. Nevertheless, successful formation of Cas9-RNP complexes was achieved for both sgRNAs. The nucleolytic activity of Tag-Cas9 (carrying the MBP-tag) and Cas9 was confirmed through *in vitro* endonuclease activity assays. The next step involve transfecting chili protoplasts with these RNP complexes to edit the chili *eIF4E1* gene.

KEYWORDS CRISPR-Cas9; Endonuclease; Recombinant protein; sgRNA Cas9-RNP

1. Introduction

As of 2019, Indonesia experienced a decrease in the production quantity of *Capsicum annuum* L. compared to 2018 (Ministry of Agriculture of the Republic of Indonesia 2019). Several provinces in Indonesia experienced a decline for more than 21.73%. This decrease was partly due to the high susceptibility of *Capsicum annuum* L. towards pathogens, such as potyvirus (da Costa et al. 2021). The transgene free method to modify this gene could be one of the solutions to increase *Capsicum annuum* L. resistance. *eIF4E1* protein is utilized by potyvirus through potyvirus genome-linked protein or VPg to replicate, so the knockout of this gene is expected to increase the resistance of *Capsicum annuum* L. towards potyvirus (Wang and Krishnaswamy 2012). VPg itself is located at the 5′ terminal of viral genomic RNA. This protein will compete with plant's m7G mRNA to bind with *eIF4E1* protein and initiate viral RNA translation (Piron et al. 2010; Moury et al. 2014). The *eIF4E1* protein is approximately 25 kDa that will bind through hydrogen bond with m7G mRNA (Tomoo et al. 2003). Once the bond is formed, *eIF4E1* will

trigger *eIF4A* and *eIF4G* to form *eIF4F* complex (Wang and Krishnaswamy 2012). The *eIF4E1* itself is also known to have an isoform called *eIF4(iso)E*. Several studies show that the presence of only one of these two isoforms in mutant plants do not affect plant growth and fertility (Duprat et al. 2002).

One of the transgene-free methods that has been widely used is the CRISPR-Cas9 system (Aliaga-Franco et al. 2019). This method is highly favored due to the minimum introduction of foreign genes to the targeted organism and easy handling (Anzalone et al. 2020). Other gene editing tools such as zinc finger nuclease (ZFN) and transcription activator-like effector nuclease (TALEN) have several disadvantages compared to CRISPR-Cas9 system, such as the troublesome design process and low working efficiency (Petersen and Niemann 2015). Clustered regularly interspaced short palindromic repeats (CRISPR) belongs to a DNA family in the genome of prokaryotes that plays a role in bacteria immune defense against bacteriophage (Aliaga-Franco et al. 2019). The CRISPR-Cas9 system is composed of Cas9 with an endonuclease

catalytic domain and specific guide-RNA that will determine the specificity of this system. Together they form Cas-RNP (Wright et al. 2016). Later this RNP will introduce a double-stranded break (DSB) and that will be repaired in the organism through non-homologous end joining (NHEJ) and resulting insertion-deletion. The inactivation of targeted gene will occur due to the sequence disruption by frameshift reading (Hsu et al. 2014; Khan et al. 2018).

The sgRNA consists of a spacer sequence that complements the targeted gene and a scaffold sequence that forms a bond with the REC domain at Cas9 (Palermo et al. 2018). This scaffold sequence, containing the crRNA and tracrRNA, is highly conserved and connected through a stem-loop secondary structure (Khan et al. 2018). For Cas9 protein to function, it must locate the protospacer adjacent motif (PAM) adjacent to the target sequence (Cribbs and Perera 2017). The complementation of the targeted sequence and sgRNA activates the Cas9 nuclease domain, breaking the hydrogen bond three bases upstream of PAM. The RuvC domain attacks the sequence identical to sgRNA, while the HNH domain attacks the complementary sequence of the sgRNA (Qi et al. 2013).

The effectiveness of the CRISPR-Cas9 system relies heavily on the design of the sgRNA to be used. This sgRNA dictates the locus of the targeted genes, with an expected GC content of 20–60% GC, including four purine residues at the 3′ terminal of the spacer sequence to enhance its binding capacity to the REC domain at Cas9 (Liang et al. 2016). Studies have demonstrated higher effectiveness of Cas9 with sgRNA targeting transcribed sequences compared to regulatory sequence (Bortesi et al. 2016; Budiani et al. 2019).

In this study, the eIF4E1 gene from chili pepper (*Capsicum annuum* L.) was used as a model for genome editing through targeted gene knockout using the CRISPR/Cas9 system. The primary objective was to validate the efficiency of the designed sgRNA and the produced Cas9 protein before advancing to *in vivo* testing. It is expected that the results of this study can be applied to develop potyvirus-resistant superior chili strains in Indonesia.

2. Materials and Methods

2.1. *Plasmids, bacterial strains, and culture media*

Escherichia coli BL21(DE3) was transformed with 4xNLS-pMJ915v2-sfGFP (pCas9) expression vector was obtained from Addgene Plasmid Repository (#88921, Jenifer Doudna deposit). The pGEM-T Easy plasmid (Promega, Madison, USA) was used to clone sgRNA and *eIF4E1* CDS. *E. coli* DH5α was used as cloning strain. Each resulting construct was confirmed by screening on LB containing 100 ppm ampicillin and colony PCR. *E. coli* BL21(DE3) was first developed by F. William Studier and Barbara A. Moffatt. In this strain cannot be found lon and ompT protease at the membrane which will support the protein expression due to the minimum degradation level

(Jeong et al. 2015). This protein is also regulated by T7 promoter that requires T7 RNA polymerase. This RNA polymerase is tightly regulated by lacUV5 promoter that is induced by IPTG (Zhang et al. 2015).

The medium used in this study was Luria-Bertani (LB) complex medium (Himedia, Maharashtra, India). LB agar medium (solid) contains the same composition as liquid LB except that 15 g/L of Bacto agar (Himedia, Maharashtra, India) was added. A final ampicillin concentration of 100 ppm was also used in LB medium when growing transformant strains.

2.2. *sgRNA design and production*

In silico sgRNA design was performed with the tools Cas-Designer (Park et al. 2015), Cas-OFFinder (Bae et al. 2014), and RNAFold (Kerpedjiev et al. 2015). The *eIF4E1* gene sequence from *Capsicuum anuum* L. (GenBank id: AF521965.1) was used as the base for designing the sgRNA. Cas-Designer was used to generate target sequences (spacers) that had appropriate parameters (GC content, PAM, and out-of-frame score). Cas-OFFinder to select sgRNAs based on the number of possible off-targets. RNAfold is used to predict the secondary structure of sgRNAs to match the general secondary structure of active sgRNAs.

The sgRNA template used was as follows (Ma et al. 2021).
5′- NNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC-3′

2.3. *Cell growth and expression of recombinant Cas9 protein in E. coli*

Inoculum was prepared by growing bacteria on LB ampicillin medium overnight. A portion of the inoculum culture was inoculated into a shake flask (5% inoculum) and grown until an OD_{600} value of 0.6–0.8 was reached. The culture was then induced by the addition of IPTG, and the culture temperature was lowered to 16 °C for 16 h. IPTG concentration optimization was performed by induction at varying isopropyl β- d-1-thiogalactopyranoside (IPTG) concentrations; 0.00, 0.25, 0.5, 0.75, and 1.00 mM. Cell samples were harvested by centrifugation (6,000 × g, 5 min, 4 °C). The pellet was dissolved in cold lysis/binding buffer (0.01 M imidazole, PBS 1×) at 100 μL for every 0.01 gram of pellet (whole-cell fraction). Cells were then lysed by sonication at 30% power, 1/3 pulse (3s on, 6s off) for 10 min. The soluble fraction (cytoplasmic crude extract) and the undissolved fraction (inclusion bodies, cell debris, contaminants) were separated by centrifugation (14,000 × g, 55 min, 4 °C). The soluble fraction was in the supernatant, while the insoluble fraction was pelleted. The pellet of the undissolved fraction is resuspended with cold lysis/binding buffer before analysis. The fractions can then be analyzed by SDS-PAGE.

2.4. Cas9 purification with IMAC

Cas9 protein purification was performed with immobilized metal affinity chromatography (IMAC), specifically His-Pur™ Ni-NTA chromatography cartridges gravity column (Thermo Fisher, Waltham, USA) equilibrated with binding buffer. The protein soluble fraction was filtered and loaded into 10 mL of column, followed by washing with wash buffer (0.02 M imidazole, PBS 1×), and elution with elution buffer (0.3 M imidazole, PBS 1×). The elution fraction containing proteins was dialyzed in dialysis buffer (400 mM Tris pH 7.5, 200 mM KCl, 10 mM $MgCl_2$, and 3% glycerol) overnight along with TEV protease treatment (NEB, Ipswich, UK) in dialysis tubes. The dialysis results were concentrated by ultrafiltration using Amicon® Ultra-4 Centrifugal Filter MWCO 100 kDa (Merck, Darmstardt, Germany). Purification results were analyzed by Bradford assay, SDS-PAGE, and densitometry. Purification results were stored at –80 °C.

2.5. Endonuclease assay substrate production

The *eIF41* CDS was obtained from *Capsicum annuum* L. leaf samples. Total RNA was isolated using TRIsureTM (Bioline, UK) and cDNA synthesis kit (TOYOBO, Osaka, Japan) was used. *eIF4E1* CDS was then amplified from the cDNA using PCR with primers targeting the gene. PCR amplified *eIF4E1* CDS was used as dsDNA substrate for Cas9 endonuclease assay.

2.6. sgRNA production

Designed synthetic sgDNA (sgRNA encoding DNA) were synthesized by Macrogen (Seoul, Korea). sgRNA was produced by *in vitro* transcription and purified using MEGAscriptTM T7 Transcription kit (Thermo Fisher, Carlsbad, USA) and purified by LiCl precipitation method. sgRNA was stored at –80 °C.

2.7. Cas9-RNP complex production and in vitro endonuclease activity assay

A test reaction volume of 20 μL was used, consisting of 1 μg Cas9, 1 μg sgRNA and 2 μL 10× Cas9 reaction buffer (0.2 M HEPES, 0.1 M $MgCl_2$, 5 mM DTT, 1.5 M KCl), and 100 ng *eIF4E1* CDS PCR product dissolved in NFW. The mixture was gently resuspended and reacted at 37 °C for 1 h. The reaction was stopped by heating at 65 °C for 10 min. The reaction results were analyzed by gel electrophoresis.

3. Results and Discussion

3.1. The design of sgRNA

The sgRNA is the easiest factor to engineer to maximize the targeting specificity of CRISPR/Cas9. CRISPR/Cas9 can be programmed by simply replacing the 20-nt spacer sequence on the sgRNA according to the desired target sequence (Anzalone et al. 2020). Characteristics of a good sgRNA include: (1) Targeting DNA sites that have a PAM (5′-NGG-3′) sequence at the downstream position (An-

ders et al. 2014); (2) has low off-target potential (Fu et al. 2013); (3) target the exon region of the target gene to increase the probability of knockout mutations; (4) has at least a duplex repeat-antirepeat secondary structure, and stem loop 1 (Jiang and Doudna 2017); (5) contains a GC spacer content of 20-80% (Schindele et al. 2020); (6) and have a high probability of frameshift (out-of-frame) mutations (> 66%) (Bae et al. 2014).

The sgRNA spacer design can be easily done using Cas-Designer web-tools (Park et al. 2015). Cas-Designer can quickly generate spacer sequences with optimal parameters based on queries of target genes and target organisms (Park et al. 2015). The selected spacer sequences can be further analysed in terms of off-target activity using Cas-OFFinder web-tools (Bae et al. 2014) and in terms of secondary structure using RNAFold (Kerpedjiev et al. 2015).

Our sgRNAs were designed to target the *eIF4E1* DNA sequence of chili (*Capsicum annuum*) based on sequences from the NCBI GenBank database (accession number AF521965.1). The off-target potential is indicated by the number of mismatch base pairing with similar sequence within the genome. This is important as Cas9 is known to tolerate up to 2-base mismatches (Anderson et al. 2015). Computation using Cas-Designer resulted in 15 Cas9 target candidates (Table 1) that were screened for optimal sgRNA parameter values. Here, we have filtered the result for zero off-target mismatch and single on-target match.

When programmable nucleases including CRISPR are used, 1-3 bp deletions or 1 bp insertions are frequently induced via the nonhomologous end-joining (NHEJ) repair pathway, whereas deletions involving microhomologies of more than 2 bases are frequently introduced via the microhomology-mediated end joining (MMEJ) pathway (Bae et al. 2014). Out-of-frame (OOF) score predicts the mutation patterns induced by the MMEJ pathway and estimates how frequently undesirable in-frame deletions occur. To maximize desirable OOF deletions in a protein-coding sequence as much as possible, target regions with high OOF scores should be chosen (Bae et al. 2014). Hence, Cas9 target with OOF score lower than 66 must be avoided. In addition, Cas9 target in the upstream region within the gene would be more favourable to produce knockout mutation. For that reason, Cas9 target in the position of 196 and 300 (Table 1) were selected and developed into two sgRNAs, $sgRNA_{196}$ and $sgRNA_{300}$.

The sgRNAs were fused with T7 promoter for high yield *in vitro* transcription (IVT) method. To ensure high yield of sgRNA, at least two guanine (G) must be added to the 5′ end of each ORF (Kuzmine et al. 2003). This addition is predicted to not interfere with the sgRNA function as RNA:DNA complementarity and binding at the 5′ PAM-distal end is not required for nuclease activity (Anderson et al. 2015). The expression cassettes of sgRNAs used in this study are dislayed in Figure 1.

The secondary structure prediction results of $sgRNA_{196}$ and $sgRNA_{300}$ (Figure 2) showed a marginal spacer interaction with the scaffold and a preserved essen-

TABLE 1 Potential Cas9 target on *Capsicum annuum eIF4E1* gene exons generated using Cas-Designer and filtered for zero mismatch.

Cas9 Target (5′ → 3′)	Position*	Exon	%GC (w/o PAM)	OOF score	Mismatch		
					0	1	2
TAAGCAATGGAAGCAGTTTCTGG	582	4	40	80	1	0	0
CGTTTCATTTGCAGCATTCTTGG	542	3	40	76.7	1	0	0
CATTTTCCATGTCCCTCCATTGG	380	2	45	76.1	1	0	0
GCTTGTTTCGATTTCGCCTCTGG	196	1	50	72.4	1	0	0
CTCCCACAACTAACTTGCTTGGG	300	2	45	72	1	0	0
AGTAGTTAGTGTCAGAGGTAAGG	498	3	40	69.7	1	0	0
GCTCCCACAACTAACTTGCTTGG	301	2	50	69	1	0	0
TTTCTGGTTTGATAATCCAGAGG	180	1	35	65.3	1	0	0
TGCTGCAAATGAAACGGCTCAGG	549	3	50	65.3	1	0	0
CAAGAATGCTGCAAATGAAACGG	543	3	35	65.2	1	0	0
TAGAGCATTCATGGACTTTCTGG	164	1	40	63.1	1	0	0
GGAGCAGTAGTTAGTGTCAGAGG	493	3	50	62.7	1	0	0
CTCCACTGTTGAAGATTTTTGGG	255	1	35	57.6	1	0	0
GACGTTTGATGAAGCTGAGAAGG	30	1	45	57.2	1	0	0
AGCATCCATTAGAGCATTCATGG	155	1	40	56.8	1	0	0

*Position corresponds to CDS nucleotide of the gene.
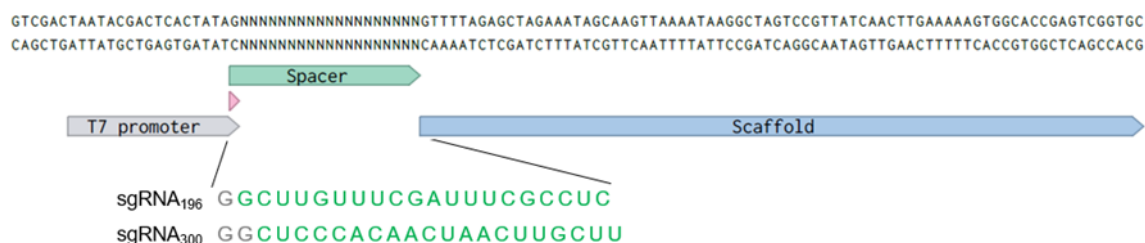
FIGURE 1 Expression cassette of sgRNA designed for T7-dependent *in vitro* transcription method. To ensure high yield of sgRNA, G nucleotides (gray font) was added to the 5′ end of each sgRNA spacer (green font) so that a total of two Gs. Image was generated by Benchling.com with modifications.
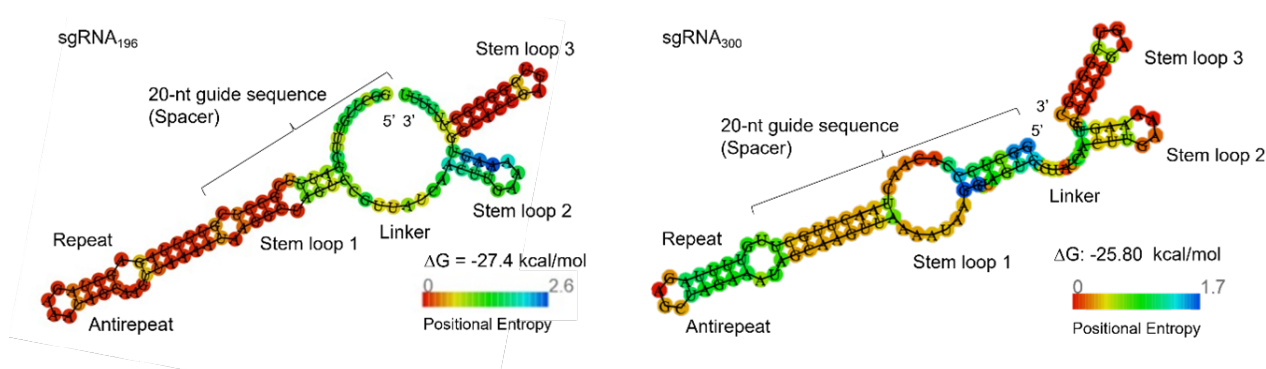
FIGURE 2 Predicted RNA secondary structure of sgRNA$_{196}$ and sgRNA$_{300}$. All theoretical sgRNA secondary structures that are important for Cas9 activity were observed in both sgRNA designs. The color indicates the positional entropy value which is inversely proportional to the stability of the structure. The RNA structure prediction was carried out using RNAFold (Kerpedjiev et al. 2015).

tial sgRNA secondary structure. All sgRNA secondary structure features including repeat-antirepeat duplex, stem loop 2, and 3 were observed in both sgRNAs (Jiang and Doudna 2017). The stem loop 2 and 3 structures showed high stability based on positional entropy values. The stem loop 1 structure was also observed, although it was not formed solely from scaffold sequences, but via the interaction between the spacer and the scaffold. These results suggest that both sgRNAs would perform well in cleaving target DNA. Graphical illustration of Cas9 cut site using sgRNA$_{196}$ and sgRNA$_{300}$ is shown in Figure 3.
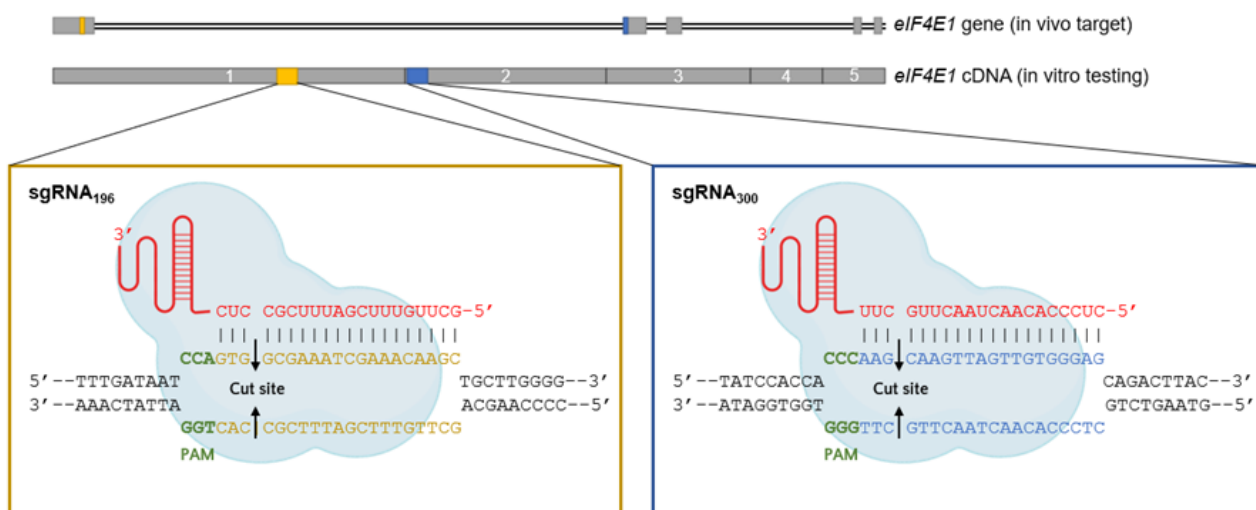
FIGURE 3 Graphical representation Cas9 cut site of the designed sgRNA$_{196}$ (yellow) and sgRNA$_{300}$ (blue) on *Capsicum annuum eIF4E1* gene and cDNA substrate.

### 3.2. Cas9 protein expression

Escherichia coli strains commonly used to produce recombinant Cas9 are BL21(DE3) and Rosetta(DE3) (Liang et al. 2018; Carmignotto and Azzoni 2019; Qiao et al. 2019). Expression of recombinant proteins with these two strains is suitable using T7 expression vectors and induction of expression with IPTG (Hayat et al. 2018). The Rosetta strain has the advantage over BL21 of being able to express rare codons better (Hayat et al. 2018). However, Rosetta(DE3) is known to produce lower Cas9 pro-

tein yields than BL21(DE3) that might be due to additional metabolic burden from pRARE plasmid expression (Carmignotto and Azzoni 2019).

Cas9 expression plasmid acquired from addgene repository (#88921) (Figure 4) was utilized to produce the Cas9 protein. This plasmid is designed for bacterial production of tagged Cas9-GFP fusion protein for eukary-
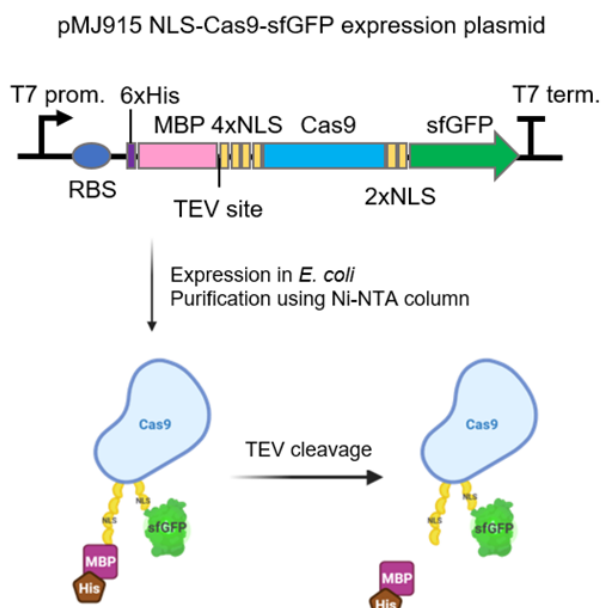
FIGURE 4 An engineered pMJ915 expression vector (Addgene Plasmid ID: 88921) was harnessed to achieve co-expression of NLS-Cas9-sfGFP in *E. coli*. The nuclear localization signal (NLS) is important for eukaryotic genome editing. The Cas9 were prepared by Ni-NTA column purification and cleavage of His-MBP tags using a TEV protease.
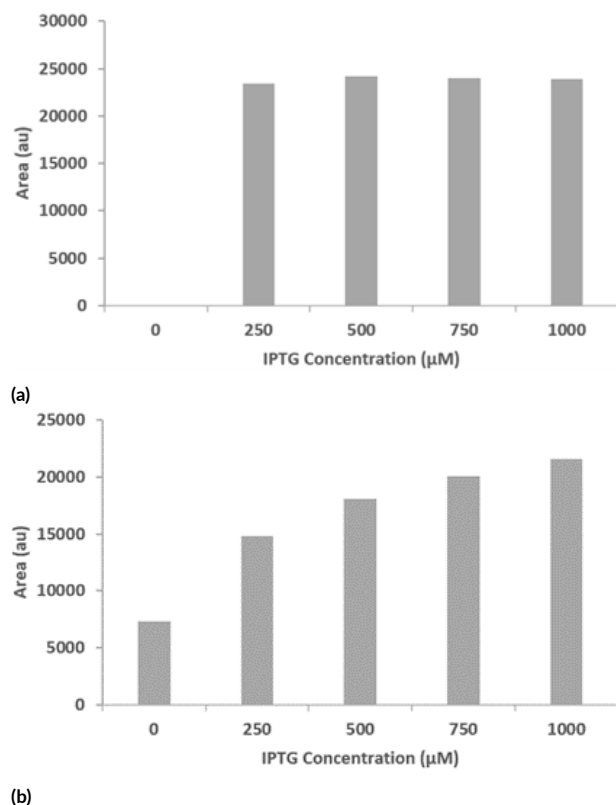
FIGURE 5 Cas9 protein band density measurement from (a) cytoplasmic fraction and (b) insoluble fraction using ImageJ showed 500 μM IPTG produced highest Cas9 band density.

otic genome editing. The Cas9 fusion protein contains MBP purification tag, 6xHis purification tag, SV40 nuclear localization signal (NLS), and green fluorescent protein (GFP) (Staahl et al. 2017). SV40 NLS sequence is a highly conserved signal peptide and enables the translocation of Cas9 into cell nucleus (Niopek et al. 2014; Groves et al. 2019; Lu et al. 2021). sfGFP acts as reporter to confirm Cas9 entering the nucleus (Dinh and Bernhardt 2011). Treatment using TEV protease cleaves purification tags from the rest of the Cas9 fusion protein.

Our first step was to evaluate the optimal IPTG concentration for Cas9 fusion protein expression using BL21(DE3) as host. The protein expression was performed in 50 mL conical tube with 15 mL of Luria-Bertani (LB) broth. The induction occurred during exponential growth phase via the addition of IPTG when OD_{600} nm of 0.6–0.8 was reached and the protein expression was performed at 18 °C for 16 hours. Five different IPTG concentrations (0, 250, 500, 750, and 1000 µM) were evaluated. The cell soluble cytoplasmic and insoluble fraction of each condition was analyzed using SDS-PAGE and densitometry. The results are presented in Figure 5.

The results obtained indicated that Cas9 protein expression at low temperature requires IPTG concentration of 500 µM for optimal expression (Figure 5a). The Cas9 protein was also observed in the insoluble fraction that increases as the IPTG concentration increases (Figure 5b). This might be due to misfolded Cas9 protein aggregation into insoluble inclusion body (Wingfield 2015; Bhatwa et al. 2021). As Cas9 protein expression rate increases, metabolic burden on *E. coli* cells rises and protein aggregations are more likely to happen (Donovan et al. 1996; Bhatwa et al. 2021). These explain why soluble Cas9 protein does not increase while the insoluble Cas9 increases when IPTG concentrations rise above 500 µM. The band density graph showed a consistent result with literature suggested IPTG concentration to induce Cas9 protein expression. The band density graph showed a consistent result with literature suggested IPTG concentration to induce Cas9 protein expression (Liang et al. 2018).

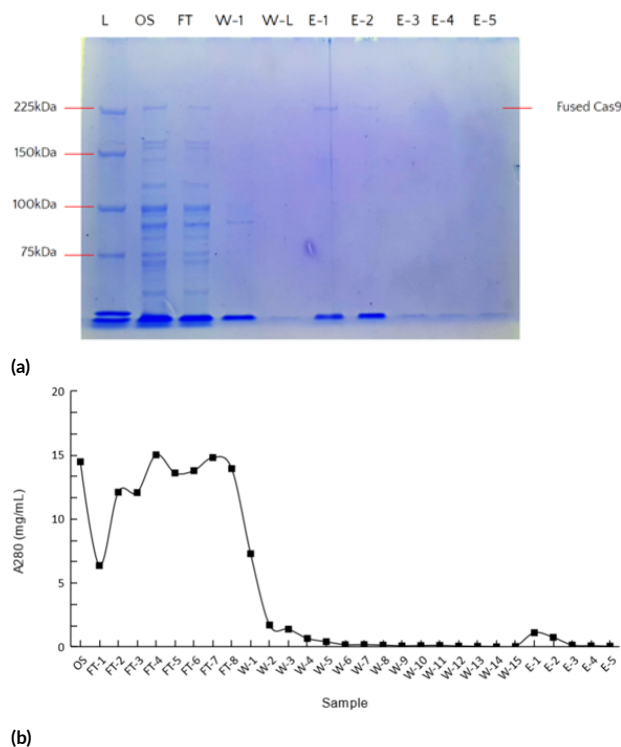### 3.3. One-step Cas9 protein purification using IMAC

Here, we attempt to yield reasonably pure Cas9 protein by using only one-step purification for efficiency. The Cas9 protein was purified from crude extract soluble fraction using IMAC (Ni-NTA resin) by taking advantage of His-tag present at the N-terminus of the Cas9 fusion protein.

The Cas9 protein expression was carried in shake flask with 100 mL LB broth following expression condition

as mentioned in the previous section and using the optimal IPTG concentration for induction. The purification was evaluated using spectrophotometry (A_{280nm}) and SDS-PAGE. The results are presented in Figure 6 and Table 2.

The gel shows the presence of a 225 kDa band in the fractions collected in the flowthrough and during the elution step, which corresponds to the Cas9 protein. These results indicate that Cas9 protein binding to the IMAC was not optimal and this explains the low overall recovery of Cas9 protein. Overloading of protein during purification and low buffering capacity of the binding buffer might be the contributing factor for the low recovery, since the purification condition was not optimized in this study.

Result (Figure 6b) indicates that *E. coli* BL21(DE3) produced plenty background proteins that were co-purified along with Cas9 protein that are barely visible in the gel due to low amounts of sample loading but become visible by closer inspection. This result is commonly seen in single purification-method, especially His-tag based IMAC (Cao and Lin 2009; Andersen et al. 2013). Most of these proteins belong to stress-responsive proteins with repetitive histidine residues. Other than that, ArnA and SlyD are two proteins that are commonly found due to certain binding affinity towards Ni^{2+} (Williams et al. 2005; Bolanos-Garcia and Davies 2006). It takes almost the same level of imidazole concentration (> 80 mM) as in the elution buffer to remove these proteins (Bolanos-Garcia and Davies 2006). The utilization of two purification method

(a)

(b)

FIGURE 6 Protein analysis of protein purification process using IMAC (Ni-NTA resin). (a) SDS PAGE analysis of protein during each purification step. (b) A_{280} (mg/mL) chromatogram of protein during each purification step (OS: Original sample, FT: Flow-through, W: Wash, E: Elution).

TABLE 2 Recovery of protein obtained during each Ni-NTA purification step.

	Total protein (mg)	Recovery (%)
Feed	115.920	—
Washing	113.879	98.239
Elution	2.051	1.769
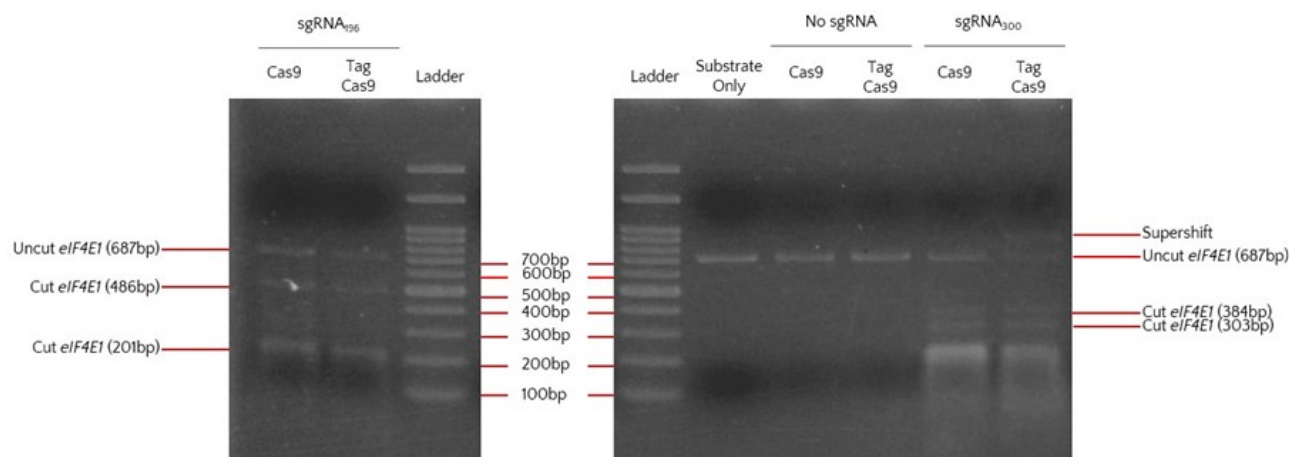
FIGURE 7 Visualisation of Cas9-RNP *in vitro* endonuclease activity assay towards *Capsicum annuum* L. *eIF4E1* CDS as the target. Two different Cas9: Cas9 and Tag-Cas9 was assayed its endonuclease activity. *eIF4E1* CDS without addition of Cas9 protein and sgRNA was used as control. Both Cas9 and Tag-Cas9 showed no cleavage activity toward *eIF4E1* CDS without the presence of sgRNA. The cleavage of *eIF4E1* CDS was performed successfully by Cas9 and Tag-Cas9 with two different sgRNAs (Agarose 2%+TBE 1×; Ladder 100 bp (Geneaid, Taiwan)).

and low background protein strain *E. coli* (LOBSTR) are highly suggested to avoid the occurrence of background proteins (Andersen et al. 2013; Flottmann et al. 2022). Furthermore, optimization of binding and wash buffer imidazole concentrations might be beneficial to enhance purity as other studies have demonstrated (Carmignotto and Azzoni 2019).

The purified Cas9 protein was then dialysed, treated with TEV protease, and concentrated by centrifugal filtration before being evaluated for its DNA cleavage activity.

3.4. Cas9 activity assay

The activity of the purified Cas9 protein produced in both LB broth was tested by *in vitro* DNA cleavage. The Cas9 protein and two different sgRNA (sgRNA$_{196}$ and sgRNA$_{300}$) containing the *C. annuum eIF4E* gene target sequence were incubated with linear dsDNA encoding the *eIF4E* CDS (687 bp). sgRNA$_{196}$ targets the substrate dsDNA into 201 bp and 486 bp fragments, while sgRNA$_{300}$ targets the dsDNA into 303 bp and 384 bp fragments. Both tag-removed (Cas9) and unremoved Cas9 (TagCas9) was evaluated. The results shown in Figure 7 indicate that both Cas9 protein did not cleave DNA in the absence of sgRNA. Unspecific Cas9 nuclease activity without the sgRNA is found only in the presence of Mn^{2+} cofactor (Sundaresan et al. 2017). Due to the lack of Mn^{2+} in our Cas9 reaction buffer, this activity does not occur.

The results showed that ribonucleoprotein complexes formed by both Cas9 and TagCas9 and both sgRNA tested were able to cleave the *eIF4E1* gene sequence. This indicates that Cas9 protein activity is highly robust as it could still performs well in the presence of protein impurities and large protein fusion. However uncut DNA was still observed after incubation which was commonly found in double-stranded oligonucleotide target (Anders and Jinek 2014; Mehravar et al. 2019). Overall, these results validated the in silico sgRNAs design scheme used to design both sgRNAs. Further *in vivo* genome editing assay in *C. annuum* is necessary to evaluate off-target activity of the designed sgRNA.

4. Conclusions

The sgRNA targeting the 196 and 300 base of *Capsicum annuum* L. *eIF4E1* was successfully designed that satisfied the prerequisite of good quality sgRNA. Cas9 Protein was successfully produced with 500 µM as optimum IPTG concentration. The Cas9-RNP complex was successfully produced and *in vitro* endonuclease activity towards *eIF4E1* of *Capsicum annuum* L. was confirmed.

Authors' contributions

KM conceived the study. All authors designed the study. JT, AP and TOST conducted the experiments and data analysis. JT and AP prepared the manuscript. SS, IT, and KM supervised the experiments and directed manuscript writing. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no conflict of interest.

References

Aliaga-Franco N, Zhang C, Presa S, Srivastava AK, Granell A, Alabadí D, Sadanandom A, Blázquez

MA, Minguet EG. 2019. Identification of transgene-free CRISPR-edited plants of rice, tomato, and *Arabidopsis* by monitoring DsRED fluorescence in dry seeds. Front. Plant Sci. 10:1150. doi:10.3389/fpls.2019.01150.

Anders C, Jinek M. 2014. *In vitro* enzymology of cas9. Methods Enzymol. 546:1–20. doi:10.1016/B978-0-12-801185-0.00001-5.

Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513(7519):569–7. doi:10.1038/nature13579.

Andersen KR, Leksa NC, Schwartz TU. 2013. Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins Struct. Funct. Bioinforma. 81(11):1857–61. doi:10.1002/prot.24364.

Anderson EM, Haupt A, Schiel JA, Chou E, Machado HB, Strezoska Ž, Lenger S, McClelland S, Birmingham A, Vermeulen A, Smith AVB. 2015. Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. J. Biotechnol. 211:56–65. doi:10.1016/j.jbiotec.2015.06.427.

Anzalone AV, Koblan LW, Liu DR. 2020. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. 38(7). doi:10.1038/s41587-020-0561-9.

Bae S, Kweon J, Kim HS, Kim JS. 2014. Microhomology-based choice of Cas9 nuclease target sites. Nat. Methods 11(7). doi:10.1038/nmeth.3015.

Bhatwa A, Wang W, Hassan YI, Abraham N, Li XZ, Zhou T. 2021. Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. Front. Bioeng. Biotechnol. 9. doi:10.3389/fbioe.2021.630551.

Bolanos-Garcia VM, Davies OR. 2006. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. Biochim. Biophys. Acta - Gen. Subj. 1760(9):1304–13. doi:10.1016/j.bbagen.2006.03.027.

Bortesi L, Zhu C, Zischewski J, Perez L, Bassié L, Nadi R, Forni G, Lade SB, Soto E, Jin X, Medina V, Villorbina G, Muñoz P, Farré G, Fischer R, Twyman RM, Capell T, Christou P, Schillberg S. 2016. Patterns of CRISPR/Cas9 activity in plants, animals, and microbes. Plant Biotechnol. J. 14(12):2203–2216. doi:10.1111/pbi.12634.

Budiani A, Nugroho IB, Sari DA, Palupi I, Putranto RA. 2019. CRISPR/Cas9-mediated knockout of an oil palm defense-related gene to the pathogenic fungus *Ganoderma boninense*. Indones. J. Biotechnol. 24(2):101–105. doi:10.22146/ijbiotech.52170.

Cao H, Lin R. 2009. Quantitative evaluation of his-tag purification and immunoprecipitation of tristetraprolin and its mutant proteins from transfected human cells. Biotechnol. Prog. 25(2):461–467.

doi:10.1002/btpr.121.

Carmignotto GP, Azzoni AR. 2019. On the expression of recombinant Cas9 protein in *E. coli* BL21(DE3) and BL21(DE3) Rosetta strains. J. Biotechnol. 306:62–70. doi:10.1016/j.jbiotec.2019.09.012.

Cribbs AP, Perera SM. 2017. Science and bioethics of CRISPR-CAS9 gene editing: An analysis towards separating facts and fiction. Yale J. Biol. Med. 90(4):625–634.

da Costa DV, Paiva CLdA, Bento CdS, Sudré CP, Cavalcanti TFM, Gonçalves LSA, Viana AP, Rodrigues R. 2021. Breeding for pepper yellow mosaic virus resistance and agronomic attributes in recombinant inbred lines of chili pepper (*Capsicum baccatum* L.) using mixed models. Sci. Hortic. (Amsterdam). 282:110025. doi:10.1016/j.scienta.2021.110025.

Dinh T, Bernhardt TG. 2011. Using superfolder green fluorescent protein for periplasmic protein localization studies. J. Bacteriol. 193(18):4984–7. doi:10.1128/JB.00315-11.

Donovan RS, Robinson CW, Click BR. 1996. Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. J. Ind. Microbiol. 16(3):145–54. doi:10.1007/BF01569997.

Duprat A, Caranta C, Revers F, Menand B, Browning KS, Robaglia C. 2002. The *Arabidopsis* eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. Plant J. 32(6):927–34. doi:10.1046/j.1365-313X.2002.01481.x.

Flottmann F, Pohl GM, Gummert J, Milting H, Brodehl A. 2022. A detailed protocol for expression, purification, and activity determination of recombinant SaCas9. STAR Protoc. 3(2):101276. doi:10.1016/j.xpro.2022.101276.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31:822–826. doi:10.1038/nbt.2623.

Groves NR, Mckenna JF, Evans DE, Graumann K, Meier I. 2019. A nuclear localization signal targets tail-anchored membrane proteins to the inner nuclear envelope in plants. J. Cell Sci. 132(7):jcs226134. doi:10.1242/jcs.226134.

Hayat SMG, Farahani N, Golichenari B, Sahebkar A. 2018. Recombinant protein expression in *Escherichia coli* (*E.coli*): What we need to know. Curr. Pharm. Des. 24(6):718–725. doi:10.2174/1381612824666180131121940.

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157(6):1262–1278. doi:10.1016/j.cell.2014.05.010.

Jeong H, Kim HJ, Lee SJ. 2015. Complete genome sequence of *Escherichia coli* strain BL21. Genome Announc. 3(2):e00134–15.

doi:10.1128/genomeA.00134-15.

Jiang F, Doudna JA. 2017. CRISPR-Cas9 structures and mechanisms. Annu. Rev. Biophys. 46:505–529. doi:10.1146/annurev-biophys-062215-010822.

Kerpedjiev P, Hammer S, Hofacker IL. 2015. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics 31(20):3377–3379. doi:10.1093/bioinformatics/btv372.

Khan S, Mahmood MS, Rahman SU, Zafar H, Habibullah S, Khan Z, Ahmad A. 2018. CRISPR/Cas9: The Jedi against the dark empire of diseases. J. Biomed. Sci. 25:29. doi:10.1186/s12929-018-0425-5.

Kuzmine I, Gottlieb PA, Martin CT. 2003. Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. J. Biol. Chem. 278(5):2819–23. doi:10.1074/jbc.M208405200.

Liang G, Zhang H, Lou D, Yu D. 2016. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Sci. Rep. 6:21451. doi:10.1038/srep21451.

Liang Z, Chen K, Zhang Y, Liu J, Yin K, Qiu JL, Gao C. 2018. Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 *in vitro* transcripts or ribonucleoproteins. Nat. Protoc. 13(3):413–430. doi:10.1038/nprot.2017.145.

Lu J, Wu T, Zhang B, Liu S, Song W, Qiao J, Ruan H. 2021. Types of nuclear localization signals and mechanisms of protein import into the nucleus. Cell Commun. Signal. 19(1):60. doi:10.1186/s12964-021-00741-y.

Ma L, Sun Y, Ruan X, Huang PC, Wang S, Li S, Zhou Y, Wang F, Cao Y, Wang Q, Wang Z, Kolomiets MV, Gao X. 2021. Genome-wide characterization of jasmonates signaling components reveals the essential role of ZmCOI1a-ZmJAZ15 action module in regulating maize immunity to Gibberella stalk rot. Int. J. Mol. Sci. 22(2):870. doi:10.3390/ijms22020870.

Mehravar M, Shirazi A, Mehrazar MM, Nazari M. 2019. *In vitro* pre-validation of gene editing by CRISPR/Cas9 ribonucleoprotein. Avicenna J. Med. Biotechnol. 11(3):259–263.

Ministry of Agriculture of the Republic of Indonesia. 2019. Data Lima Tahun Terakhir Sektor Hotrikultura Indonesia [Data for the Last Five Years of the Indonesian Hotriculture Sector]. Ministry of Agriculture of the Republic of Indonesia, Jakarta.

Moury B, Charron C, Janzac B, Simon V, Gallois JL, Palloix A, Caranta C. 2014. Evolution of plant eukaryotic initiation factor 4E (*eIF4E*) and potyvirus genome-linked protein (VPg): A game of mirrors impacting resistance spectrum and durability. Infect. Genet. Evol. 27:472–80. doi:10.1016/j.meegid.2013.11.024.

Niopek D, Benzinger D, Roensch J, Draebing T, Wehler P, Eils R, Di Ventura B. 2014. Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. Nat. Commun. 5:4404.

doi:10.1038/ncomms5404.

Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, Doudna JA, McCammon JA. 2018. Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. Q. Rev. Biophys. 51:e91. doi:10.1017/S0033583518000070.

Park J, Bae S, Kim JS. 2015. Cas-Designer: A web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics 31(24):4014–4016. doi:10.1093/bioinformatics/btv537.

Petersen B, Niemann H. 2015. Molecular scissors and their application in genetically modified farm animals. Transgenic Res. 24(3):381–396. doi:10.1007/s11248-015-9862-z.

Piron F, Nicolaï M, Minoïa S, Piednoir E, Moretti A, Salgues A, Zamir D, Caranta C, Bendahmane A. 2010. An induced mutation in tomato *eiF4E* leads to immunity to two potyviruses. PLoS One 5(6):e11313. doi:10.1371/journal.pone.0011313.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. Cell 152(5):1173–1183. doi:10.1016/j.cell.2013.02.022.

Qiao J, Li W, Lin S, Sun W, Ma L, Liu Y. 2019. Co-expression of Cas9 and single-guided RNAs in *Escherichia coli* streamlines production of Cas9 ribonucleoproteins. Commun. Biol. 2:161. doi:10.1038/s42003-019-0402-x.

Schindele P, Wolter F, Puchta H. 2020. CRISPR guide RNA design guidelines for efficient genome editing. Methods Mol. Biol. 2166:331–342. doi:10.1007/978-1-0716-0712-1_19.

Staahl BT, Benekareddy M, Coulon-Bainier C, Banfal AA, Floor SN, Sabo JK, Urnes C, Munares GA, Ghosh A, Doudna JA. 2017. Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. Nat. Biotechnol. 35(5):431–434. doi:10.1038/nbt.3806.

Sundaresan R, Parameshwaran HP, Yogesha SD, Keilbarth MW, Rajan R. 2017. RNA-independent DNA cleavage activities of Cas9 and Cas12a. Cell Rep. 21(13):3728–3739. doi:10.1016/j.celrep.2017.11.100.

Tomoo K, Shen X, Okabe K, Nozoe Y, Fukuhara S, Morino S, Sasaki M, Taniguchi T, Miyagawa H, Kitamura K, Miura KI, Ishida T. 2003. Structural features of human initiation factor 4E, studied by X-ray crystal analyses and molecular dynamics simulations. J. Mol. Biol. 328(2):365–383. doi:10.1016/S0022-2836(03)00314-0.

Wang A, Krishnaswamy S. 2012. Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. Mol. Plant Pathol. 13(7):795–803. doi:10.1111/j.1364-3703.2012.00791.x.

Williams GJ, Breazeale SD, Raetz CR, Naismith JH. 2005.

Structure and function of both domains of ArnA, a dual function decarboxylase and a formyltransferase, involved in 4-amino-4-deoxy-L-arabinose biosynthesis. J. Biol. Chem. 280(24):23000–8. doi:10.1074/jbc.M501534200.

Wingfield PT. 2015. Overview of the purification of recombinant proteins. Curr. Protoc. Protein Sci. 80:6.1.1–6.1.35. doi:10.1002/0471140864.ps0601s80.

Wright AV, Nuñez JK, Doudna JA. 2016. Biology and applications of CRISPR systems: Harnessing nature's toolbox for genome engineering. Cell 164(1-2):29–44. doi:10.1016/j.cell.2015.12.035.

Zhang Z, Kuipers G, Niemiec Ł, Baumgarten T, Slotboom DJ, de Gier JW, Hjelm A. 2015. High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG. Microb. Cell Fact. 14(1):142. doi:10.1186/s12934-015-0328-z.