3. Design a single guide RNA (sgRNA)

Overview

In bacteria Cas9 is guided to a specific region of DNA by a crRNA molecule, which is complementary to the target DNA sequence. However, the crRNA cannot bind directly to Cas9. Instead, a region of the crRNA forms a duplex with another RNA molecule called trans-activating crispr RNA (tracrRNA) and it is the tracrRNA molecule that binds Cas9 (see figure 2a). Molecular biologists have adapted this system into a single RNA molecule that performs both functions (DNA-targeting and Cas9-binding). This synthetic RNA is called a single guide RNA (sgRNA) (see figure 2b).

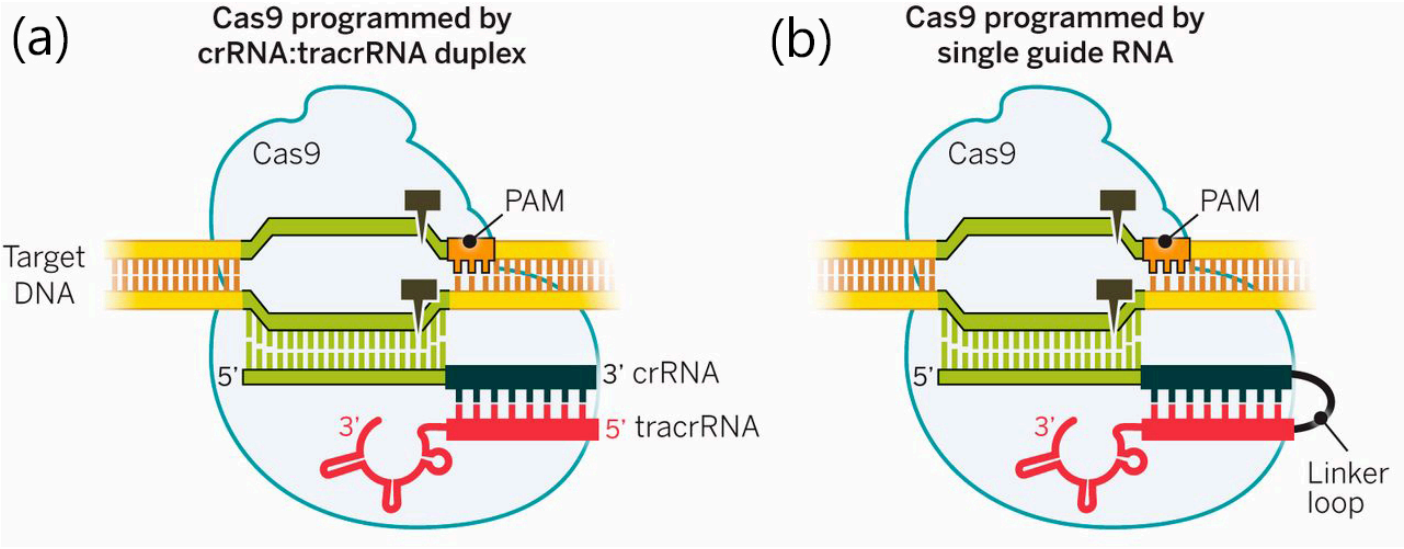


Figure 2: Cas9 targeting by RNA molecules. (a) In bacteria, Cas9 is targeted to a specific DNA sequence by a duplex of crRNA (targets the DNA) and tracrRNA (binds Cas9). **(b)** In molecular biology applications, a single guide RNA (sgRNA) incorporates both of these functions in one molecule. (Figure taken from Doudna & Charpentier, 2014).

When searching for target DNA sequences, Cas9 first binds a specific DNA sequence called a Protospacer Adjacent Motif (PAM). Once bound to a PAM sequence, Cas9 checks the adjacent DNA sequence for complementarity to the sgRNA (or crRNA). If the two are complementary, Cas9 will generate a DSB in the DNA. If they are not complementary, Cas9 will move on to another PAM sequence. PAM sequences are specific to the species and endonuclease (for *S. pyogenes* Cas9, the PAM is **5'−NGG−3'**). This means that, when you are designing a synthetic sgRNA, you need to choose a DNA sequence that is adjacent to a PAM sequence (NGG).

Although the functional sgRNA is an RNA molecule, RNA is quite difficult to work with for a molecular biologist. It is relatively unstable (compared to DNA) and we don't have as many molecular tools available to allow us to synthesise it in large amounts. When we want to express an RNA molecule in a cell, a common strategy is to clone the DNA coding sequence for the RNA into a plasmid that will express the RNA (this is what we did in our experiment).

Today you will design your own sgRNA. Ideally, you would do this exercise in a computer lab (doing database searches on a tablet might not be the easiest activity). You have a keyboard in your drawer that should make using the tablets easier (make sure that the number on your keyboard matches the number on the tablet! They are all individually linked!) And you are welcome to use your own laptop, if you have it with you.

Finding a DNA sequence

You can search for DNA sequences in several online databases, including NCBI, Ensembl, or SGD (the database you used in session 1). As we want to target a yeast gene, SGD is probably the best database as it is a specific *S. cerevisiae* database. For today's task, I have downloaded the sequence for you:

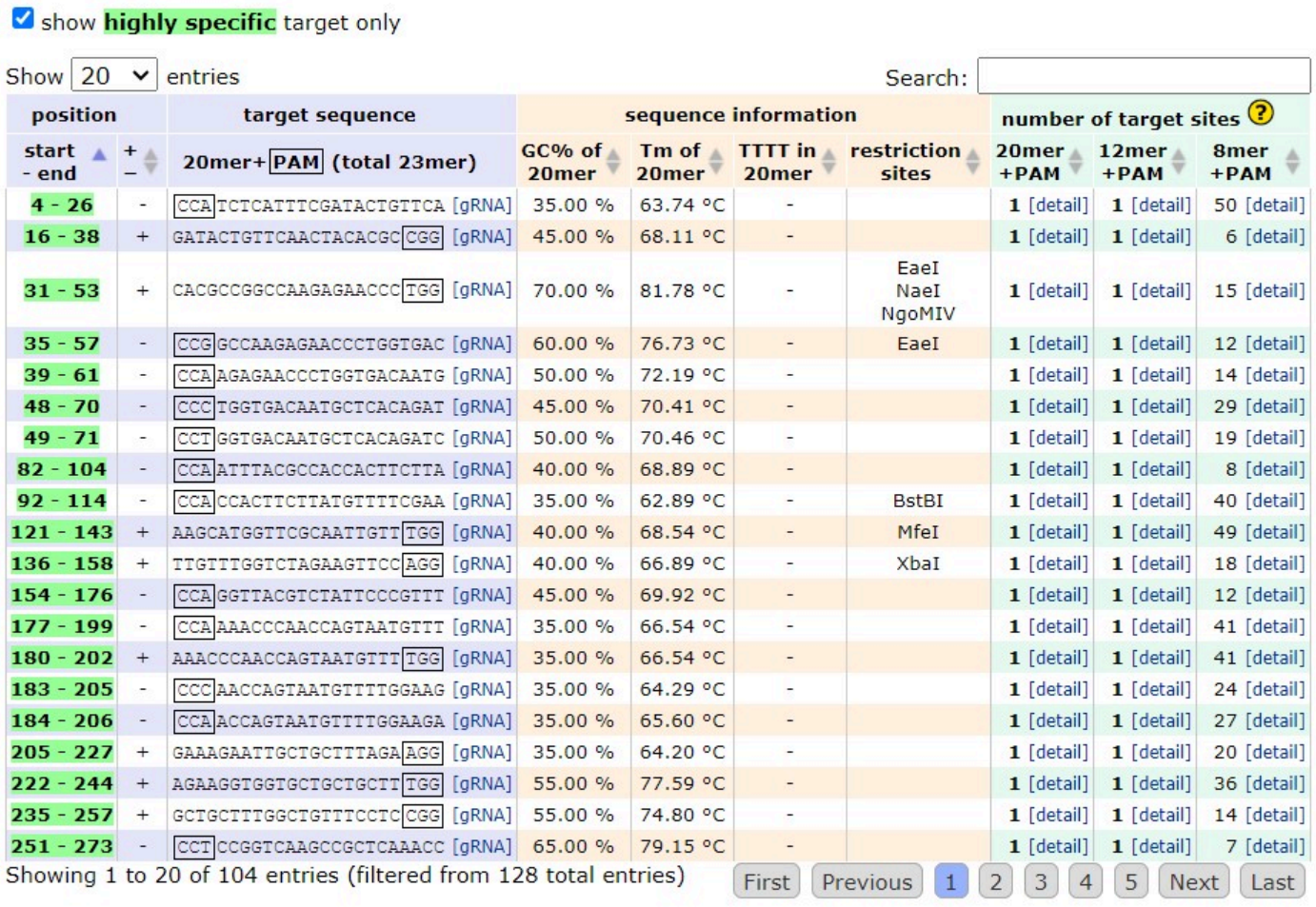
Sequence of the *MET15* gene

ATGCCATCTCATTTCGATACTGTTCAACTACACGCCGGCCAAGAGAACCCTGGTGACAATGCTCACAGATCCAGAGCTGTACCAATTTACGCCACCACTTCTT
ATGTTTTCGAAAACTCTAAGCATGGTTCGCAATTGTTTGGTCTAGAAGTTCCAGGTTACGTCTATTCCCGTTTCCAAAACCCAACCAGTAATGTTTTGGAAGAA
AGAATTGCTGCTTTAGAAGGTGGTGCTGCTGCTTTGGCTGTTTCCTCCGGTCAAGCCGCTCAAACCCTTGCCATCCAAGGTTTGGCACACACTGGTGACAAC
ATCGTTTCCACTTCTTACTTATACGGTGGTACTTATAACCAGTTCAAAATCTCGTTCAAAAGATTTGGTATCGAGGCTAGATTTGTTGAAGGTGACAATCCAGA
AGAATTCGAAAAGGTCTTTGATGAAAGAACCAAGGCTGTTTATTTGGAAACCATTGGTAATCCAAAGTACAATGTTCCGGATTTTGAAAAAATTGTTGCAATT
GCTCACAAACACGGTATTCCAGTTGTCGTTGACAACACATTTGGTGCCGGTGGTTACTTCTGTCAGCCAATTAAATACGGTGCTGATATTGTAACACATTCTG
CTACCAAATGGATTGGTGGTCATGGTACTACTATCGGTGGTATTATTGTTGACTCTGGTAAGTTCCCATGGAAGGACTACCCAGAAAAGTTCCCTCAATTCTC
TCAACCTGCCGAAGGATATCACGGTACTATCTACAATGAAGCCTACGGTAACTTGGCATACATCGTTCATGTTAGAACTGAACTATTAAGAGATTTGGGTCCA
TTGATGAACCCATTTGCCTCTTTCTTGCTACTACAAGGTGTTGAAACATTATCTTTGAGAGCTGAAAGACACGGTGAAAATGCATTGAAGTTAGCCAAATGGT
TAGAACAATCCCCATACGTATCTTGGGTTTCATACCCTGGTTTAGCATCTCATTCTCATCATGAAAATGCTAAGAAGTATCTATCTAACGGTTTCGGTGGTGTC
TTATCTTTCGGTGTAAAAGACTTACCAAATGCCGACAAGGAAACTGACCCATTCAAACTTTCTGGTGCTCAAGTTGTTGACAATTTAAAGCTTGCCTCTAACTT
GGCCAATGTTGGTGATGCCAAGACCTTAGTCATTGCTCCATACTTCACTACCCACAAACAATTAAATGACAAAGAAAAGTTGGCATCTGGTGTTACCAAGGA
CTTAATTCGTGTCTCTGTTGGTATCGAATTTATTGATGACATTATTGCAGACTTCCAGCAATCTTTTGAAACTGTTTTCGCTGGCCAAAAACCATGA

Identify potential guide RNAs

We are going to use an online software program to help us design a gRNA today. The software we are using is quite basic (more flashy ones exist) but it does make it easier to understand the design process (in my opinion). You will need to know the basic steps of how to design a gRNA.

1. Go to the CRISPRdirect website (http://crispr.dbcls.jp). Ideally, open this in a new window so that you can read these instructions too (annoyingly, Lt will not let me specify this so you will need to do it manually).
2. Delete the sample seqeunce from the box and paste in the *MET15* DNA sequence (from the box above). There are a few parameters you need to check/change:
    1. Make sure it says NGG in the PAM box (this is the appropriate PAM for *S. pyogenes* Cas9, which we are using in our experiments).
    2. Change specificity check from Human to Budding Yeast.
3. Now click design. You should get a list of results (you might need to scroll down to see them). Check the box that says highly specific target only.
4. You should now have a list of potential gRNA sequences that looks like the screenshot below (figure 3).



☑ show highly specific target only

Show 20 entries Search:

position		target sequence	sequence information				number of target sites		
start - end	+ –	20mer+PAM (total 23mer)	GC% of 20mer	Tm of 20mer	TTTT in 20mer	restriction sites	20mer +PAM	12mer +PAM	8mer +PAM
4 - 26	-	CCA TCTCATTTCGATACTGTTCA [gRNA]	35.00 %	63.74 °C	-		1 [detail]	1 [detail]	50 [detail]
16 - 38	+	GATACTGTTCAACTACACGC CGG [gRNA]	45.00 %	68.11 °C	-		1 [detail]	1 [detail]	6 [detail]
31 - 53	+	CACGCCGGCCAAGAGAACCC TGG [gRNA]	70.00 %	81.78 °C	-	EaeI NaeI NgoMIV	1 [detail]	1 [detail]	15 [detail]
35 - 57	-	CCG GCCAAGAGAACCCTGGTGAC [gRNA]	60.00 %	76.73 °C	-	EaeI	1 [detail]	1 [detail]	12 [detail]
39 - 61	-	CCA AGAGAACCCTGGTGACAATG [gRNA]	50.00 %	72.19 °C	-		1 [detail]	1 [detail]	14 [detail]
48 - 70	-	CCC TGGTGACAATGCTCACAGAT [gRNA]	45.00 %	70.41 °C	-		1 [detail]	1 [detail]	29 [detail]
49 - 71	-	CCT GGTGACAATGCTCACAGATC [gRNA]	50.00 %	70.46 °C	-		1 [detail]	1 [detail]	19 [detail]
82 - 104	-	CCA ATTTACGCCACCACTTCTTA [gRNA]	40.00 %	68.89 °C	-		1 [detail]	1 [detail]	8 [detail]
92 - 114	-	CCA CCACTTCTTATGTTTTCGAA [gRNA]	35.00 %	62.89 °C	-	BstBI	1 [detail]	1 [detail]	40 [detail]
121 - 143	+	AAGCATGGTTCGCAATTGTT TGG [gRNA]	40.00 %	68.54 °C	-	MfeI	1 [detail]	1 [detail]	49 [detail]
136 - 158	+	TTGTTTGGTCTAGAAGTTCC AGG [gRNA]	40.00 %	66.89 °C	-	XbaI	1 [detail]	1 [detail]	18 [detail]
154 - 176	-	CCA GGTTACGTCTATTCCCGTTT [gRNA]	45.00 %	69.92 °C	-		1 [detail]	1 [detail]	12 [detail]
177 - 199	-	CCA AAACCCAACCAGTAATGTTT [gRNA]	35.00 %	66.54 °C	-		1 [detail]	1 [detail]	41 [detail]
180 - 202	+	AAACCCAACCAGTAATGTTT TGG [gRNA]	35.00 %	66.54 °C	-		1 [detail]	1 [detail]	41 [detail]
183 - 205	-	CCC AACCAGTAATGTTTTGGAAG [gRNA]	35.00 %	64.29 °C	-		1 [detail]	1 [detail]	24 [detail]
184 - 206	-	CCA ACCAGTAATGTTTTGGAAGA [gRNA]	35.00 %	65.60 °C	-		1 [detail]	1 [detail]	27 [detail]
205 - 227	+	GAAAGAATTGCTGCTTTAGA AGG [gRNA]	35.00 %	64.20 °C	-		1 [detail]	1 [detail]	20 [detail]
222 - 244	+	AGAAGGTGGTGCTGCTGCTT TGG [gRNA]	55.00 %	77.59 °C	-		1 [detail]	1 [detail]	36 [detail]
235 - 257	+	GCTGCTTTGGCTGTTTCCTC CGG [gRNA]	55.00 %	74.80 °C	-		1 [detail]	1 [detail]	14 [detail]
251 - 273	-	CCT CCGGTCAAGCCGCTCAAACC [gRNA]	65.00 %	79.15 °C	-		1 [detail]	1 [detail]	7 [detail]

Showing 1 to 20 of 104 entries (filtered from 128 total entries) First Previous 1 2 3 4 5 Next Last

Figure 3: Potential gRNA target sequences in *MET15*.

Choosing an appropriate gRNA

There are over 100 possible gRNAs in this list! How are we going to choose the best one? First, let's make sure we understand the information in the table above. Each column has a title, which tells us what properties of the gRNA can be found in that column and the columns are grouped into four sections (see figure 4).

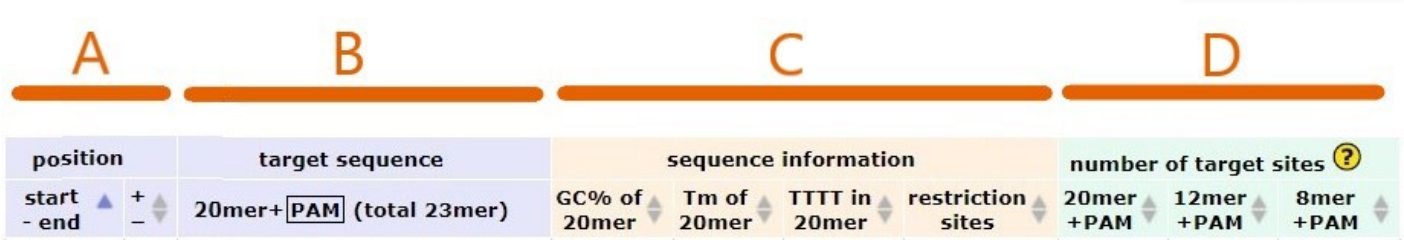


Figure 4: Parameters for selecting gRNAs.

Look at the separate sections (labelled A–D in figure 4) and answer the questions below (clicking on the little yellow question marks on the CRISPRdirect website might provide you with some useful information!)

(Section A): What do these columns tell us about each gRNA?

(Section B): The PAM sequence (NGG in *S. pyogenes* Cas9) is indicated by the boxed nucleotides. The remaining 20 nucleotides are the gRNA, which will target *MET15*. Why is the PAM at different ends of some of these sequences?

(Section C): What is this section telling us about each gRNA?

(Section D): What is this section telling us about each gRNA?

Now we know what the numbers in the columns mean, we can use them to help us choose a good gRNA. By default, the list is initially sorted by the start position of the gRNA in the gene (first column). You can use the arrows at the top of each column to sort the gRNAs using more useful parameters.

Here are some guidelines to help you sort and choose a gRNA (some of this will be done by the programme already, but it is still important information for you to know):

- Stick to a maximum length of 20 nucleotides (a **20mer**). Longer gRNA sequences have more off-target effects.
- The PAM sequence is **not** part of the 20mer gRNA. The **PAM sequence is in the genome**.
- Aim for a GC content of between **50–75%**.
- Avoid homopolymers (for examples, AAAA, GGGG, etc.)
- It doesn't matter whether you target the sense or antisense strand. For this exercise, I recommend that you pick two sgRNA sequences (one that targets each strand) to make sure you are comfortable with reverse complementing DNA and understanding which strand a gRNA will anneal to.
- For our experiment, it does not matter which part of the *MET15* gene you target. If you wanted to knock out a gene, the 5' end of a gene is usually the best place to target.
- Minimise the possibility of off-target effects by choosing a gRNA that is highly specific.

Use these guidelines to pick two gRNAs (one + and one –). Whilst these parameters are helpful when choosing a gRNA, in reality, we would choose more than one "good" gRNA sequence and screen them all for their efficacy in a real experiment.

Construct the single guide RNA (sgRNA)

The gRNA sequence(s) you have chosen will target Cas9 to the desired place in the genome (the yeast *MET15* gene). However, the gRNA sequence alone cannot bind Cas9. So now you need to add a scaffold sequence to the gRNA, which will mediate binding to Cas9 (see figure 5). Together, these sequences form an sgRNA.

Figure 5: Structure and domains of an sgRNA sequence. The sgRNA interacts with Cas9 via the scaffold sequence at the 3' end of the sgRNA (green line). The 5' end of the sgRNA (green letters) is complementary to the target DNA sequence (shown in bold). Adjacent to the target DNA sequence is the PAM sequence (blue). Once bound, Cas9 induces a DSB as indicated by the orange arrowheads. (This is a ***MET15*** sgRNA and was designed on https://design.synthego.com/#/).

The sequence of the scaffold RNA is shown in the table below as both an RNA and a DNA seqeunce.

	Scaffold sequence for sgRNA
RNA	GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
DNA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT

Use your chosen gRNA sequences and the scaffold sequence from above to assemble the whole sgRNA (I would recommend using copy and paste; but it is OK to type if you prefer). Things to remember when you do this:

- It is fine to work with the DNA sequences. The sgRNA functions as an RNA molecule but we would usually clone the DNA coding sequence. (RNA is also fine if you prefer. But don't mix the two!)
- The gRNA goes at the 5' end of the sgRNA.
- The scaffold RNA goes at the 3' end of the sgRNA.
- The PAM sequence is **not** part of the sgRNA (it is only in the genome).
- For the gRNA labelled as –, you will need to reverse complement the target sequence (ask if you are not sure why).

It might seem a bit pointless to do this manually! It is important that you understand the different domains of an sgRNA, the function of these domains, and how the sgRNA anneals to the target DNA sequence.

Fill in the tables below with your sgRNA sequences and the relevant parameters (you can delete the text in the right-hand column; it is just there to help you get started).

Once you have done this manually, you should check that it is correct. You can do this using CRISPRdirect (the online programme). Click on this popup of instructions for checking the sgRNA.

sgRNA 1

Sequence of 20mer gRNA (5'-3') **[+PAM]**	[this is an example - delete and paste in yours] 5' - ACGTGTTCAGCTACCTATCT**TGG** - 3'
Target nucleotides	[nucleotides of the *MET15* gene]
GC content	
Tm	
Binds to	[coding strand / non-coding strand. This is related to the + or – in the second column but can be confusing. ASk if you are unsure]
Full sgRNA sequence (5'-3')	gRNA sequence + scaffold sequence

Sequence of sgRNA 1 from CRISPRdirect to check. This should be identical. If it's not, ask a demonstrator to help.

sgRNA 2

Sequence of 20mer gRNA (5'-3') **[+PAM]**	
Target nucleotides	
GC content	
Tm	
Binds to	
Full sgRNA sequence (5'-3')	

Sequence of sgRNA 2 from CRISPRdirect to check. This should be identical. If it's not, ask a demonstrator to help.
