

## Vejledende svar til opgaver

# CRISPR/Cas9 – Den Genteknologiske Revolution

### Teoretiske Spørgsmål (Grundteori)

#### Opgave 1 – Vejledende svar:

- a. Cas9 (CRISPR associeret protein 9)
- b. gRNA (guide RNA), eller sgRNA (single chimeric guide RNA)
- c. 5'-NGG-3'
- d. 20 nukleotider
- e. NHEJ (non-homologous end joining)
- f. 3. og 4. basepar efter PAM-sekvensen
- g. 3'-TCGCATAC-5'
- h. 5'-AGCGUAUG-3'
- i. HDR (homology directed repair)

#### Opgave 2 – Vejledende svar:

##### a. Dobbeltstrengsbrud og reparationsmekanismer

1) Når begge DNA-strengene overskæres således, at der haves to nye ender. Tilstanden er kritisk for cellen, da denne åbning af DNA-strukturen tillader mutationer.

2) Ved NHEJ samles enderne af dobbeltstrengsbruddet. Reparationen kan være fejlfri, men der kan også opstå mutationer, i form af deletioner eller insertioner, der kan ødelægge læserammen (frame-shift mutationer). Ved HDR anvendes en DNA-skabelon med homologe ender til at reparere dobbeltstrengsbruddet, hvormed der kan indsættes nye DNA-sekvenser i bruddet.

3) Først genkender Cas9 DNA-sekvensen på det givne sted, hvorefter ét dobbeltstrengsbrud dannes. NHEJ kan effektivt reparere bruddet og genoprette den originale DNA-sekvens. Denne reparerede DNA-sekvens kan genkendes igen af Cas9 og dermed dannes et nyt dobbeltstrengsbrud. Denne proces gentager sig indtil NHEJ laver en fejl i reparationsprocessen, hvormed DNA-sekvensen muteres til en ny sekvens, som Cas9 ikke kan genkende. Mutationen kan være en lille insertion eller deletion, og forårsage en frame-shift mutation.

## b. Cas9 og identifikation

- 1) gRNA er et stykke RNA, som Cas9 kan binde og anvende til at genkende DNA-sekvenser med. En 20 basepar lang sekvens i 5'-enden af gRNA bestemmer DNA-sekvensen, som kan genkendes. DNA-sekvenser genkendes, hvis de er komplementære til de 20 basepar i gRNA (under tilstedeværelsen af en PAM-sekvens).
- 2) PAM-sekvensen er en DNA-sekvens, der genkendes af selve Cas9-proteinet. Den er 5'-NGG-3' for Cas9 fra *S. pyogenes*. Det er et krav, at PAM-sekvensen er tilstede for, at Cas9 kan identificere en specifik DNA-sekvens med sit gRNA og danne et dobbeltstrengsbrud.
- 3) Først finder Cas9 det givne DNA, hvorefter Cas9 genkender PAM-sekvensen med sit PAM-interagerende domæne. DNA-strengene skiller sig ad ved bindingen af Cas9, hvorefter DNA-sekvensen sammenlignes med den 20 basepar lange gRNA-sekvens. Succesfuld Watson-Crick baseparring danner en gRNA-DNA-binding, som betyder, at DNA-sekvensen er genkendt. De første basepar efter PAM-sekvensen er vigtigst for genkendelsen, mens mismatches kan tillades længere nede (mod 5'-enden af gRNA). Nukleaserne aktiveres nu, hvorved der dannes et dobbeltstrengsbrud mellem 3. og 4. basepar efter PAM-sekvensen. RuvC-nukleasen kløver i DNA-strengen med PAM-sekvensen, og HNH-nukleasen kløver i DNA-strengen med den genkendte sekvens. Cas9 giver herefter slip og efterlader et åbent dobbeltstrengsbrud i DNA'et.

## c. Molekylære strategier og genfunktioner

- 1) Multiplexing er når man anvender flere forskellige gRNA, hvormed man kan ramme flere DNA-sekvenser på samme tid. Dette tillader, at man kan foretage mere kompliceret genmodificering. Eksempelvis kan man modificere flere sekvenser i forskellige gener eller klippe større DNA-sekvenser ud.
- 2) Ved en insertion dannes ét enkelt dobbeltstrengsbrud ud fra ét stykke gRNA. I bruddet indsættes en DNA-sekvens ud fra en indført skabelon ved HDR. Der fjernes intet af den originale DNA-sekvens, men tilføres kun ny DNA-sekvens. Dette gælder ikke for en erstatning. Ved en erstatning dannes to dobbeltstrengsbrud ud fra to stykker gRNA, hvormed der altså anvendes multiplexing. Den fjernede sekvens erstattes med en ny sekvens via en indført skabelon ved HDR.
- 3) En loss-of-function mutation medfører, at genets tilhørende protein får reduceret funktion, fjernet den oprindelige funktion eller bliver udtrykt i svagere grad.

4) En gain-of-function mutation medfører, at genets tilhørende protein får ændret den oprindelige funktion, tilføjet nye funktioner eller bliver udtrykt i stærkere grad.

#### d. Leveringmetoder

1) Protein-baseret levering. Både DNA- og RNA-baseret levering kræver translation af Cas9-kodende mRNA.

2) DNA-baseret levering. Uden transskription kan Cas9-kodende mRNA og gRNA ikke produceres fra plasmidet.

#### e. Komplikationer

1) Off-target effekter er farlige, fordi de er uforudsete ændringer af genomet, som kan forekomme i sekvenser, der minder om de eftersøgte sekvenser. Man ønsker ikke, at Cas9 fejlagtigt skal danne dobbeltstrengsbrud i andre DNA-sekvenser, som er vigtige eller har grundlæggende funktioner.

Eksempelvis ved genterapeutisk anvendelse af CRISPR/Cas9, ville det være en alvorlig situation, hvis der blev dannet dobbeltstrengsbrud i andre gener. Det kunne måske ligefrem forværre situationen.

2) *En længere PAM-sekvens*: lavere risiko for off-target effekter, da der er et højere krav til sekvensen, der skal genkendes. Jo mere specifik en sekvens der kræves, jo lavere sandsynlighed er der for, at den findes flere steder i genomet eller at andre sekvenser minder om den.

*En kortere genkendelsessekvens (<20 nukleotider) på gRNA*: højere risiko for off-target effekter, da der mindre krav til specificiteten af sekvensen, der skal genkendes.

*Genmodificering i et ekstremt stort genom*: højere risiko for off-target effekter, da der er flere muligheder for, at der findes sekvenser, der minder om hinanden i et større genom.

*Ved brug af mange flere gRNA til at ramme forskellige steder i samme celle (multiplexing)*: højere risiko for off-target effekter, da der foretages mange forskellige indgreb, der hver især kan resultere i hver deres forskellige off-target effekter.

### Opgave 3 - Vejledende svar:

a. PAM-sekvensen for Cas9 er 5'-NGG-3', og er placeret som vist herunder:

5' - AACAAAGTACATCACTAGATGATCTAAGAGGTC AATAACACACTCTAAGATGATACTAACTAATTATC - 3'  
3' - TTGTTCA TGTAGTGATCTACTAGATTCTCCAGTTATTGTGTGAGATTCTACTATGATTGATTAATAG - 5'

b. De 20 nukleotider, som Cas9 vil kunne genkende med sgRNA, ligger i direkte forlængelse af PAM-sekvensens 5'-retning, men på den modsatte streng:

5' - AACAAAGTACATCACTAGATGATCTAAGAGGTC AATAACACACTCTAAGATGATACTAACTAATTATC - 3'  
3' - TTGTTCA TGTAGTGATCTACTAGATTCTCCAGTTATTGTGTGAGATTCTACTATGATTGATTAATAG - 5'

c. Den komplementære RNA-sekvens findes, så denne kan anvendes i sgRNA. Denne er komplementær til den markerede DNA-sekvens. Den er lavet af RNA, hvormed der haves uracil (U) fremfor thymin (T):

DNA-sekvens som genkendes: 3' - TGTAGTGATCTACTAGATTC - 5'  
Tilsvarende sgRNA-sekvens: 5' - ACAUCACUAGAUGAUCUAAG...

d. Dobbeltstrengsbruddet dannes mellem det 3. og 4. basepar efter PAM-sekvensen, i den genkendte sekvens:

5' - AACAAAGTACATCACTAGATGATCT | AAGAGGTC AATAACACACTCTAAGATGATACTAACTAATTATC - 3'  
3' - TTGTTCA TGTAGTGATCTACTAGA | TTCCTCAGTTATTGTGTGAGATTCTACTATGATTGATTAATAG - 5'

## Mammut DNA (Grundteori)

### Vejledende svar:

a. Vi ved, at PAM-sekvensen, der genkendes af Cas9 fra *S. pyogenes*, er givet som 5'-NGG-3'. Her huskes det at 'N' betyder en hvilken som helst nukleotid. Dermed tjekkes begge strenge for, om de indeholder denne sekvens. Det ses, at der haves 4 PAM-sekvenser i alt, markeret med rødt på nedenstående sekvens:

```
5' - AAAGGCAGCTACTAATCTAAGATGTGCCTTGATGAGCACCTCAAGGC - 3'
3' - TTTCCGTCGATGATTAGATTCTACACGGAAGTACTCGTGGAGTTCCG - 5'
```

b. Nu tages der udgangspunkt i den sidste PAM-sekvens, hvor vores sgRNA vil blive designet ud fra. Denne er markeret med rødt herunder. (se afsnit 'Design af sgRNA')

```
5' - AAAGGCAGCTACTAATCTAAGATGTGCCTTGATGAGCACCTCAAGGC - 3'
3' - TTTCCGTCGATGATTAGATTCTACACGGAAGTACTCGTGGAGTTCCG - 5'
```

For at designe sgRNA, skal vi huske på nogle vigtige informationer. Vi ved, at vi skal finde 20 nukleotider, som vi kan indsætte i vores sgRNA. Vi ved, at sekvensen, der genkendes ligger i direkte forlængelse af PAM-sekvensens 5'-retning. Dermed har vi indskærpet det interessante område:

```
5' - AAAGGCAGCTACTAATCTAAGATGTGCCTTGATGAGCACCTCAAGGC - 3'
3' - TTTCCGTCGATGATTAGATTCTACACGGAAGTACTCGTGGAGTTCCG - 5'
```

Den næste vigtige information er, at PAM-sekvensen og den sekvens, der skal identificeres, ligger på hver sin streng. Dermed har vi nu fundet de sekvenser, som Cas9 skal binde sig til:

```
5' - AAAGGCAGCTACTAATCTAAGATGTGCCTTGATGAGCACCTCAAGGC - 3'
3' - TTTCCGTCGATGATTAGATTCTACACGGAAGTACTCGTGGAGTTCCG - 5'
```

Nu kan de 20 nukleotider blot oversættes til den komplementære version på RNA form.

Den komplementære sekvens kan aflæses direkte af DNA-sekvensen på den anden streng, hvor man udskifter T med U, eller man kan manuelt oversætte den med Watson-Crick baseparringsreglerne.

De 20 nukleotider der genkendes er: 3' - CACGGAAGTACTCGTGGAGT - 5'

Den komplementære sekvens er: 5' - GTGCCTTGATGAGCACCTCA - 3'

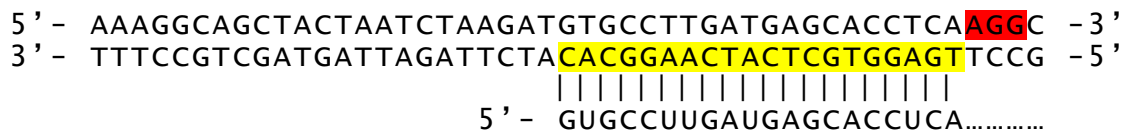
RNA-sekvensen er: 5' - GUGCCUUGAUGAGCACCUC A - 3'

Dermed haves nu de 20 nukleotider, der kan indsættes i sgRNA-skabelonen (se Design af sgRNA: Figur 9):

```
5' - GUGCCUUGAUGAGCACCUC A - 3'
```

c. Det Cas9 styret af gRNA vil identificere PAM-sekvensen og de 20 nukleotider på den modsatte streng, som angivet herunder. Det er hertil Cas9 vil danne sin binding. (Se afsnittet 'Beskrivelse af Cas9-proteinet og identifikation af en specifik DNA-sekvens')

Identifikationen, der udføres af Cas9 med sgRNA'et, giver følgende interaktion:



Nukleaserne i Cas9 klipper i DNA-strengene mellem det 3. og 4. basepar efter PAM-sekvensen, hvormed vi forventer at få det følgende dobbeltstrengsbrud:



d. Vi forventer ikke at kunne genbruge vores sgRNA, da mutationen sidder midt i PAM-sekvensen. Denne sekvens er ekstremt vigtig for aktiviteten af Cas9, da det er denne der indleder hele identifikationsprocessen. Uden 5'-NGG-3' PAM-sekvensen, vil Cas9 ikke kunne genkende området, selvom de 20 nukleotider i sgRNA'et stemmer perfekt overens. (Se 'Beskrivelse af Cas9-proteinet og identifikation af en specifik DNA-sekvens')

e. Vi forventer, at vi godt kan bruge vores sgRNA alligevel, da mutationen sidder i det 20. basepar efter PAM-sekvensen. Det kan ses, at PAM-sekvensen er bevaret og at de efterfølgende 19 basepar stemmer overens med sgRNA'et. Det er de første basepar efter PAM-sekvensen, der skal være komplementære til sgRNA, mens de efterfølgende godt kan have små afvigelser (hen mod 5'-enden af gRNA). Da mutationen befinder sig på den sidste og mindst krævende position, forventes der teoretisk set stadigvæk Cas9 aktivitet. (Se 'Beskrivelse af Cas9-proteinet og identifikation af en specifik DNA-sekvens')