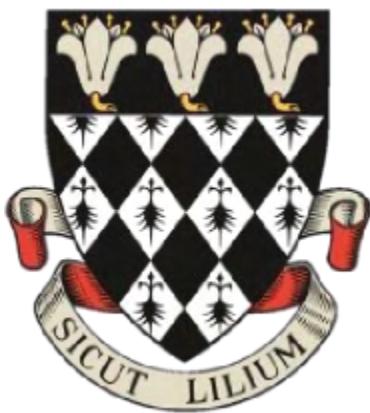


# **Design and Analysis of CRISPR Class 2, Type II, V, and VI Systems as Effective Treatments for Viral Infections**

by

Luqi Wang, Magdalen College School Oxford



Cowley PI, Oxford OX4 1DZ

23/May/2022

Word count: 5589

(Excluding references, footnotes for figures, tables, the  
appendix, titles, and the index)

## **Index:**

<b>1 Abstract .....</b>	<b>3</b>
1.1 Introduction .....	4
1.2 The CRISPR/Cas System .....	5
<b>2 How Do CRISPR Mediated Antiviral Therapies and Vaccines Work? .....</b>	<b>8</b>
<b>3 General Design of CMAVT and CMAVV .....</b>	<b>13</b>
3.1 Target Host Factors or Viral Genome? .....	13
3.2 crRNA selection .....	14
<b>4 crRNA and Cas Nuclease Expression in Target Cells .....</b>	<b>19</b>
4.1 Design of DNA Plasmid Expression Vector .....	20
4.2 Delivering the Expression Vectors .....	22
<b>5 Potential uses of CMAVT and CMAVV Outside of the Human Body .....</b>	<b>25</b>
5.1 CMAVT and CMAVV in <i>Bombyx mori</i> .....	25
5.2 CMAVT and CMAVV in Plants .....	27
5.3 Controlling the Spread of Viral Diseases Using CMAVT and CMAVV in <i>Aedes aegypti</i> .....	28
<b>6 Effectiveness of CMAVT and CMAVV in Suppressing Viral Infections .....</b>	<b>33</b>
<b>7 Conclusion .....</b>	<b>35</b>
<b>8 Appendix(A) .....</b>	<b>36</b>
<b>9 Appendix(B) .....</b>	<b>37</b>
<b>10 Supplementary Materials .....</b>	<b>40</b>
<b>11 Abbreviations List .....</b>	<b>40</b>
<b>12 References .....</b>	<b>41</b>

## **Abstract**

We lack effective vaccines for numerous deadly viruses such as *Zaire ebolavirus*, HIV-1, and so on. We are also constantly under the threat of new emerging viruses such as the SARS-CoV-2 (severe acute respiratory syndrome coronavirus-2) pandemic in 2019, and a recent monkeypox virus outbreak in May 2022. To combat these problems, the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system has emerged as a promising technology for antiviral therapy. It has shown its potential to be more effective than traditional vaccines (including mRNA vaccines) due to its fast activation and strong suppression of viral load. In this paper, I have made a summary of how the CRISPR/Cas system works and hypothesized how CRISPR mediated antiviral therapy (CMAVT) and CRISPR mediated antiviral vaccine (CMAVV) could be applied to the different types of viruses using specific viruses as examples. I have also constructed a set of instructions to follow for selecting target sequences leading to the construction of an effective crRNA. Furthermore, I have constructed a potential crRNA targeting the RdRp (RNA-dependent RNA-polymerase) mature peptide sequence in the highly conserved ORF1ab (Open Reading Frame1ab) gene section of SARS-CoV-2 using my instructions. I also summarised the 3 different ways that the crRNA and the relevant nuclease could be expressed in the cell which then led to the details of DNA plasmid expression vector design in which I discussed how the DNA plasmid vectors for expression and vector cloning should be constructed. Then I looked into possible delivery mechanisms for the expression vectors. I have also explored the other potential uses of CMAVT and CMAVV outside of the human body, namely in plants and animals (*Bombyx mori*). In addition, I looked into controlling the spread of diseases using CMAVT and CMAVV associated with gene drives in *Aedes aegypti* and designed the DNA plasmid expression vectors for it. Throughout this paper, I have assessed and discussed the effectiveness and the potential of CMAVT and CMAVV in the future.

Key Words: CRISPR/Cas, Antiviral therapy, vaccine, crRNA design, Cas9, Cas12, Cas13.

## Introduction

Viral infections are a problem that has troubled humans for decades. CDC estimates that flu has resulted in 9mil – 41mil illnesses, 140,000 – 710,000 hospitalizations, and 12,000 – 52,000 deaths annually between 2010 and 2020<sup>1,2</sup>. Countless other viruses such as HIV-1, SARS-CoV-2, and much more are affecting countless people worldwide. Vaccinations have been offered as a solution to these viral infections, and are the promising solutions to many viral diseases so far. However, many deadly viruses lack effective vaccines, such as *Zaire ebolavirus*, HIV-1 and the Marburg virus<sup>3</sup>. Vaccines are ineffective when they fail to induce B lymphocytes to rearrange their DNA segments in a way necessary to manufacture the needed antibodies to combat that specific virus. Furthermore, some viruses perform antigenic switching frequently to achieve an antigenic escape, a strategy adopted by the Influenza Virus<sup>4</sup>. Many viruses also lack an effective treatment against them, mainly due to latency phases that massively decrease the efficacy of treatments (e.g., the highly active antiretroviral therapy for HIV<sup>5</sup>).

Table 1.1: Estimated Flu Disease Burden, by Season — the United States, 2010-11 through 2019-2020 Flu Seasons, adapted from  
1

Season	Symptomatic Illnesses		Medical Visits		Hospitalizations		Deaths	
	Estimate	95% UI	Estimate	95% UI	Estimate	95% UI	Estimate	95% UI
<u>2015-2016</u>	24,000,000	(20,000,000 – 33,000,000)	11,000,000	(9,000,000 – 15,000,000)	280,000	(220,000 – 480,000)	23,000	(17,000 – 35,000)
<u>2016-2017</u>	29,000,000	(25,000,000 – 45,000,000)	14,000,000	(11,000,000 – 23,000,000)	500,000	(380,000 – 860,000)	38,000	(29,000 – 61,000)
<u>2017-2018</u>	41,000,000	(35,500,000 – 53,000,000)	21,000,000	(18,000,000 – 27,000,000)	710,000	(560,000 – 1,100,000)	52,000	(3,000 – 95,500)
<u>2018-2019</u>	29,000,000	(25,000,000 – 40,000,000)	17,000,000	(11,500,000 – 18,500,000)	380,000	(300,000 – 66,000)	28,000	(19,000 – 97,000)
Preliminary estimates*	Estimate	95% UI	Estimate	95% UI	Estimate	95% UI	Estimate	95% UI
<u>2019-2020*</u>	35,000,000	(30,000,000 – 49,000,000)	16,000,000	(14,000,000 – 22,000,000)	380,000	(312,000 – 630,000)	20,000	(18,000 – 80,000)

The CRISPR/Cas system could have the potential to resolve all of the problems stated above. Currently, it is being explored mainly on its potential for genetic editing.

However, as an effective, adaptive immune defence mechanism, the CRISPR/Cas system should be explored more on how it could be adapted into a treatment for viral infections.

## **The CRISPR/Cas System**

The clustered regularly interspaced short palindromic repeats (CRISPR) are a family of DNA sequences found in the genomes of prokaryotic organisms such as bacteria and archaea<sup>6</sup>. Together with the Cas (CRISPR-associated proteins) system they make up an adaptive immune defence mechanism<sup>6-9</sup> and are found in approximately 50% of sequenced bacterial genomes and nearly 90% of sequenced archaea<sup>6</sup>.

There are 2 classes and multiple types. However, only class 2 type II, type V and type VI are relevant and useful in the context of this paper since they can target all types of common viral genomes ranging from double-stranded DNA (dsDNA), single-stranded RNA (ssRNA), to the transcriptomes of double-stranded RNA (dsRNA)<sup>10-13</sup>. Class 2 comprises a diverse set of single-effector proteins<sup>14</sup> which was believed to be the product of competitive coevolution of CRISPR-Cas systems with different evolving viruses and associated anti-CRISPR proteins<sup>15</sup>. For type V and VI systems, it consists of a Cas nuclease in complex with a crRNA that is made of a user-definable spacer sequence which can cleave a complementary target nucleic acid (DNA or RNA)<sup>16</sup> and a 30 / 36 nucleotide direct repeat which forms a scaffold that is essential for Cas12 and Cas13 function<sup>17,18</sup> (Table 2.1).

Class 2, type II systems are characterized by the Cas9 effector nuclease and a dual RNA-guide (sgRNA), made of a trans-activating crRNA (tracrRNA) and a crRNA (Table 2.1).

Effector protein (class 2)	Cas9 (type II)	Cas12 (type V)	Cas13 (type VI)
crRNA/gRNA length (nucleotides)	42 <sup>19</sup>	42–44 <sup>20</sup>	60–66 <sup>21</sup>
tracrRNA length (nucleotides)	75–100 <sup>22</sup>	N/A	N/A
Direct repeat length (nucleotides)	N/A	19, 20 <sup>20</sup>	30, 36 <sup>18</sup>
crRNA spacer length (nucleotides)	17-20 <sup>23</sup>	23-25 <sup>20</sup>	28–30 <sup>21</sup>
Optimal spacer length with very high knockdown activity (nucleotides)	20 <sup>19,24</sup>	21 <sup>25</sup>	22 <sup>26,27</sup>

Table 2.1 Summary of crRNA / gRNA and spacer lengths. Made by author. Some optimal crRNA spacer lengths have been truncated and found to have a very high knockdown activity.

The Cas9-sgRNA complex recognizes a Guanine-rich 3' end-located protospacer adjacent motif (PAM) on a dsDNA target. PAM is a 2-6 base-pair DNA sequence

immediately following the DNA sequence targeted by the Cas9 nuclease<sup>28</sup>. PAM recognition leads to the complementary annealing of the sgRNA spacer sequence and the DNA and the activation of the nuclease to induce a blunt-ended double-stranded break (DSB) 3-4 nucleotides upstream of the PAM sequence on the target (Figure 1.1)<sup>16</sup>. Cas9 will not successfully bind to or cleave the target DNA sequence if it is not followed by the PAM sequence, thus limiting target sequences that can be selected.<sup>28-31</sup>

Class 2 type VI systems contain the programmable single-effector Cas13 (Cas a, b, c, and d) that target ssRNA<sup>32</sup>. The Cas 13 nuclease lacks PAM requirements, some orthologs are activated upon the protospacer flanking site (PFS) recognition<sup>31</sup>. In other words, Cas13 has no targeting sequence constraints, such as a PFS or PAM, and no motif preference surrounding the target adenine<sup>32</sup> (Figure 1.1). The Cas-13-crRNA complex not only expresses cis-cleavage (on-target) activity but also expresses high trans-cleavage (target-independent) activity with surrounding ssRNAs<sup>33</sup>.

Class 2, type V system is mainly represented by a group of effector proteins / Nuclease RuvC that lacks HNH (a metal-ion-dependent nuclease domain that cleaves the DNA strand complementary to the RNA guide)<sup>13,34</sup>. Cas12 (Cas12a, or Cpf1 and Cas12b) uses a single RNA-guide and targets dsDNA. The recognition of a T-rich 5'end-located PAM enables crRNA pairing with a dsDNA target. The complex expresses cis-cleavage activity and expresses high trans-cleavage activity<sup>33</sup>. It produces a staggered-ended DSB on the target (Figure 1.1, a visualization of the Cas systems).

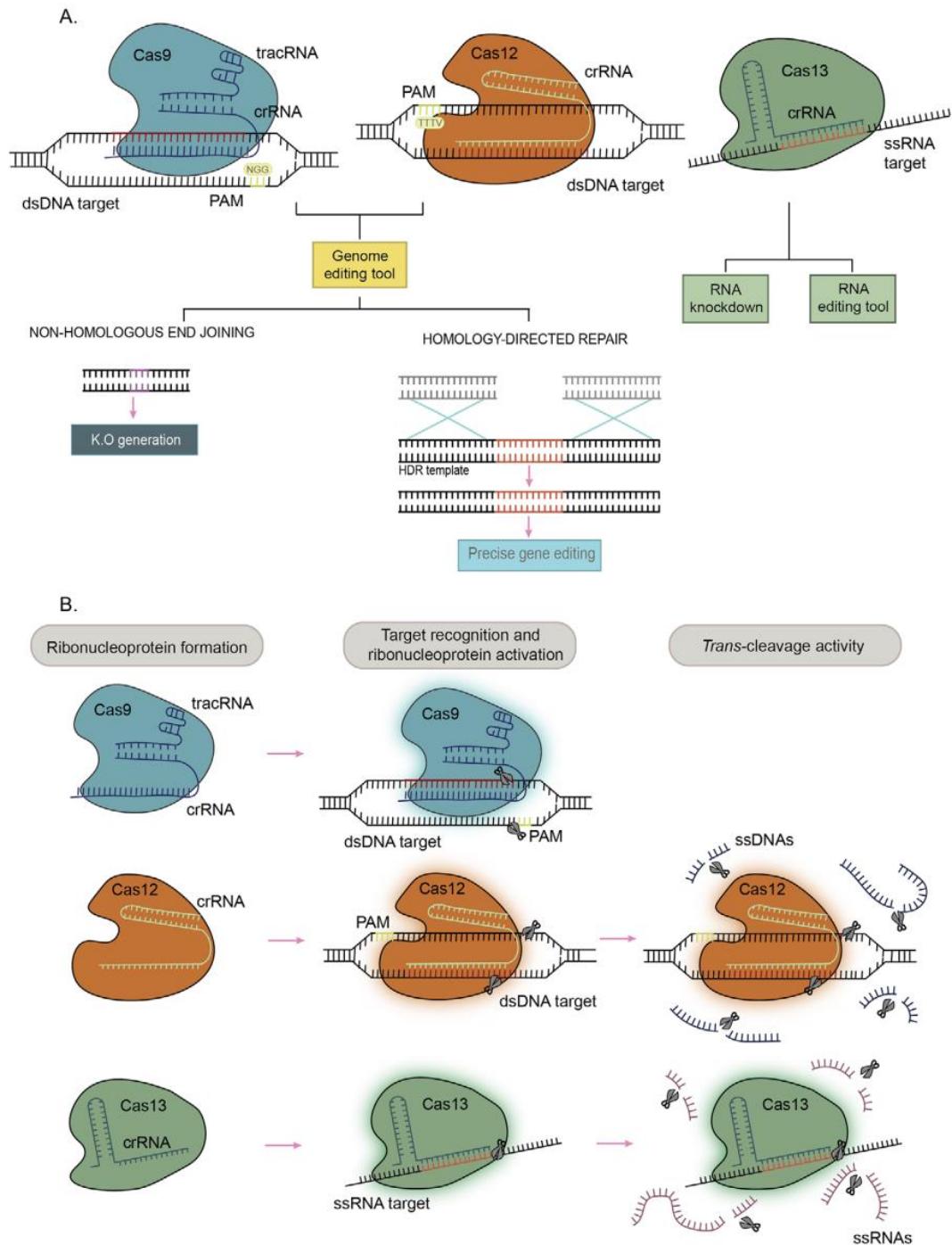


Figure 1.1 – adapted figure from<sup>35</sup>, a simplified figure of the function of CRISPR class 2 type II, V, and VI effector proteins, Cas9, Cas12, and Cas13 respectively, and their effects on their target DNA or RNA.

Effector protein	Cas9	Cas12	Cas13
Target	dsDNA	dsDNA	ssRNA
trans-Cleavage	-	dsDNA/ssDNA	ssRNA
PAM	G rich	T rich	-
PFS	-	-	Ortholog dependant
Guide type	Dual (tracRNA and crRNA)	Single (crRNA)	Single (crRNA)

Table 3.1 – adapted from<sup>35</sup>, a basic summary of the CRISPR Cas class 2 type II, V, and VI system

With the basics of effector proteins and their targets clarified, it now becomes quite apparent how these systems can be used to combat viral infections. CRISPR class 2 type II and type V systems characterized by the Cas9, and Cas12 nucleases respectively, could be used to target dsDNA viruses or retroviruses which have a dsDNA intermediate during their replication cycle. In contrast, the CRISPR class 2 type VI system characterized by the Cas13 nuclease can be used to target the viral genome of ssRNA viruses and retroviruses and all viral transcriptomes of all viruses. The precise mechanisms of these ideas would be thoroughly explored and explained using specific viruses in the next section.

## How Do CRISPR Mediated Antiviral Therapies and Vaccines work?

In this section, we will explore how CRISPR can be applied to have antiviral effects which could be used as an antiviral therapy and potentially be used as vaccines. In this paper, we will explore 2 general types of strategies (specific examples using viruses presented in this section are hypothesized by the author) that can be altered slightly to adapt to different viruses:

1. CRISPR/Cas Mediated Antiviral Therapy (CMAVT)
2. CRISPR/Cas Mediated Viral Vaccine (CMVV)

There are 4 main categories of viral genomes:

1. dsDNA (Belongs to group VIII or I In the Baltimore classification system<sup>36</sup>)  
-
2. ssDNA (Belongs to group II In the Baltimore classification system<sup>36</sup>)
3. dsRNA (Belongs to group III In the Baltimore classification system<sup>36</sup>)
4. ssRNA:  
- positive-strand RNA (+) ssRNA (Belongs to group IV In the Baltimore

classification system<sup>36</sup>)

- negative-strand RNA (-) ssRNA (Belongs to group V In the Baltimore classification system<sup>36</sup>)

However, we will only be focusing on dsDNA, dsRNA, and ssRNA since these viral genomes are the most commonly found in deadly viruses that infect humans.<sup>37-43</sup>

For each type of viral genome and specific type of virus, there are specific and multiple different mechanisms by which the CMAVT can be used to target the viral replication strategies<sup>44</sup> that can ultimately prevent entirely or the correct packaging and assembly of new viruses as shown in papers<sup>45,46</sup>. HIV, a type of lentivirus<sup>47</sup> will be explored separately as a model example for the family Retroviridae<sup>48</sup> (retroviruses).

For positive-strand ssRNA viruses, the SARS-CoV-2 virus will be used as the model example.

Three stages in the SARS-CoV-2 virus replication process could potentially be targeted by the CRISPR Cas 13 nucleases that ultimately prevent viral packaging and formation, prevents viral release, or produces a non-functional virus. (Figure 3.1)

1. The positive-sense, viral ssRNA that is released upon viral entry is translated into viral enzymes such as RNA-dependent RNA polymerase (RdRp), helicase, capping enzymes, NTPase<sup>49</sup>, and viral polymerase proteins (3CLpro/PLpro for SARS-CoV-2). Targeting this would block gene expression which results in the lack of necessary proteins for replicating the viral genome and thus stops replication into negative sense.
2. Targeting multiple important viral mRNAs stops essential proteins from being transcribed, thus preventing viral packaging and assembly, making a non-functional virus.
3. Targeting the final replicated positive-sense viral ssRNA that will be the genome for new viruses.

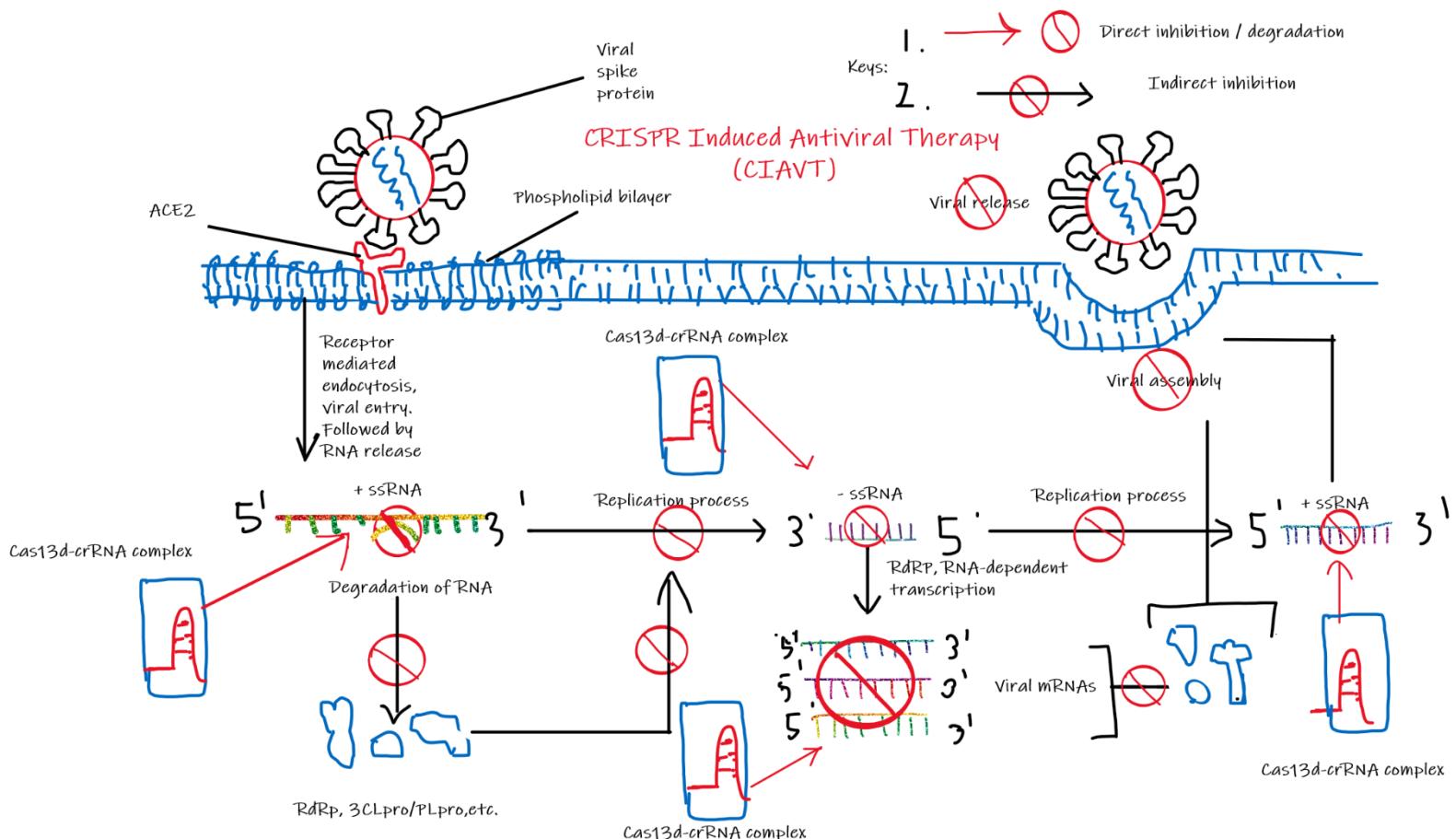


Figure 3.1 Stages of Viral Replication Processes in Positive-Strand RNA Viruses Targeted Using CMAVT. (By author)

SARS-CoV-2 used as an example of positive-strand RNA viruses in this figure. This figure shows a summary of all possible targets of CMAVT that will ultimately prevent the synthesis or release of viruses in the infected cells. Process: 1. Identification of viral spike glycoprotein (S-glycoprotein) by ACE2 (angiotensin-converting enzyme 2) facilitates viral entry by receptor-mediated endocytosis<sup>50</sup> 2. The use of Cas13d-crRNA complex to target various viral ssRNA strands in various replication stages, ultimately stopping the progression of further replication processes and thus inhibiting productions of new viral genomes and essential viral proteins leading to 3. Stopping viral assembly thus stopping viral release from the host cell.

For negative-strand RNA viruses, there are 3 general targets for the CMAVT. (Figure 3.2)

1. The negative-sense, viral ssRNA. Released upon viral entry. Targeting this would block mRNA transcription, and also inhibits the viral genome replication. Thus, viral proteins could not be translated and new viral genomes cannot be made.
2. The mRNA transcriptomes could also be a target that stops viral proteins from being made which ultimately stops viral assembly.

3. The positive-sense, viral ssRNA (the antigenome). Targeting this would inhibit viral progeny genome replication. Which results in empty protein coats.

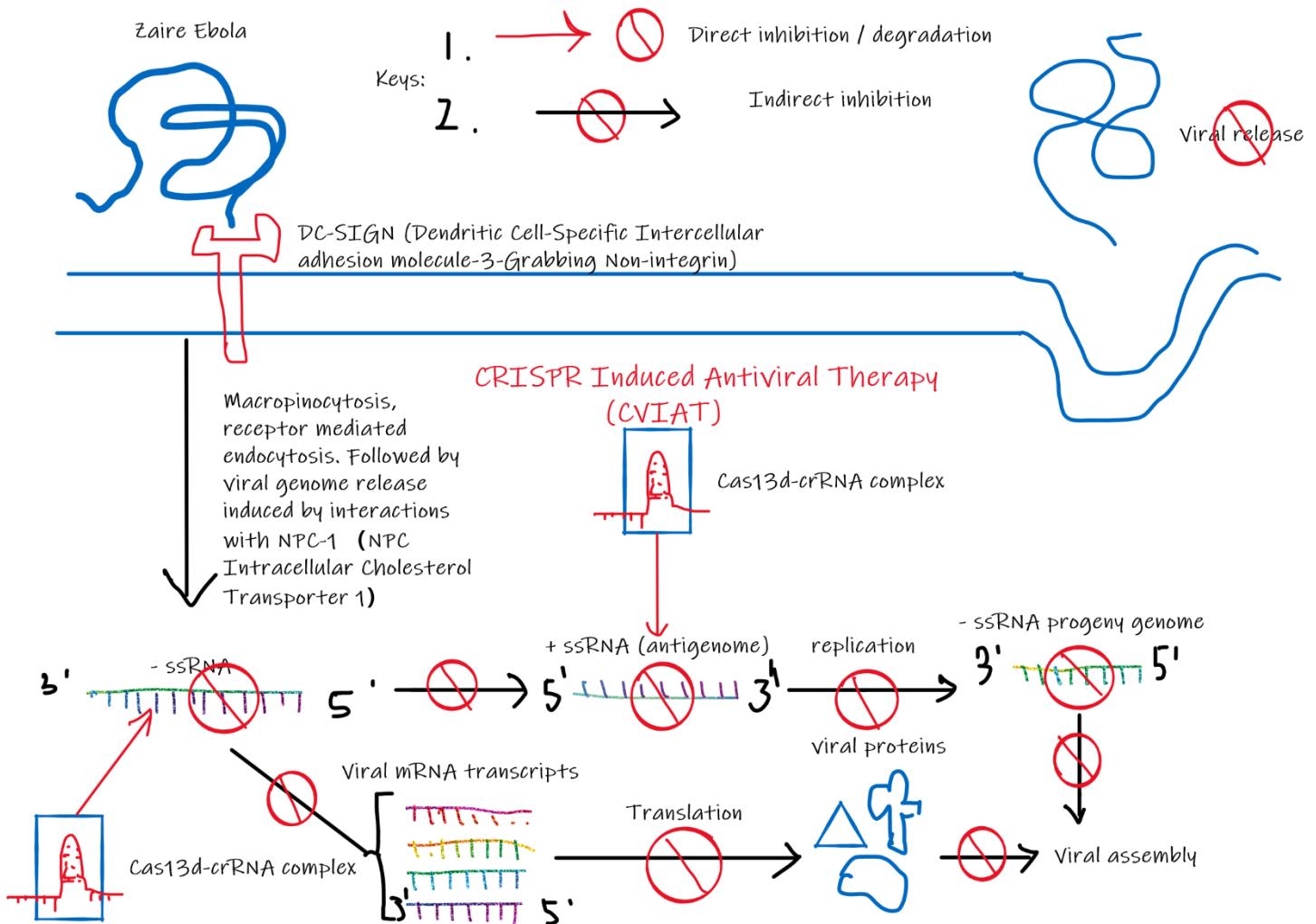


Figure 3.2 Stages of Viral Replication Processes in Negative-Strand RNA Viruses Targeted Using CMAVT. (By author)

Zaire Ebola Virus used as an example of negative-strand ssRNA viruses in this figure. This figure shows a summary of all possible targets of CMAVT that will ultimately prevent the synthesis or release of viruses in the infected cells. Process: 1. A Zaire Ebola virus enters the cell by micropinocytosis facilitated by DC-SIGN<sup>51</sup>, 2. Viral genome release by interactions with NPC-1<sup>52</sup>, 3. The use of Cas13d-crRNA complex to target various viral ssRNA strands in various replication stages, ultimately stopping the progression of further replication processes and thus inhibiting productions of new viral genomes and essential viral proteins leading to 4. Stopping viral assembly thus stopping viral release from the host cell.

For dsDNA viruses, there are 2 general targets for the CMVAT and CMAVV: (Figure 3.3)

1. The dsDNA viral genome released into the nucleus. This would be the optimal target, since triggering a double-stranded break on this target would inhibit all the

mRNA transcriptomes dependent on the viral dsDNA, but also would decrease the likelihood of viral DNA integration risk into the host genome. It can also degrade the viral genome if it were already integrated into the host genome.

2. The viral mRNA transcriptomes that code for the Immediate early, early, and late proteins. This would inhibit the production of virally encoded DNA pol II (if present in the viral genome) which inhibits replication of the viral genome, and also inhibit the production of the late proteins which many are essential for successful viral assembly and release.

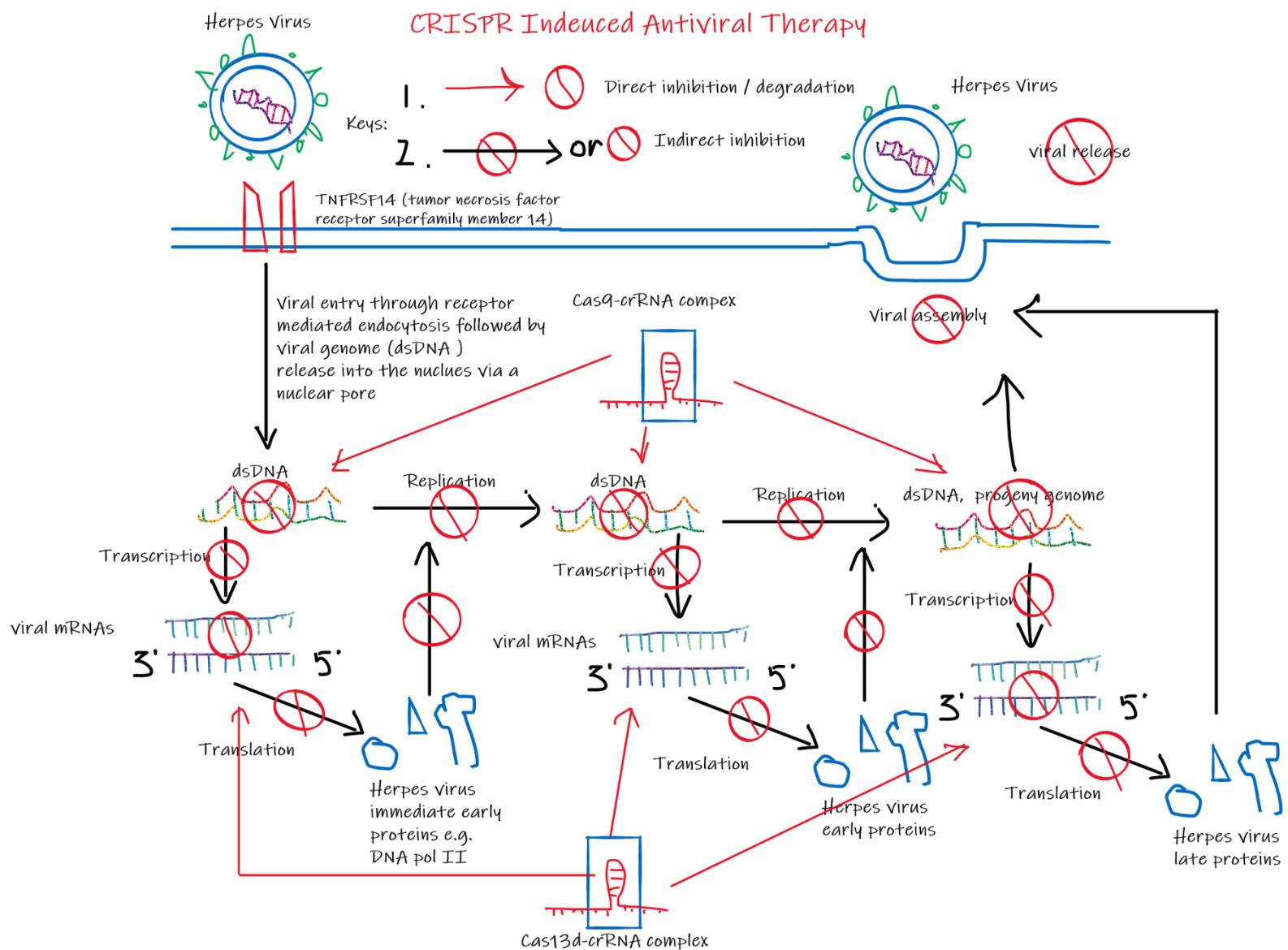


Figure 3.3 Stages of Viral Replication Processes dsDNA Viruses Targeted Using CMAVT. (By author)

The Herpes Virus is used as an example of dsDNA viruses in this figure. This figure shows a summary of all possible targets of CMAVT that will ultimately prevent the synthesis or release of viruses in the infected cells.

Process: 1. A Herpes Virus enters the cell by micropinocytosis facilitated by TNFRSF14<sup>53</sup>

2. The use of Cas9-crRNA and Cas13d-crRNA complexes to target various viral dsDNA strands and transcriptomes in various replication stages, ultimately stopping the progression of further replication processes and thus inhibiting productions of new viral genomes and essential viral proteins (immediate early<sup>54</sup>, early<sup>54</sup> and late proteins<sup>55</sup>) leading to 3. Stopping viral assembly thus stopping viral release from the host cell.

Overall, the way that CMAVV and CMAVT function is basically the same as described in the figures. The major difference is the expression vectors that will be explored later in this paper. With CMAVV we would want DNA integration into the host genome allowing the cell to express the Cas nuclease and guide RNA to ensure lifetime immunity to targeted viruses. On the other hand, with CMAVT it is more of a measure to control the virus once a patient has been infected thus, we would want the expression vector to express the Cas nuclease and guide RNA for a short period (5-10 days) and avoid DNA integration into the host genome.

## **General Design of CMAVT and CMAVV**

Excluding other possible factors, the effectiveness of this antiviral approach would ultimately be based on a careful selection of the target sequence, thus leading to the design of an effective crRNA.

One of the major advantages of the CMAVT and CMAVV systems described in this paper is that the viral sequence target is not limited to the antigens, the target sequence could be any fraction of genetic code in the viral genome which offers a variety of selections and is a hard counter to antigenic switching, given that the target sequence is in the antibody coding region. However, we don't have to target viral elements, for some viruses we can choose to target host factors<sup>46</sup> which could be much more effective.

## **Target Host Factors or Viral Genome?**

There are 2 approaches when designing the CMAVT and CMAVV. The more straightforward approach is to target important, and highly conserved viral genome sequences, their mRNA transcript or target other intermediates such as the dsDNA of retroviruses. Thus, stopping correct replication of its genome<sup>45</sup> or stopping transcription in essential proteins in the process of packaging.

Alternatively, we could target host factors. For the Human Immunodeficiency Virus type -1 (HIV-1), the co-receptors CCR5 and CXCR5 gene knockout (Figure 4.1) prevent HIV-1 from entering helper-T cells<sup>56,57</sup> which is an active area of research.

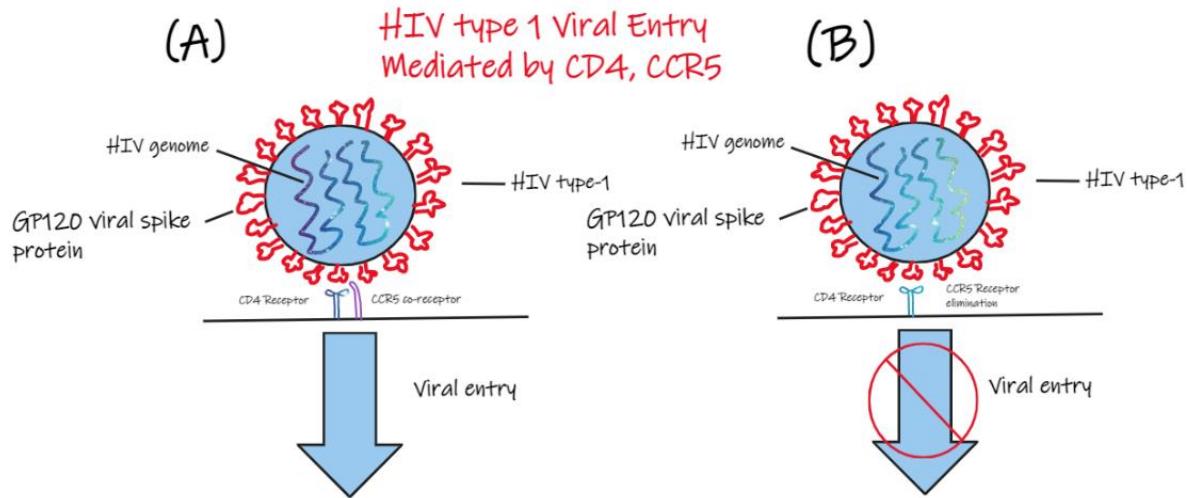


Figure 4.1 HIV Entry Mediated by CD4 and CXCR5. (By author)

(A) shows an HIV entering a Helper T cell, mediated by the binding of GP120 viral glycoprotein to a CD4 receptor<sup>58,59</sup> and a CCR5 co-receptor<sup>60</sup>. (B) shows a viral entry of HIV into a Helper T cell blocked out due to a lack of the co-receptor CCR5 that has been eliminated<sup>61,62</sup>.

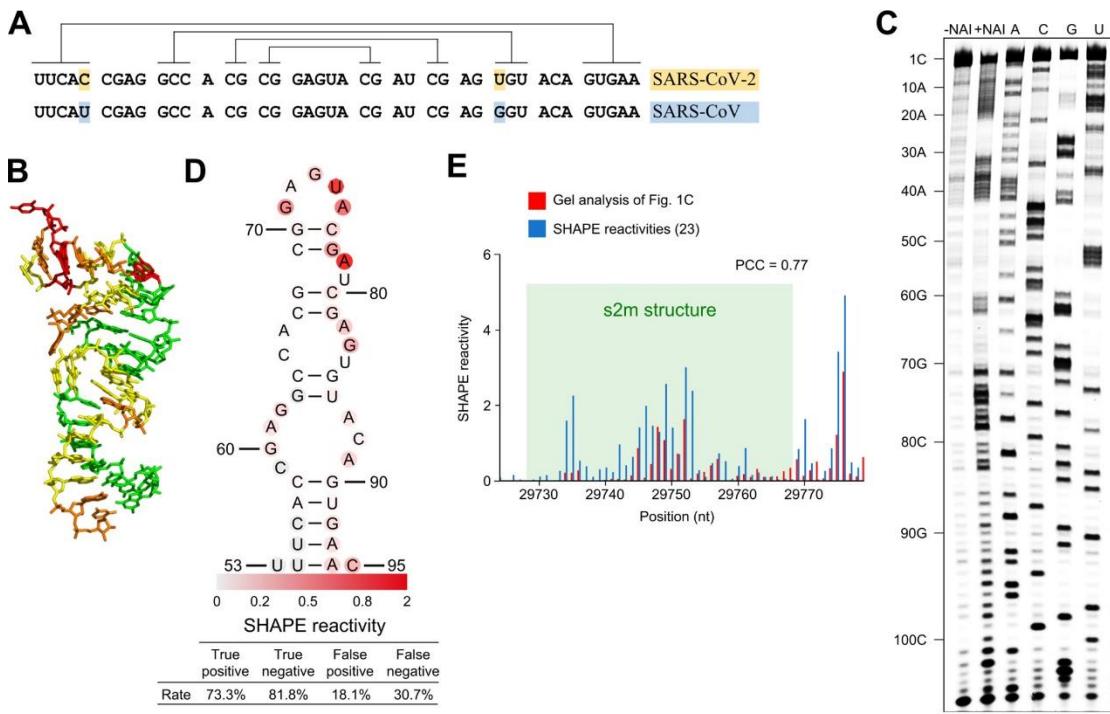
The receptor CD4, gene knockout also prevents HIV entry into helper T cells, but considering the significance of its role in adaptive immunity, it is not considered a therapeutic option. Researchers have used Cas9 to perform gene knockout of the CCR5 gene in numerous studies<sup>63,64</sup>. However, the side-effects of performing these gene knock-offs are still relatively unknown and due to the off-target activity of the nucleases, there are risks of tampering other genes. Furthermore, gene-editing in humans is still under a massive ethical debate<sup>65–68</sup>.

Therefore, at this moment the best option is to target viral genomes since it offers a much more straightforward, effective, and low-risk solution when compared to targeting host factors.

## crRNA Selection

The selection of the target gene and thus the design of the crRNA is a very important process that directly affects the efficacy of the CMAVT and CMAVV. Using SARS-CoV-2 as an example here, in order to design an effective crRNA, the crRNA must satisfy the following criteria:

Firstly, the crRNA must target highly conserved regions of the viral genome to ensure the crRNA would still work on viral mutants and new emerging strains<sup>69</sup>. A highly conserved region of the viral genome also indicates its importance in the viral replication cycle which would ensure the effectiveness of CMAVT targeting that conserved region. For example, the Coronavirus' s2m element that is able to dimerize into a thermodynamically stable duplex conformation<sup>70</sup>. Furthermore, it is a secondary structure motif identified in the 3' untranslated region (3'UTR) of many astroviruses, some picornaviruses, noroviruses, equine rhinoviruses, and multiple coronavirus genomes<sup>71,72</sup>, which means it is present in the genome of SARS-CoV-2 and all its variants since there are no species-level losses of s2m<sup>73</sup>.



**F**

Figure 5.1 (adapted from<sup>74</sup>) s2m is a conserved structural element in the SARS-CoV-2 genome. (A) Sequence alignment of the s2m element in the 3' UTRs of SARS-CoV-2 and SARS-CoV. Lines indicate base-pairing regions within the element. Highlighted bases show single-point mutations. (B) Crystal structure of the SARS-CoV s2m element (adapted from PLoS Biology<sup>75</sup>). (C) SHAPE chemical probing of the 3' UTR of SARS-CoV-2. RNA was denatured and refolded in the presence of 100 mM K<sup>+</sup> and 0.5 mM Mg<sup>2+</sup> and then incubated with NAI (+NAI channel) or DMSO control (-NAI channel). NAI modification was detected by reverse transcription stalling and gel-based analysis. Sequencing lanes were generated by adding ddT (for A), ddG (for C), ddC (for G), and ddA (for U) when reverse transcription was performed. (D) Annotation of SHAPE signal on the s2m structure. (E) Bar plot showing the reactivities of structural profiling of gel-based analysis in the study of<sup>74</sup> (C) and the SHAPE-MaP experiment described by<sup>76</sup>. The s2m structure is highlighted by shading in green.

Gene sequencing also revealed that 2 of the more highly conserved regions (Figure 5.2) contain the replication complex: RNA-dependent RNA polymerase (RdRp) gene<sup>77-80</sup>, and the nucleocapsid (N) gene at the 3' end of the genome, which encodes the capsid protein for viral packaging<sup>81</sup>. In contrast, the ORF3a gene section has a very low conservation percentage which indicates a very high mutation rate which suggests

its insignificance in the viral replication cycle. Thus, ORF3a would not be an ideal area of the SARS-CoV-2 genome to target.

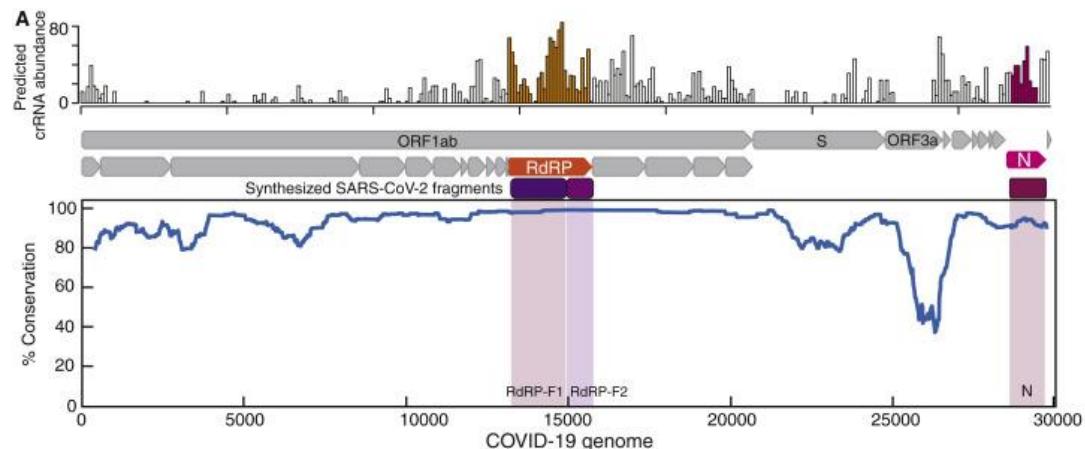


Figure 5.2 adapted from<sup>82</sup>. Alignment of 47 patient-derived SARS-CoV-2 sequences with SARS-CoV and MERS-CoV. Top: a predicted abundance of crRNAs that can target SARS-CoV-2 genomes and SARS or MERS. Middle: annotation of genes in the SARS-CoV-2 genomes, along with conserved regions chosen to be synthesized into the SARS-CoV-2 reporters (magenta and purple). Bottom: percentage of conservation between aligned viral genomes.

Secondly, the crRNA must not contain any poly-Uracil sequences (more than 3) which can suppress crRNA expression<sup>82</sup> in the cell.

Thirdly, the crRNA must not have any potential off-target bindings in the human transcriptome, otherwise, human transcriptomes are at risk of being degraded by Cas nucleases which could cause severe side effects in the human body.

To prevent off-target bindings with the human transcriptome, Abbott et al. suggested selecting crRNAs that have 2 or more mismatches in the human transcriptome<sup>82</sup>. However, I would disagree with their idea. In my opinion, I would argue that instead of checking the number of mismatches, we should instead check for the annealing temperatures of the crRNA and target strand, since nucleotides G and C are more likely to pair, and are more stable than A and T (Figure 5.3). Furthermore, mismatches on the 2 ends of the sequence wouldn't affect the guide RNA binding to the target as much as mismatches in the middle of the sequence (Figure 5.4). Thus, in my opinion, any crRNA spacer sequence selected should have an annealing temperature of at least 10°C lower or with any human transcriptome to minimize off-target activity. For annealing temperature calculations, see appendix (C)

Fourthly, the target sequence and the complementary crRNA should have an annealing temperature of at least 55°C, this ensures a more spontaneous pairing between the crRNA and the target.

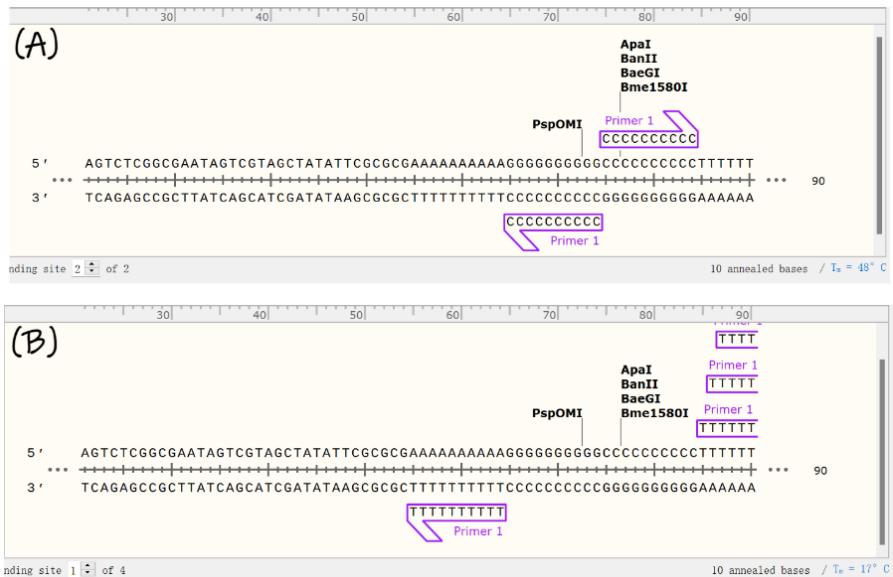


Figure 5.3 Comparison of the Annealing Temperatures of 10 annealed C-G base pairs, and T-A base pairs  
Figure made by author using SnapGene®.

- (A) The annealing temperature for 10 C-G base pairs is  $48^\circ\text{C}$
- (B) The annealing temperature for 10 T-A base pairs is  $17^\circ\text{C}$ , significantly lower.

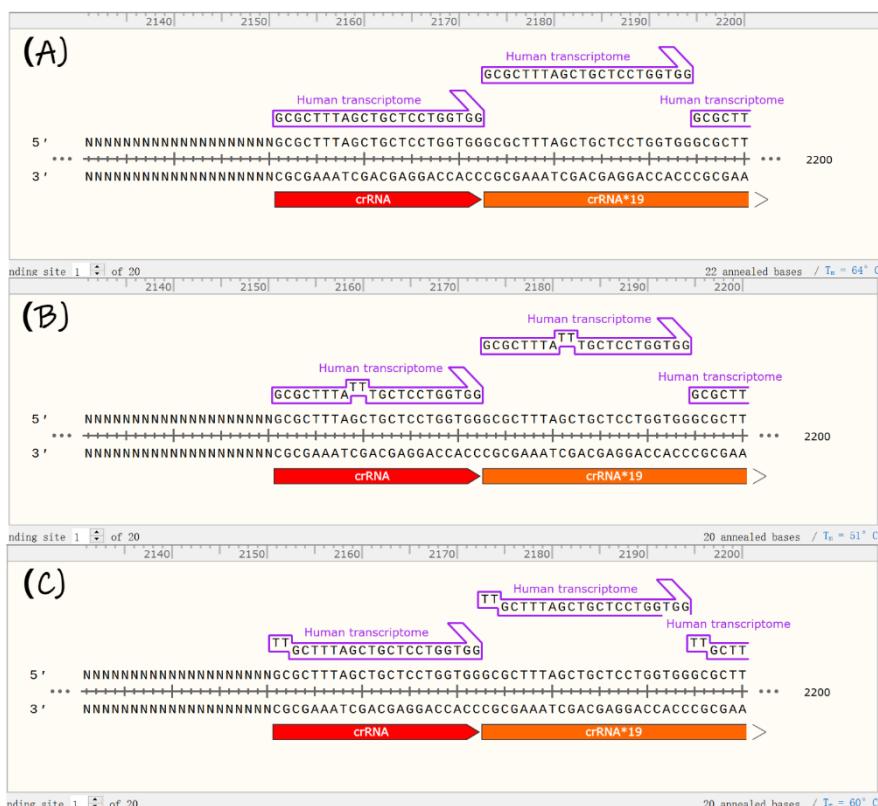


Figure 5.4 Comparison of a hypothetical crRNA made by the author and Human Transcriptome Annealing Temperatures, made by the author in SnapGene®

- (A) Perfect match between crRNA and Human Transcriptome has an annealing temperature of  $64^\circ\text{C}$ , 2 mismatches at points 2159, and 2160 (right in the middle of the human transcriptome) bring down annealing temperature significantly to  $51^\circ\text{C}$
- (B) However, 2 mismatches at points 2150, and 2151 (right at the start) only bring the annealing temperature down a little, to  $60^\circ\text{C}$ . This means the crRNA is still quite likely to bind to that sequence.

Fifthly, the GC percentages of the crRNA spacer sequence should be within the range of 40%–60% for highest cleavage efficiency<sup>83</sup>.

Sixthly, the target sequence of the crRNA should ideally be a perfect match with other strains of the virus along with SARS-CoV and MERS-CoV (Middle East respiratory syndrome–related coronavirus) since they are of the same genus<sup>84</sup>. However, 1 or 2 mismatches should be tolerable since there are some flexibilities, given that not every position in the guide RNA needs to match the target sequence<sup>31,85</sup>. But as stated earlier if the mismatches occur in the middle of the sequence, it would significantly decrease the likelihood of the binding between the crRNA and the target sequence as opposed to having the mismatches at the ends of the target sequence. Therefore, it would be ideal to ensure that the annealing temperature of the chosen crRNA and the target sequence in SARS-CoV and MERS-CoV will not be lower than 2°C when compared to the annealing temperature of the perfect match between the crRNA and the target sequence in SARS-CoV-2.

Lastly, the spacer length of the crRNA should ideally be one of the optimal lengths shown in (Table 2.1). For example, the optimal crRNA spacer length for Cas9 is 20 nucleotides.

Using the instructions suggested above, I have designed a potential crRNA that targets the RdRp mature peptide sequence in the highly conserved ORF1ab gene section of SARS-CoV-2 (Figure 5.5). For detailed steps of design, see appendix (B)

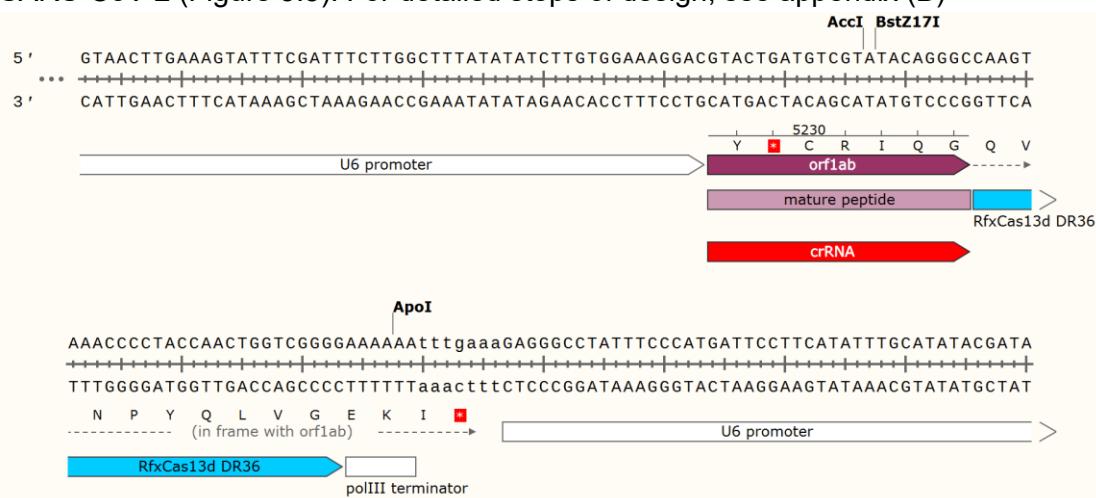


Figure 5.4 crRNA designed by the author using SnapGene®, crRNA spacer sequence: GUACUGAUGUCGUAUACAGGGC (annealing temperature of target sequence: 57°C).

Now that we have explored the 6 most important points in selecting a target sequence and thus an effective crRNA, we would need to understand how we can express the crRNA and the relevant nucleases in the target cells.

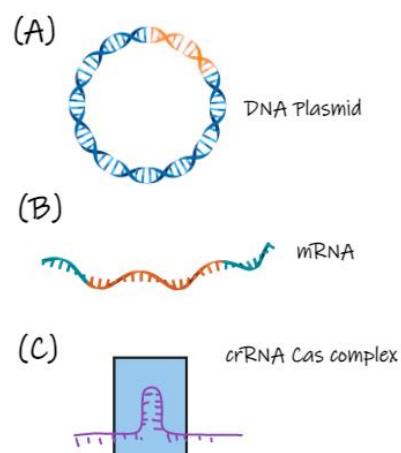
## **crRNA and Cas Nuclease Expression in Target Cells**

For CMAVT and CMAVV to work, we would need cells to express or contain the crRNA and the relevant Cas nucleases. In general, there are 3 ways (Figure 6.1).

Figure 6.1 Different Expression Vectors for crRNA and Cas nuclease (Made by author)

- (A) crRNA and Cas nuclease expression vector as a DNA plasmid. (B) crRNA and Cas nuclease expression vector as an mRNA strand. (C) crRNA and Cas nuclease delivered as a single complex

### **Expression vectors for crRNA and Cas nucleases**



1. A DNA plasmid can be used as the expression vector for the crRNA and Cas nucleases. Due to the double helix structure, the expression vector would have a long half-life and be quite stable which is advantageous. However, DNA integration would be a risk in terms of CMAVT but would be desirable in terms of CMAVV.
2. mRNAs could also be used as the expression vector. However, it is very unstable and has a very short half-life which makes it less appealing as an expression vector.
3. Directly delivering Cas nuclease crRNA complex is a straightforward strategy but it has a short half-life and its large size presents a challenge in delivering them efficiently and cost-effectively when applied in vivo.

Therefore, the most desirable and suitable expression vector for the crRNA and Cas nucleases is a DNA plasmid. After exploring the different possible expression vectors, we now have the challenge of designing the expression vector so that it will express the crRNA and Cas nuclease efficiently inside the cell.

## Design of DNA Plasmid Expression Vector

For expressing the crRNA and Cas nuclease(s) (using RfxCas13d as an example) it would be optimal to use 2 different DNA plasmids.

There are 2 general scenarios for designing the expression vector, one is for conducting a study not in humans in which a fluorescent protein should be used, such as mCherry, which Abbott et al. used<sup>82</sup>. A strong promoter should be included such as CAG, CMV, or even better the human EF - 1 $\alpha$  promoter that has the highest promoter activity and normally is more stable in long - term culture<sup>86</sup>, then after the promoter sequence, the Cas nuclease gene should follow. (Figure 7.1) For the crRNA/guide RNA expression plasmid, the promoter used should be an RNA polymerase III promoter such as the H1 promoter or U6 promoter. However, the U6 promoter should be used since it has been proved to be stronger than the H1 promoter in vitro and in vivo<sup>87</sup>.

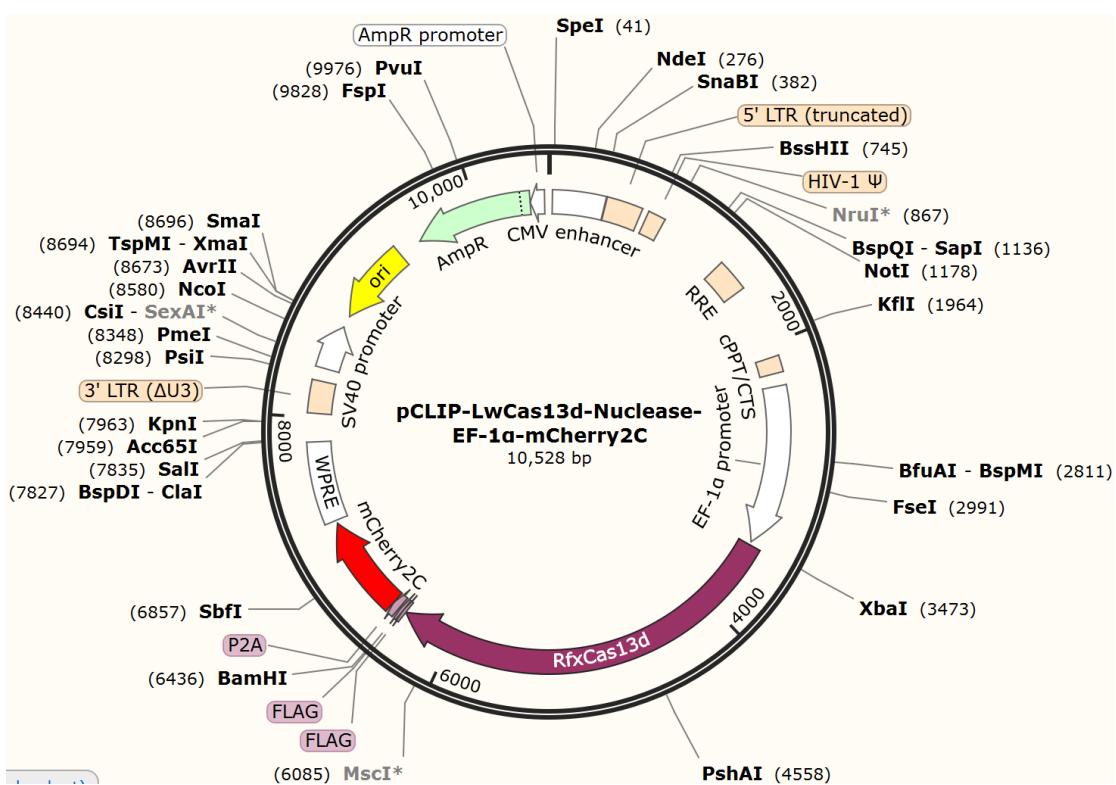


Figure 7.1 pCLIP-LwCas13d-Nuclease-EFS-mCherry2C, designed by the author using pCLIP-Cas9-Nuclease-EFS-ZsGreen vector backbone provided by transOMIC<sup>88</sup>. An EF-1 $\alpha$  core promoter followed by RfxCas13d gene<sup>27</sup>. Then with 2 FLAG® epitope tags<sup>89</sup> followed by mCherry2C (variant with lower cytotoxicity)<sup>90</sup>, then in between FLAG® and mCherry2C a P2A self-cleaving sequence<sup>91</sup> to separate the mCherry2C polypeptide<sup>90</sup> from RfxCas13d polypeptide<sup>27</sup>. This plasmid vector is optimised for lentiviral transduction. For detailed reasoning of design, see appendix (A).

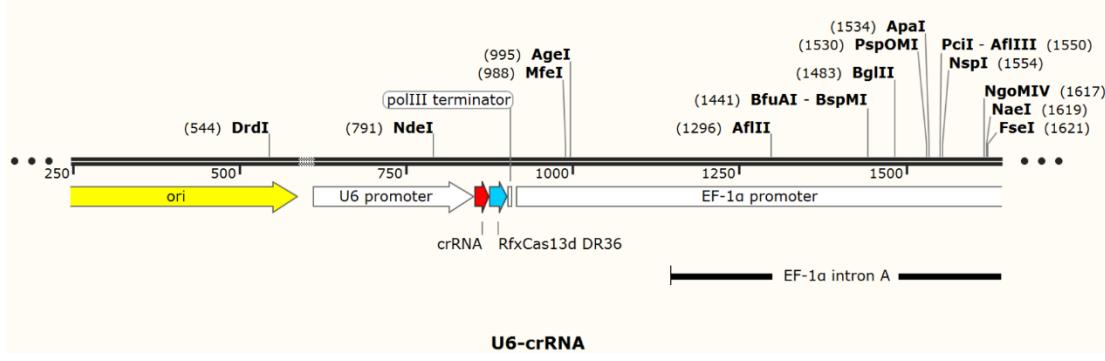


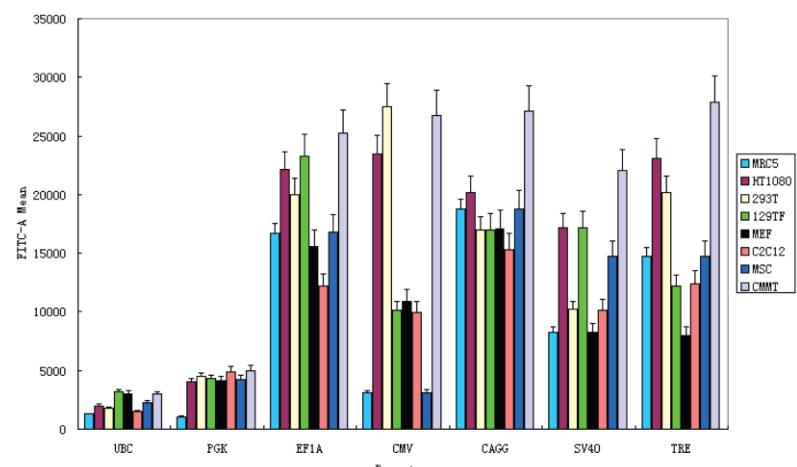
Figure 7.2 U6-crRNA-RfxCas13d DR36, designed by author using SnapGene®.

A crRNA expression vector using a strong U6 promoter and EF-1 $\alpha$  promoter. U6 promoter followed by crRNA binding sequence followed by a RfxCas13d DR36 that is required for the formation and functionality of the Cas13d crRNA complex ensuring binding and degrading of the selected target sequence<sup>27</sup>. Transcription halted by an RNA polymerase III transcription terminator sequence. For detailed reasoning of design, see appendix (A).

The second scenario is for the CMAVT and CMAVV described earlier in this paper. Under this scenario, a fluorescent protein would be unnecessary. Furthermore, there should be a change in the selection of the promoter sequence. Two different promoters should be used. One promoter should be a stable promoter to ensure a long-lasting, low but steady expression of the crRNA and Cas nuclease(s). The other should be a strong promoter with a very high promoter activity that is activated by transcription factors released when the virus infects a cell. Using HIV-1 as an example, the weak promoter selected should be either UBC or PGK as it had very low expression rates in mammalian cell lines<sup>92</sup> (Figure 7.3). However, PGK promoter should be used since it is stronger than UBC and its strength is more uniform across the 8 different mammalian cell lines<sup>92</sup> (Figure 7.3), (Figure 7.4).

Figure 7.3 Adapted from<sup>92</sup> Flow cytometry measurement of GFP fluorescence in eight mammalian cell types transduced with lentiviral vectors carrying GFP reporter driven by promoters of interest. Six mammalian constitutive promoters were tested, along with the doxycycline-inducible TRE promoter at maximal induction.

Note: FTTC-A: Fourier Transform Traction Cytometry-A



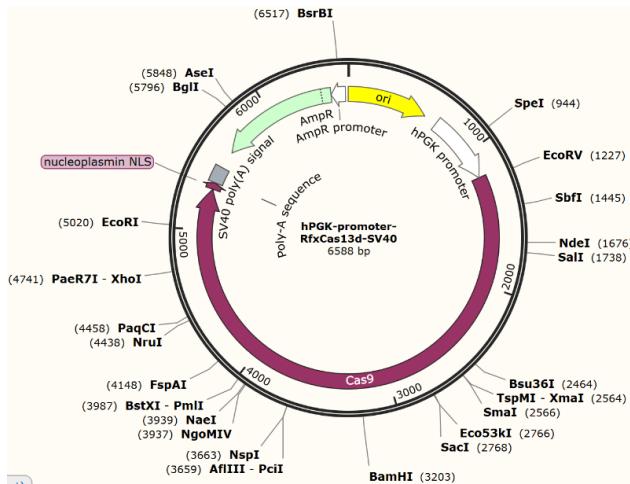


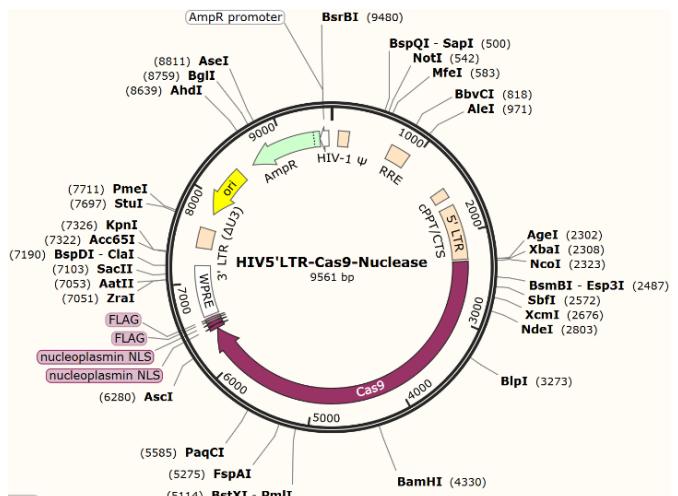
Figure 7.4 hPGK promoter Cas9 SV40, made by author using SnapGene®.

Using a weak ubiquitous human PGK promoter to drive low, but steady and ubiquitous expression of Cas9 in human cells  
For detailed reasoning of design, see appendix (A).

For the strong viral inducible promoter, I would select the HIV-1 5' LTR (Long Terminal Repeat) promoter that plays an essential role in driving viral transcription and productive infection<sup>93–95</sup> (Figure 7.5) and can only be activated when HIV-1 Tat (Trans-Activator of Transcription) regulatory protein binds to the TAR sequence in the 5' LTR<sup>96–98</sup>, which drastically enhances the efficiency of viral transcription<sup>97,99</sup>.

Figure 7.5 HIV-1 5' LTR-Cas9-Nuclease made by author using SnapGene®. An HIV-1 inducible Cas9 expression DNA plasmid using HIV-1 5' LTR as promoter sequence with HIV-1 3' LTR as secondary promoter sequence and transcription terminator sequence. Optimised for lentiviral transduction. Nucleoplasmin NLS<sup>100</sup> used to transport Cas9 nuclease into the nucleus, sorting out HIV-1 genome integration.

For detailed reasoning of design, see appendix (A).



For the expression of the sgRNA a U6 promoter will be used (Figure 7.6 and 7.7)

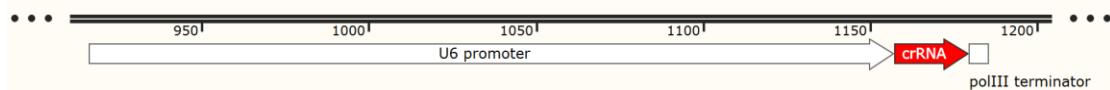


Figure 7.6 Cas9 crRNA 42 nucleotides, using U6 promoter and a pol III terminator sequence.

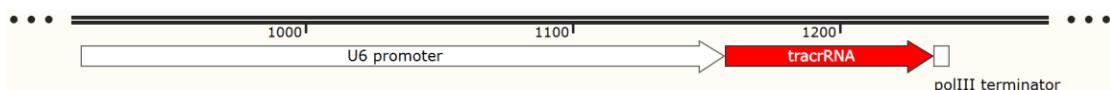


Figure 7.7 Cas9 tracrRNA 78 nucleotides long, using U6 promoter and pol III terminator sequence.

With the design of the expression vectors explored, we face the issue of delivering the expression vectors into the desired cells in an efficient and non-toxic manner.

## **Delivering the Expression Vectors**

The practicality of the CMAVT and CMAVV greatly relies on delivering the expression vectors efficiently, and effectively with low toxicity in the body and this poses a great challenge. Delivering the expression vectors in vitro and ex vivo are simple, easy, straightforward, and yields great results. For example, lipid or polymer-based transfection<sup>101</sup>, electroporation<sup>102</sup>, etc. However, these techniques that yield great results in vitro or ex vivo do not work in vivo. The in vivo delivery of the expression vectors is the real challenge.

So far, the most common delivery mechanism is by using viruses. Viruses do not have the capacity to carry Cas nuclease guide RNA complexes but can carry a few DNA plasmids. Here is a table for the most commonly used viral vectors (Table 4.1). One of the major drawbacks of using viral vectors in vivo is that the viral vectors are extremely likely to cause an immune response which would limit their potential as a delivery mechanism to cells. However, adeno-associated virus vectors (Table 4.1 first row) are proved to have a low immunogenicity<sup>103,104</sup>, have yielded good results in gene therapy<sup>105</sup>, and are widely used, thus it is a promising candidate for delivering CRISPR expression vectors. The second major drawback is that one specific viral vector cannot be used multiple times since the body would have acquired immune memory which draws a limit to the number of times viral vectors can be used to deliver in one single patient. Furthermore, most viruses show tropism towards certain cell types. For example, SARS-CoV-2 targets ciliated and AT2 cells in the airway and alveolar regions<sup>106</sup>. However, in certain scenarios, it would make sense to have viral vectors that show a strong tropism to certain cell types. For example, when delivering CMAVV or CMAVT to combat SARS-CoV-2, it would make sense to use a viral vector that shows strong tropism to ciliated and AT2 cells in the airway and alveolar regions<sup>106</sup>. Or when targeting HIV-1, a viral vector that shows high tropism for Helper-T cells, T lymphocytes, or cells that express a high level of CD4 would be ideal, since HIV-1 entry is mediated by CD4<sup>107</sup>.

Viral Vector	Genetic Material	Packaging Capacity	NIH Risk Groups	CDC Recommended Biosafety Level (in vitro)	CDC Recommended Biosafety Level (in vivo)	Recommended Practice in a Clinical Setting	Host Range	Main Limitations	Main Advantages	Disinfection	References
Adeno-associated virus (all serotypes)	ssDNA	<5 kb	I	BSL-1 <sup>b</sup>	ABSL-1	Universal/standard precautions or BSL-2 practices	Broad host range; infective for many cell types, including neurons	Small packaging capacity	Nonimmunogenic; nonpathogenic	Freshly prepared 0.5% sodium hypochlorite is recommended disinfectant. Alcohol is not an effective disinfectant against AAV	1, 76, 107
Adenovirus (all serotypes)	dsDNA	8 kb (replication defective)	2	BSL-2	ABSL-2 <sup>c</sup>	Universal/standard precautions or BSL-2 practices	Broad host range, infective for many cell types	May initiate strong inflammatory response	Efficient transduction of most tissue	Freshly prepared 0.5% sodium hypochlorite is recommended disinfectant. Alcohol is not an effective disinfectant against adenovirus.	53, 68, 107
		30 kb (helper dependent)									
Retrovirus	RNA	8 kb	2	BSL-2	ABSL-2 <sup>d</sup>	Universal/standard precautions or BSL-2 practices	Broad host range	Only transduces dividing cells; integration might induce oncogenesis in some applications	Persistent gene transfer in dividing cells	Freshly prepared 0.5% sodium hypochlorite is recommended disinfectant.	19, 26, 82
Lentivirus	RNA	8 kb	3 (for HIV)	BSL-2	ABSL-2 <sup>d</sup>	Universal/standard precautions or BSL-2 practices	Ecotropic, amphotropic	Can transduce both dividing and nondividing cells; integration might induce oncogenesis in some applications	Persistent gene transfer in dividing cells	Freshly prepared 0.5% sodium hypochlorite is recommended disinfectant.	31, 82, 108
Herpesviral vector (herpes simplex virus-1)	dsDNA	40 kb (replication defective)	2	BSL-2	ABSL-2 <sup>d</sup>	Universal/standard precautions or BSL-2 practices	Broad host range	May initiate strong inflammatory response; transient gene expression in cells other than neurons	Large packaging capacity; strong tropism for neurons	Freshly prepared 0.5% sodium hypochlorite is recommended disinfectant.	9, 12, 91, 115
		150 kb (amplicon)									

<sup>a</sup>Risk classification differs between countries; only the US classification is presented here without taking into account the transgene. RG1: agents that are not associated with disease in healthy adult humans. RG2: agents that are associated with human disease that is rarely serious and for which preventive or therapeutic interventions are available. RG3: agents that are associated with serious or lethal human disease for which preventive or therapeutic interventions may be available (high individual risk but low community risk). RG4: agents that are associated with serious or lethal human disease for which preventive or therapeutic interventions are not usually available (high individual risk and high community risk). AAV, adeno-associated viruses; CDC, Centers for Disease Control and Prevention; ds, double-stranded; kb, kilobases; NIH, National Institutes of Health; RG, risk group; ss, single-stranded.

<sup>b</sup>BSL-1 refers to the containment level based on parent virus risk group. However, most procedures involving the handling and manipulation of the viral vectors are done at BSL-2 laboratories to protect cell cultures and viral stocks from contamination.

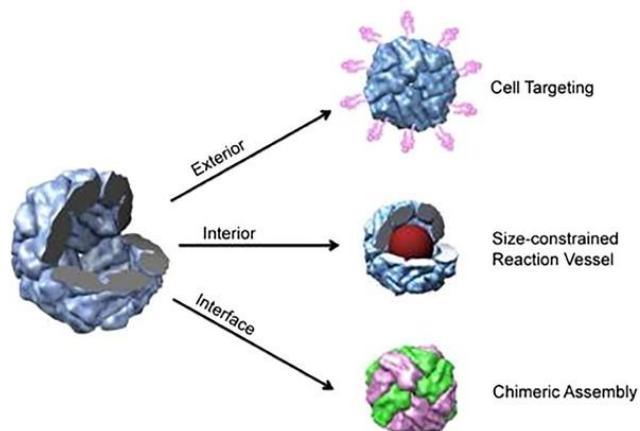
<sup>c</sup>ABSL-2: Animals after 7 days may be housed at ABSL-1 labs.

<sup>d</sup>ABSL-2: Animals after 48 hours may be housed at ABSL-1 labs.

Table 4.1 A Summary of all Commonly Used Viral Vectors and Their Relative Facts. Table adapted from<sup>108</sup>

Though adeno-associated virus vectors yield good results in gene therapy, it would be ideal to have effective non-viral delivery mechanisms which don't elicit an immune response, such as nanotechnology. A range of different nanoparticles have been extensively studied and have shown their in vivo applicability<sup>109,110</sup>. First, protein nanocages recently have provoked a great excitement in the field of drug delivery, they have the capacity to carry the CRISPR expression vectors and can target cells/tissues with high specificity<sup>111,112</sup>. (Figure 8.1)

Figure 8.1 Protein cage modifications at distinct interfaces, adapted from<sup>113</sup>



Second, lipid-based nanoparticles such as exosomes have also shown great applicability in delivering CRISPR plasmid expression vectors *in vivo*<sup>114</sup>. Furthermore, a study more related to CRISPR's antiviral potential by Kaushik et al. used magnetically guided, magnetic nanoparticles to eradicate HIV-1 infection across the blood-brain barrier<sup>115</sup>, introducing inorganic nanoparticles as a strong candidate for delivering the CRISPR Cas system.

Considering the drawbacks of viral vectors, and the promising results yielded from both organic and non-organic nanoparticles it is quite clear that nanotechnology is likely going to be the ideal delivery mechanism of the plasmid expression vectors *in vivo*.

## **Potential uses of CMAVT and CMAVV Outside of the Human Body**

CMAVT and CMAVV described in this paper aren't limited to the human body. It has multiple uses in other areas such as agriculture and animal husbandry to reduce financial loss due to viral infections in crops and cattle. Furthermore, since those transgenic animals and crops are easier to produce<sup>116</sup> and less ethically debated, CMAVT and CMAVV in crops and cattle would definitely look promising in the near future.

### **CMAVT and CMAVV in *Bombyx mori***

Liu Y et al. had constructed a CRISPR Cas 9 antiviral system that is induced by Baculovirus and targets *Bombyx mori* nucleopolyhedrovirus (BmNPV) the cause of 80% of financial loss in silkworm production<sup>117</sup>. They targeted late expression factors 1 and 3 which are essential for viral DNA replication<sup>117</sup>. Dong et al. were also working on this topic<sup>118</sup>, for their expression vector design they used the 39k promoter which had the highest BmNPV-induced transcriptional activity<sup>119</sup> followed by the Cas9 gene. They also included a U6 promoter followed by four guide RNA sequences targeting factors 1 and 3. Overall, their design of the expression vectors was in line with what this paper suggested (Figure 9.1).



Figure 9.1 A short section of the 2 expression vectors made by Liu Y et al. Adapted from<sup>118</sup>

The left is a section for the expression vector for the Cas9 nuclease using a BmNPV induced 39k promoter, then a fluorescent protein EGFP using a 3XP3 promoter. The right is for the expression of the guide RNA using a strong U6 promoter, using an RNA polymerase III transcription terminator sequence at the end of the sgRNA. A fluorescent protein DsRed is also used.

The system was rapidly activated (Figure 9.3) when the silkworm was infected and showed considerably higher resistance to BmNPV infection than the wild-type silkworm<sup>117</sup>, (Figure 9.2) demonstrating CMAVT and CMAVV's in vivo applicability and its antiviral potential.

In the study of Dong et al.<sup>118</sup> targeted viral proteins have also been significantly suppressed. (Figure 9.4)

Figure 9.2 Right, resistance to BmNPV infection in transgenic and wild-type silkworms. Adapted from Liu et al.<sup>117</sup>

Survival rate analysis of Transgenic type (TG) and Wild type (WT) (A) Control, uninfected (B)  $1 \times 10^5$  OBs (occlusion bodies)/larva (C)  $1 \times 10^6$  OBs/larva, and (D)  $1 \times 10^7$  OBs/larva. Values are the means of three independent replicates with 30 larvae in each replicate. Statistical analysis of mortality was conducted 10 days after inoculation with OBs.

Note: OBs are a measure of viral load.

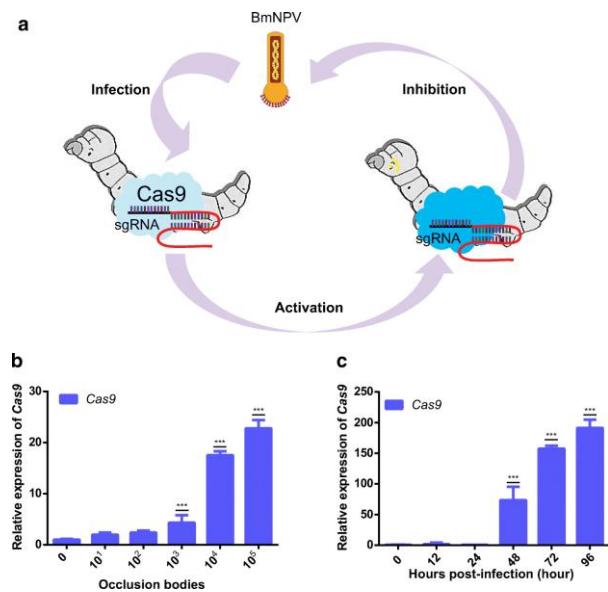
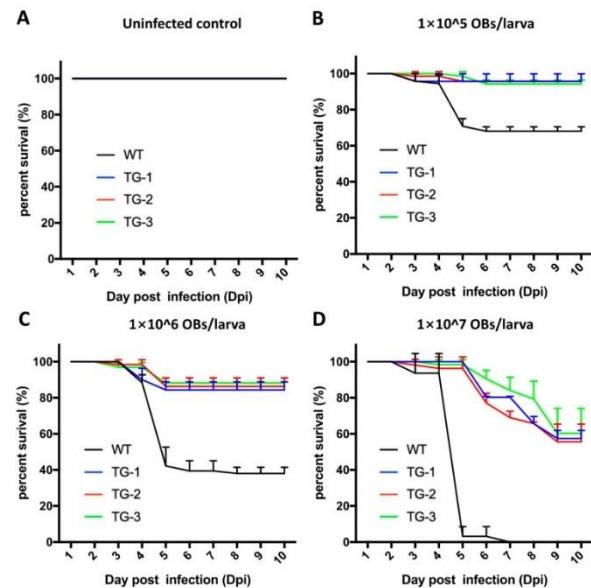
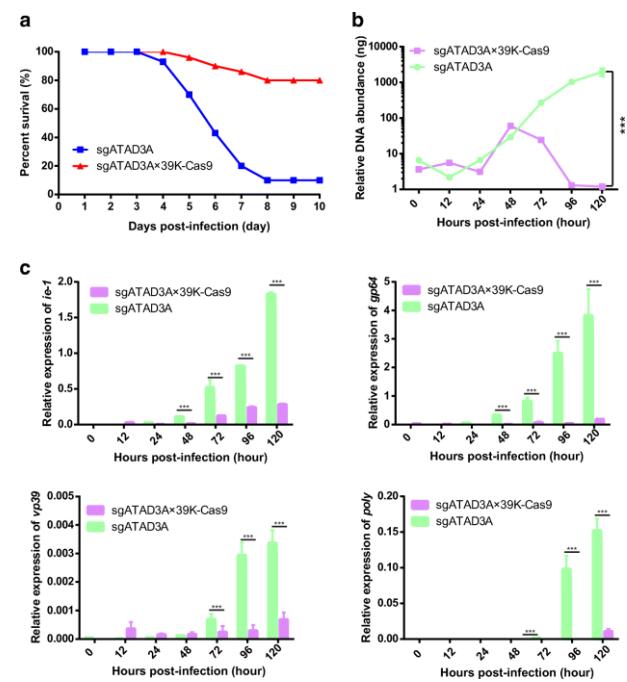


Figure 9.3 Left, adapted from Dong et al.<sup>118</sup> (a)  
Shows the summary steps of the antiviral system  
(b) Shows the trend of relative expression of Cas9 over OBs.  
(c) Shows the trend of relative expression of Cas9 over time (post-infection).  
Note: OBs are a measure of viral load.

Figure 9.4 Right, adapted from Dong et al.<sup>118</sup> (a) The percentage survival of *Bombyx mori* over time with blue as control not expressing the Cas9 nuclease.

(b) Relative DNA abundance over time with green as control  
(c) The relative expression of the 4 different viral genes: ie-1, gp64, vp39, and poly. The expression of these 4 different viral genes was very successfully suppressed (shown in purple). Green is a control group.



## CMAVT and CMAVV in Plants

Extensive studies of controlling plant viruses using the CRISPR/Cas system had also shown promising results in reducing yield loss of agriculturally important plant crops<sup>120–123</sup> (Table 5.1).

Virus/viruses	Plant	Target (Viral/host)	GM/Transgene free	References
BSCTV	<i>Nicotiana benthamiana</i> and <i>Arabidopsis thaliana</i>	IR, CP, and Rep	GM	Ji et al., 2015 <sup>124</sup>
BeYDL	<i>N. benthamiana</i>	LIR and Rep/RepA	GM	Baltes et al., 2015 <sup>125</sup>
TYLCV, BCTV, and MeMV	<i>N. benthamiana</i>	IR, CP, and Rep	GM	Ali et al., 2015 <sup>126</sup>
CLCuKoV, TYLCV 2.3, TYLCSV, MeMV, BCTV-Logan, BCTV-Worland	<i>N. benthamiana</i>	IR, CP, and Rep	GM	Ali et al., 2016 <sup>127</sup>
TuMV	<i>A. thaliana</i>	Host factor eIF(iso)4E	Transgene free	Pyott et al., 2016 <sup>128</sup>

Virus/viruses	Plant	Target (Viral/host)	GM/Transgene free	References
CVYV, ZYMV, and PRSMV	<i>Cucumis sativus</i>	Host factor <i>eIF4E</i>	Transgene free	Chandrasekaran <i>et al.</i> , 2016 <sup>129</sup>

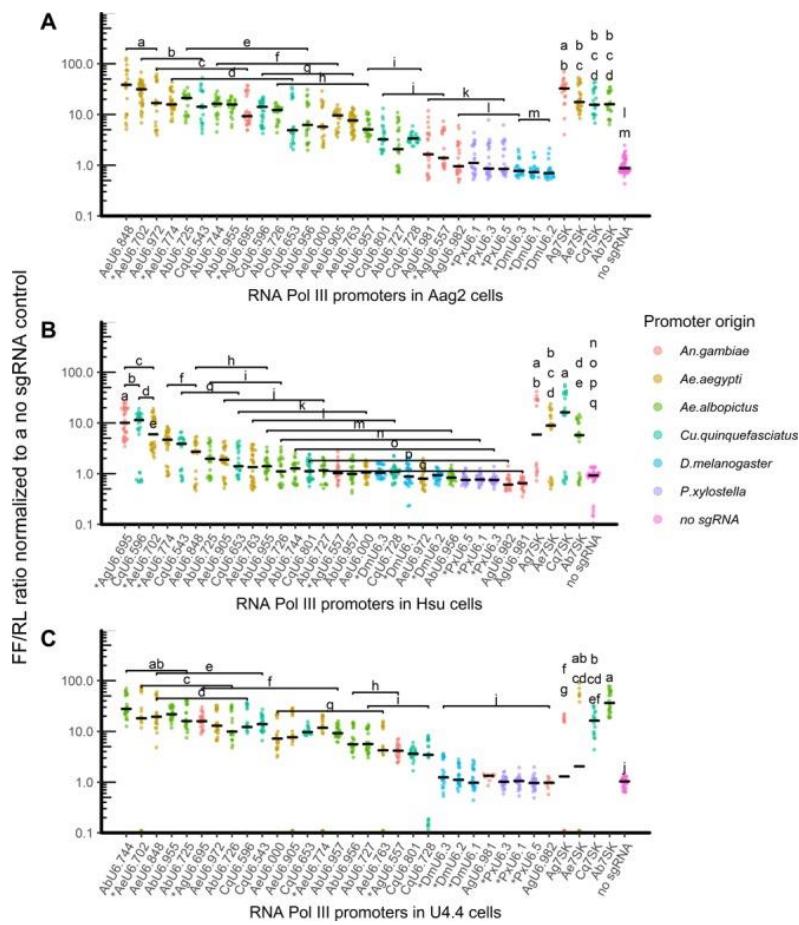
Table 5.1 Summary of some studies using CRISPR/Cas9 to mediate viral resistance in plants, adapted from Zaidi et al.<sup>130</sup> Genetically Modified (GM), Beet severe curly top virus (BSCTV), Bean yellow dwarf virus (BeYDV), Tomato yellow leaf curl virus (TYLCV), Beet curly top virus (BCTV), Merremia mosaic virus (MeMV), Cotton leaf curl Kokhran virus (CLCuKoV), Tomato yellow leaf curl Sardinian virus (TYLCSV), Turnip mosaic virus (TuMV), Cucumber vein yellowing virus (CVYV), Zucchini yellow mosaic virus (ZYMV), Papaya ring spot mosaic virus (PRSMV), intergenic region (IR), coat protein (CP), replication-associated protein (Rep), long intergenic region (LIR).

## **Controlling the Spread of Viral Diseases Using CMAVT and CMAVV in *Aedes aegypti***

CMAVV's usage is not only limited to commercial uses in humans, cattle, and crops. It can also be applied to control certain viral diseases that are transmitted by common vectors such as the mosquito *Aedes aegypti*, which is a known vector of Arthropod-borne viruses (arboviruses), including Zika (ZIKV), dengue (DENV), chikungunya (CHIKV), and yellow fever viruses (YFV) that are emerging global health threats which have recently exhibited resurgence in prevalence<sup>131</sup>. *A. aegypti* being a highly invasive species, it has adapted to urban environments<sup>132</sup>. Its preference of feeding on humans and high vectorial capacity makes it a great target for controlling the arboviruses<sup>131</sup>. Currently, *A. aegypti* populations are controlled by the application of insecticides namely DDT<sup>133</sup>. However, DDT resistance developed quickly<sup>134</sup> and its negative effects persisted in the environment<sup>135–137</sup>. Therefore, using pesticides to control the disease by killing viral vectors is not a sustainable and reliable method. Thus currently, novel mosquito control strategies are based on genetic manipulation of mosquitoes, asserting gene drives thorough the population that aim to achieve either 1. Population suppression, eradicating the mosquito population by inducing lethal alleles or sterility. Or 2. Population replacement, replacing the wild-type mosquito with genetically engineered mosquitos that have little or no vectorial capacity. For population suppression of *A. aegypti* or mosquito virus vectors in general, their population-level eradication is likely to cause major effects up and down the food chain and damage ecosystems<sup>138,139</sup>. Therefore, Population replacement would be the optimal choice. Using *A. aegypti* as an example, we will explore the details of how we can genetically modify them to limit their vectorial capacity for the ZIKV (+ ssRNA) to achieve population replacement. A gene drive would need to be associated with 2 expression vectors that will be integrated into the genome of *A. aegypti*. One expression vector should contain the Cas13d nuclease with a strong promoter that has high promoter activity during the adult stage (Figure 9.6) which means strong promoters during early and mid-development such as UbL40 promoter and polyubiquitin promoter<sup>140</sup> should

not be used. Therefore, I shall choose the hsp83 gene promoter which drives strong gene expression throughout development and during adulthood, in nearly all cells, as shown by Developmental RNA-seq analyses<sup>141,142</sup>. Hsp83 promoter activity also increases under heat shock, In *A. aegypti*, a 1-hr heat shock at 42°C produced a ten-fold increase in Aeahsp83 RNA levels in adults compared to 23°C controls<sup>143</sup>, this feature of the hsp83 promoter would be quite useful since the population spread of *A. aegypti* is in hot areas which the heat could give frequent, small surges in the promoter activity. As for the crRNA / guide RNA expression vector (Figure 9.7), the *A. aegypti* U6 promoter<sup>144</sup> and 7SK promoter<sup>145</sup> are all strong candidates. However, the 7SK promoter would be more ideal since there was a trend toward 7SK promoter sequences having stronger activity levels than U6 promoter sequences<sup>145</sup> (Figure 9.5).

Figure 9.5 adapted from Anderson et al.<sup>145</sup> dCas9-VPR assay in vitro. Ratios of FF/RL luciferase normalized to a no-sgRNA control are shown. Promoters are organized by median relative activation within U6 and 7SK promoter categories, and the colours denote the promoter origin by species. Lowercase letter groupings denote significant differences at  $P < 0.05$  following post hoc analysis. Each point represents one well of a 96-well plate, with at least eight replicate wells transfected in at least three replicate experiments.



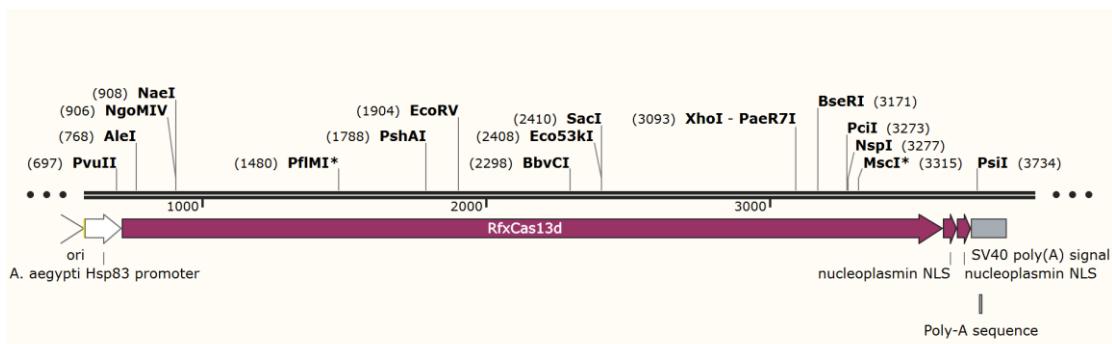


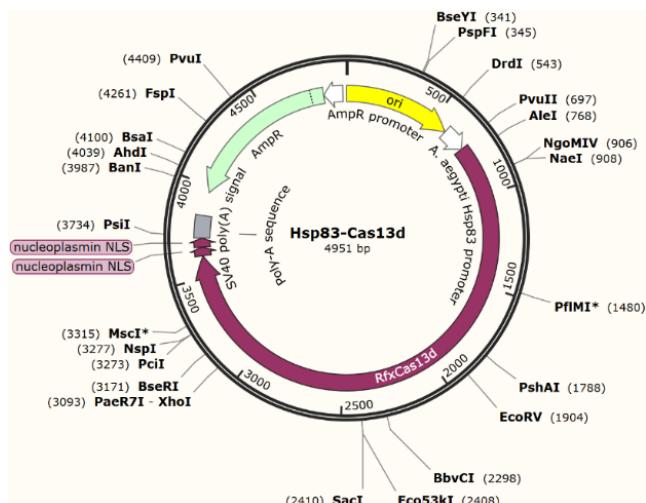
Figure 9.6 Expression vector to be integrated into *Aedes aegypti* made by author using SnapGene®. Using a strong *Aedes aegypti* hsp83 promoter followed by the RfxCas13d gene<sup>27</sup> followed by 2 nucleoplasmmin NLS<sup>100</sup> and an SV40 poly(A) signal to enhance translation. SV40 poly(A) signal could be replaced with a 3' UTR.

For detailed reasoning of design, see appendix (A).



Figure 9.7 Expression vector to be integrated into *Aedes aegypti* made by author using SnapGene®. Using a strong 75K promoter<sup>145</sup> followed by the crRNA sequence and RfxCas13d DR36<sup>27</sup> that is required for the formation and functionality of the Cas13d crRNA complex, ensuring the binding and degrading of the selected target sequence. Transcription halted by an RNA polymerase III transcription terminator sequence. The entire expression vector contains 4 repeats of the promoter, crRNA, RfxCas13d DR36<sup>27</sup> and pol III terminator sequence. For detailed reasoning of design, see appendix (A).

Figure 9.8.1 Plasmid for vector cloning optimized for E. coli (made by author using SnapGene®) for the hsp83-Cas13d expression vector (Figure 9.6). AmpR promoter with AmpR gene inducing ampicillin resistance. Ori sequence for E. coli.  
For detailed reasoning of design, see appendix (A).



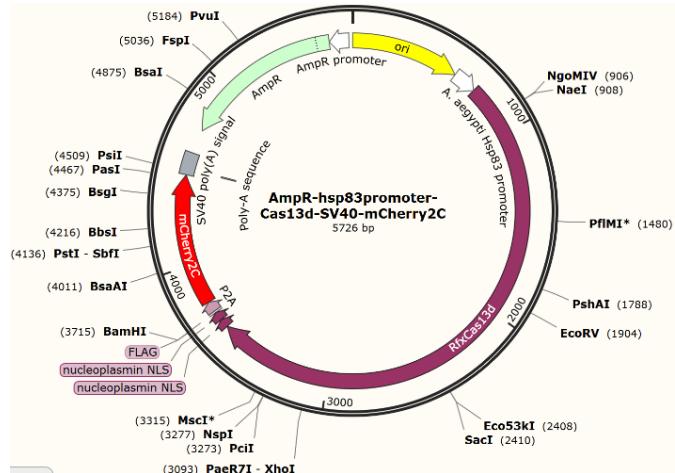


Figure 9.9 Plasmid for vector cloning optimized for *E. coli* (made by author using SnapGene®) for the 7SK-crRNA expression vector (Figure 9.7) AmpR promoter with AmpR gene inducing ampicillin resistance. Ori sequence for *E. coli*. For detailed reasoning of design, see appendix (A).

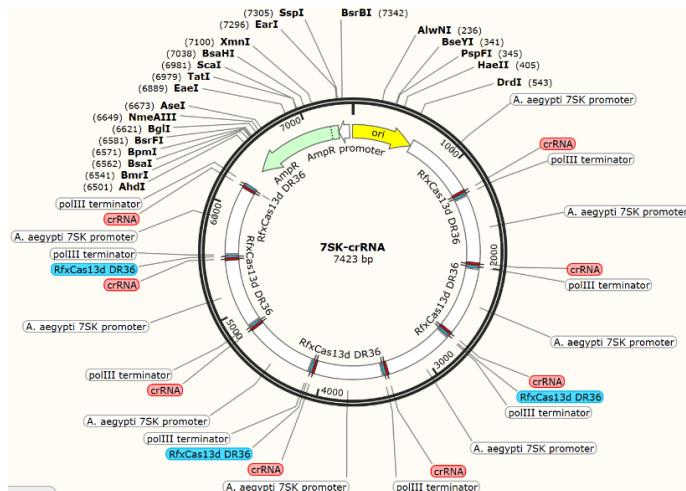


Figure 9.8.2 Alternative version of the vector cloning plasmid outlined in figure 9.8.1. with mCherry2C fluorescent protein<sup>90</sup> added to allow RfxCas13d expression levels to be monitored by monitoring the fluorescent intensity of mCherry (red). P2A self-cleaving sequence<sup>91</sup> added in between RfxCas13d<sup>27</sup> sequence and mCherry2C<sup>90</sup>, separating RfxCas13d and mCherry2C. For detailed reasoning of design, see appendix (A).

A study done by Tng et al. had demonstrated the applicability of the CRISPR/Cas system in vector disease control by targeting chikungunya virus (CHIKV) using Cas13b in *A. Aegypti*<sup>146</sup>, their 2 RNA guides were designed against the non-structural protein 2 (nsP2) region of CHIKV, which was expressed in the mosquito using viral reporters (Figure 9.10). The reporter firefly luciferase and nanoluciferase activity were shown to be suppressed (Figure 9.10) (Figure 9.11), indicating Cas13b's success in targeting and degrading the CHIKV viral sequences in *A. aegypti*.

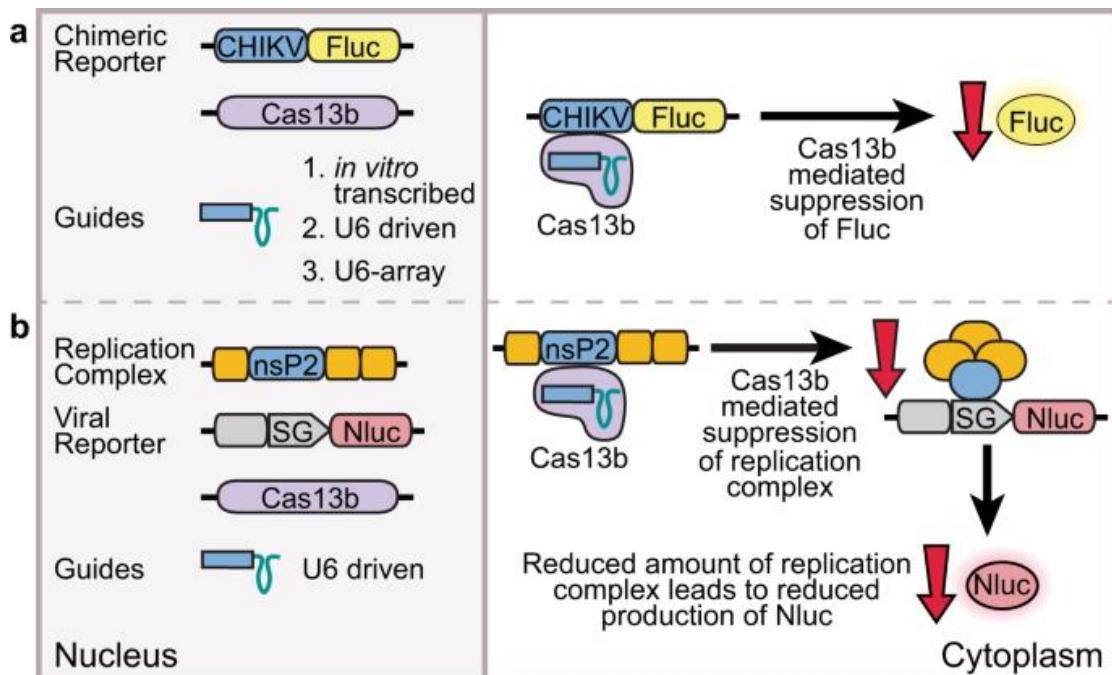


Figure 9.10 Adapted from Tng et al.<sup>146</sup>. In vitro-transcribed or U6-driven guides were used to assess if Cas13b can target viral sequences in (a) directly targeted chimeric reporter, or (b) a viral reporter responding to targeted viral replicase.

Note: CHIKV: chikungunya virus sequence, Fluc: firefly luciferase, U6: Ae. aegypti U6-3 promoter, SG: CHIKV subgenomic promoter, Nluc: nanoluciferase.

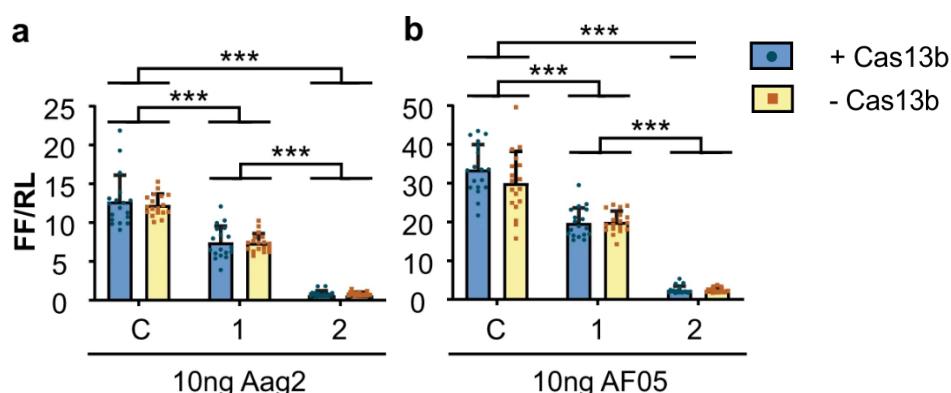


Figure 9.11 adapted from<sup>146</sup>. Suppression of Firefly Luciferase induced luminance indicating the suppression of the target viral protein elements. Surprisingly, the data had shown Cas13b independent suppression of the viral protein elements as shown in yellow. This Cas13b independent suppression might be due to RNA interference, which had not been encountered before in other similar studies.

Note: C: non-targeting control, 1: guide 1 crRNA, 2: guide 2 crRNA

This study by Tng et al.<sup>146</sup> shows the potential of the CRISPR/Cas system to limit the spread of viral disease by mosquito vectors. However, the fact that this study was conducted *ex vivo*, using *A. aegypti* cell lineages, and the fact that artificial viral reporters for CHIKV were used instead of the actual virus limits this study from being strong evidence. Nevertheless, the effectiveness of the CRISPR/Cas system in degrading viral sequences in mosquito cells had been demonstrated.

## Effectiveness of CMAVT and CMAVV in Suppressing Viral Infections

Many studies have confirmed the effectiveness of CMAVT and CMAVV in suppressing viral infections. Abbott T et al.<sup>82</sup> using their PAC-MAN strategy had suppressed SARS-CoV-2 in human lung epithelial cells (A549 cell line). Reporter expressed fluorescence and RNA abundance both greatly suppressed by a mean of around 85% using G4 and G6 crRNA. (Figure 10.1)

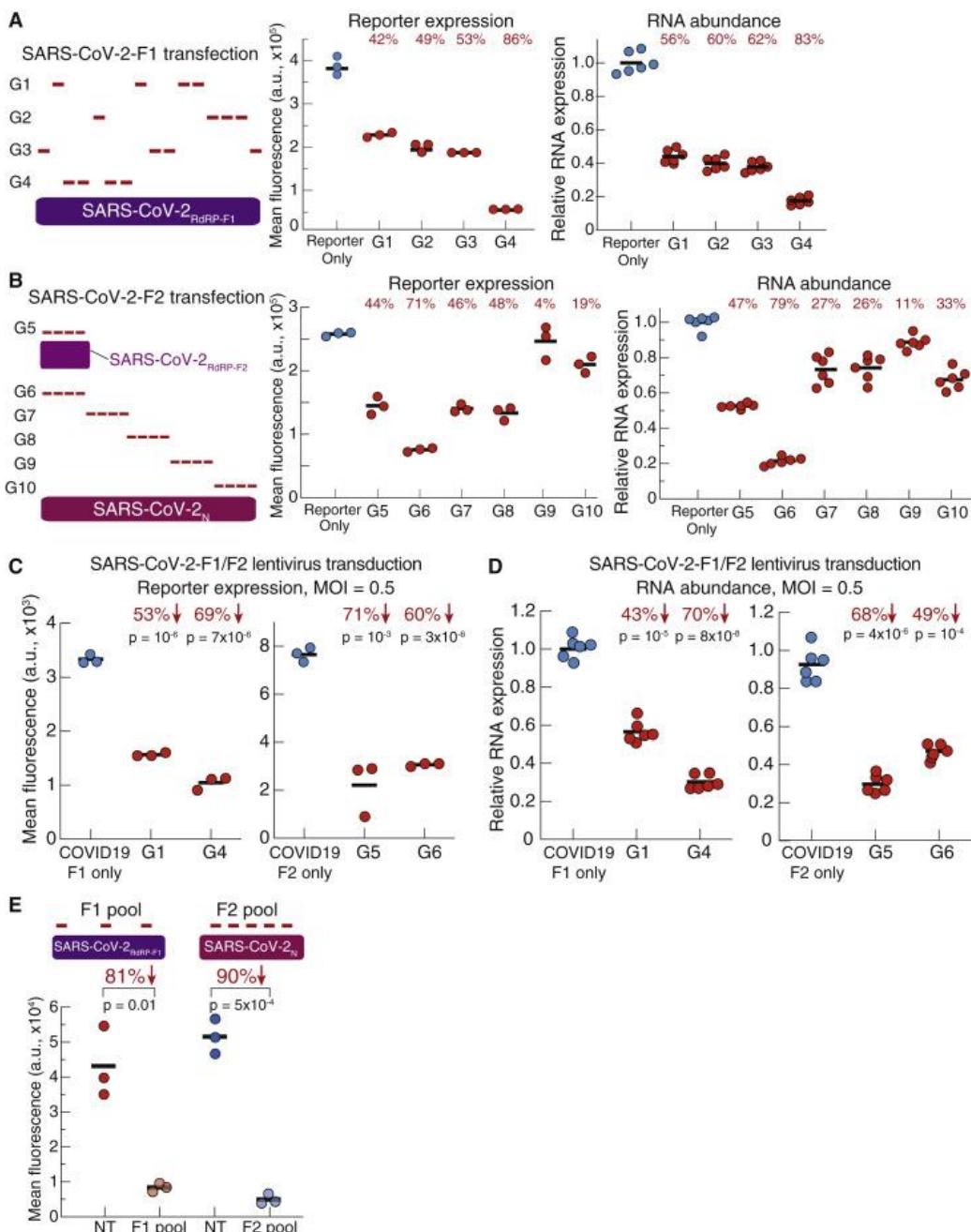


Figure 10.1 PAC-MAN Can Inhibit SARS-CoV-2 Reporters, adapted from<sup>82</sup>

(A and B) Left: schematics of pools of crRNAs targeting transfected (A) SARS-CoV-2-F1 or (B) SARS-CoV-2-F2 reporters. Middle: GFP expression as measured by flow cytometry. Right: mRNA abundance as measured

by quantitative real-time PCR. Relative RNA expression is calculated by normalizing to the reporter-only sample.

(C) GFP expression as measured by flow cytometry when SARS-CoV-2-F1 (left) or SARS-CoV-2-F2 (right) is delivered via lentiviral transduction. p values for G1 ( $p = 10^{-6}$ ) and G4 ( $p = 7 \times 10^{-6}$ ) are relative to SARS-CoV-2-F1 only, while p values for G5 ( $p = 0.001$ ) and G6 ( $p = 3 \times 10^{-6}$ ) are relative to SARS-CoV-2-F2 only. MOI = 0.5.

(D) mRNA abundance is measured by quantitative real-time PCR when SARS-CoV-2-F1 (left) or SARS-CoV-2-F2 (right) is delivered via lentiviral transduction. Relative RNA expression is calculated by normalizing to the reporter-only sample. p values for G1 ( $p = 10^{-5}$ ) and G4 ( $p = 8 \times 10^{-8}$ ) are relative to SARS-CoV-2-F1 only, while p values for G5 ( $p = 4 \times 10^{-6}$ ) and G6 ( $p = 1 \times 10^{-4}$ ) are relative to SARS-CoV-2-F2 only. MOI = 0.5.

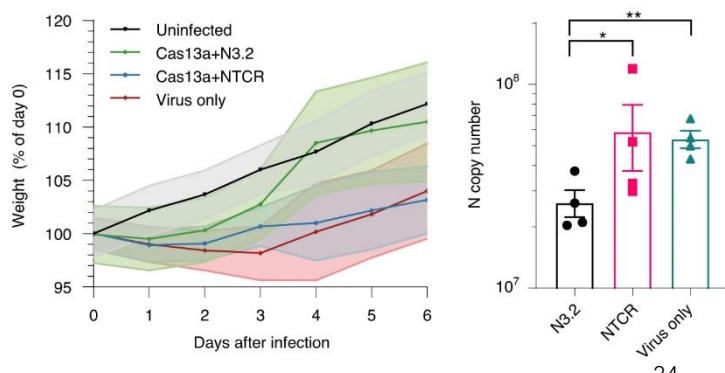
(E) Top: schematic of pools of crRNAs tiling across SARS-CoV-2-F1 and SARS-CoV-2-F2 reporters. Bottom: GFP expression levels as measured by flow cytometry. Red, SARS-CoV-2-F1 reporter; blue, SARS-CoV-2-F2 reporter. NT, non-targeting crRNAs. p values are relative to the NT samples;  $p = 0.01$  for the F1 pool and  $p = 5 \times 10^{-4}$  for the F2 pool.

Though the results gained suggest that CMAVT and CMAVV are effective in suppressing viral infections, the fact that this experiment was performed *ex vivo* using cells of the A549 cell line limits the reliability of this study in exploring the effectiveness of CMAVT and CMAVV in humans. In addition, genetically edited cells that express viral reporters which contained a fraction of the viral genome were used instead of the actual virus, which again limits this study, but does not affect its reliability.

Regarding the effectiveness of CMAVT and CMAVV in *Bombyx mori*, the two studies done by Dong et al.<sup>118</sup> and Liu et al.<sup>117</sup> had shown a great suppression of the 4 targeted viral protein elements (Figure 9.4) and a higher survival rate of *Bombyx mori* (Figure 9.2). Considering the fact that in both studies, the experiments were conducted *in vivo*, using real viruses, we can safely conclude that CMAVT and CMAVV are effective in *Bombyx mori* and will likely be effective *in vivo* in most other animals, including humans as well.

Since, lab mice and hamsters are common test subjects that are related to humans, an *in vivo* study indicating the effectiveness of CMAVT and CMAVV in them would to some degree, suggest the effectiveness of CMAVT and CMAVV in humans. Luckily, Blanchard et al. had already demonstrated that mRNA-encoded Cas13a in rodents suppresses influenza and SARS-CoV-2 infections (Figure 10.2)<sup>147</sup>.

Figure 10.2 adapted from<sup>147</sup>. The left graph shows the change of weight in the rodents over days after infection. The right graph shows the Lung viral loads from hamsters 6 days after infection ( $n = 4$ ). Note: N3.2 is the targeting guide crRNA. NTCR = non-targeting crRNA control. Data represent



mean N copy number  $\pm$  s.e.m. Brown–Forsythe and Welch ANOVA with Dunnett's multiple comparisons on log-transformed data, where \*\*P = 0.0016 and \*P = 0.0198.

Since the weight change might be due to multiple different factors, the line graph that shows the change in weight (Figure 10.2, left) is on its own insufficient to prove CMAVT and CMAVV's effectiveness. However, the bar chart (Figure 10.2, right) indicating a significant suppression of viral load in the lungs reinforces the weight change line graph, making the data more reliable. Therefore, we can safely conclude that CMAVT and CMAVV are effective in mice and hamsters, indicating that CMAVT and CMAVV may also be effective in humans.

## **Conclusion**

Firstly, as outlined in this paper, CRISPR class 2 type II, type V and type VI systems can target all types of viruses in the world and can target all stages of the viral replication cycle which ultimately inhibits viral packaging and release, greatly minimizing the viral load inside the body if not eradicating it completely. However, CMAVT and CMAVV cannot directly target the genome of dsDNA viruses since we lack a Cas nuclease that can target dsDNA which is a small drawback, but we can still target its transcriptome.

Secondly, with the target sequence correctly selected as suggested in this paper, CMAVT and CMAVV would still most likely be effective against any new or mutant strains of the virus and antigenic switching does not affect the efficacy. Furthermore, CMAVT and CMAVV have great flexibility. It can be used to target any newly emerging pathogenic viruses in the future as long as we can quickly identify and select potential target sequences in the viral genome using the criteria outlined in this paper as a guide, we can then alter the crRNA spacer sequence to match, which is an easy and very quick process, much faster than traditional vaccine developments.

Thirdly, nanotechnology had many key breakthroughs recently, allowing the delivery of the expression vectors to be cheaper, specific, efficient, and with an increased capacity. However, making the delivery process even cheaper and increasing the capacity further would be ideal.

Fourthly, many studies either in vitro, ex vivo, or in vivo have demonstrated that CMAVT and CMAVV greatly suppress viral infection rates, viral load, and viral genome expressions inside cells and reduce mortality rates, which again demonstrates its potential. However, the effectiveness of CMAVT and CMAVV for humans have only been tested using human cell lines, and the effectiveness demonstrated in lab mice and hamsters do not necessarily mean a correlation with the effectiveness of CMAVT and CMAVV in humans. Thus, a few in vivo studies and randomized controlled clinical trials in humans is essential.

In conclusion, CRISPR class 2 type II, type V and type VI systems are an effective treatment against viral infections in crops and animals, with its high flexibility, and strong suppression of viral load and infection rate which significantly decreases mortality rates (*Bombyx mori*). However, we cannot safely conclude its effectiveness in humans yet due to the lack of experimental data, but the efficacy demonstrated using human cell lines and rodents show its great potential.

## **Appendix (A)**

Some features used in the construction of the DNA plasmids will be explained here.

1. AmpR: Ampicillin resistance gene, produces  $\beta$ -lactamase, providing ampicillin resistance for vector cloning in bacteria. Type: CDS
2. P2A: 2A peptide from porcine teschovirus-1 polyprotein, self-cleaving, used to separate 2 peptides/proteins, usually put in between a protein of interest and a fluorescent protein such as GFP, used to monitor gene expression levels. Type: CDS
3. nucleoplasmin NLS: bipartite nuclear localization signal from nucleoplasmin, targets the protein for transport into the nucleus, mainly associated with the Cas9 nuclease. Type: CDS
4. FLAG: FLAG® epitope tag, followed by an enterokinase cleavage site, used for protein purification from mammalian expression systems. Type: CDS
5. AmpR promoter: Ampicillin resistance gene promoter, promotes the expression of the AmpR gene. Type: promoter for RNA pol II
6. EF-1 $\alpha$  promoter: strong constitutive promoter for human elongation factor EF-1 $\alpha$ . Used to drive strong Cas nuclease gene expression in studies. Type: promoter for RNA pol II
7. U6 promoter: RNA polymerase III promoter for U6 snRNA. Type: promoter for RNA pol III
8. Hsp83 promoter: *Aedes. aegypti* heat shock protein 83 promoter. Type: promoter for RNA pol II
9. hPGK promoter: human phosphoglycerate kinase 1 promoter, with weak promoter activity, use described earlier in this paper. Type: promoter for RNA pol II
10. 39k promoter: *Bombyx mori* nucleopolyhedrovirus inducible promoter. Type: promoter for RNA pol II
11. A. aegypti 7SK promoter: strong 7SK promoter derived from the mosquito A. aegypti. Type: promoter for RNA pol III
12. CMV enhancer: human cytomegalovirus immediate-early enhancer, used to promote gene expression. Type: enhancer
13. Ori: high-copy-number ColE1/pMB1/pBR322/pUC origin of replication, used for vector cloning in bacteria. Type: rep\_origin
14. 5' LTR: 5' long terminal repeat (LTR) from HIV-1, an HIV-1 inducible promoter activated by Tat<sup>97,98</sup>. Type: LTR
15. 3' LTR ( $\Delta$ U3): self-inactivating 3' long terminal repeat (LTR) from HIV-1. Acts as a

- secondary promoter sequence when 5' LTR is inactivated and also acts as a transcription terminator. Type: LTR
16. RRE: The Rev response element (RRE) of HIV-1 allows for Rev-dependent mRNA export from the nucleus to the cytoplasm. Type: misc\_feature
  17. cPPT/CTS: central poly-purine tract and central termination sequence of HIV-1. Type: misc\_feature
  18. WPRE: woodchuck hepatitis virus posttranscriptional regulatory element, increases transgene expression from a variety of viral vectors<sup>148</sup>. Type: misc\_feature
  19. HIV-1 Ψ: packaging signal of human immunodeficiency virus type 1. Type: misc\_feature

## **Appendix (B)**

Detailed process in designing the crRNA for targeting SARS-CoV-2.

Firstly, I chose to target the RdRp gene of the highly conserved ORF1ab section. Then I constructed an array of all possible 22 nucleotide spacer sequences based on the RdRp gene sequence using the SARS-CoV-2 genome map<sup>149</sup> provided by SnapGene®. Then I immediately filtered out sequences with poly-U inside the sequence and sequences with a GC percentage of either under 40% or over 60%, this ensures a high cleavage efficiency on the target sequence. I then calculated the annealing temperatures and filtered out any that were lower than 55 °C. All the processes described above was done by a program coded by myself in the coding language Python using Python 3.8.5 32-bit (system). To access the code, click on the following link: [PYTHON CODE](#). Note: equation of calculating annealing temperature is

$$T_m (\text{°C}) = 81.5 + 0.41(\%GC) - (675/N)^{150}$$

where %GC is the percentage of G and C nucleotides in the oligo and N is the length of the oligo given in nucleotides.

Using the python program, it had selected 852 crRNA spacer sequences out of 2774 possible spacer sequences that can be made out of the 2795 nucleotides in the SARS-CoV-2 RdRp gene (Figure 11.1) ([crRNA Array File](#))

```
UAUGUACACACCGCAUACAGUC
AUGUACACACCGCAUACAGUCU
UGUACACACCGCAUACAGUCUU
GUACACACCGCAUACAGUCUUA
UACACACCCGAUACAGUCUUAC
ACACACCCGAUACAGUCUUACA
CACACCCGAUACAGUCUUACAG
Target gene: SARS-CoV-2 RdRp
Number of nucleotides in SARS-CoV-2 RdRp gene:
2795
Total number of crRNA spacer sequences possible for targeting SARS-CoV-2 RdRp:
2774
Total number of selected crRNAs:
852
PS C:\Users\路奇\OneDrive\桌面\CRISPR resource\find_crRNA> []
```

Figure 11.2, results gained from the python program, made by author.

Secondly, For each of the selected sequences ([Multi-query BLAST, FASTA format file](#)) I conducted a BLAST®<sup>151</sup> (blastn suite) against the complete genome sequence of *Homo sapiens*, provided in the database of NCBI (BLAST®<sup>151</sup> [DATA FILE-1](#)). Then I

compared the BLAST®<sup>151</sup> results of all the sequences. I found that the sequence: **GUACUGAUGUCGUUAUACAGGGC** had on average, a very low sequence of alignment maximum score with the *Homo sapiens* genome (28.322) calculated using the BLAST®<sup>151</sup> data (Figure 11.2). It had a high number of mismatches.

Sequences producing significant alignments:										
Description	Scientific Name	Common Name	Max Taxid	Total Score	Query cover	E value	Per cent identity	Len	Accession	
Homo sapiens family with sequence similarity 230 member I...	Homo sapiens	human	9606	30.2	30.2	68%	29	100.00	1598	NR_165490.1
Homo sapiens family with sequence similarity 230 member I...	Homo sapiens	human	9606	30.2	30.2	68%	29	100.00	1495	NR_165489.1
Homo sapiens family with sequence similarity 230 member I...	Homo sapiens	human	9606	30.2	30.2	68%	29	100.00	1505	NR_165488.1
Homo sapiens family with sequence similarity 230 member I...	Homo sapiens	human	9606	30.2	30.2	68%	29	100.00	1578	NR_110539.2
Homo sapiens chromosome 22 clone CH17-429E15, complete sequence	Homo sapiens	human	9606	30.2	30.2	68%	29	100.00	220256	AC255379.1
Homo sapiens isolate CHM13 chromosome 22	Homo sapiens	human	9606	30.2	901	81%	29	100.00	51324926	CP068256.2
Homo sapiens DNA, chromosome 22, nearly complete genome	Homo sapiens	human	9606	30.2	338	81%	29	100.00	46684173	AP023482.1
Homo sapiens hypothetical LOC388882, mRNA (cDNA clone MGC:4630...)	Homo sapiens	human	9606	30.2	30.2	68%	29	100.00	1626	BC036910.1
Homo sapiens genomic DNA, chromosome 22q11.2, clone KB1269D1	Homo sapiens	human	9606	30.2	30.2	68%	29	100.00	186421	AP000344.1
Homo sapiens family with sequence similarity 230-like...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	4002	NR_166643.1
Homo sapiens family with sequence similarity 230 member D...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	2984	NR_136570.2
Homo sapiens family with sequence similarity 230 member E...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	3174	NR_165635.1
Homo sapiens family with sequence similarity 230 member H...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	3106	NR_136559.2
Homo sapiens family with sequence similarity 230 member J...	Homo sapiens	human	9606	28.2	141	63%	114	100.00	3530	NR_165634.1
Homo sapiens family with sequence similarity 230 member F...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	3191	NR_136571.2
Homo sapiens family with sequence similarity 230 member A...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	3186	NR_136560.2
Homo sapiens family with sequence similarity 230 member B...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	3202	NR_165629.1
Homo sapiens family with sequence similarity 230 member B...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	3464	NR_165621.1
Homo sapiens uncharacterized LOC107987394 (LOC107987394)...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	3282	NR_165501.2
Homo sapiens uncharacterized LOC107987394 (LOC107987394)...	Homo sapiens	human	9606	28.2	112	63%	114	100.00	3309	NR_165500.2
Homo sapiens gamma-glutamyl carboxylase (GGCX), transcript...	Homo sapiens	human	9606	28.2	28.2	63%	114	100.00	7385	NM_001142269.4
Homo sapiens gamma-glutamyl carboxylase (GGCX), transcript...	Homo sapiens	human	9606	28.2	28.2	63%	114	100.00	7556	NM_000821.7
Homo sapiens unk zinc finger (UNK), transcript variant 1, mRNA	Homo sapiens	human	9606	28.2	28.2	81%	114	94.44	3890	NM_001080419.3
Homo sapiens chromosome 22 clone CH17-5019, complete sequence	Homo sapiens	human	9606	28.2	112	63%	114	100.00	229172	AC277850.1
Homo sapiens chromosome 22 clone CH17-369L23, complete sequence	Homo sapiens	human	9606	28.2	112	63%	114	100.00	225459	AC277836.1
Homo sapiens chromosome 22 clone CH17-291K13, complete sequence	Homo sapiens	human	9606	28.2	197	63%	114	100.00	205795	AC277824.1
Homo sapiens chromosome 22 clone CH17-232B17, complete sequence	Homo sapiens	human	9606	28.2	225	63%	114	100.00	245081	AC277820.1
Homo sapiens chromosome 22 clone CH17-113F14, complete sequence	Homo sapiens	human	9606	28.2	56	63%	114	100.00	21931	AC277816.1
Homo sapiens chromosome 22 clone CH17-137E11, complete sequence	Homo sapiens	human	9606	28.2	112	63%	114	100.00	199817	AC277806.1
Homo sapiens chromosome 22 clone CH17-137E11, complete sequence	Homo sapiens	human	9606	28.2	20	63%	114	100.00	147617	AC277804.1

Figure 11.2, BLAST®<sup>151</sup> data for the highlighted sequence against *Homo sapiens* genome

Then, within that BLAST®<sup>151</sup> result, I then looked through the aligned sequence that had the highest maximum score, which was 30.2 in the *Homo sapiens* family with sequence similarity 230 member I (FAM230I), transcript variant 4, long non-coding RNA (Figure 11.3) (BLAST®<sup>151</sup> [DATA FILE-1](#)).

Alignments:

```
>Homo sapiens family with sequence similarity 230 member I (FAM230I), transcript variant 4, long non-coding RNA
Sequence ID: NR_165490.1 Length: 1598
Range 1: 667 to 681

Score:30.2 bits(15), Expect:29,
Identities:15/15(100%), Gaps:0/15(0%), Strand: Plus/Minus

Query 5 TGATGTCGTATACAG 19
|||||||:::|||:::|
Sbjct 681 TGATGTCGTATACAG 667
```

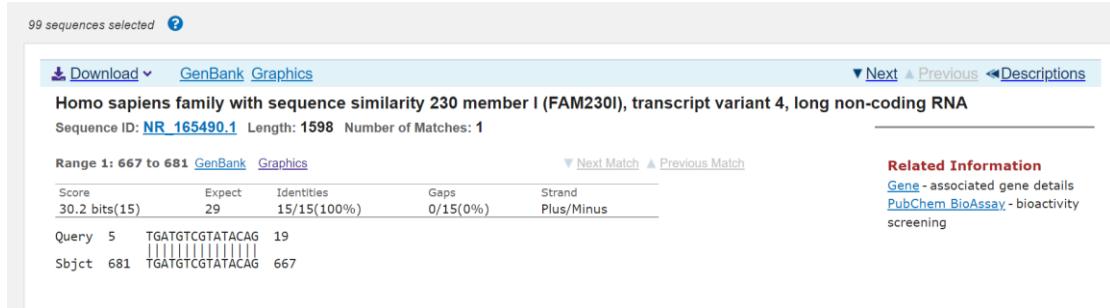


Figure 11.3, BLAST®<sup>151</sup> data for the highlighted sequence against *Homo sapiens* genome

A perfect match of 15 nucleotides, with 7 mismatches at the 5' end between the crRNA sequence and the *Homo sapiens* sequence. The annealing temperature of this mismatch pairing was significantly lower (more than 20 °C lower) than the perfect match. Therefore, the crRNA should have minimal off-target activity on the human transcriptome.

Thirdly, I used the sequence of **GUACUGAUGUCGUUAUACAGGGC** and performed a BLAST®<sup>151</sup> against SARS-CoV, MERS-CoV (Figure 11.4) ([DATA FILE-2](#)).

Sequences producing significant alignments:										
Description	Scientific Name	Common Name	Max Taxid	Total Score	Query cover	E value	Per cent	Accession	Len	Accesion
SARS coronavirus Tor2 isolate Tor2/FP1-10895, complete genome	SARS coronav... NA	SARS coronav... NA	227984	36.2	91.2	100%	0.002	95.45	29646	JX163928.1
SARS coronavirus Tor2 isolate Tor2/FP1-10851, complete genome	SARS coronav... NA	SARS coronav... NA	227984	36.2	91.2	100%	0.002	95.45	29646	JX163927.1
SARS coronavirus isolate Tor2/FP1-10912, complete genome	SARS coronav... NA	SARS coronav... NA	227984	36.2	91.2	100%	0.002	95.45	29646	JX163926.1
SARS coronavirus Tor2 isolate Tor2/FP1-10895, complete genome	SARS coronav... NA	SARS coronav... NA	227984	36.2	91.2	100%	0.002	95.45	29646	JX163925.1
SARS coronavirus Tor2 isolate Tor2/FP1-10851, complete genome	SARS coronav... NA	SARS coronav... NA	227984	36.2	91.2	100%	0.002	95.45	29646	JX163924.1
SARS coronavirus Tor2 isolate Tor2/FP1-10912, complete genome	SARS coronav... NA	SARS coronav... NA	227984	36.2	91.2	100%	0.002	95.45	29646	JX163923.1
SARS coronavirus Tor2, complete genome	SARS coronav... NA	SARS coronav... NA	227984	36.2	91.2	100%	0.002	95.45	29751	NC_004718.3
SARS coronavirus isolate CFB/SZ/94/03, complete genome	Severe acute... NA	Severe acute... NA	2901879	36.2	91.2	100%	0.002	95.45	29739	AY545919.1
SARS coronavirus isolate HC/GZ/32/03, complete genome	Severe acute... NA	Severe acute... NA	2901879	36.2	91.2	100%	0.002	95.45	29737	AY545918.1
SARS coronavirus isolate HC/GZ/81/03, complete genome	Severe acute... NA	Severe acute... NA	2901879	36.2	91.2	100%	0.002	95.45	29739	AY545917.1
SARS coronavirus isolate HC/SZ/266/03, complete genome	Severe acute... NA	Severe acute... NA	2901879	36.2	91.2	100%	0.002	95.45	29721	AY545916.1
SARS coronavirus isolate HC/SZ/DM1/03, complete genome	Severe acute... NA	Severe acute... NA	2901879	36.2	91.2	100%	0.002	95.45	29714	AY545915.1
SARS coronavirus isolate HC/SZ/79/03, complete genome	Severe acute... NA	Severe acute... NA	2901879	36.2	91.2	100%	0.002	95.45	29723	AY545914.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	24.3	97.6	95%	8.3	90.00	30483	MG987421.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	24.3	79.3	95%	8.3	90.00	30484	MG987420.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30123	MN723544.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30120	MN723543.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30120	MN723542.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	29994	OL622036.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	29994	OL622035.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30103	MW545528.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30103	MW545527.1
Middle East respiratory syndrome-related coronavirus strain...	Middle East ... NA	Middle East ... NA	1335626	22.3	40.6	54%	33	100.00	20189	MT818221.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	97.6	77%	33	100.00	30118	MT225607.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30118	MT225606.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30118	MT225605.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30118	MT225604.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30118	MT225603.1
Middle East respiratory syndrome-related coronavirus isolate...	Middle East ... NA	Middle East ... NA	1335626	22.3	79.3	59%	33	100.00	30110	MT225602.1

Figure 11.3, BLAST®<sup>151</sup> data for the highlighted sequence against SARS-CoV and MERS-CoV.

The sequence only had one mismatch with SARS-CoV, which was ideal (Figure 11.4). Showing that the sequence is highly conserved ([DATA FILE-2](#))

#### SARS coronavirus Tor2 isolate Tor2/FP1-10895, complete genome

Sequence ID: [JX163928.1](#) Length: 29646 Number of Matches: 4

Range 1: 13410 to 13431 GenBank Graphics					<a href="#">▼ Next Match</a>	<a href="#">▲ Previous Match</a>
Score	Expect	Identities	Gaps	Strand		
36.2 bits(18)	0.002	21/22(95%)	0/22(0%)	Plus/Plus		
Query 1	GTACTGATGTCGTATAACAGGGC	22				
Sbjct 13410	GTACTGATGTCGTCTACAGGGC	13431				

Figure 11.4, BLAST®<sup>151</sup> data for the highlighted sequence against SARS-CoV

Thus, I have selected this crRNA sequence of **GUACUGAUGUCGUUAUACAGGGC**.

## **Supplementary Material**

All of the supplementary materials should be available here: <https://github.com/Luqi-Wang/Waynflete-supplementary-materials-Luqi-Wang>

## **Abbreviations List**

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats  
Cas: CRISPR-associated proteins  
CMAVT: CRISPR/Cas Mediated Antiviral Therapy  
CMAVV: CRISPR/Cas Mediated Antiviral Vaccine  
SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus - 2  
MERS-CoV: Middle East Respiratory Syndrome – Related Coronavirus  
HIV-1: Human Immunodeficiency Virus – 1  
Tat: Trans-Activator of Transcriptome  
RdRp: RNA-dependent RNA-polymerase  
ORF: Open Reading Frame  
dsDNA: double-stranded DNA  
dsRNA: double-stranded RNA  
ssRNA: single-stranded RNA  
sgRNA: Dual RNA-guide  
tracrRNA: trans-activating crRNA  
PAM: Protospacer Adjacent Motif  
PFS: Protospacer Flanking Site  
DSB: Double-stranded Break  
BmNPV: *Bombyx mori* nucleopolyhedrovirus  
ACE2: Angiotensin-converting enzyme 2  
NPC-1: NPC Intracellular Cholesterol Transporter – 1  
DC-SIGN: Dendritic Cell-Specific Intracellular Adhesion Molecule-3-Grabbing Non-integrin  
TNFRSF14: Tumour Necrosis Factor Receptor Superfamily Member 14

## **References**

1. Centres for Disease Control and Prevention. Burden of Influenza. *Centres for Disease Control and Prevention* <https://www.cdc.gov/flu/about/burden/index.html> (2021).
2. Iuliano, A. D. *et al.* Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet* **391**, 1285–1300 (2018).
3. Suschak, J. J. & Schmaljohn, C. S. Vaccines against Ebola virus and Marburg virus: recent advances and promising candidates. *Hum Vaccin Immunother* **15**, 2359–2377 (2019).
4. Wikipedia. Antigenic variation. *Wikipedia* [https://en.wikipedia.org/wiki/Antigenic\\_variation](https://en.wikipedia.org/wiki/Antigenic_variation) (2022).
5. Shafer, R. W. & Vuitton, D. A. Highly active antiretroviral therapy (HAART) for the treatment of infection with human immunodeficiency virus type 1. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **53**, 73–86 (1999).
6. Barrangou, R. The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Curr Opin Immunol* **32**, 36–41 (2015).
7. Redman, M., King, A., Watson, C. & King, D. What is CRISPR/Cas9? *Arch Dis Child Educ Pract Ed* **101**, 213–5 (2016).
8. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–12 (2007).
9. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–5 (2008).
10. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579–86 (2012).
11. Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202 (2015).
12. Makarova, K. S. *et al.* Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18**, 67–83 (2020).
13. Makarova, K. S., Zhang, F. & Koonin, E. v. SnapShot: Class 2 CRISPR-Cas Systems. *Cell* **168**, 328–328.e1 (2017).
14. Pyzocha, N. K. & Chen, S. Diverse Class 2 CRISPR-Cas Effector Proteins for Genome Engineering Applications. *ACS Chem Biol* **13**, 347–356 (2018).
15. Koonin, E. v., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67–78 (2017).
16. Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–62 (2016).
17. Lin, L. *et al.* Engineering the Direct Repeat Sequence of crRNA for Optimization of FnCpf1-Mediated Genome Editing in Human Cells. *Molecular Therapy* **26**, 2650–2657 (2018).
18. Smargon, A. A. *et al.* Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell* **65**, 618–630.e7 (2017).
19. Khan, F. J., Yuen, G. & Luo, J. Multiplexed CRISPR/Cas9 gene knockout with simple crRNA:tracrRNA co-transfection. *Cell & Bioscience* **9**, 41 (2019).

20. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–71 (2015).
21. Huynh, N., Depner, N., Larson, R. & King-Jones, K. A versatile toolkit for CRISPR-Cas13-based RNA manipulation in Drosophila. *Genome Biology* **21**, 279 (2020).
22. Chylinski, K., le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol* **10**, 726–37 (2013).
23. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279–284 (2014).
24. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (1979)* **337**, 816–821 (2012).
25. Tu, M. *et al.* A “new lease of life”: FnCpf1 possesses DNA cleavage activity for genome editing in human cells. *Nucleic Acids Res* **45**, 11295–11304 (2017).
26. Xu, C. *et al.* Programmable RNA editing with compact CRISPR-Cas13 systems from uncultivated microbes. *Nature Methods* **18**, 499–506 (2021).
27. Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665–676.e14 (2018).
28. Shah, S. A., Erdmann, S., Mojica, F. J. M. & Garrett, R. A. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol* **10**, 891–9 (2013).
29. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–7 (2014).
30. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* **155**, 733–740 (2009).
31. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–21 (2012).
32. Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).
33. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
34. Wright, A. v., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature’s Toolbox for Genome Engineering. *Cell* **164**, 29–44 (2016).
35. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722–36 (2015).
36. Baltimore, D. Expression of animal virus genomes. *Bacteriol Rev* **35**, 235–41 (1971).
37. Smith, G. L. & McFadden, G. Smallpox: anything to declare? *Nature Reviews Immunology* **2**, 521–527 (2002).
38. Jacob, S. T. *et al.* Ebola virus disease. *Nature Reviews Disease Primers* **6**, 13 (2020).
39. Fontela, C. *et al.* Trends and causes of mortality in a population-based cohort of HIV-infected adults in Spain: comparison with the general population. *Scientific Reports* **10**, 8922 (2020).
40. McLaren, P. J. & Fellay, J. HIV-1 and human genetic variation. *Nature Reviews Genetics* **22**, 645–657 (2021).
41. di Paola, N., Sanchez-Lockhart, M., Zeng, X., Kuhn, J. H. & Palacios, G. Viral genomics in Ebola virus research. *Nature Reviews Microbiology* **18**, 365–378 (2020).
42. Wei, H. *et al.* Deep-sequencing of Marburg virus genome during sequential mouse

- passaging and cell-culture adaptation reveals extensive changes over time. *Scientific Reports* **7**, 3390 (2017).
43. Sohrabi, F. *et al.* Role of plasmonics in detection of deadliest viruses: a review. *Eur Phys J Plus* **136**, 675 (2021).
  44. Lee, C. CRISPR/Cas9-Based Antiviral Strategy: Current Status and the Potential Challenge. *Molecules* **24**, (2019).
  45. Fareh, M. *et al.* Reprogrammed CRISPR-Cas13b suppresses SARS-CoV-2 replication and circumvents its mutational escape through mismatch tolerance. *Nature Communications* **12**, 4270 (2021).
  46. Koujah, L., Shukla, D. & Naqvi, A. R. CRISPR-Cas based targeting of host and viral genes as an antiviral strategy. *Semin Cell Dev Biol* **96**, 53–64 (2019).
  47. National Institutes of Health. International Committee on Taxonomy of Viruses (2002). *National Institutes of Health* <https://talk.ictvonline.org/taxonomy/> (2002).
  48. National Institutes of Health. International Committee on Taxonomy of Viruses (2002). *National Institutes of Health* <https://talk.ictvonline.org/taxonomy/> (2002).
  49. Tao, Y. J. & Ye, Q. RNA virus replication complexes. *PLoS Pathog* **6**, e1000943 (2010).
  50. Yang, J. *et al.* Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor. *Nat Commun* **11**, 4541 (2020).
  51. Alvarez, C. P. *et al.* C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. *J Virol* **76**, 6841–4 (2002).
  52. Wang, H. *et al.* Ebola Viral Glycoprotein Bound to Its Endosomal Receptor Niemann-Pick C1. *Cell* **164**, 258–268 (2016).
  53. Allen, S. J. *et al.* Interactions between herpesvirus entry mediator (TNFRSF14) and latency-associated transcript during herpes simplex virus 1 latency. *J Virol* **88**, 1961–71 (2014).
  54. Murphy, E., Vaníček, J., Robins, H., Shenk, T. & Levine, A. J. Suppression of immediate-early viral gene expression by herpesvirus-coded microRNAs: Implications for latency. *Proceedings of the National Academy of Sciences* **105**, 5453–5458 (2008).
  55. Gruffat, H., Marchione, R. & Manet, E. Herpesvirus Late Gene Expression: A Viral-Specific Pre-initiation Complex Is Key. *Front Microbiol* **7**, 869 (2016).
  56. Hütter, G. *et al.* Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med* **360**, 692–8 (2009).
  57. Claireaux, M. *et al.* Low CCR5 expression protects HIV-specific CD4+ T cells of elite controllers from viral entry. *Nature Communications* **13**, 521 (2022).
  58. McDougal, J. S. *et al.* Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **231**, 382–5 (1986).
  59. Maddon, P. J. *et al.* The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**, 333–48 (1986).
  60. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–59 (1998).
  61. Liu, Z. *et al.* Genome editing of the HIV co-receptors CCR5 and CXCR4 by CRISPR-Cas9 protects CD4+ T cells from HIV-1 infection. *Cell Biosci* **7**, 47 (2017).
  62. Lopalco, L. CCR5: From Natural Resistance to a New Anti-HIV Strategy. *Viruses* **2**, 574–600 (2010).

63. Hou, P. *et al.* Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection. *Sci Rep* **5**, 15577 (2015).
64. Bogerd, H. P., Kornepati, A. V. R., Marshall, J. B., Kennedy, E. M. & Cullen, B. R. Specific induction of endogenous viral restriction factors using CRISPR/Cas-derived transcriptional activators. *Proc Natl Acad Sci U S A* **112**, E7249–56 (2015).
65. National Human Genome Research Institute. What are the Ethical Concerns of Genome Editing? National Human Genome Research Institute <https://www.genome.gov/about-genomics/policy-issues/Genome-Editing/ethical-concerns> (2017).
66. Mary Todd Bergman, H. C. Perspectives on gene editing. *The Harvard Gazette* <https://news.harvard.edu/gazette/story/2019/01/perspectives-on-gene-editing/> (2019).
67. Almeida, M. & Ranisch, R. Beyond safety: mapping the ethical debate on heritable genome editing interventions. *Humanities and Social Sciences Communications* **9**, 139 (2022).
68. Rothschild, J. Ethical considerations of gene editing and genetic selection. *J Gen Fam Med* **21**, 37–47 (2020).
69. Hussein, M., Andrade Dos Ramos, Z., Berkhout, B. & Herrera-Carrillo, E. In Silico Prediction and Selection of Target Sequences in the SARS-CoV-2 RNA Genome for an Antiviral Attack. *Viruses* **14**, (2022).
70. Imperatore, J. A. *et al.* Highly conserved s2m element of SARS-CoV-2 dimerizes via a kissing complex and interacts with host miRNA-1307-3p. *Nucleic Acids Research* **50**, 1017–1032 (2022).
71. Jonassen, C. M., Jonassen, T. O. & Grinde, B. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J Gen Virol* **79** ( Pt 4), 715–8 (1998).
72. Robertson, M. P. *et al.* The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol* **3**, e5 (2005).
73. Gilbert, C. & Tengs, T. No species-level losses of s2m suggests critical role in replication of SARS-related coronaviruses. *Scientific Reports* **11**, 16145 (2021).
74. Lulla, V. *et al.* Targeting the Conserved Stem Loop 2 Motif in the SARS-CoV-2 Genome. *Journal of Virology* **95**, (2021).
75. Robertson, M. P. *et al.* The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol* **3**, e5 (2005).
76. Manfredonia, I. *et al.* Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res* **48**, 12436–12452 (2020).
77. Martin, R. *et al.* Genetic conservation of SARS-CoV-2 RNA replication complex in globally circulating isolates and recently emerged variants from humans and minks suggests minimal pre-existing resistance to remdesivir. *Antiviral Res* **188**, 105033 (2021).
78. Nesterenko, P. A. *et al.* HLA-A\*02:01 restricted T cell receptors against the highly conserved SARS-CoV-2 polymerase cross-react with human coronaviruses. *Cell Rep* **37**, 110167 (2021).
79. Aftab, S. O. *et al.* Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a

- potential therapeutic drug target using a computational approach. *Journal of Translational Medicine* **18**, 275 (2020).
80. Martin, R. *et al.* Genetic conservation of SARS-CoV-2 RNA replication complex in globally circulating isolates and recently emerged variants from humans and minks suggests minimal pre-existing resistance to remdesivir. *Antiviral Res* **188**, 105033 (2021).
  81. Tilocca, B. *et al.* Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect* **22**, 188–194.
  82. Abbott, T. R. *et al.* Development of CRISPR as an Antiviral Strategy to Combat SARS-CoV-2 and Influenza. *Cell* **181**, 865–876.e12 (2020).
  83. Liu, X. *et al.* Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Sci Rep* **6**, 19675 (2016).
  84. Abdelrahman, Z., Li, M. & Wang, X. Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses. *Frontiers in Immunology* **11**, (2020).
  85. Wu, X., Kriz, A. J. & Sharp, P. A. Target specificity of the CRISPR-Cas9 system. *Quant Biol* **2**, 59–70 (2014).
  86. Wang, X. *et al.* The EF-1 $\alpha$  promoter maintains high-level transgene expression from episomal vectors in transfected CHO-K1 cells. *J Cell Mol Med* **21**, 3044–3054 (2017).
  87. Mäkinen, P. I. *et al.* Stable RNA interference: comparison of U6 and H1 promoters in endothelial cells and in mouse brain. *J Gene Med* **8**, 433–41 (2006).
  88. transOMIC. pCLIP-Cas9-Nuclease-EFS-ZsGreen. *SnapGene* 5.2.1 <https://www.snapgene.com> (2020).
  89. Hopp, T. P. *et al.* A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology* **6**, 1204–1210 (1988).
  90. Valbuena, F. M. *et al.* A photostable monomeric superfolder green fluorescent protein. *Traffic* **21**, 534–544 (2020).
  91. SnapGene. P2A gene sequence. *SnapGene* 5.2.1 [www.snapgene.com](http://www.snapgene.com) (2022).
  92. Liu, Q., Liu, M. & Wu, W. Strong/Weak Feature Recognition of Promoters Based on Position Weight Matrix and Ensemble Set-Valued Models. *Journal of Computational Biology* **25**, 1152–1160 (2018).
  93. Tripathy, M. K., Abbas, W. & Herbein, G. Epigenetic regulation of HIV-1 transcription. *Epigenomics* **3**, 487–502 (2011).
  94. Jeeninga, R. E. *et al.* Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *J Virol* **74**, 3740–51 (2000).
  95. Schiralli Lester, G. M. & Henderson, A. J. Mechanisms of HIV Transcriptional Regulation and Their Contribution to Latency. *Mol Biol Int* **2012**, 614120 (2012).
  96. Talmadge, J. E. & Cowan, K. H. Gene Therapy in Oncology. in *Abeloff's Clinical Oncology* 493–507.e4 (Elsevier, 2014). doi:10.1016/B978-1-4557-2865-7.00031-X.
  97. Debaisieux, S., Rayne, F., Yezid, H. & Beaumelle, B. The ins and outs of HIV-1 Tat. *Traffic* **13**, 355–63 (2012).
  98. Turk, G. *et al.* Higher transactivation activity associated with LTR and Tat elements from HIV-1 BF intersubtype recombinant variants. *Retrovirology* **3**, 14 (2006).
  99. An, D. S. *et al.* An inducible human immunodeficiency virus type 1 (HIV-1) vector which

- effectively suppresses HIV-1 replication. *J Virol* **73**, 7671–7 (1999).
- 100. SnapGene. Nucleoplasmin NLS sequence. *SnapGene 5.2.1* www.snapgene.com (2022).
  - 101. Wahane, A. *et al.* Role of Lipid-Based and Polymer-Based Non-Viral Vectors in Nucleic Acid Delivery for Next-Generation Gene Therapy. *Molecules* **25**, (2020).
  - 102. Potter, H. Transfection by electroporation. *Curr Protoc Mol Biol Chapter 9*, Unit 9.3 (2003).
  - 103. Sun, J. Y., Anand-Jawa, V., Chatterjee, S. & Wong, K. K. Immune responses to adeno-associated virus and its recombinant vectors. *Gene Therapy* **10**, 964–976 (2003).
  - 104. Ronzitti, G., Gross, D.-A. & Mingozzi, F. Human Immune Responses to Adeno-Associated Virus (AAV) Vectors. *Front Immunol* **11**, 670 (2020).
  - 105. Verdera, H. C., Kuranda, K. & Mingozzi, F. AAV Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer. *Mol Ther* **28**, 723–746 (2020).
  - 106. Hou, Y. J. *et al.* SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* **182**, 429–446.e14 (2020).
  - 107. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: cell binding and entry. *Cold Spring Harb Perspect Med* **2**, (2012).
  - 108. Ghosh, S., Brown, A. M., Jenkins, C. & Campbell, K. Viral Vector Systems for Gene Therapy: A Comprehensive Literature Review of Progress and Biosafety Challenges. *Applied Biosafety* **25**, 7–18 (2020).
  - 109. Wei, T., Cheng, Q., Min, Y.-L., Olson, E. N. & Siegwart, D. J. Systemic nanoparticle delivery of CRISPR-Cas9 ribonucleoproteins for effective tissue specific genome editing. *Nature Communications* **11**, 3232 (2020).
  - 110. Lee, K. *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nature Biomedical Engineering* **1**, 889–901 (2017).
  - 111. Bhaskar, S. & Lim, S. Engineering protein nanocages as carriers for biomedical applications. *NPG Asia Materials* **9**, e371–e371 (2017).
  - 112. Molino, N. M. & Wang, S.-W. Caged protein nanoparticles for drug delivery. *Current Opinion in Biotechnology* **28**, 75–82 (2014).
  - 113. Uchida, M., Kang, S., Reichhardt, C., Harlen, K. & Douglas, T. The ferritin superfamily: Supramolecular templates for materials synthesis. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1800**, 834–845 (2010).
  - 114. McAndrews, K. M. *et al.* Exosome-mediated delivery of CRISPR/Cas9 for targeting of oncogenic KrasG12D in pancreatic cancer. *Life Sci Alliance* **4**, (2021).
  - 115. Kaushik, A. *et al.* Magnetically guided non-invasive CRISPR-Cas9/gRNA delivery across blood-brain barrier to eradicate latent HIV-1 infection. *Scientific Reports* **9**, 3928 (2019).
  - 116. Baddeley, H. J. E. & Isalan, M. The Application of CRISPR/Cas Systems for Antiviral Therapy. *Frontiers in Genome Editing* **3**, (2021).
  - 117. Liu, Y. *et al.* Construction of Baculovirus-Inducible CRISPR/Cas9 Antiviral System Targeting BmNPV in Bombyx mori. *Viruses* **14**, (2021).
  - 118. Dong, Z. *et al.* Establishment of a baculovirus-inducible CRISPR/Cas9 system for antiviral research in transgenic silkworms. *Applied Microbiology and Biotechnology* **102**, 9255–9265 (2018).

119. Cao, M.-Y. *et al.* Screening and optimization of an efficient *Bombyx mori* nucleopolyhedrovirus inducible promoter. *J Biotechnol* **231**, 72–80 (2016).
120. Cao, Y., Zhou, H., Zhou, X. & Li, F. Control of Plant Viruses by CRISPR/Cas System-Mediated Adaptive Immunity. *Front Microbiol* **11**, 593700 (2020).
121. Khatodia, S., Bhatotia, K. & Tuteja, N. Development of CRISPR/Cas9 mediated virus resistance in agriculturally important crops. *Bioengineered* **8**, 274–279 (2017).
122. Khan, Z. A., Kumar, R. & Dasgupta, I. CRISPR/Cas-Mediated Resistance against Viruses in Plants. *International Journal of Molecular Sciences* **23**, 2303 (2022).
123. Ji, X., Zhang, H., Zhang, Y., Wang, Y. & Gao, C. Establishing a CRISPR-Cas-like immune system conferring DNA virus resistance in plants. *Nat Plants* **1**, 15144 (2015).
124. Ji, X., Zhang, H., Zhang, Y., Wang, Y. & Gao, C. Establishing a CRISPR-Cas-like immune system conferring DNA virus resistance in plants. *Nat Plants* **1**, 15144 (2015).
125. Baltes, N. J. *et al.* Conferring resistance to geminiviruses with the CRISPR–Cas prokaryotic immune system. *Nature Plants* **1**, 15145 (2015).
126. Ali, Z. *et al.* CRISPR/Cas9-mediated viral interference in plants. *Genome Biol* **16**, 238 (2015).
127. Ali, Z., Ali, S., Tashkandi, M., Zaidi, S. S.-E.-A. & Mahfouz, M. M. CRISPR/Cas9-Mediated Immunity to Geminiviruses: Differential Interference and Evasion. *Sci Rep* **6**, 26912 (2016).
128. Pyott, D. E., Sheehan, E. & Molnar, A. Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-free *Arabidopsis* plants. *Mol Plant Pathol* **17**, 1276–88 (2016).
129. Chandrasekaran, J. *et al.* Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol Plant Pathol* **17**, 1140–53 (2016).
130. Zaidi, S. S.-E.-A., Tashkandi, M., Mansoor, S. & Mahfouz, M. M. Engineering Plant Immunity: Using CRISPR/Cas9 to Generate Virus Resistance. *Front Plant Sci* **7**, 1673 (2016).
131. Wilder-Smith, A. *et al.* Epidemic arboviral diseases: priorities for research and public health. *Lancet Infect Dis* **17**, e101–e106 (2017).
132. Powell, J. R. & Tabachnick, W. J. History of domestication and spread of *Aedes aegypti*-a review. *Memorias do Instituto Oswaldo Cruz* **108 Suppl 1**, 11–7 (2013).
133. Schliessmann, D. J. Initiation of the *Aedes aegypti* eradication programme of the USA. *Bull World Health Organ* **36**, 604–9 (1967).
134. Abedi, Z. H. & Brown, A. W. A. Peritrophic Membrane as Vehicle for DDT and DDE Excretion in *Aedes aegypti* Larvae. *Ann Entomol Soc Am* **54**, 539–542 (1961).
135. Porter, R. D. & Wiemeyer, S. N. Dieldrin and DDT: effects on sparrow hawk eggshells and reproduction. *Science* **165**, 199–200 (1969).
136. J. Routt Reigart, M. D. and J. R. R. M. D. , M. P. H. *Recognition and Management of Pesticide Poisonings*. (United states Environmental Protection Agency, 2013).
137. Beyer, W. N. & Gish, C. D. Persistence in Earthworms and Potential Hazards to Birds of Soil Applied DDT, Dieldrin and Heptachlor. *The Journal of Applied Ecology* **17**, 295 (1980).
138. Collins, C. M., Bonds, J. A. S., Quinlan, M. M. & Mumford, J. D. Effects of the removal or reduction in density of the malaria mosquito, *Anopheles gambiae* s.l., on interacting predators and competitors in local ecosystems. *Med Vet Entomol* **33**, 1–15 (2019).

139. Fang, J. Ecology: A world without mosquitoes. *Nature* **466**, 432–434 (2010).
140. Anderson, M. A. E., Gross, T. L., Myles, K. M. & Adelman, Z. N. Validation of novel promoter sequences derived from two endogenous ubiquitin genes in transgenic *Aedes aegypti*. *Insect Mol Biol* **19**, 441–9 (2010).
141. Akbari, O. S. *et al.* The developmental transcriptome of the mosquito *Aedes aegypti*, an invasive species and major arbovirus vector. *G3 (Bethesda)* **3**, 1493–509 (2013).
142. Matthews, B. J. *et al.* Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature* **563**, 501–507 (2018).
143. Zhao, L., Becnel, J. J., Clark, G. G. & Linthicum, K. J. Expression of AeaHsp26 and AeaHsp83 in *Aedes aegypti* (Diptera: Culicidae) larvae and pupae in response to heat shock stress. *J Med Entomol* **47**, 367–75 (2010).
144. Rozen-Gagnon, K., Yi, S., Jacobson, E., Novack, S. & Rice, C. M. A selectable, plasmid-based system to generate CRISPR/Cas9 gene edited and knock-in mosquito cell lines. *Sci Rep* **11**, 736 (2021).
145. Anderson, M. A. E. *et al.* Expanding the CRISPR Toolbox in Culicine Mosquitoes: In Vitro Validation of Pol III Promoters. *ACS Synth Biol* **9**, 678–681 (2020).
146. Tng, P. Y. L. *et al.* Cas13b-dependent and Cas13b-independent RNA knockdown of viral sequences in mosquito cells following guide RNA expression. *Communications Biology* **3**, 413 (2020).
147. Blanchard, E. L. *et al.* Treatment of influenza and SARS-CoV-2 infections via mRNA-encoded Cas13a in rodents. *Nature Biotechnology* **39**, 717–726 (2021).
148. Higashimoto, T. *et al.* The woodchuck hepatitis virus post-transcriptional regulatory element reduces readthrough transcription from retroviral vectors. *Gene Therapy* **14**, 1298–1304 (2007).
149. SnapGene. SARS-CoV-2 Genome Sequence. *SnapGene* [https://www.snapgene.com/resources/coronavirus-resources/?resource=SARS-CoV-2\\_\(COVID-19\)\\_Genome](https://www.snapgene.com/resources/coronavirus-resources/?resource=SARS-CoV-2_(COVID-19)_Genome) (2022).
150. Jeltsch M. AnnealingTemperature Equation. *jeltsch.org* [https://jeltsch.org/annealing\\_temperature](https://jeltsch.org/annealing_temperature) (2006).
151. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).

