



UPPSALA
UNIVERSITET

Unikt immunsystem hos bakterier och arkéer

- clustered regularly interspaced short palindromic repeats

Annica Löfling

Independent Project in Biology
Självständigt arbete i biologi, 15 hp, höstterminen 2010
Institutionen för biologisk grundutbildning, Uppsala universitet

Unikt immunsystem hos bakterier och arkeer

- clustered regularly interspaced short palindromic repeats

Annica Löfling

Självständigt arbete i biologi 2010

Sammandrag

Alla behöver vi skydda oss mot våra fiender. Hos bakterier och arkeer finns en del system som ger dem skydd mot transduktion och konjugering. Virus är den organism på vår jord som finns i högst antal och de har en otrolig anpassningsförmåga samt påverkar andra organismer väldigt lätt. Virus tillsammans med plasmider styr den mikrobiella världens evolution genom att utföra genetisk överföring och styra sina värdars fitness. Konkurrens leder till utveckling, utveckling som vi vill ta vara på eftersom det i många fall påverkar oss i relativt hög grad.

Nyligen upptäcktes ett nytt försvarssystem i bakterier och arkeer mot virus och plasmider. Till en början drogs paralleller med RNAi som är en försvarsmekanism hos eukaryoter, men ju mer man lärde sig om detta nya, desto mer insåg man att det var unikt.

CRISPR - clustered regularly interspaced short palindromic repeats är precis vad det heter. Ett område med regelbundna repeterande sekvenser, åtskilda av sekvenser med varierat ursprung och betydelse. Första gången man såg detta i ett genom var av en slump men det fångade genast intresset hos många forskare. Ganska snart insåg man att de sekvenser som var varierade hade sitt ursprung från framförallt virus och plasmider och att de kunde ge ett skydd till organismen i fråga. CRISPR lokus består utav en ledarsekvens, med en promotorliknande funktion, ett kluster med repeterade sekvenser åtskilda om vartannat av så kallade spacers som är identiska med sekvenser från virus och plasmider. Nedströms (vanligen) om detta kluster finns ett varierat antal *cas*-proteiner (CRISPR associated proteins) med olika funktioner, men som är avgörande för hela CRISPRs funktion. Idag är den molekylära mekanismen bakom funktionen inte helt fastställd, men det finns förståelse för principerna och systemets syfte. Då en prokaryot angrips kan den inkorporera en speciell bit av det främmande DNA:t i CRISPR. Nästa gång organismen angrips av samma virus eller plasmid känner den igen dessa med hjälp av att den kan matcha det inkommande DNA:t med den sekvens som finns inkorporerat i CRISPR (spacer). På detta sätt hinner bakterien eller arkeén skydda sig mot en infektion. Det fina med detta system är att sekvenserna finns kvar så organismen har ett skydd om och om igen, tills viruset lyckas ändra sin sekvens så den inte längre matchar.

Det finns stora förhoppningar om CRISPRs framtida användningsområde. Idag utnyttjar man systemet inom mejeriindustrin, framförallt vid framställning av mjölk och ost. Inom sjukvården hoppas man kunna utnyttja att CRISPR även skyddar mot HGT (horisontell genöverföring), och på så sätt minska antibiotikaresistans som effektivt överförs på detta sätt.

Inledning

Katt och råtta. Jägare och älg. Predator och byte. Ni känner nog igen förhållandet. Det har funnits sedan urålders tider, vän eller fiende. Givetvis är även bakterier och arkéer ständigt utsatta för samma typ av fara; angrepp av virus och fager. Även de behöver kunna skydda sig, precis som vi eukaryoter kan. Till hjälp har vi vårt oerhört utvecklade och välfungerande immunsystem med bland annat antikroppar. Hur går då dessa små men viktiga organismer till väga för att överleva i kampen mot sina fiender? Bakterier har fler varianter på fagresistens; vid adsorption kan de undvika fagen genom att förändra eller helt ta bort receptorerna alternativt någon annan form av barriär mellan fag och bakterie (Hyman & Abedon 2010). Så kallad restriktion sker då bakterien blivit infekterad och den tar död på fagen men överlever själv (inkluderat CRISPR) och vid abortiv resistans dör både fagen och bakterien (Karginov & Hannon 2010, Hyman & Abedon 2010).

På senare år har man studerat ett nytt, oväntat system i dessa små organismer. Ett system som kan skydda bakterier och arkéer mot allt yttre DNA som överförs till dem. CRISPR - clustered regularly interspaced short palindromic repeats, är ett system med hög anpassningsförmåga, en form av minnesfunktion likt vårt immunsystem och med precision. Systemet återfinns i 40 % av alla bakterier och 90 % av alla arkéer. När en population infekteras dör största delen av den ut, dock finns det ibland några överlevande individer, vilket oftast är tillräckligt för att det snart skall uppstå en ny population. CRISPR är inte särskilt effektivt, men vanligt förekommande. De flesta bakterier och arkéer dör fortfarande vid en infektion, men de överlevande har haft ett bra skydd och kan föra sina gener vidare.

HGT - Horisontell genöverföring - är ett vanligt förekommande fenomen bland bakterier och arkéer. Gener överförs nästintill obegränsat, på gott och ont. Principerna är enkla; upptag av fritt DNA (transformation), överföring av gener in i cellen via plasmider (konjugering) eller virus (transduktion) är de olika varianter som finns. Överföring av gener orsakar problem för bakterier av den enkla anledningen att de kan bli infekterade och dö. För att undvika detta använder de sig utav de tidigare nämnda resistensmetoderna. För oss människor finns det dock ett större problem med just HGT, i och med att det är så effektivt, sker ofta och enkelt, så kan överföring av gener orsaka stora problem. Ett välkänt exempel på detta är antibiotikaresistans, gener som orsakar resistans mot olika antibiotika överförs flitigt och på så vis blir resistansen större och vi får allt svårare att bekämpa sjukdomar. Tänk om man kunde utnyttja bakteriens eget försvar mot genöverföring och så minska spridningen av resistensgener?

Idag har man bitvis en relativt god bild av hur systemet fungerar, framförallt i vissa organismer. Det var i *Streptococcus thermophilus* som man för första gången kunde ge en förklaring till vad systemet gjorde och hur det såg ut och sedan dess finner man varianter på systemet i de organismer man funnit CRISPR, men principen är densamma.

Varför skiljer det sig då så mycket mellan bakterier och arkéer på hur ofta man ser detta system? Vad har bakteriofager och plasmider för skydd emot detta? Hur fungerar det egentligen? Kan vi utnyttja detta system till någonting som vi människor kan ha nytta utav? Dessa frågor samt kringliggande detaljer kommer att tas upp och diskuteras i kommande sidor.

Upptäckten av CRISPR

Det var av en ren slump som forskare först upptäckte CRISPR. En grupp forskare, Ishino mfl, höll 1987 på med kloning av en gen i *Escherichia coli*. Omedelbart nedströms om den gen de intresserade sig för upptäcktes en klunga repeterade gener blandat med icke repetitiva gener av samma längd. De sekvenserade dessa gener och detta blev första gången som CRISPR observerades. Efter detta fann fler grupper liknande genuppsättningar i andra bakterier och arkéer

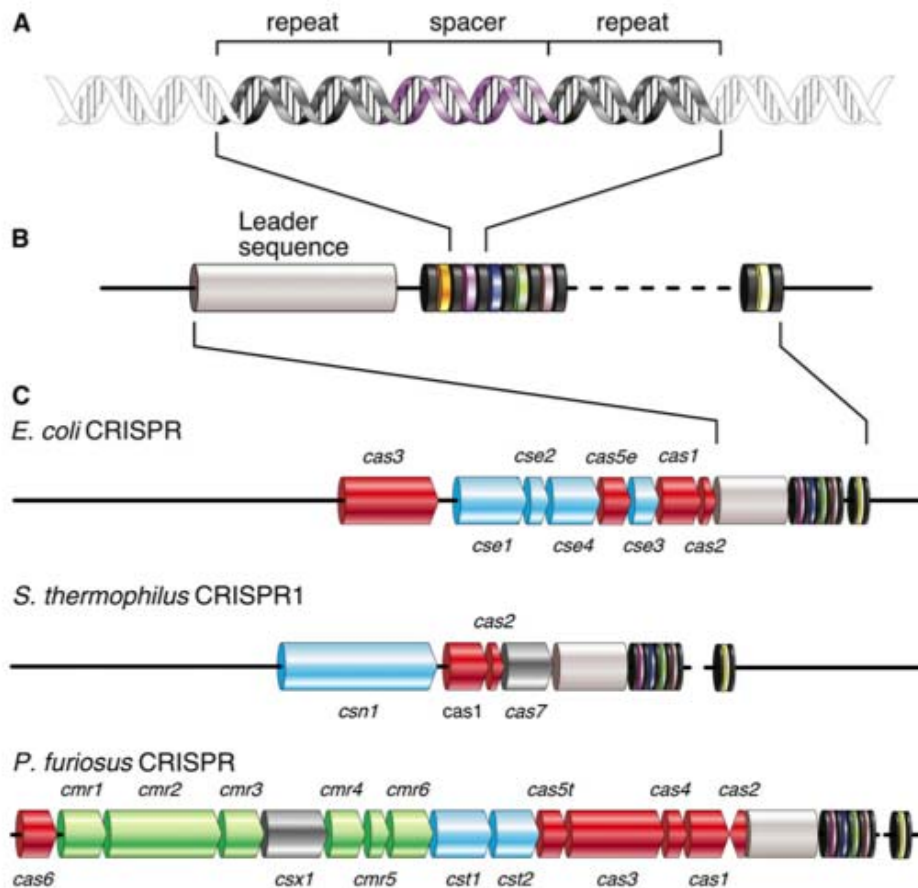
och snart kändes CRISPR igen som en familj repeterande gener med okänd funktion (Horvath & Barrangou 2010).

År 2002 myntades namnet CRISPR utifrån strukturen på hela systemet, som typiskt ser ut som i figur 1. Några år senare stod det klart att CRISPR verkade ha en immunitetsfunktion med en resistansbildande anpassning. Detta visades i *S. thermophilus* som kom att bli den första modellorganismen för detta system. Samma forskargrupp (Ishino m.fl) visade även att sekvenserna mellan repetitiva sekvenser, så kallade spacers, var identiska med fag-DNA (s.k. protospacer) och dessa spacers var nödvändiga för att kunna motarbeta fagerna (Horvath & Barrangou 2010). Även *cas*-genernas huvudroll påvisades. Nyligen bekräftades liknande funktioner för konjugering med plasmider (Horvath & Barrangou 2010).

Idag vet man syftet med CRISPR, vilka komponenter som är delaktiga - dock ej dess specifika funktion - och man har kunnat lägga upp hypoteser om vad CRISPR skulle kunna användas till. Man har hittat systemet i ca 40 % av alla bakterier och nära 90 % av alla arkeer. Varför det bara är hälften så vanligt i bakterier som i arkeer finns det ingen riktig förklaring till. Men man tror att det kan ha att göra med att de flesta bakteriestammar man studerar är laboriestammar. Dessa har inte levt i sin naturliga miljö och alltså inte blivit utsatta för virus eller fager lika frekvent som i naturen och CRISPR kan därmed ha selekterats bort hos dessa stammar (Marraffini & Sontheimer 2010a).

CRISPR loci anatomi

CRISPR är precis vad förkortningen säger; Clustered regularly interspaced short palindromic repeats. Korta sekvenser som repeteras i varierat antal och storlek. Antalet kan vara alltifrån ett till hundratals (Marraffini & Sontheimer 2010a) medan storleken varierar mellan 23-50 nukleotider (Marraffini & Sontheimer 2010a). Sekvenserna är palindromiska, vilket innebär att de läses lika från båda hållen- jämför ANNA. De repeterade sekvenserna skiljs åt av icke repetitiva sekvenser med varierat ursprung med ungefär samma storlek, så kallade spacers (figur 1). Spacers kommer huvudsakligen från plasmider eller bakteriofager och utgör målet för ingripande av CRISPR. Intill detta område med repetitiva och icke repetitiva delar finns en leadersekvens, normalt sett uppströms (3'-änden i en DNA molekyl) om den repeterande klungan, CRISPR. Leadersekvensen är väl bevarad mellan olika loci, till skillnad från andra gener i CRISPR systemet. Den består av ett 100-tal baspar, saknar läsram, har en fixerad riktning och är alltid lokaliserad på en sida om CRISPR. Vad som dock är intressant är att leadersekvensen har en promotorliknande funktion, vilket skulle förklara dess placering uppströms om CRISPR (Marraffini & Sontheimer 2010a). Nedströms finns ett antal proteiner, CRISPR associated proteins, så kallade *cas*-proteiner (figur 1). Här ser man en stor variation mellan olika CRISPR system i olika bakterier och arkeer. Vissa av proteinerna har bevarats genom de flesta systemen, medan andra bara verkar finnas hos vissa organismer. Dock fyller de olika proteinerna i princip samma funktion hos alla bakterier och arkeer även om det är många proteiner vars funktion fortfarande är okänd. Dessa proteiner är ansvariga för CRISPR aktiviteten och trots att det saknas detaljerad kunskap om dem så är det säkert att de fyller en avgörande funktion i systemet.



Figur 1. CRISPR loci anatomi. **A,B:** Leadersekvens uppströms om repetitiva sekvenser (svart) och spacers (färg). **C:** Cas-proteinerna varierar mellan arter, men de erhåller liknande funktioner hos alla organismer. Huvudcasproteiner är röda, subtypspecifika proteiner blåa, RAMP moduler gröna och oklassificerade proteiner är gråa. (Karginov & Hannon 2010) med tillstånd från upphovsrättsinnehavaren.

Sekundärstruktur

De repetitiva sekvenserna är som sagt palindromiska vilket ger en antydning om att dessa sekvenser kan bilda sekundära strukturer. Detta stärks ytterligare utav att de repeterade sekvenserna under transkriptionen dessutom befinner sig i en RNA intermediär, och som vi vet kan RNA lätt bilda sekundära strukturer. Kunin och hans forskargrupp testade detta 2006 genom att mäta stabiliteten i olika strukturer från CRISPR loci och de sekundära strukturerna man funnit visade sig vara väldigt stabila. De utförde också andra undersökningar med bland annat mutationer som bevisade att den sekundära strukturen var både stabil och viktig för CRISPR funktionen. (Kunin m.fl 2006)

cas-proteinerna

Alla CRISPR system innefattar även de så kallade cas-proteinerna, förutom ett undantag hos *Thermoplasma acidophilum* (Marraffini & Sontheimer 2010a). De utgör tillsammans ett 40-tal genfamiljer, inkluderat både huvudfamiljer och vissa subtyper av cas-proteiner. Alla cas-proteiner varierar i riktning och ordning, dock kan man se att det är mindre skillnader mellan närmare besläktade arter samt inom arter. Huvudcas-genfamiljerna utgörs av Cas 1-4 (Makarova m.fl 2007) där framförallt Cas 1 och 2 återfinns i nästan alla CRISPR loci. Senare forskning har också kommit med nya huvudcas-gener, cas 5 och 6, dessa har dock inte blivit helt accepterade av alla inom området. Ett problem har varit namngivningen då olika forskargrupper jobbat med just dessa proteiner och gett dem lite olika namn fast man talar om samma protein, vilket har lett till viss förvirring (Karginov & Hannon 2010). Det kommer i första hand vara Hales namngivning som

gäller då det talas om *cas*-proteiner i denna uppsats, vid eventuell förvirring finns tabell 1 att gå tillbaka till.

Förutom dessa större proteiner som återfinns på flertalet ställen finns dessutom alla de minst lika viktiga proteiner men som saknar någon form av homologi till varandra, och som bara hittats i något enstaka CRISPR system. Här är det mer vanligt förekommande att man döpt dem efter vilken organism de är relaterade till, till exempel för en *cas* subtyp i *E.coli* blir är namnet Cse 1 (Karginov & Hannon 2010). Det har även visat sig finnas en koppling mellan *cas*-subtyperna och CRISPR repeatsen. Man hittar ofta klungor av CRISPR repeats tillsammans med vissa specifika *cas*-subtyper, och dessa olika genuppsättningar verkar vara viktiga i systemet för att *cas*-proteiner skall kunna binda till repeaterna och förbereda dessa områden i loci i förberedelsen för nya spacers (Kunin m.fl 2006). De proteiner som kodas utav *cas*-generna kan vara RNA-bindande proteiner, endo- och exonukleaser, helikaser och polymeraser.

RAMPs

RAMPs - Repeat-associated mysterious proteins. Ytterligare en komponent I CRISPR systemet. Denna proteinfamilj karaktäriseras utav en G-rik ögla (Guaninrik loop på DNA kedjan) på c-terminal änden (COOH-ände på peptidkedjan) (Makarova m.fl 2006). Återfinns i alla genom där CRISPR finns och kan väl där befinna sig både i närheten av CRISPR eller en bit ifrån (Karginov & Hannon 2010). RAMPs bildar ofta stabila komplex med ett polymeras plus några okända gener i olika CRISPR system, så kallade pol-cassettes (av vissa även nämnt som RAMP-modul) (Makarova m.fl 2006). Pol-cassettes påträffas med en något högre frekvens bland termofila organismer (Makarova m.fl 2006). I tabell 1 finns en överskådlig bild över alla proteiner delaktiga i CRISPR aktiviteten.

Tabell 1. Proteiner i CRISPR systemet med känd funktion. (Karginov & Hannon 2010) med tillstånd från upphovsrättsinnehavaren

Name (Haft et al., 2005)	Alternative Name(s)	Function/Activity	Model System
cas1		Acquisition of new spacers/ss/dsRNA endonuclease	<i>P. aeruginosa</i>
cas2		ssRNA endonuclease	<i>S. solfataricus</i> and others
cas3		crRNA-guided degradation of invading NAs	<i>E. coli</i>
cas6		Endonucleolytic cleavage of pre-crRNA	<i>P. furiosus</i>
csn1	cas5	Phage resistance using existing spacers	<i>S. thermophilus</i>
–	cas7	Acquisition of new spacers?	<i>S. thermophilus</i>
–	SSO0454	Specific binding of CRISPR repeat DNA	<i>S. solfataricus</i>
Cascade Complex		Endonucleolytic Cleavage of Pre-crRNA	<i>E. coli</i>
cse1	casA		
cse2	casB		
cse3	casE	Catalytic subunit	
cse4	casC		
cas5e	casD		
RAMP Module Complex		crRNA-Guided Endonucleolytic Cleavage of RNA Targets	<i>P. furiosus</i>
cmr1			
cmr2			
cmr3			
cmr4			
cmr5		Dispensable for activity	
cmr6			

Mekanism och biologisk funktion

Vad som verkligen ansågs som mest spektakulärt när man först hittade CRISPR under en genomsekvensering var det varierande DNA:t som fanns emellan de upprepade sekvenserna. Att dessa dessutom visade sig ha sitt ursprung till stor del från bakteriofager och plasmider gav forskarna en modell av att CRISPR var ett slags försvarssystem gentemot inkommande DNA. Vad som krävs för denna modell är att den nya informationen kan komma in i CRISPR, vilket sålunda också skulle betyda att dessa spacers borde säga oss något om organismens forna livsmiljö, vilka bakteriofager och plasmider den stött på och så vidare.

Valet av DNA som inkorporeras, de så kallade spacers, är inte slumpmässigt. På den del av det främmande DNA:t som tränger in i cellen finns en markör, PAM (protospacer adjacent motif). Denna lilla sekvens ligger strax utanför den egentliga viktiga sekvensen och blir Cas 1 proteinets mål när den skall klyva det främmande DNA:t, figur 2. Antalet spacers som kan inkorporeras är inte

obegränsat, och studerar man olika CRISPR loci ser man en stor polymorfism. Framförallt syns skillnader i den främre delen närmast leadern, vilket också understryker att spacers förmodligen fylls på just därifrån. Tittar man på det äldre materialet i CRISPR är det mindre skillnader mellan organismer där, då detta material är mer konserverat och ofta inaktuellt i den bemärkelsen att sekvensen inte längre matchar något virus eller plasmid.

Mekanismen delas vanligen in i tre steg; integrering av nya spacers, uttryck och bildandet av crRNAs och själva ingripandet av CRISPR. Idag vet man mest om mellansteget, just hur crRNA (CRISPR RNA) bildas, hur dessa små sekvenser används för att guida under ett kommande ingripande. Man tror sig vara ganska säker på vilka proteiner som är inblandade också, bland annat ett komplex av cas-gener, kallat cascade (Horvath m.fl 2008). Inkorporeringen av nytt DNA, nya spacers, sker troligtvis med hjälp av Cas 1 som klyver det invaderade DNAt och sätter in det närmast leadern. Under ingripandet sker en igenkänning av det främmande DNA och infektionen kan på något sätt blockeras. DNA har visat sig vara måltavlan, men det finns även en organism med RNA som måltavla.

Inkorporering av nya spacers

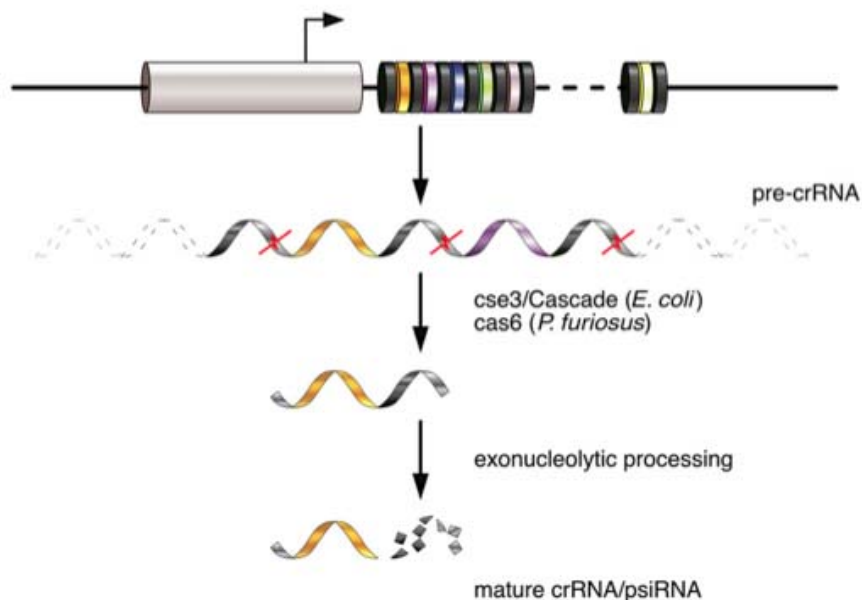
Främmande DNA kommer in i bakterien via en bakteriofag (transduktion). Cas 1 (*E.coli*) känner igen det främmande DNAt och klyver detta som därefter kan inkorporeras mellan repeats vid leader änden. Den molekylära mekanismen bakom detta är ännu inte känd. Det har visat sig att inkorporering av nya spacers alltid är koncentrerad till den övre delen av CRISPR loci (Barrangou m.fl 2007). CRISPR uppvisar en hög polymorfism, där en högre aktivitet i loci bidrar till ett högre antal nya spacers men också fler bortfall. Detta leder till en stor diversitet mellan olika arter men också mellan individer (Horvath m.fl 2008). En högre diversitet, ett större antal varierade spacers, ger en högre resistans hos organismen (Barrangou m.fl 2007). Storleken på en spacer varierar mellan 26-32 nt. Vanligast är att endast en spacer åt gången sätts in, men man har även sett att upp till fyra stycken har lagts till på samma gång (Marraffini & Sontheimer 2010a). Den nyligen tillagda sekvensen måste vara identisk med en region på den angripande fagens DNA, så kallad protospacer (Deveau m.fl 2008). Valet av sekvenser som inkorporeras är inte slumpmässigt. Denna sekvens känns troligen igen utav en annan kort sekvens, protospacer adjacent motif, PAM (figur 3b). Denna korta sekvens, bestående av mindre än 10 baser varierar mellan arter, dels var den är placerad i förhållande till protospacern och vilka baser den innehåller. Förutom att vara viktig för inkorporering är sekvensen viktig för virusförsvar mot CRISPR, då en enda mutation i denna sekvens gör det omöjligt för bakterien eller arkéen att känna igen det inkommande DNAt (Karginov & Hannon 2010).

Genom att "förvara" information om sina angripare kan CRISPR ingripa väldigt fort när faran väl är framme. Se figur 3 a. Det finns tre huvudsakliga teorier om hur inkorporeringen går till, den första är att CRISPR fångar upp främmande DNA när det kommer in i cellen och skapar en immunitet mot viruset och sedan klarar sig från en infektion. Problemet med detta är att virus är specialister på att snabbt döda och så fort de kommer in i en cell stänger de av nödvändiga mekanismer, framförallt de väldigt aggressiva. Om nu virus dödar cellen så fort, hur skall den hinna bli immun under denna tid? En alternativ hypotes är att det finns en annan mekanism som dödar eller åtminstone förhindrar viruset från att angripa cellen och CRISPR hinner då inkorporera DNA och utveckla en immunitet. Ett tredje alternativ är att CRISPR tar upp DNA från icke-funktionsdugliga virus, många virus har mutationer av olika slag som försämrar deras funktion, och på så sätt kan bakterien ta upp DNA från viruset i lugn och ro och vara immun vid en annan attack. Detta skulle vara som när vi vaccinerar oss, då vi får en dos av ett skadat virus, men som ändå gör att vårt immunsystem känner igen det nästa gång vi stöter på det, så kallat immunologiskt minne.

Bearbetning av crRNA

När Hela CRISPR loci inkluderat repeats och spacers transkriberas och vi får en pre-crRNA. Denna klyvs sedan i bitar, så kallade crRNA, av cascade-komplexet med endonukleär funktion (klyver DNA). Pre-crRNA klyvs på olika ställen, men vanligt är att repeatsen i sekvensen bildar sekundära strukturer och att klyvningen sker i närheten av dessa. De kan också försvåra för enzymer att hitta rätt klyvningsställe. Cascade-komplexet består av olika Cas-proteiner. Dessa crRNA (CRISPR RNA) används sedan vid igenkänning av något inkommande DNA, som en slags guide för hela försvaret. Om det främmande DNA:t matchar kan det därefter angripas. (Brouns m.fl 2008)

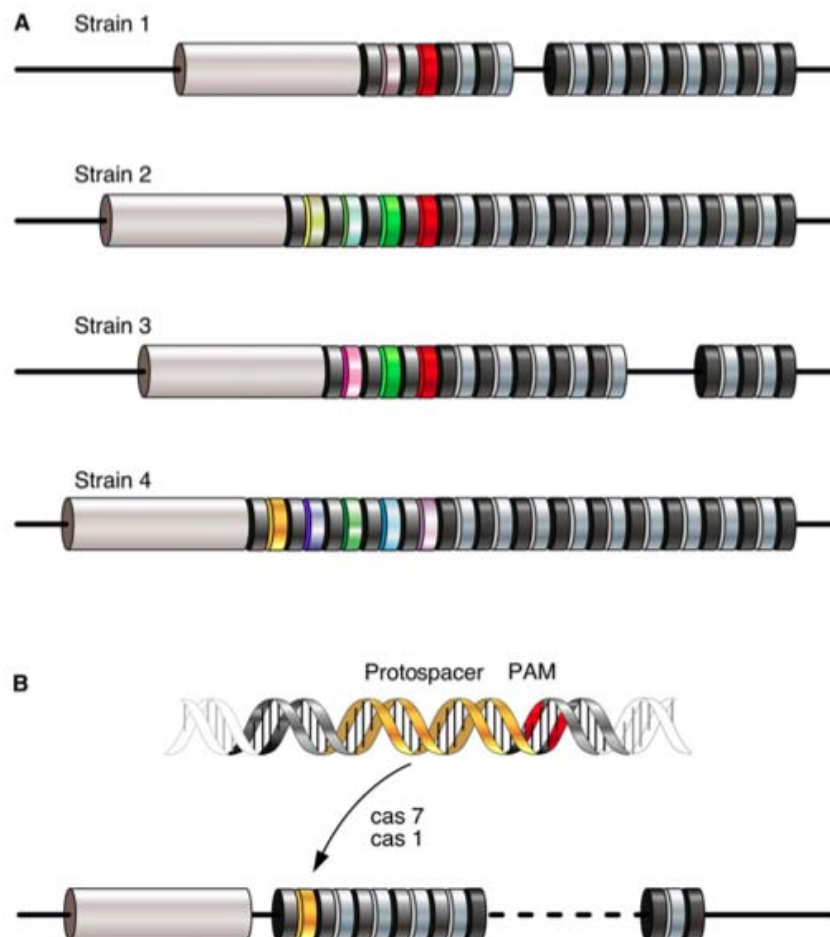
I *E.coli* har forskare identifierat åtta stycken gener som kodar för viktiga proteiner i systemet (Brouns m.fl 2008). Dessa proteiner är; Cas 3 som troligtvis är en kombination av ett nukleas - bryter ned nukleinsyror och ett helikas - utför negativ supercoiling, Cas 1 som troligen är ett integras - rekombinering mellan två specifika sekvenser i DNA (Snyder & Champness 2007). Cas 2, ett endonukleas - enzym som bryter fosfordiesterbindningar i nukleinsyror (Björn L.O m.fl 2009). Ytterligare fem stycken; CasABCDE som tillsammans bildar ett cascade-komplex (CRISPR-associated complex for antiviral defense). Försök som gjordes på *E.coli* visade tydligt att Cascade komplexet var ansvarig för klyvningen utav pre-crRNA till små crRNA. De visade också att CasE är det enda protein som är helt nödvändigt för klyvningen, utan detta sker ingen bearbetning av crRNA. CasE hör till den största och mest spridda proteinfamiljen inom CRISPR - RAMP (Brouns m.fl 2008). Forskargruppen kunde också hitta exakta klyvningsställen för crRNA, dessa visade sig vara exakt två nt från steem-loopopen uppströms. Denna sekundära struktur kan orsaka problem för casE, men för att undvika detta tas ytterligare någon nukleotid bort så sekvensen inte kan återgå i en sekundär struktur (Brouns m.fl 2008). En crRNA med sekundär struktur kan ses i figur 4.



Figur 2. Bearbetning av crRNA. Transkription av CRISPR området till pre-crRNA som därefter klyvs av cas-proteiner till små crRNA. (Karginov & Hannon 2010) med tillstånd från upphovsrättsinnehavaren.

CRISPR ingripande

Det främmande DNA:t känns igen och på något sätt blockeras en fortsatt infektion. Man misstänker att leadern kan vara inblandad, eftersom spacers inkorporeras just efter leadersekvensen så bör den ha någon med det att göra (Karginov & Hannon 2010). Att DNA är måltavlan har man flera starka bevis för, även om man inte har kunnat utesluta RNA som måltavla (Marraffini & Sontheimer 2008, Hale m.fl 2009).



Figur 3. A. Inkorporering av spacers är koncentrerat åt ett håll. Detta syns i flera arter. **B.** Spacers inkorporeras från en protospacer som känns igen med en kort sekvens, PAM. (Karginov & Hannon 2010) med tillstånd från upphovsrättsinnehavaren.

Spacers - bakteriens historia

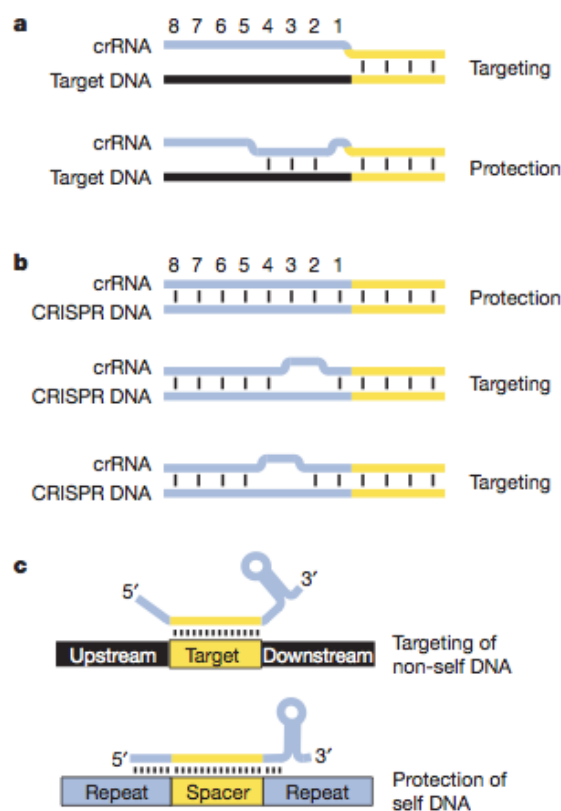
Nya spacers inkorporeras, gamla tas bort. I CRISPR finns en historia, en inblick i prokaryotens miljö. Genom att studera spacers i en organism kan man få reda på vilka virus och plasmider den råkat ut för och utifrån detta få reda på hur bakterien levde, i vilken typ av miljö (Horvath m.fl 2008). Men också tvärtom, man kan få reda på mer om viruset i fråga genom att studera den infekterade bakterien. Det handlar mycket om symbios, mellan virus och olika arter av bakterier samt arkéer.

Autoimmunitet

Marraffini och Sontheimer kunde 2010 bekräfta genom sina försök på *Staphylococcus epidermis* hur CRISPR undviker att angripa sig själv. De började med att testa om CRISPR överhuvudtaget verkade kunna angripa sig själv eller om det var en omöjlighet. De klonade in tre kända spacers

(matchar aktiva fager) från vildtyp av *S.epidermis* in i en plasmid, vilket innebär att plasmiden har tre märkta DNA i sig som vi vet bör matcha i en cell. Konjugering mellan plasmiden och vildtypcell samt en cell med muterat CRISPR ledde inte till någon skillnad i effektivitet (till skillnad från en vanlig *nes*-protospacer plasmid). Betyder detta att dessa targets undankom CRISPR ingripande?

Ett annat försök som byggde på att man visste att en del av repeaterna följde med spacers vid märkning av DNA visade att repeaterna utgör CRISPRs skydd mot sig själv. Det är den lilla delen av den repeat uppströms om en spacer som är avgörande. En crRNA består utav spacers samt en del repeat. När denna basparar (basparning enligt Watson och Crick) mot ett främmande DNA kommer enbart spacer sekvensen att matcha och CRISPR systemet kommer att ingripa. Skulle dock denna lilla sekvens, på 5' änden av crRNA ungefär -4 till -2 nukleotider uppströms, skulle den baspara med DNA sekvensen skulle ingenting hända (Se figur 4). Systemet är på så sätt skyddat mot autoimmunitet. Detta autoimmunitetsskydd utmanas enkelt genom att felpara denna sekvens, och inte mer än två felparningar klarar det förrän skyddet av CRISPR loci upphävs. (Marraffini & Sontheimer 2010 b)



Figur 4. Hur CRISPR undviker att angripa sig själv har att göra med hur crRNA och inkommande DNA matchar. (Marraffini A, Sontheimer Erik J. b. 2010) med tillstånd från upphovsrättsinnehavaren (Reprinted by permission from Macmillan Publishers Ltd: Nature (Marraffini A, Sontheimer Erik J. 2010b), copyright 2010)

DNA som måltavla

Bevisen för att DNA är måltavlan under CRISPR ingripande är starka. Det försök som bäst visar på detta gjordes på *S.epidermis* av Marraffini och Sontheimer 2010. I *S.epidermis* finns *spc1*, en spacer som matchar *nickase*-genen, denna förekommer i de allra flesta plasmider som angriper *S.epidermis*. Det första tecknet på att DNA är måltavlan är orienteringen på *spc1*, vilket tydligt indikerar på att den är identisk med *Nickase* och inte komplementär som den varit om det vart RNA som var målet. För att testa detta närmare förde man in en self-splicing intron i *nes* på en plasmid (Marraffini & Sontheimer 2008). En självklyvande intron är en parasitiskt DNA element som inte stör värden, eftersom den själv vinner på att ha en levande värd. Den har en förmåga att klippa ut

sig själv ur ett mRNA före translation, detta för att rädda värden om dess närvaro stört någon gen (Snyder & Champness 2007). Vad man såg var att konjugering skedde med samma effektivitet som tidigare, vilket antyder att ett skadat mRNA inte spelar någon roll, huvudsaken är att DNAt är helt för att crRNA skall kunna känna igen sekvensen. Ytterligare ett test gjordes för att försäkra sig om att detta var fallet, då sattes en skadad intron in istället, en intron som förlorat sin förmåga att klippa ut sig själv. Denna gång följde skadan som intronen orsakar på mRNA med till DNA eftersom det hela translateras, och ett felaktigt DNA ger låg aktivitet hos *nickase*-genen (Marraffini & Sontheimer 2008). Slutsatsen man kan dra från detta är att DNA krävs för att spacers skall matcha och CRISPR aktiveras. Försöken utesluter inte att RNA kan finnas som måltavla, men talar för DNA, iallafall i *S.epidermis*.

Nyligen publicerades en ny artikel i Nature med starka bevis på att DNA är måltavlan i *S.thermophilus* (Garneau m.fl 2010). De gjorde *in vivo* tester som visade att *cas 5* är ett nödvändigt protein för CRISPRs ingripande, att DNA är måltavlan både för plasmider och bakteriofager. Dessa har samma klyvningsställe, 3 baser från protospacer startpunkten. För klyvning krävs protospacer specificitet samt korrekt orientering (Garneau m.fl 2010). Det är mycket möjligt att klyvning av DNA sker i de flesta CRISPR system, men det krävs mer forskning kring detta innan man kan fastställa det (Marraffini & Sontheimer 2010c).

RNA som måltavla

DNA verkar vara måltavlan i de flesta organismer med det finns belegg för att även RNA är måltavlan i vissa fall. Det är i *Pyrococcus furiosus* som forskare har kunnat visa detta. Hale och hans kollegor visade 2009 att samma små RNA sekvenser, crRNA, som normalt guidar CRISPR försvaret har liknande funktion i *P.furiosus* dock med den skillnaden att de har inkommande RNA som mål och det är denna som klyvs (Hale m.fl 2009).

P.furiosus är en hypertermofil arkée vars genom innehåller 200 crRNA uppdelat i 7 olika CRISPR loci och ungefär 29 *cas*-proteiner med varierad funktion. Den undersökning Hale och hans kollegor gjorde testade om crRNA i *P.furiosus* klyver komplementära RNA (Hale m.fl 2009). Hela CRISPR sekvensen transkriberas, och RNAt klyvs sedan i mindre RNA sekvenser som guidar CRISPR-casförsvaret baserat på att de basparar med inkommande nukleinsyror. I *P.furiosus* har man hittat två olika crRNAs; en med 39 nt längd och en sekvens som visat sig vara något vanligare med en längd på 45 nt. De har samma markör på 5'-ändan men skiljer sig åt på 3'-ändan, vilket leder till att de RNAt kommer att klyvas på olika ställen. Ett antal *cas*-proteiner och RAMP-moduler är involverade i klyvningsprocessen, crRNAs uppgift är att leda proteinerna till rätt RNA och hitta rätt position på denna. Skillnaden mellan att RNA är måltavlan jämfört med DNA som måltavla är att crRNA söker efter en komplementär sekvens medan crRNA som söker DNA söker efter en identisk sekvens.

Försvar mot CRISPR

Virus och plasmider har snabbt utvecklat försvar mot detta immunsystem. Den enorma specificiteten hos CRISPR underlättar för dessa att hitta sätt att undvika att infektionen hindras av bakterien eller arkéen själv. En enda singelmutation, borttagning eller insättning i protospacer sekvensen är nog för att crRNA inte skall känna igen denna och ge ett gott försvar (Deveau m.fl 2008).

Evolution av bakteriofager

Virus styr det mikrobiella samhälle de lever i genom att begränsa deras värdars fitness och genom att utföra genetiskt utbyte (Anderson & Banfield 2008). Virus är den mikrob som det finns störst antal av i hela biosfären och bakterier och arkéer utvecklar ständigt nya försvar för att kunna skydda sig mot dessa erövrare. Samtidigt som prokaryoterna hittar sätt att komma undan virus och

förhindra dem på olika sätt utvecklar virus i sin tur system att komma runt sina värdars försvar. På detta sätt sker en ständig evolution av både värden och viruset (Marraffini & Sontheimer 2010a).

I den naturliga miljön ser det ut som om genetisk överföring är det favoriserade sättet för virus att undkomma CRISPR. Jämfört med mutationer i protospacer eller PAM, som visserligen kan vara enklare att få till, så minskar risken för att övriga proteiner att skadas. Dessutom rensas inte ett helt genom ut, utan enstaka sekvenser (motiv) som känns igen av spacers hos viruset förändras eller tas bort. Detta bibehåller virusets diversitet men gör det samtidigt omöjligt för bakterien eller arkeén att skydda sig (Anderson & Banfield 2008). Anderson och Banfield undersökte detta närmare med hjälp av metagenomiska data från olika viruspopulationer plockade från två olika biofilmer. Bakterierna sekvenserades och man hittade ett antal CRISPR med repeats som man undersökte närmare. Från 6044 stycken spacersekvenser fann man 2348 stycken unika sekvenser. Matchning mot flera databaser visade att 40% av dessa spacers hade ursprung från virus, plus en del plasmider och transposoner. Varje CRISPR system som hittas tillhör sin unika cell, vilket innebär att varje cell har skydd mot flera viruspopulationer (Anderson & Banfield 2008). Inom en population är det också en stor variation på nukleotidnivå, under försöket hittades inte två identiska spacers på mer än 25 nt! Allt detta innebär att det finns en enorm diversitet hos dessa organismer, mycket med hjälp utav homolog rekombination (Anderson & Banfield 2008). Rekombination är ett väldigt viktigt fenomen för evolution av den här typen av organismer, både virus, bakterier och arkéer. Utan det hade kanske inte mycket skett, men nu på grund utav det försvårar det för oss att kunna utnyttja, undersöka och skydda oss mot dessa små organismer. Det är som sagt inte bara för oss det blir mer komplicerat, för dem själva också som måste utvecklas ständigt i snabb takt för att överleva i den miljön de lever i.

Danisco och den första upptäckten

CRISPR används faktiskt redan idag vid produktion av mejeriprodukter. Det är företaget Danisco som använder detta vid framställning av sina produkter. Danisco är världsledande företag inom ingredienser i mat, enzymer och bio-baserade lösningar. ”Genom att använda naturens eget material, forskning och den kunskap vi besitter kan vi designa och leverera bio-baserade ingredienser för nyttigare och tryggare produkter.” Så uttrycker Danisco själva sitt arbete, och tar man ett titt på deras hemsida, bläddrar förbi historia, finns år 2007 med följande rubrik; ”Scientific breakthrough.” Detta var alltså året då mekanismen CRISPR blev lite mer förstådd och kunde användas i praktiken. En utav de främsta forskarna inom området kommer från Danisco, vilket också är anledningen till att just de använder sig av denna metod så tidigt.

Den rapport Barrangou och hans medarbetare lämnade 2007 var den första att visa CRISPR ingripande, inkorporering av nya spacers i svar mot angripande virus och att det också är nödvändiga *cas*-gener inblandade (Marraffini & Sontheimer 2010a). Försöket utfördes på en inom mejeriindustrin väl använd vildtyp av *S.thermophilus*. *S.thermophilus* är en av de viktigaste mjölksyrebakterierna. Det är en G+C fattig grampositiv bakterie lämplig för storskalig mjölkfermentering eftersom varje stam har egenskaper som tillämpar sig för olika typer av jäsning (Deveau m.fl 2008).

CRISPR loci i *S.thermophilus* undersöktes, och man såg vissa skillnader i typ av spacers och storleken på de sekvenserna. Vilka spacers som finns i bakterien är starkt korrelerat till känslighet mot en viss fag. Vildtypen utsattes för några, i mejeriindustrin, vanliga fager. Vad det visade var givetvis att de bakterier som överlevde hade inkorporerat en ny sekvens, spacer, med motsvarighet från fagen. Nästa steg blev att undersöka vad som hände då denna sekvens som matchade en fag så väl, om den togs bort blev då bakterien känslig igen? Visst blev den det. Samma sak även vid ett motsatt försök. Det var nu solklart att en kopia av fagens DNA i bakteriens CRISPR gav ett försvar mot denna. Vad man även gjorde experiment på var *cas*-gener, genom att ta bort några utav dem

kunde man också påverka sensitiviteten mot fager. Detta gällde dock inte vilken *cas*-gen som helst, vilket antyder att de har olika roll i CRISPR systemet. (Barrangou m.fl 2007)

Inte lika vanligt i bakterier?

Varför har man bara funnit CRISPR i 40 % av alla bakterier, men i hela 90 % av alla arkéer? Frågan ställdes redan tidigt i arbetet, men har under skrivandets gång inte getts något riktigt svar. De teorier som kommit upp senare har kommit från Magnus Lundgren som var på CRISPR konferens i oktober månad i år. Frågan om varför det ser ut såhär hade tagits upp och svaret är att man egentligen inte vet. Det man kan undra är om siffrorna faktiskt stämmer. Arkéer är kända och intressanta för att de lever i extrema miljöer, men de finns givetvis i alla slags miljöer. Kanske har man plockat arkéer från de extrema miljöerna, eftersom det är dessa som man känner mest till och har ett intresse för, men kan detta ha lett till en favorisering som ger oss ett felaktigt värde? Har vi omedvetet valt arkéer som har ett CRISPR system? Paris-sud 11 universitet i Paris har gjort en databas som uppdateras regelbundet när något nytt genom sekvenseras för att hitta CRISPR. Den 11:e november i år hade 88 stycken arkéer undersökts, varav 75 hade CRISPR lokus. 1093 bakterier hade undersökts, men endast i 502 av dessa hittade man ett CRISPR loci. Bakom dessa siffror finns en del mörkertal, där man inte kan säga om det finns eller inte. Men på något sätt talar siffrorna för sig själva, även om det finns en förklaring till varför det skiljer sig så mycket mellan arkéer och bakterier som kan minska skillnaden, så är siffrorna så pass åtskilda att det måste vara så att det är vanligare i arkéer.

En möjlig förklaring som en av arkéeforskarna lade upp var att lytiska virus inte är lika vanligt hos arkéer som för bakterier. De är alltså inte utsatta för lika aggressiva virus och denna skillnad kan generera fler CRISPR, eftersom det kan vara enklare för arkéer att använda systemet. Sen finns det också ett evolutionärt perspektiv, som alltid. Är det lönsamt att ha ett immunsystem som CRISPR? Fördelen är självklart att organismen blir immun mot virus, men det finns också risken att det går snett och cellen dödar sig själv. Dessutom är det kostsamt och så vidare.

Som sagt, oavsett vad som ligger bakom skillnaden får vi nog acceptera att det är så det ser ut, och just nu vet vi inte varför.

CRISPR och framtiden

Om framtiden vet man inte mycket ännu. Visst har man sina drömmar på flera håll, men än så länge är det bara drömmar och hypoteser som kräver mer forskning och kunskap. Idag används CRISPR redan inom mejeribranschen men det finns andra applikationer. Genom att manipulera CRISPR skulle det kunna användas mot spridning av antibiotikaresistans på sjukhus eftersom det har visat sig kunna hindra konjugering. CRISPR är inte begränsat till försvar mot virus, försök har tydligt visat att CRISPR är effektivt mot HGT (Horizontal gene transfer). Marraffini och Sontheimer (2008) utförde försök på *Staphylococcus epidermis*. Inom sjukvården är HGT ett stort problem med bland annat *S.epidermis* eftersom bakterien snabbt sprider MRSA och VRSA (methicillin- och vancomycin-resistans) via HGT. *S.epidermis* har visat sig ha en spacer (*spc 1*) som är homolog med *nickase*-genen (*nes*) som är en gen som finns i alla konjugerande plasmider. För att undersöka om CRISPR kunde förhindra HGT muterades *nes* och frekvensen överföring kontrollerades. Ytterligare försök där repeats och spacers hos *S.epidermis* togs bort gav samma slutsats; att CRISPR förhindrar HGT genom att matcha sekvenser från plasmiden. Försöken visade också att leader sekvensen är nödvändig för att CRISPR skall kunna fungera och ge ett skydd samt att DNA är måltavlan även här (Marraffini & Sontheimer 2008). Lyckas man manipulera CRISPR försvaret kan man bekämpa den otroligt svåra spridningen av antibiotikaresistans på sjukhus, men det är inte lösningen på problemet. Vi måste komma ihåg att det är en nackdel för cellen att inte vara antibiotikaresistent, och i en miljö med antibiotika kommer den att dö vilket gör att de celler som är resistent lätt kan ta

över. Mutationer uppstår också lätt vilket gör att vi alltid kommer få resistenta bakterier även i en odling med CRISPR som förhindrar överföringen från plasmider.

Hos *P. furiosus* klyvs märkt RNA specifikt av ett crRNP komplex (komplex med crRNA). Detta skulle kunna användas in vitro för att klyva RNA molekyler. Kan man dessutom hitta andra crRNP komplex med endonukleär DNA aktivitet kan man uppgradera sig till att specifikt klyva DNA-molekyler in vitro. Vilket skulle vara en väldigt god applikation inom molekylärbiologin. (Maraffini & Sontheimer 2010a)

Med tanke på vilken stor mängd *cas*-proteiner man hittat i de olika CRISPR systemen, finns det förmodligen också ett enormt utbud av olika funktioner bland dessa. När man så småningom kan botanisera bland dessa kommer det upptäckas fler användningsområden för CRISPR. Vad som krävs är mer forskning, som ger oss mer kunskap om vad vi egentligen har för möjligheter, men det huvudsakliga användningsområdet för oss med CRISPR är ändå att kunna göra viktiga bakterier immuna.

Diskussion

Visst kvarstår en hel del frågetecken innan CRISPR är någon som vi helt förstår. Mekanismen, vilka proteiner är inblandade? Hur ser dessa ut, vad har de för struktur? Vad skiljer systemet mellan arter? Är dessa skillnader relevanta för CRISPRs funktion? Hur stoppas egentligen infektionen? Vi vet att det händer, men vad händer? Kan vi förvänta oss ett sådant här system i alla bakterier och arkéer, trots att vi idag inte funnit det i alla? Hur snart kan vi börja utnyttja CRISPR för vår egen vinning? Är det så bra som vi tror, eller bättre? Kan vi stoppa spridningen av antibiotika resistans? Kan man tänka sig fler användningsområden än de som redan föreslagits?

Ja frågorna är många, och alla kommer inte att lösas i ett nafs. Att bakteriofager driver utvecklingen av bakterier, och att bakteriers utveckling i sin tur driver bakteriofagers utveckling kan det göra det svårt att använda systemet mot antibiotikaresistans spridning? Eller kommer bakterier alltid att ligga steget före, virusen följer den riktning CRISPR antar? Jag skulle inte tro att det skulle kunna komma att bli ett problem, nog för att virus förändrar sig snabbt, men hade detta varit fallet hade CRISPR nog inte funnits kvar idag. Prokaryoter hade inte behållit CRISPR om det inte hade varit lönsamt för dem. Vi människor upptäckte CRISPR nyligen, men första gången vi såg det var 1987 och det har förmodligen funnits längre än så. Så finns det kvar idag finns det kvar imorgon.

Studier på *S. Thermophilus* har visat att vid en infektion av bakteriofager inkorporerar celler nya spacers och blir immuna med en frekvens på 1 på miljonen (Garneau m.fl 2010). Denna låga siffra stämmer inte direkt överens med vad Andersson och Banfields studier (2008), där man hittat en större andel spacers i de bakterier man studerat. En på miljonen är visserligen en siffra baserat på en enskild art, medan Andersson och Banfield studerade flera arter ur en biofilm och har gjort rena DNA studier. Hur som helst, hur liten siffran än är så är det tillräckligt för att bakterien skall kunna återuppbygga en population. Det krävs inte många celler för att få en enorm mängd bakterier, det känner vi alla igen från de laborationer vi utfört.

En del forskare vill gärna lyfta fram att CRISPR skulle kunna vara lösningen på antibiotikaresistansproblemet. Visst låter det fantastiskt, men som diskuterades tidigare så är förväntningarna lite väl höga då det finns en del naturliga problem som uppstår vilka gör att det inte är möjligt. Visst kan CRISPR hjälpa oss att förhindra spridning till viss del i vissa fall, men det är ingen långsiktigt eller storskalig lösning, tyvärr. Som tur är kan CRISPR vara användbart i flera andra fall, som också kommer att vara viktiga för oss.

Tack

Ett enormt stort tack till Magnus Lundgren, min externa handledare som hjälpte mig otroligt mycket redan från start. Jag önskar honom och hans kollegor all lycka till med sitt fortsatta arbete med CRISPR! Jag vill också tacka Hannon & Kurginov som gett mig lov att använda alla (!) deras bilder och tabeller i mitt arbete, och för att jag fick svar samma kväll. Marraffini & Sontheimer skall också ha tack för att jag får låna deras bild. Tack också till alla er som gett välbehövlig återkoppling på mitt arbete.

Referenser

- Andersson A. F, Banfield J. F. 2008. Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. *Science* **320**: 1047-1050.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero Dennis A, Horvath P. 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**: 1709-1712.
- Björn L.O, Endell P.H, Meurling P, Pelger S, Ståhl S. 2005. Biologisk ordlista. Studentlitteratur, Sverige Lund.
- Brouns S, Jore M, Lundgren M, Westra E., Slijkhuis R, Snijders A, Dickman M, Makarova K, Koonin E, Van der Oost J. 2008. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321**: 960-964.
- Danisco A/S. 2010. About us. WWW-hemsida:
http://www.danisco.com/wps/wcm/connect/www/corporate/about_us. Information läst 2010-11-23.
- Deveau H, Barrangou R, Garneau J.E, Laboné J, Fremaux C, Boyaval P, Romero D, Horvath P, Moineau S. 2008. Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus* *Journal of Biology* **190**: 1390-1400.
- Garneau J, Dupuis M-E, Villion M, Romero D, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán A, Moineau S. 2010. The CRISPR/Cas Bacterial Immune System cCeaves Bacteriophage and Plasmid DNA. *Nature* **468**: 67-72.
- Horvath P, Barrangou R. 2010. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* **327**: 167-170.
- Horvath P, Romero D, Coûté-Monvoisin A-C, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R. 2008. Diversity, Activity, and Evolution of CRISPR Loci in *Streptococcus thermophilus*. *Journal of Bacteriology* **190**: 1401-1412.
- Hyman P, Abedon S.T. 2010. Bacteriophage Host Range and Bacterial Resistance. *Advances in Applied Microbiology* **70**: 217-248.
- Karginov F, Hannon G. 2010. The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. *Molecular Cell* **37**: 7-19.
- Kunin V, Sorek R, Hugenholtz P. 2007. Evolutionary Conservation of Sequence and Structures in CRISPR Repeats. *Genome Biology* **8**: 1-7.

Makarova S, Grishin V, Shabalina A, Wolf I, Koonin E. V. 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* **1:7**: 1-26.

Marraffini A, Sontheimer Erik J. 2010a. CRISPR Interference: RNA-Directed Adaptive Immunity in Bacteria and Archaea. *Nature Reviews Genetics* **11**, 181-190

Marraffini A, Sontheimer Erik J. 2008. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science* **322**: 1843-1845.

Marraffini A, Sontheimer Erik J. 2010b. Self Versus Non-Self Discrimination During CRISPR RNA-Directed Immunity. *Nature* **463**: 568-572.

Marraffini A, Sontheimer Erik J. 2010c. Slicer for DNA. *Nature* **468**: 45-46.

Paris -sud 11 université. 2010. CRISPR Database. <http://crispr.u-psud.fr/crispr/> Hämtat 2010-12-03.

Snyder L, Champness W. 2007. *Molecular Genetics of Bacteria*. 3:e upplagan. ASM press, Washington DC.