CRISPR/Cpf1 – Vårt nästa stora verktyg för genmanipulationer?

Ida Isolehto

Självständigt arbete i biologi 2016

Sammandrag

Upptäckten av de prokaryota immunsystemen CRISPR/Cas har haft en stor inverkan på utvecklingen av bioteknologiska verktyg för genmodifiering. CRISPR-systemen är anpassningsbara och fungerar *in vivo* genom att spara sekvenser från invaderande DNA som sedan används för att snabbt känna igen och eliminera en upprepad infektion. CRISPR-proteinerna har förmågan att klyva DNA och styrs till målsekvensen av ett RNA som härstammar från en infekterande bakteriofag eller plasmid. Cas9 är det mest välstuderade CRISPR-systemet och har utvecklats till ett bioteknologiskt verktyg med många användningsområden. Främst används CRISPR/Cas9 för att slå ut gener men också för att göra punktmutationer eller insättningar och raderingar av sekvenser ur genomet. Trots att CRISPR har sitt ursprung i prokaryoter fungerar CRISPR/Cas9 i både prokaryota och eukaryota celler. CRISPR-systemet CRISPR/Cpf1 har liknande egenskaper som CRISPR/Cas9 och tillhör samma klass av CRISPR-proteiner men skiljer sig från Cas9 i flera avseenden. Cpf1 binder till DNA vid andra sekvenser, har ett klyvningsmönster som resulterar i överhängande ändar, systemet är avsevärt mindre än CRISPR/Cas9 och Cpf1 har en helt unik proteinstruktur. Sammantaget gör detta att Cpf1 har lovande egenskaper för att utvecklas till ett kompletterande verktyg till Cas9.

Inledning

Precis som människor är mikroorganismer utsatta för diverse infektioner, av både virus (så kallade bakteriofager) och andra bakterier. Följaktligen har skyddande immunförsvar evolverat fram även hos dem. Ett sådant är CRISPR (Clustered Regularly Interspaced Repeats) som tillsammans med Cas (CRISPR associerade proteiner) återfinns hos ungefär hälften av alla bakterier och nästan alla arkéer (Horvath *et al.* 2010). Vid en infektion av en bakteriofag eller en annan bakterie utsätts cellen för främmande genetiskt material i form av viralt DNA (och i vissa fall RNA) eller bakteriella plasmider. CRISPR-systemen sparar fragment av det infekterande DNA:t i den egna kromosomen (Jinek *et al.* 2012). Fragmenten kan sedan användas som ett ”minne” från tidigare infektioner så att en upprepad infektion snabbt kan elimineras. CRISPR-biblioteket är uppbyggt av palindroma, repetitiva sekvenser separerade av dessa DNA-fragment (Bolotin *et al.* 2005). Den engelska termen för fragmenten är ”spacers” vilket skulle kunna översättas till mellanrum. Namngivningen av dessa skedde innan dess funktion var känd och beskriver hur de separerar de repetitiva sekvenserna. Detta säger dock inget om dess funktion varför de fortsättningsvis kommer refereras till som igenkänningssekvenser. Vid en upprepad infektion används igenkänningsekvensen för att känna igen det främmande DNA:t. Uppströms om CRISPR-biblioteket kodas Cas-proteinerna och vissa ickekodande RNA (Figur 1). Det finns flera olika Cas-proteiner i CRISPR-loci som har olika funktioner i systemet. Cas-proteinerna behövs både för insamlingen av nya igenkänningssekvenser och för att bryta ned främmande DNA (Marrafini & Sontheimer 2010).

Figur 1. Struktur för CRISPR-loci. Markerat i svart är de repetitiva sekvenserna som separerar de olika igenkänningssekvenserna (gult, blått, grönt, rött). Uppströms om CRISPR-biblioteket kodas Cas-proteinerna. Röda pilar indikerar riktningen för transkriptionen. Figuren är omarbetad med tillstånd från [iGEM Uppsala 2016] (http://2016.igem.org/Team:Uppsala/Project/CRISPR.) copyright (2016).

Mekanism och klassificering av CRISPR/Cas-system

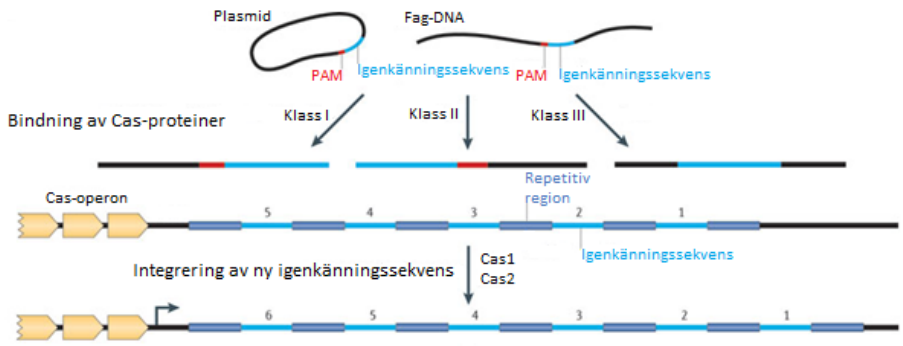
Eftersom CRISPR-systemen återfinns hos så många arter finns en stor diversitet av system som alla delar samma funktion men skiljer sig något i mekanism och struktur (Marrafini & Sontheimer 2010). Systematiseringen av CRISPR-system har flera nivåer. Först delas de in i klasser som huvudsakligen baseras på om de har enkla endonukleas (protein som klyver DNA) eller om dess endonukleas bildar komplex och hur de komplexen verkar (Makarova *et al.* 2011a). Klasserna delas sedan vidare in i typer utefter systemets funktionella egenskaper och evolutionära ursprung. CRISPR-systemens mekanism för att mediera immunitet kan delas in i tre olika stadier: Insamling av nya igenkänningssekvenser, uttryck och bearbetning och slutligen interaktion med målsekvensen (Makarova *et al.* 2015).

Insamling av igenkänningssekvenser

CRISPR-systemet är ett dynamiskt och anpassningsbart immunsystem. Vid infektioner samlar CRISPR-systemet hela tiden in nya igenkänningssekvenser vilket gör att systemet anpassas efter vilka infektioner cellen blir utsatt för.

Hos CRISPR-system av klass I och II är insamlingen av igenkänningssekvenser beroende av en PAM-sekvens i det invaderande DNA:t (Figur 2) (Makarova *et al.* 2011b). PAM-sekvensen består av en kort nukleotidsekvens bredvid målsekvensen och tros vara ett sätt att skilja mellan främmande DNA och cellens egna genom (Anders *et al.* 2014). Vid en infektion känns PAM-sekvensen i det invaderande genomet igen av Cas-proteinet som plockar upp en ny igenkänningssekvens intill PAM-sekvensen. Till skillnad från system av klass I och II kräver upplockningen av igenkänningssekvenser hos system av klass III ingen PAM-sekvens (Makarova *et al.* 2011b).

Om en potentiell igenkänningssekvens hittas klyvs denna (vanligtvis till en längd av ungefär 30 nukleotider) och inkorporeras i CRISPR-biblioteket av Cas1 och Cas2 (Figur 2) så att den sedan kan uttryckas och användas vid en upprepad infektion (Figur 3). När en ny igenkänningssekvens sätts in i CRISPR-biblioteket sker det alltid närmast Cas-operonet och den repetitiva regionen dubbleras så att bibliotekets struktur bevaras (Makarova *et al.* 2011b).

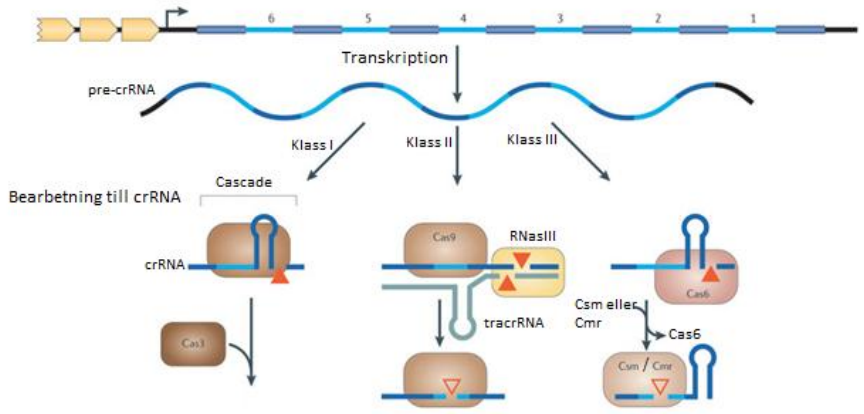


Figur 2. Insamling av igenkänningssekvenser. Insamlingen av igenkänningssekvenser (ljusblått) medieras av Cas-proteiner i de olika klasserna av CRISPR-system, exakt vilka Cas-proteiner är dock ännu ej klarlagt. System av klass I och II är beroende av en PAM-sekvens (rött) i det invaderande DNA:t till skillnad från system av klass III. Cas-proteinerna binder till det invaderande DNA:t och med hjälp av Cas1 och Cas2 integreras en ny igenkänningssekvens mellan de repetitiva regionerna (blått) i CRISPR-biblioteket. Insättningen sker i änden närmast Cas-operonet. Figuren är omarbetad med tillstånd från Macmillan Publishers Ltd: [NATURE REVIEWS MICROBIOLOGY] (Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., & Charpentier, E., Evolution and classification of the CRISPR-cas systems) copyright (2011).

Uttryck och bearbetning

CRISPR-biblioteket uttrycks till ett långt prekursor transkript (pre-crRNA) som innehåller både igenkänningssekvenser och de repetitiva regionerna. Pre-crRNA:t bearbetas sedan på olika vis av de olika klasserna av CRISPR-system för att skapa det mogna crRNA:t som kan styra endonukleaset till målsekvensen (Figur 3). Bearbetningen kan antingen utföras av komplex eller av enkla endoribonukleaser, och i vissa fall med hjälp av ett ytterligare RNA (Marrafini & Sontheimer 2010).

I klass I bildar flera av Cas-proteinerna ett komplex som kallas för Cascade. I detta komplex ingår Cas6 som klyver pre-crRNA:t till moget crRNA. Efter bearbetningen rekryteras Cas3 till Cascadekomplexet (Makarova *et al.* 2011b). Hos system av klass II uttrycks förutom pre-crRNA även ett transaktiverande CRISPR-RNA (tracrRNA). TracrRNA:t binder till pre-crRNA:t vilket i sin tur möjliggör bindning av Cas9. När bindningen mellan Cas9, tracrRNA och pre-crRNA har skett kan Cas9 tillsammans med RNasIII generera det mogna crRNA:t (Deltcheva *et al.* 2011). I system av klass III utförs bearbetningen precis som i klass I av Cas6, men efter bearbetningen överförs crRNA:t till ett annat Cas-complex (Csm eller Cmr)(Makarova *et al.* 2011b).

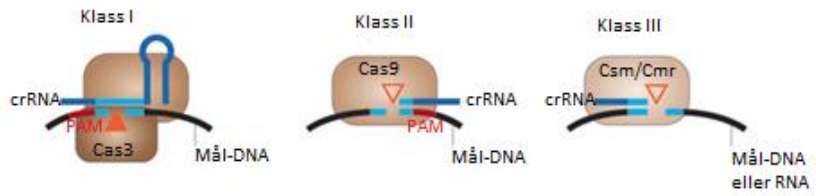


Figur 3. Uttryck och bearbetning. CRISPR-biblioteket uttrycks till ett pre-crRNA som innehåller igenkänningssekvensen (ljusblått) och repetitiv region (mörkblått). Genom klyvning av pre-crRNA (klyvningspositioner utmärkta med pilar) genereras det mogna crRNA:t. I system av klass I bearbetas pre-crRNA av Cascadekomplexet och det mogna crRNA:t fortsätter att vara bundet till komplexet. I detta stadium rekryteras Cas3 till Cascadekomplexet. I klass II bearbetas pre-crRNA:t av Cas9 och RNasIII med hjälp av tracrRNA. I klass III bearbetas det av Cas6 varpå crRNA:t överförs till Csm eller Cmr. Figuren är omarbetad med tillstånd från Macmillan Publishers Ltd: [NATURE REVIEWS MICROBIOLOGY] (Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., & Charpentier, E., Evolution and classification of the CRISPR-cas systems) copyright (2011).

Interaktion med målsekvensen

Eftersom crRNA:t innehåller igenkänningssekvensen som ursprungligen kommer från en tidigare infektion kan det efter bearbetningen styra endonukleaset till målsekvensen i det invaderande genomet. Styrningen sker genom att igenkänningssekvensen i crRNA:t basparar med den komplementära sekvensen i det invaderande DNA:t (Horvath *et al.* 2010). System av klass I och II är båda beroende av en PAM-sekvens intill målsekvensen (Makarova *et al.* 2011a). Basparningen mellan crRNA:t och målsekvensen möjliggör för endonukleaset att klyva målsekvensen och på så vis eliminera hotet som det invaderande DNA:t utgör.

Hos klass I är det Cas3-subenheten i Cascade-komplexet som utför klyvningen medan det hos klass II är Cas9 (Figur 4). System av klass III särskiljer sig dels genom att vissa system har förmågan att klyva både DNA och RNA, och dels genom att de inte är beroende av en PAM-sekvens. Klyvningen i klass III-system utförs av Csm- eller Cmr-komplexet. Eftersom systemet inte använder sig av en PAM-sekvens förlitar det sig istället på strukturen i CRISPR-biblioteket för att skilja mellan sitt eget och det invaderande genomet (Makarova *et al.* 2011b).



Figur 4. Interaktionen mellan endonukleaset och målsekvensen. I system av klass I klyvs målsekvensen (svart) av Cas3, i klass II av Cas9 och i klass III av Csm/Cmr. Klyvningen (indikerat med pilar) är beroende av en PAM-sekvens (rött) i system av klass I och klass II. Figuren är omarbetad med tillstånd från Macmillan Publishers Ltd: [NATURE REVIEWS MICROBIOLOGY] (Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., & Charpentier, E., Evolution and classification of the CRISPR-cas systems) copyright (2011).

CRISPR/Cas9

En typ av system, CRISPR/Cas9, har sedan upptäckten fått särskilt stor uppmärksamhet på grund av dess stora användbarhet inom bioteknologin. Man kan nämligen utforma egna system som med hjälp av Cas9 klyver DNA vid en önskad position i valfritt genom (Hsu *et al.* 2014). De enda komponenter som behövs för att klyva i genomet med hjälp av Cas9 är endonukleaset i sig, en igenkänningssekvens transkriberad till crRNA, ett transaktiverande tracrRNA och en målsekvens med en intilliggande PAM-sekvens (Jinek *et al.* 2012). För att förenkla användningen av systemet som bioteknologiskt verktyg har man skapat en fusion mellan crRNA och tracrRNA så att de bildar en enkel enhet, ett så kallat sgRNA. SgRNA:t kan designas till att matcha vilken sekvens man önskar så den enda restriktionen för var man kan välja att klyva i genomet är huruvida det finns en intilliggande PAM-sekvens eller inte (Jinek et al 2012).

Modulariteten och specificiteten i CRISPR/Cas-systemen har öppnat upp en helt ny värld av möjligheter inom biotekniken. Användningsområdena för CRISPR-baserade verktyg är sannerligen omfattande och sträcker sig från basforskning, bioteknik och medicin till framtida genterapi. Sedan upptäckten av Cas9 och dess stora användbarhet har man isolerat och karaktäriserat många olika Cas-proteiner från olika organismer i jakten på fler system som kan öka effektiviteten eller komplettera Cas9. Ett sådant system som har visat lovande egenskaper är CRISPR/Cpf1 (Fonfara *et al.* 2016). Denna litteraturstudie fokuserar på detta nyupptäckta CRISPR-system och undersöker hur det skiljer sig i struktur och mekanism från det numera välstuderade CRISPR/Cas9-systemet. Möjliga användningsområden där Cpf1 kan vara mer fördelaktig att använda än Cas9 diskuteras även.

CRISPR-system av klass II

År 2015 publicerades upptäckten av Cpf1, ett Cas-protein som har liknande verkan som Cas9 men med en del andra egenskaper. Cpf1 känner igen andra PAM-sekvenser och har ett annat klyvningsmönster. Till skillnad från Cas9 som klyver nära PAM-sekvensen och lämnar trubbiga ändar klyver Cpf1 långt nedströms om PAM-sekvensen och lämnar överhängande ändar (Zetsche *et al.* 2015). Detta klyvningsmönster gör att Cpf1 har potential att vara mer effektiv än Cas9 för att användas till ändringar i genomet baserat på homolog rekombination. CRISPR/Cpf1 är dessutom det mest minimalistiska CRISPR system man hittills upptäckt,

både på grund av att dess guidande RNA är avsevärt kortare och på grund av att det själv bearbetar sitt crRNA och varken är beroende av ett tracrRNA eller cellens egna RNaser (Fonfara *et al.* 2016).

Klassificering av Cpf1

Både Cas9 och Cpf1 är CRISPR-system av klass II. Detta innebär att de har ett enda nukleas som utför klyvningen till skillnad från klass I och III-system som karaktäriseras av komplex av flera proteiner (Makarova *et al.* 2015). Trots att de har ganska liknande egenskaper finns det också stora skillnader mellan Cpf1 och Cas9. Efter indelningen i klasser delas systemen vidare in i typer. CRISPR/Cas9 är ett klass II- typ II-system, men CRISPR-locus för Cpf1 liknar strukturellt mer de loci som hittas hos klass I och III än hos klass II. Medan CRISPR-locus hos Cas9 innehåller både ett tracrRNA och ett pre-crRNA har Cpf1-locus bara ett pre-crRNA (Deltcheva *et al.* 2011, Fonfara *et al.* 2016). Dessutom har Cpf1 endast en aktiv nukleasdomän som är känd sedan tidigare, till skillnad från Cas9 som har två välkaraktäriserade aktiva nukleasdomäner (Zetsche *et al.* 2015, Makarova *et al.* 2011b). Strukturen och uppbyggnaden av både locus och det aktiva proteinet antyder att Cpf1 har vitt skilda verkningsmekanismer än Cas9 (Makarova *et al.* 2015). Man kan därav dra slutsatsen att Cas9 och Cpf1 är evolutionärt skilda trots att de har vissa funktionella likheter. Därför har Cpf1 kategoriserats som en ny typ av klass II endonukleaser, typ V.

Igenkänningssekvensen

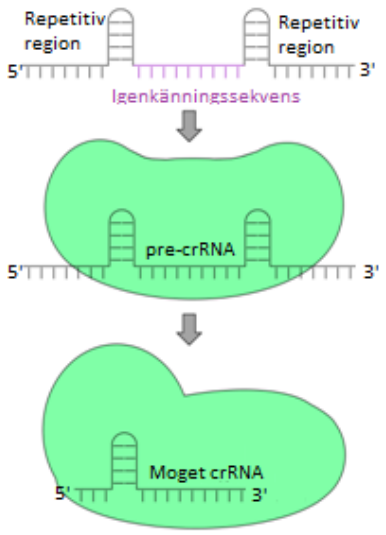
För att igenkänningssekvensen ska kunna styra Cas-proteinet till rätt målsekvens måste den först transkriberas till pre-crRNA och sedan bearbetas till moget crRNA. Pre-crRNA:t innehåller både igenkänningssekvensen och de flankerande repetitiva sekvenserna (Zetsche *et al.* 2015). Längden och strukturen av det mogna crRNA:t ligger devis till grund för klassifikationen av Cpf1. Det mogna crRNA:t för Cpf1 är 42-44 nukleotider (nt) långt och inkluderar i 5’- änden19 nt repetitiv sekvens följt av 23-25 nt med ursprung i igenkänningssekvensen i 3’-änden, vilket liknar uppbyggnaden hos klass I- och III-system (Fonfara *et al.* 2016). Typ II CRISPR/Cas-system, som Cas9 tillhör, har istället omvänd ordning på igenkänningssekvens och repetitiv region så att igenkänningssekvensen sitter i 5’-änden och den repetitiva sekvensen i 3’-änden (Makarova *et al.* 2011). Längden på crRNA:t skiljer sig också något i typ II-system jämfört med system av typ V då de har en igenkänningssekvens på mellan 20 och 24 nt och en repetitiv region på 22 nt.

PAM-sekvensen

Cas-proteinerna guidas till målsekvensen av crRNA, men enbart en matchning mellan crRNA:t och målsekvensen räcker inte för att klyvning ska ske i system av klass II (som både Cas9 och CPF1 tillhör). Nukleasaktiviteten är beroende av att det finns en PAM-sekvens intill målsekvensen. Cpf1 känner igen thymidinrika PAM-sekvenser lokaliserade uppströms om igenkänningssekvensen (Zetsche *et al.* 2015), till skillnad från Cas9 som känner igen guaninrika PAM-sekvenser nedströms om igenkänningssekvensen (Deltcheva *et al.* 2011). PAM-sekvenserna skiljer sig mellan olika CRISPR-system men de kan också variera något mellan samma system hos olika arter. Till exempel är PAM-sekvensen för Cpf1 från *Francicella Novicida* 5’-TTN-3’ medan den från *Acidaminococcus sp* är 5’-TTTN-3’. Det finns en viss flexibilitet i igenkänningen av PAM-sekvenserna och för Cpf1 har den mittersta thymidinbasen störst betydelse för igenkänningen (Zetsche *et al.* 2015, Fonfara *et al.* 2016). Specificiteten för olika PAM-sekvenser och lokaliseringen av dessa förklaras av hur nukleasets aktiva domäner är strukturerade.

Kärnsekvensen
Om det finns en komplementär målsekvens och rätt PAM-sekvens kan endonukleaset klyva DNA:t. Ju högre komplementaritet det är mellan igenkänningssekvensen och målsekvensen desto högre effektivitet får klyvningen, men klyvning kan ske trots att sekvenserna inte är fullt komplementära till varandra. Sekvensen närmast PAM-sekvensen måste dock vara fullt komplementär till igenkänningssekvensen for att endonukleaset ska kunna binda in till DNA:t. Denna del av igenkänningssekvensen kallas kärnsekvensen och är olika lång för olika CRISPR-system. Kärnsekvensen för Cpf1 är 5-8 nt lång och lokaliserad i 5'-änden (Zetsche *et al.* 2015) medan den för Cas9 är 13 nt i 3'-änden (Jinek *et al.* 2012). Cpf1 kan tolerera enstaka icke-komplementära nukleotider utanför kärnsekvensen medan dubbla nästan helt reducerar aktiviteten (Kim *et al.* 2016, Kleinstiver *et al.* 2016). Detta innebär att Cpf1 är mycket specifik för den tilltänkta målsekvensen.

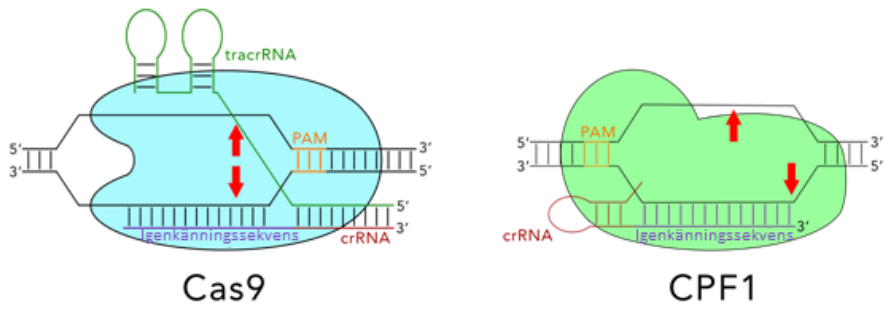
Bearbetning av pre-crRNA
Som tidigare nämnt transkriberas igenkänningssekvenserna i CRISPR-loci till pre-crRNA som måste bearbetas (Figur 3). De flankande repetitiva sekvenserna omvandlas genom klyvning till det mogna crRNA:t. Cas 9 kräver transkription av både pre-crRNA och tracrRNA som binder till varandra och tillsammans med RNas III bearbetar pre-crRNA:t till moget crRNA (Deltcheva *et al.* 2011). Cpf1 däremot behöver inget tracrRNA eller RNas för att bearbeta sitt crRNA (Zetsche *et al.* 2015). Istället bearbetar nukleaset själv sitt eget pre-crRNA genom att först klyva det uppströms från 3'-hårnålsstrukturen och sedan vidare bearbeta det genom en ännu ej kartlagd mekanism. I 5'-änden behålls den repetitiva sekvensen så att det mogna crRNA:t kvarhåller en stamloop (Figur 5). Baserat på nukleotidsekvensen förutsågs denna bilda sekundärstrukturen av en hårnål, men analys av kristallstrukturen har visat att den snarare bildar en pseudoknut (Zetsche *et al.* 2015, Yamano *et al.* 2016, Gao *et al.* 2016). Denna sekundärstruktur refereras till som 5'-handtaget. Bindningen mellan Cpf1 och crRNA:t sker via 5'-handtaget och är beroende av både struktur och sekvens (Fonfara *et al. 2016).*



Figur 5. Bearbetning av pre-crRNA. Överst visas pre-crRNA med de repetitiva regionerna markerade i svart och igenkänningssekvensen i lila. Cpf1-proteinet (grönt) binder till pre-crRNA vilket leder till en konformationsändring. Pre-crRNA:t klyvs uppströms om 3'-sekundärstrukturen och genererar ett moget crRNA med ett 5'-handtag. Figuren är omarbetad med tillstånd från [iGEM Uppsala 2016] (http://2016.igem.org/Team:Uppsala/Project/CRISPR.) copyright (2016).

Klyvningsmönster

En aspekt som har stor betydelse är att klyvningsmönstret skiljer sig åt mellan Cpf1 och Cas9 (Figur 5). Cas9 klyver den komplementära strängen till igenkänningssekvensen tre baspar uppströms om PAM-sekvensen. Den ickekomplementära strängen klyvs en eller flera gånger vid 3 till 8 baspar uppströms om PAM-sekvensen och bearbetas sedan av exonukleaser så att trubbiga ändar lämnas (Garneau *et al.* 2010). Viktigt att notera är att Cas9 alltså klyver målsekvensen inom kärnsekvensen. Cpf1 däremot klyver DNA-strängarna vid olika positioner nedströms om PAM-sekvensen, vilket är utanför kärnsekvensen. Den komplementära strängen klyvs vid 23 nt och den ickekomplementära vid 18 nt nedströms från PAM-sekvensen. Eftersom Cpf1 klyver de båda strängarna vid olika positioner lämnar den överhängande ändar (Figur 5). Klyvningen resulterar i fyra till fem nukleotiders överhäng i 5'-ändorna av strängarna (Zetsche *et al.* 2015). Cpf1 kan likt Cas9 klyva både linjärt och övertvinnat DNA (Jinek *et al.* 2012, Zetsche *et al.* 2015).



Figur 5. Klyvningsmönstret hos Cas9 och Cpf1. Endonukleasen styrs till målsekvensen av crRNA (rött) genom basparning med igenkänningssekvensen (lila). Cas9 kräver även tracrRNA (grönt) och klyver målsekvensen 3 nt uppströms om PAM-sekvensen (orange) vid samma position på båda DNA-strängarna vilket resulterar i trubbiga ändar. Cpf1 klyver DNA-strängarna i målsekvensen vid 18 respektive 23 nt nedströms om PAM-sekvensen (orange) och lämnar överhängande ändar. Klyvningspositionerna är markerade med pilar. Figuren är omarbetad med tillstånd från [iGEM Uppsala 2016] (http://2016.igem.org/Team:Uppsala/Project/CRISPR) copyright (2016).

Proteinstruktur

CRISPR-proteinernas struktur avgör dess egenskaper både vad gäller klyvningsmönster och PAM-specificitet. Cas9 har två aktiva nukleasdomäner som liknar HNH- och RuvC-endonukleaser (Makarova *et al.* 2011). Hos Cas9 klyver HNH-domänen den komplementära strängen medan RuvC-domänen klyver den ickekomplementära. Via inaktivering av dessa kan man skapa ett Cas9 som bara klyver den ena eller andra strängen, ett så kallat nickas. Man kan också inaktivera båda nukleasdomänerna och på så vis skapa ett endonukleas som binder till målsekvensen men helt saknar nukleasaktivitet (Jinek *et al.* 2012). Sådana modifierade endonukleaser kan vara användbara för att reglera uttryck av specifika gener genom att binda till regulatoriska regioner.

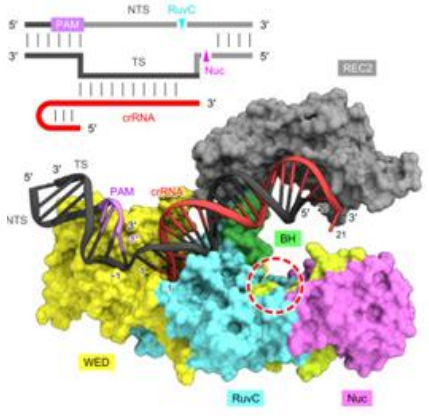
I stora drag finns en strukturell likhet mellan Cas9 och Cpf1 då de båda består av två lober som förbinds av en helixbrygga och bildar en aktiv kanal runt cr-RNA:t. Cpf1 har likt Cas9 en RuvC-liknande nukleasdomän (Figur 6). Trots att deras strukturer liknar varandra finns ingen större homologi mellan proteinernas sekvenser utöver den som återfinns i RuvC-domänen (Yamano *et al.* 2016). Den första studien av Cpf1 antog att denna stod för klyvningen av båda strängarna (Zetsche *et al.* 2015). Om RuvC-domänen inaktiveras förlorar Cpf1 helt sin endonukleasaktivitet till skillnad från Cas9 som då bara klyver den ena DNA-strängen. Detta resultat låg till grund för antagandet att RuvC-domänen skulle klyva båda strängarna vilket i sin tur förklarades med att Cpf1 skulle verka som en dimer. Denna hypotes styrktes i studien av att Cpf1 vid rening hade en molekylvikt på ca 300 kDa (ungefär dubbla molekylvikten) (Zetsche *et al.* 2015). Senare studier har dock indikerat att Cpf1 verkar som en monomer då de observerat en molekylvikt på 187 kDa, vilket också styrkts av studier av kristallstrukturen (Fonfara *et al.* 2016, Gao *et al.* 2016, Yamano *et al.* 2016). Istället föreslås hypotesen att Cpf1 har en andra nukleasdomän som inte är känd sedan tidigare och som klyver den ena strängen men är beroende av RuvC-domänen (Fonfara *et al.* 2016, Gao *et al.* 2016, Yamano *et al.* 2016). Cpf1 har även en aktiv RNas-domän som står för bearbetningen

av pre-crRNA (Fonfara *et al.* 2016). För aktivitet kräver de närvaro av katjoner av magnesium eller kalcium (Fonfara *et al.* 2016).

Genom att studera kristallstrukturen har man bekräftat att Cpf1 bildar en struktur av två lober bestående av en alfahelix-igenkänningslob (REC) och en nukleaslob (NUC) som skapar en positivt laddad aktiv kanal runt crRNA:t och mål-DNA:t (Figur 6) (Gao *et al.* 2016, Yamano *et al.* 2016). Loberna binds samman av en helixbrygga som interagerar med socker-fosfatryggraden i mål-DNA:t. REC-loben består av två domäner, REC1 och REC2, och NUC-loben av RuvC-domänen samt domänerna WED, PI och Nuc vilka har unika domänstrukturer som aldrig tidigare observerats utanför Cpf1-familjen (Figur 6)(Gao *et al.* 2016, Yamano *et al.* 2016). Domänerna WED och PI hos Cpf1 är inte strukturellt närbesläktade med, men har liknande funktion som, WED- och PI-domänerna hos Cas9 (Anders *et al.* 2014, Gao *et al.* 2016, Nishimasu *et al.* 2015, Yamano *et al.* 2016).

PI-domänen bestämmer PAM-specificiteten och interagerar med PAM-sekvensen både genom att justera igenkänningssekvensen till rätt position i förhållande till målsekvensen och genom att fungera som ett topoisomeras (Anders *et al.* 2014). Topoisomerasfunktionen kommer av att PI-domänen binder till PAM-sekvensen vilket får DNA:strängen att vridas. Detta leder till att bindningen mellan DNA-strängarna i målsekvensen destabiliseras och DNA:t börjar lindas upp. Upplindningen av DNA:t möjliggör basparning mellan målsekvensen och igenkänningssekvensen och ett RNA-DNA-heteroduplex bildas. Basparningen destabiliserar ytterligare bindningen mellan målsekvensen och den icke-komplementära strängen och lindar ytterligare upp DNA:t. Detta förklarar varför missmatchningar inom kärnsekvensen (närmast PAM-sekvensen) inte tolereras för att klyvning skall ske (Anders *et al.* 2014). När igenkänningssekvensen basparar med målsekvensen undergår Cpf1 förändringar i konformationen (Gao *et al.* 2016, Yamano *et al.* 2016).

WED-domänen deltar tillsammans med REC-loben i att känna igen 5’-handtaget i crRNA:t så att crRNA:t kan bilda komplex tillsammans med Cpf1(Yamano *et al.* 2016). RuvC-domänen och Nuc-domänen innehåller de aktiva klyvningssätena (Figur 6). Nuc-domänen klyver den komplementära DNA-strängen i målsekvensen (Gao *et al.* 2016, Yamano *et al.* 2016). Inaktivering av Nuc-domänen resulterar i ett nickas som bara klyver den ickekomplementära strängen, medan inaktivering av RuvC-domänen helt inaktiverar klyvningsaktiviteten. Detta tyder på att RuvC först måste klyva den ickekomplementära strängen för att Nuc ska kunna klyva den andra strängen (Gao *et al.* 2016, Yamano *et al.* 2016).



Figur 6. 3D-struktur för Cpf1 bundet till crRNA. För att ge en enklare överblick visas ej REC1- och PI-domänerna. REC- och NUC-loberna binds samman av en helixbrygga (grönt). WED-domänen (gul) binder till 5'-handtaget i crRNA (röd) tillsammans med REC2 (grå).. Målsekvensen är markerad i svart och crRNA i rött och PAM-sekvensen i lila. Nuc-domänen (lila) och RuvC-domänen (turkos) har de aktiva sätena som klyver DNA. Det aktiva centret markerat med en ring. Under basparningen med målsekvensen genomgår proteinet strukturella förändringar som möjliggör interaktion mellan målsekvensen och de aktiva sätena. Ovan strukturen visas en schematisk bild av basparningen mellan målsekvensen och crRNA:t med klyvningspositionerna markerade med pilar. I grått visas de sekvenser som inte är synliga i strukturbilden. Figuren är återpublicerad från [CELL (CAMBRIDGE)] (volume 165, Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., & Slaymaker, I. M,. Crystal structure of Cpf1 in complex with guide RNA and target DNA, sidorna 949-962.) copyright (2016), med tillåtelse från Elsevier.

Diskussion

Utvecklingen inom forskning och bioteknink går ständigt framåt och ju mer vi lär oss desto fler nya upptäckter kan vi också göra. CRISPR/Cas-systemen är ett utmärkt exempel på en sådan upptäckt som i sin tur lett till en mycket större förståelse för andra system. Genom att använda sig av CRISPR/Cas-system kan man nu på ett mycket mer effektivt sätt än tidigare utforska gener och deras funktioner. Cas9 har redan kommit att bli ett oumbärligt tillskott i den genetiska verktygslådan och med vidare forskning har Cpf1 potential att kanske till och med överträffa Cas9 inom vissa användningsområden.

Användningsområden för Cas9

För att förenkla användningen av Cas9 har man skapat en fusion mellan tracrRNA och crRNA så att de bildar ett enkelt sgRNA som guidar Cas9 till målsekvensen. I sgRNA:t är igenkänningssekvensen och hårnålsstrukturen i 3'-änden av crRNA:t fuserat till 5'-änden av tracrRNA:t (Jinek *et al.* 2012, Cong *et al.* 2013). Igenkänningssekvensen kan designas fritt för att klyva i genomet vid önskade positioner. Till följd av klyvningen skapas ofta mutationer som förändrar läsramen vilket kan leda till ett trunkerat protein eller inaktivering av genen i fråga beroende på var i genen man väljer att klyva. Detta är en smidig metod som används till att studera funktionen hos olika gener.

Cas9 kan också användas för att göra både stora och små raderingar ur genomet. Genom att använda två olika igenkänningssekvenser kan man klippa ut en DNA-sekvens (Cong *et al.* 2013). På så vis kan man till exempel ta bort en hel gen eller en specifik domän ur ett protein. Man kan också använda sig av homologibaserade mekanismer för att göra olika typer av förändringar i genomet. Då förser man cellen med en DNA-sekvens som har homologa ändar

till positionen varvid man klipper. Cellen kan då använda denna DNA sekvens som templat för att laga brottet i genomet. På detta sätt kan man göra punktmutationer, insättningar och raderingar (Cong *et al.* 2013).

Användningsområden för Cpf1

Eftersom Cpf1 har andra egenskaper än Cas9 har systemet potential att användas som ett kompletterande verktyg men har också potential att överträffa Cas9 inom vissa användningsområden. Cpf1 har isolerats från ett flertal organismer och alla Cpf1-nukleaser fungerar inte för klyvning i humanceller. Å andra sidan fungerar inte heller alla Cas9 för detta ändamål. Två olika Cpf1 har visat liknande effektivitet som Cas9 för klyvning i humana celler, Cpf1 från *Acidaminococcus sp* och *Lachnospiraceae bacterium* (Zetsche *et al.* 2015, Kim *et al.* 2016). Eftersom Cas-proteinerna tillåter en varierande grad av missmatchningar innebär det att de kan klyva på fler ställen i genomet än det önskade. Detta är problematiskt eftersom klyvningen ofta resulterar i mutationer, vilket man optimalt vill undvika på andra positioner än i sin målsekvens. Cpf1 har visat sig vara mycket specifik för klyvning i humana celler med markant lägre aktivitet utanför målet än Cas9 (Nishimasu *et al.* 2015, Kim *et al.* 2016, Kleinstivier *et al.* 2016).

De olika klyvningsmönstren och positionerna hos Cas9 och Cpf1 gör också att de producerar olika mutationer efter klyvning. Cpf1 har en längre frekvens av insättning och radering av enstaka nukleotider än Cas9 men en högre frekvens av punktmutationer (Kim *et al.* 2016). Detta betyder att Cas9 är mer effektiv för att slå ut gener eller skapa trunkerade proteiner medan Cpf1 är mer effektiv för att generera mutationer som bevarar uttrycket av genen. Likt Cas9 kan Cpf1 muteras för att skapa ett inaktivt nukleas som bara binder till målsekvensen eller ett nickas som bara klyver den ena DNA-strängen. Dock kan man enbart skapa ett Cpf1-nickas som klyver den icke-komplementära strängen på grund av att Nuc-domänens nukleasaktivitet är beroende av RuvC-domänen. Genom att inaktiva nukleaser och designa sgRNA som binder till regulativa regioner kan man reglera genuttryck, vilket är användbart när man studerar geners funktioner. SgRNA:t kan levereas till cellen antingen som PCR-amplikon eller i en plasmid. I humanceller är resultatet mer effektivt och specifikt om man använder sig av en plasmid. Detta kan möjligtvis förklaras av att risken för felaktiga sgRNA är större vid PCR-amplifiering än vid uttryck från en plasmid. Om sgRNA är inkorrekta minskar specificiteten då det leder det till klyvning utanför målsekvensen (Kim *et al.* 2016).

Det kanske allra mest intressanta med Cpf1 är dess klyvningsmönster som har lovande egenskaper för genmodifiering både via ickehomolog sammanfogning och via homolog rekombination. Eftersom Cpf1 lämnar överhängande ändar skulle man kunna använda Cpf1 för att klistra in DNA-sekvenser i genomet med hjälp av mekanismen för icke-homolog sammanfogning genom att designa sekvenser med matchande överhängande ändar. Cpf1 skulle också kunna användas för att än mer effektivt göra modifikationer via homolog rekombination. Detta används Cas9 till just nu men effektiviteten är låg på grund av att reparationsmekanismerna som baseras på homolog rekombination inte är lika vanligt förekommande som de för ickehomolog sammanfogning. Det innebär att vid de flesta klyvningar kommer brottet att lagas via ickehomolog sammanfogning. Det är därför svårt att åstadkomma modifikationer med hög effektivitet via homolog rekombination. Det faktum att Cpf1 klyver bortom kärnsekvensen gör att även om klyvningen skapar en mutation kan Cpf1 fortsätta att klyva i genomet tills dess att den önskade sekvensen har rekombinerat in. För att Cpf1 inte ska fortsätta att klyva efter att den önskade sekvensen har rekombinerat in kan man designa sekvensen som används som templat med punktmutationer som förändrar PAM-sekvensen eller kärnsekvensen. Hos Cas9 leder själva klyvningen till en förändring i

kärnsekvensen vilket får som konsekvens att Cas9 inte kan fortsätta att klyva. Man har alltså bara ett klyvningstillfälle på sig att rekombinera in den önskade sekvensen om man använder sig av Cas9 medan man med Cpf1 bör få flera chanser (Zetsche *et al.* 2015). Detta betyder att teoretiskt bör Cpf1 vara mer effektiv för genmodifiering via homolog rekombination. Utöver detta öppnar också den thymidinrika PAM-sekvensen upp för ännu fler potentiella målsekvenser och kan vara ett komplement till Cas9:s guaninrika PAM-sekvens (Zetsche *et al.* 2015).

Att kunna göra omfattande genmodifieringar på så många ställen i genomet som Cpf1 tillåter skulle vara ett mycket användbart verktyg inom många typer av medicinsk, genetisk och molekylärbiologisk forskning. Dess specificitet och effektivitet verkar även lovande för utvecklingen av nya genterapier. För att göra genetiska modifieringar i humana genom på ett säkert sätt är det synnerligen viktigt att modifieringarna är både specifika och effektiva. Om vi lyckas utveckla verktyg som kan leva upp till detta skulle samhällsnyttan vara mycket betydande då genetiska sjukdomar i dagsläget ofta är allvarliga och svårbehandlade. Genterapi består just nu främst av att förse cellen med en fungerande kopia av den defekta genen, men om vi kan använda oss av CRISPR/Cas-system skulle man istället permanent reparera den genetiska defekten. Den stora innebörden i detta är att patienten slipper upprepade dyra behandlingar och livslånga medicineringar, vilket givetvis är fördelaktigt för patientens livskvalitet men också ur ett samhällsekonomiskt perspektiv. Just nu sker pågående forskning kring att implementera Cas9-baserade system som behandlingsmetod för ett brett spektrum av genetiska sjukdomar, och om Cpf1 kan vara en än mer effektiv och säker metod återstår att se. Sammanfattningsvis kan man konstatera att det behövs mer forskning kring Cpf1 och en lämplig infallsvinkel bör vara att undersöka om hypotesen kring dess effektivitet för genmodifiering baserad på homolog rekombination stämmer.

Tack

Till Magnus Eklund för vägledning och goda råd, Björn Greijer, Freja Sörenby, Karin Vickberg och Mårten Ljungberg för konstruktiv kritik under uppsatsens utformande, till iGEM Uppsala 2016 som väckte mitt intresse för Cpf1 och slutligen Federico Turco som varit ett bollplank under min process att förstå mekanismerna bakom Cpf1.

Referenser

Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature **513**: 569-573.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151**: 2551–2561.

Cong L, Ran FA, Cox D, Lin S, Barretto R. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science **339**: 819-823.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature **471**: 602-607.

Fonfara I, Richter H, Bratovic M, Le Rhun A, Charpentier E. 2016. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature **532**: 517-521.

Gao P, Yang H, Rajashankar KR., Huang Z, Patel DJ. 2016. Type V CRISPR-cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. Cell research **26**: 901-913.

Garneau JE, Dupuis M, Villion M, Romero DA, Barrangou R. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468**: 67-71.

Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327**: 167-170

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. Cell **157**: 1262-1278

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337**: 816-821.
Kim D, Kim J, Hur JK, Been KW, Yoon S. 2016. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nature biotechnology **34**: 863-868.
Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM. 2016. Genome-wide specificities of CRISPR-cas Cpf1 nucleases in human cells. Nature biotechnology **34**: 869-874.
Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 2006. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biology direct DOI: 10.1186/1745-6150-1-7
Makarova KS, Aravind L, Wolf YI, Koonin EV. 2011a. Unification of cas protein families and a simple scenario for the origin and evolution of CRISPR-cas systems. Biology direct DOI: 10.1186/1745-6150-6-38
Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E. 2011b. Evolution and classification of the CRISPR-cas systems. Nature reviews Microbiology **9**:467-477.
Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA. 2015. An updated evolutionary classification of CRISPR-cas systems. Nature reviews Microbiology **9**:467-477.
Marraffini LA, Sontheimer EJ. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nature reviews Genetics **11**: 181-190
Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B. 2015. Crystal structure of staphylococcus aureus Cas9. Cell **162**:1113-1126.
Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM. 2016. Crystal structure of Cpf1 in complex with guide RNA and target DNA. Cell **165**:949-962.
Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-cas system. Cell **163**:759-771.

[CRISPR/Cpf1-Vårt nästa stora verktyg för genmanipulationer?]: Etisk bilaga

Ida Isolehto
Självständigt arbete i biologi 2016

Genmanipuleringar av humana genom

Användningsområdena för CRISPR/Cas-system är breda men den huvudsakliga forskningen riktas mot att ta fram verktyg för genmanipulering i humanceller med syftet att utveckla nya metoder för genterapi. Dagens genterapi går ut på att förse celler med en extra kopia av genetiskt material som kompenserar för den genetiska defekten, men med hjälp av CRISPR/Cas kan man istället permanent korrigera den genetiska defekten. Men är det moraliskt riktigt att göra permanenta förändringar i det mänskliga genomet?

En viktig aspekt för att besvara frågan är att undersöka i vilket syfte detta görs och att väga nyttan mot vilka risker som det innebär. De största riskerna med CRISPR/Cas-baserade metoder är framförallt att låg specificitet kan leda till att förändringen sker på andra än den önskade positionen i genomet. Men om vi kan utveckla verktyg med hög specificitet blir riskerna också ganska små i relation till nyttan. Att permanent reparera en genetisk defekt skulle innebära att patienten slipper upprepade behandlingar, livslånga medicineringar och minskat lidande för individen. Det finns redan idag många behandlingsmetoder där man gör omfattande, permanenta, kroppsliga förändringar såsom till exempel organtransplantationer, och så länge vi talar om behandling av somatiska celler (som ej nedärvs) bör CRISPR/Cas-baserad behandling utvärderas på samma sätt som vilken annan medicinsk behandling som helst. Ett begrepp som ofta florerar i diskussioner kring genmanipulationer är ”naturlighet” och det anses inte ”naturligt” att göra förändringar i genomet. ”Naturlighet” är ett svårdefinierat begrepp som ofta bygger på en subjektiv upplevelse av vad som anses ”naturligt”. Det är ändå intressant att fråga sig hur det kan anses mer ”naturligt” att till exempel förflytta en njure från en människa till en annan än att byta ut enstaka molekyler i en cell. Kanske känns det mindre ”naturligt” eftersom vi inte kan observera förändringen med blotta ögat. Å andra sidan använder vi oss redan idag av mediciner som altererar kemiska strukturer i våra kroppar vars verkan vi inte kan observera mer än i abstrakta termer.

Så skall vi göra detta? Om vi kan utveckla metoder för att permanent behandla genetiska sjukdomar bör det självklara svaret vara ja! Man kan också vända på frågan, om vi *kan* bota dessa sjukdomar, är det moraliskt riktigt att avstå? Nej skulle jag svara, men det går inte att undkomma att det uppkommer följdfrågor såsom vem som har rätt till behandling och på vilka grunder och hur den genetiska informationen skall skyddas vilket behöver diskuteras och regleras.

Forskningsetik

I mitt arbete har jag endast använt mig av recenserade källor för att säkerställa en hög vetenskaplig kvalitet. Jag har gjort mitt yttersta för att citera forskares upptäckter på ett ärligt sätt och ge en så rättvisande bild som möjligt av deras arbeten.

Formaterat: Teckensnitt:Times New Roman

Formaterat: Svenska (Sverige)