

SYSTEM INTERFERENCYJNEGO RNA BAKTERII (CRISPR-CAS) I JEGO ROLA W OBRONIE PRZED INFEKCJĄ FAGOWĄ

Adam Jaworski, Anita Dobrowolska*

Zakład Genetyki Drobnoustrojów Uniwersytet Łódzki 91-237 Łódź, ul. Banacha 12/16

Wpłynęło w październiku 2008 r.

1. Rys historyczny. 2. Zróżnicowanie strukturalne systemu CRISPR w świecie bakterii. 3. System CRISPR-CAS analogiem systemu interferencyjnego RNA eukriontów. 4. Możliwości praktycznych zastosowań systemu CRISPR-CAS

Organization of the bacterial interference RNAi-like system (CRISPR-CAS) and its role against bacteriophages

Abstract: Clustered, regularly interspaced short palindromic repeats (CRISPR) are present in the genomes of many different species of bacteria and archaea. They are composed of direct repeats (DRs) that are separated by non-repetitive spacers. Sets of CRISPR sequences together with associated proteins encoded by a group of *cas* genes are involved in resistance to bacteriophages. The structure and the antiviral activity of this new resistance mechanism in bacteria and archaea as well as the possibilities of its present and future applications are discussed in this article.

1. History of CRISPR research. 2. Structural features of CRISPR system. 3. CRISPR as an analog of eukaryotic interference RNAi system. 4. Present and future possibilities of application

Słowa kluczowe: geny i białka grupy CAS, nowy system obrony bakterii przed bakteriofagami, trakty powtórzonych sekwencji DNA (CRISPR)

Key words: CAS (CRISPR associated sequences), CRISPR (clustered, regularly interspaced short palindromic sequences), new bacterial system against bacteriophages

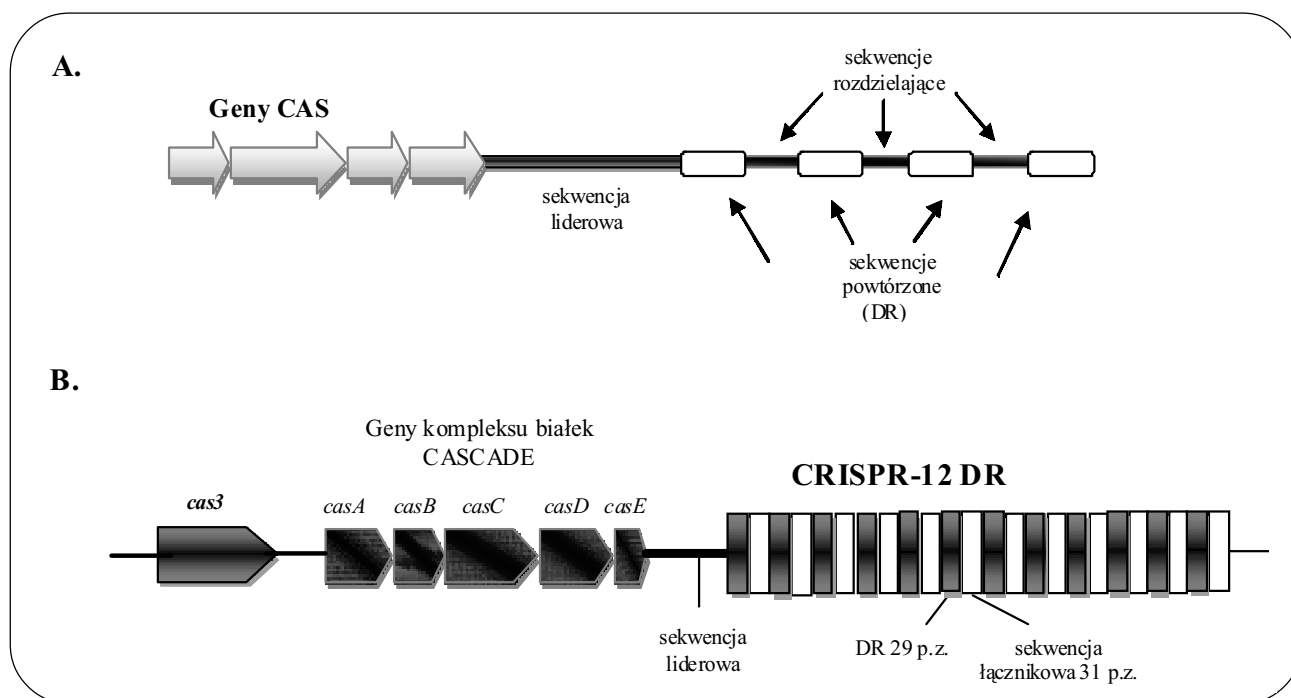
1. Rys historyczny

Szlak zgrupowanych, regularnie rozmieszczonych, krótkich sekwencji palindromowych CRISPR (ang. *clustered, regularly interspaced short palindromic repeats*) zidentyfikowano po raz pierwszy w genomie *E. coli* w 1987 roku [16]. Okazało się, że sekwencja ta, przylegająca bezpośrednio do genu *iap* kodującego syntezę alkalicznej fosfatazy, złożona jest z 14 prostych powtórzeń (ang. *direct repeats* – DRs) zbudowanych z 29 nukleotydów, rozdzielonych unikatowymi sekwencjami łącznikowymi (ang. *spacers*) o długości 33 nukleotydów [20, 27]. Całe szlaki CRISPR mogą być zbudowane, w zależności od gatunku bakterii, nawet z 250 takich elementów (Rys. 1). W latach 90-tych ubiegłego wieku zidentyfikowano i scharakteryzowano ciągi sekwencji CRISPR w genomach *Mycobacterium tuberculosis* [15], *Haloferax mediterranei* [24], *Methanocaldococcus jannaschii* [6] i *Termotoga maritima* [28]. Sekwencje CRISPR są szeroko rozpowszechnione w świecie bakterii, do roku 2007 zidentyfikowano je w genomach 232 gatunków, w tym w 40% zsekwencjonowanych genomach prokariotów właściwych i aż w 90% genomach archeonów [2, 10, 12, 13, 25].

W latach 2002–2006 doniesiono o identyfikacji nieznanych dotąd genów, nazwanych genami *cas* (ang. *CRISPR associated sequences*), z racji ich położenia w sąsiedztwie traktów CRISPR, z którymi połączone są przez sekwencję liderową (ang. *leader sequence*). Początkowo w analizowanych genomach zidentyfikowano cztery geny *cas* [17,18], ale dalsze analizy przeprowadzone dla dużej liczby różnych gatunków bakterii pozwoliły zidentyfikować około 45 różnych takich genów [13, 22, 23]. Okazało się także, że niektóre z genów rodziny *cas* nie występują w genomach gatunków bakterii pozbawionych sekwencji CRISPR [20], co sugerowało, iż geny te spełniają jakąś funkcję biologiczną, ściśle związaną z obecnością i funkcją systemu CRISPR.

Od ponad 10 lat trwają dyskusje na temat funkcji biologicznych spełnianych przez elementy CRISPR. Sugerowano, że wzrost liczby kopii tego elementu u *Haloferax volcanii* zaburza proces prawidłowej segregacji chromosomów do komórek potomnych (ang. *replicon partitioning*) [24]. Tej obserwacji nie udało się jednak potwierdzić w badaniach segregacji chromosomów u *Mycobacterium tuberculosis* [17, 18]. Nie wykluczano również możliwości, że sekwencje CRISPR są ruchomymi elementami genetycznymi

* Autor korespondencyjny: Zakład Genetyki Drobnoustrojów Uniwersytet Łódzki 91-237 Łódź, ul. Banacha 12/16, e-mail: anidobro@biol.uni.lodz.pl



Rys. 1. Ogólny schemat organizacji bakteryjnego systemu CRISPR-CAS (A) oraz schemat organizacji systemu CRISPR-CAS *Escherichia coli* (B) (dodatkowe informacje w tekście)

z racji ich obecności w genomach niektórych bakterii w wielu kopiach [17, 18]. Formułowano także opinie o ich udziale w systemach naprawy DNA, ponieważ okazało się, że białka kodowane przez wiele genów systemu CRISPR zawierają specyficzne domeny oddziałujące z DNA [22]. W 2005 roku pojawiły się pierwsze hipotezy na temat ich udziału w nadawaniu komórkom bakterii odporności na infekcje fagami, plazmidami koniugacyjnymi i transpozonami [3, 26, 30]. Okazało się bowiem, że sekwencje łącznikowe często zawierają fragmenty DNA fagów lub plazmidów; obserwowano również odwrotnie proporcjonalną korelację pomiędzy wrażliwością na infekcje fagami a liczbą powtarzających się sekwencji łącznikowych w CRISPR. Pod koniec 2007 roku ukazały się prace, w których potwierdzono doświadczalnie, że fragmenty DNA fagów, wbudowane w czasie infekcji fagowej do CRISPR jako nowe sekwencje łącznikowe, nadają komórkom trwałą odporność na kolejną infekcję tym fagiem [1, 8]. We wrześniu 2008 potwierdzono jednoznacznie, że bakteryjny system CRISPR funkcjonuje analogicznie do systemu interferencyjnego RNA w komórkach organizmów eukariotycznych [4, 37].

Rekonstrukcję modelowego systemu CRISPR-CAS *E. coli*-*fagλ* oraz genetyczną i biochemiczną charakterystykę jego ważnej aktywności biologicznej należy uznać za duże osiągnięcie naukowe w mikrobiologii o charakterze poznawczym, a także za ważne odkrycie dla praktyki przemysłowej, wykorzystującej bakterie w procesach biotechnologicznych, narażonych na groźne infekcje fagowe.

2. Charakterystyka strukturalnego zróżnicowania systemu CRISPR w świecie bakterii

Identyfikacja systemów CRISPR u co raz większej liczby gatunków bakterii właściwych i archeonów oraz dokładna analiza ich budowy wskazały, że generalny plan budowy tych powtórzonych szlaków jest konserwatywny w całym świecie bakterii. Jednakże poszczególne elementy tego systemu są dość znacznie zróżnicowane, co dotyczy zarówno prostych powtórzeń jak i sekwencji łącznikowych oraz genów *cas* (Rys. 1A). Wielkość sekwencji powtórzonych, w zależności od gatunku bakterii, waha się od 24 do 47 nukleotydów, a wielkość sekwencji łącznikowych od 26 do 72 nukleotydów. Zróżnicowana jest także liczba powtórzeń w traktach CRISPR, od 2 u niektórych gatunków bakterii właściwych do nawet 249 u *Verminephrobacter eisenia* [11, 12, 17, 18]. Ponadto, niektóre gatunki bakterii zawierają 1 kopię CRISPR w określonym *locus* genomu, inne jak np. *M. jannaschii*, aż 18 w różnych regionach genomu [6]. Podobne zróżnicowanie dotyczy liczby zidentyfikowanych genów *cas*, od kilku do 20, a nawet więcej [13].

Analiza stopnia homologii prostych powtórzeń dla wielu różnych gatunków bakterii właściwych i archeonów pozwoliła dotychczas wyróżnić wśród nich 12 grup. Najliczniejsza grupa obejmuje sekwencje palindromowe o wielkości 5–7 nukleotydów [17, 18]. Odkrycie, że sekwencje palindromowe, wraz z sekwencjami łącznikowymi, podlegają transkrypcji, a syntetyzowany mRNA może w tych rejonach tworzyć

strukturę pętli (*stem-loop secondary structure*), było szczególnie ważne dla rozważań na temat funkcji biologicznej systemu CRISPR [20, 21, 34, 35]. Dalsze szczegółowe analizy dowiodły, że wiele prostych powtórzeń charakteryzuje się obecnością konserwatywnej sekwencji GAAA(C/G) zlokalizowanej na końcu 3', prawdopodobnie odpowiedzialnej za wiązanie specyficznych białek komórkowych, co dodatkowo wskazywało na istotną funkcję biologiczną CRISPR [20].

Poszczególne nukleotydowe sekwencje łącznikowe, rozdzielające kolejne proste powtórzenia w określonym trakcie CRISPR, charakteryzują się bardzo dużą różnorodnością. Szczegółowe analizy sekwencji tych elementów, przeprowadzone dla dużej liczby CRISPR różnych gatunków bakterii, ujawniły ich duże pokrewieństwo do DNA fagów i innych ekstrachromosomalnych elementów genetycznych [3, 21, 26]. Badania przeprowadzone dla 4500 sekwencji łącznikowych u 67 gatunków prokariotów dowiodły, że sekwencje nukleotydowe jedynie 88 z nich (2%) wykazywały podobieństwo do znanych sekwencji zdeponowanych w światowych bazach danych, wśród których 60% stanowiły sekwencje znanych fagów i plazmidów. Udało się także ustalić, po pierwsze, że sekwencje łącznikowe zostały przeniesione do traktów CRISPR z kodujących lub niekodujących nici DNA fagów, i po drugie, że tego typu sekwencje obficie występują w różnych regionach genomów wielu znanych fagów [1, 21, 26, 30]. Wyniki ostatnio opublikowanych prac wskazują, że w sąsiedztwie sekwencji fagowych (przy ich końcu 3'), homologicznych do sekwencji łącznikowych CRISPR, znajdują się motywy AGAA lub GGNG, które prawdopodobnie są miejscami rozpoznania lub cięcia dla produktów systemu CRISPR [1, 8]. Można sądzić, że w miarę identyfikacji oraz poznawania sekwencji nukleotydowych genomów nieznanymi dotychczas fagów i plazmidów, bytujących w hodowlanych i niehodowlanych mikroorganizmach środowiskowych, możliwym będzie ustalenie pochodzenia coraz większej liczby sekwencji łącznikowych systemu CRISPR. Potwierdzeniem są opublikowane niedawno dane dla bakterii mlekowych *Lactococcus lactis*. W szczepach tego gatunku, noszących wiele różnych fagów i plazmidów o znanej sekwencji nukleotydowej, udało się ustalić fagowe lub plazmidowe pochodzenie ponad 40% sekwencji łącznikowych obecnych w ich systemach CRISPR [3].

Kolejnym elementem składowym systemu CRISPR jest sekwencja liderowa o wielkości do 550 par zasad, zlokalizowana bezpośrednio przy końcu 5' pierwszego powtórzenia (Rys. 1). Sekwencja ta jest bogata w pary AT i nie zawiera, podobnie jak proste powtórzenia, otwartych ramek odczytu; podzielone są natomiast zdania na temat stopnia jej konserwatywności w traktach CRISPR różnych gatunków prokariotów [17, 19,

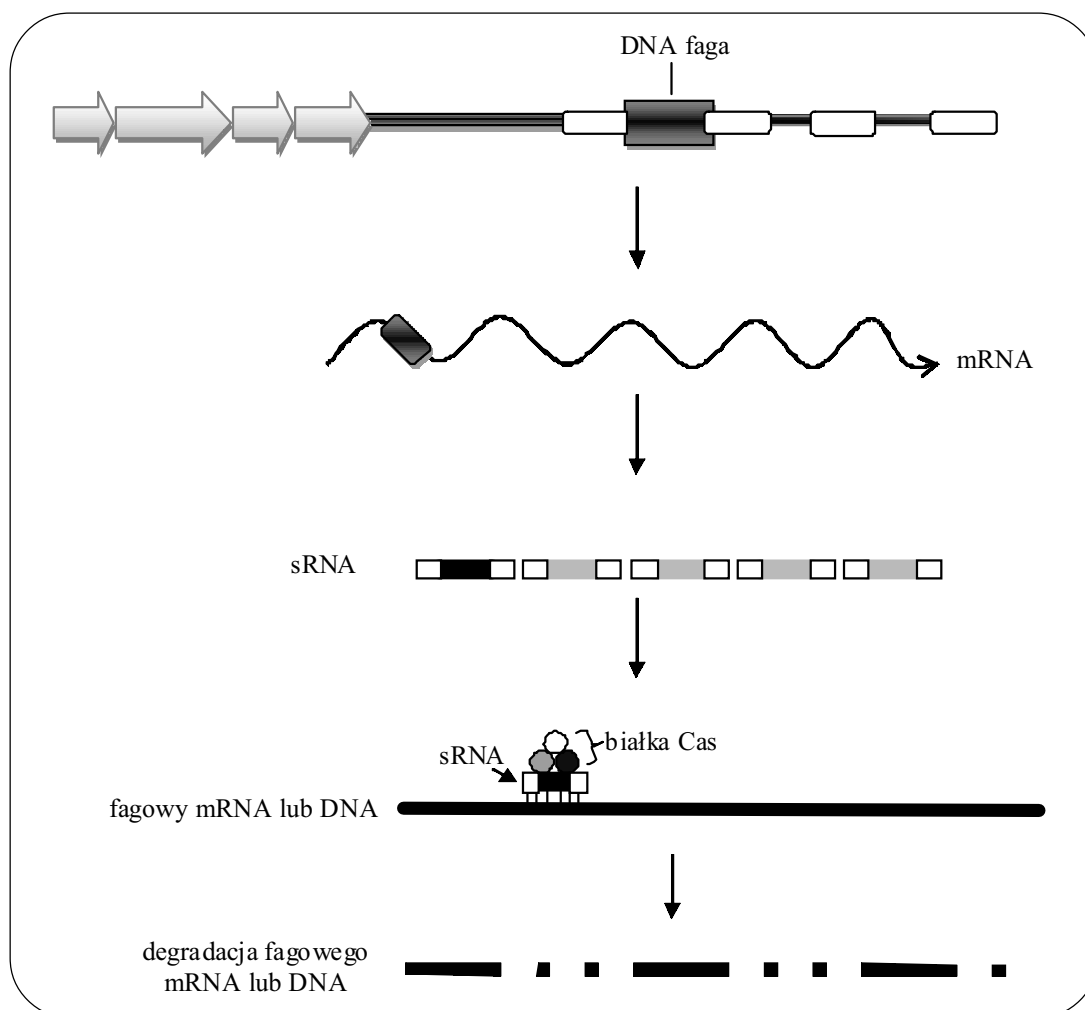
21, 31]. Interesującym jest odkrycie, że wbudowanie w dany trakt CRISPR nowej jednostki, proste powtórzenie-sekwencja łącznikowa (*DR-spacer*), następuje zawsze pomiędzy sekwencją liderową a pierwszym powtórzeniem, co wskazuje na jej funkcję akceptorową w tym procesie. Z racji lokalizacji tej sekwencji sugeruje się również jej funkcję jako promotora dla transkrypcji całego ciągu sekwencji CRISPR [1, 30, 34, 35].

Bezpośrednio przy końcu 5' sekwencji liderowej leżą niedawno odkryte, a ostatnio lepiej scharakteryzowane geny *cas* (Rys. 1). Najogólniej można stwierdzić, że rodzina genów *cas* obejmuje obecnie 7–8 różnych typów, a wśród nich najbardziej rozpowszechnionych jest 6, od *cas1* do *cas6* [13, 33]. W białkach kodowanych przez geny *cas* zidentyfikowano ważne funkcjonalne domeny jak egzo- i endonukleaz, helikaz, a także domeny odpowiedzialne za wiązanie RNA i DNA oraz domeny zaangażowane w regulację transkrypcji [17, 18]. Warto w tym miejscu podkreślić, że sekwencja nukleotydowa genu *cas1* (kod dostępu COG1518) jest powszechnie wykorzystywana jako uniwersalny marker dla badań poznawczych bakteryjnych systemów CRISPR [9, 13, 23]. Inną rodziną białek, syntetyzowanych wyłącznie przez bakterie zawierające szlaki CRISPR, są białka związane prawdopodobnie z funkcją tych systemów. Nie jest znana, jak dotychczas, rola biologiczna tych białek, zwanych RAMP (*repeat associated mysterious protein*), wiadomo jednakże, że geny kodujące syntezę tych białek nie są zlokalizowane w bezpośrednim sąsiedztwie ciągów sekwencji CRISPR [13, 22].

3. System CRISPR-CAS analogiem systemu interferencyjnego RNA eukariotów

Wyniki najnowszych prac, opublikowanych w ubiegłym roku [1, 8] wywołały żywą dyskusję wśród specjalistów na temat funkcji biologicznych systemu CRISPR oraz mechanizmu jego działania w obronie bakterii przed infekcją fagami, a także w regulacji innych procesów komórkowych [20]. Badania przeprowadzono na modelu *Streptococcus thermophilus*, gatunku bakterii powszechnie stosowanych w przemyśle mleczarskim do produkcji jogurtu i serów.

W wyniku infekcji jednego ze szczepów dwoma różnymi fagami udało się wyselekcjonować z hodowli 9 mutantów bakterii opornych na dalszą infekcję tymi fagami. Genomowy DNA tych mutantów poddano sekwencjonowaniu, koncentrując szczególną uwagę na szlaku CRISPR. Stwierdzono, że każdy z tych mutantów w czasie infekcji nabył od 1 do 3 nowych sekwencji łącznikowych, wbudowanych precyzyjnie pomiędzy sekwencję liderową i pierwszy DR. Co więcej, okazało się, że nabyte, nowe sekwencje łącznikowe pochodzą z fagów użytych do infekcji. Dalsze doświadczenia



Rys. 2. Molekularny mechanizm funkcjonowania układu CRISPR-CAS w obronie komórek bakteryjnych przed infekcją fagową (szczegółowe wyjaśnienia w tekście)

oraz dokładniejsze analizy molekularne dowiodły, po pierwsze, że u mutantów opornych na nową infekcję fagową nabyta sekwencja łącznikowa wykazuje 100% homologii z odpowiednim regionem DNA użytego faga, i po drugie, że w przypadku gdy w nabytej sekwencji łącznikowej pojawia się od 1 do kilku mutacji (substytucji), to komórka pozostaje wciąż wrażliwa na infekcję. Ostatecznym potwierdzeniem były wyniki doświadczeń, w których nową, nabytą sekwencję łącznikową mutantu opornego na infekcję wklonowano w ten sam *locus* szczepu wrażliwego i stwierdzono, że nabył on oporność na badanego faga (Rys. 2). I odwrotnie, delekcja tej sekwencji w szczepie opornym na infekcję czyniła go ponownie wrażliwym. Nieoczekiwanie okazało się, że niewielka populacja izolowanych w czasie doświadczeń fagów zachowała zdolność infekowania szczepów *S. thermophilus*. Udało się jednak dowieść, że w DNA części z tych fagów pojawiły się mutacje w rejonie tej sekwencji, która w czasie infekcji jest przenoszona do CRISPR bakterii, zaś w niektórych innych nabyta sekwencja wykazywała,

co prawda, 100% homologię z odpowiednim regionem DNA faga, ale mutacje zidentyfikowano w motywie AGAA przylegającym do jej końca 3'. Wyniki te potwierdzają wcześniejsze sugestie o roli tego motywu w rozpoznawaniu przez faga traktu CRISPR. Ustalenie organizacji systemu CRISPR na poziomie DNA nie wyjaśniało mechanizmu jego działania, w tym udziału określonych białek komórkowych, w szczególności białek CAS i ich specyficznych funkcji. Autorzy wyżej cytowanych prac w jednym z wyselekcjonowanych szczepów opornych na infekcję fagową zainaktywowali gen *cas5*, w drugim zaś gen *cas7*. W pierwszym przypadku inaktywacja genu, kodującego syntezę endonukleazy, doprowadziła do utraty oporności na infekcję fagową, w drugim zaś przypadku inaktywacja genu o bliżej nieznannej funkcji, nie miała wpływu na zmianę oporności szczepu na ponowną infekcję tym samym fagiem, ale osłabiła jego zdolność do nabywania oporności na infekcję innym fagiem. Postawiono wniosek, że produkt tego nieznanego genu jest odpowiedzialny za wbudowywanie do traktu CRISPR określ-

lonych fragmentów DNA faga jako nowych, kolejnych sekwencji łącznikowych [1, 8, 13]. Co prawda, molekularny mechanizm odpowiedzialny za aktywność biologiczną układu CRISPR jest jeszcze daleki od pełnego wyjaśnienia, ale w świetle wyżej omówionych wyników, a także nieco wcześniejszych danych [34, 35], wskazujących, że u szczepów z rodzaju *Archaeoglobus* i *Sulfolobus* trakty CRISPR podlegają transkrypcji w całości, z jednego promotora, można było zaproponować ogólny model dla wyjaśnienia funkcjonowania całego systemu. Wykazano, że transkrypty podlegają w komórkach enzymatycznej obróbce z wytworzeniem cząsteczek małych sRNA (*small RNA*), o wielkości równej jednej sekwencji DR plus sekwencja łącznikowa, a trawienie transkryptu ma miejsce w środku każdej jednostki DR. Stąd, końcowe produkty sRNA zawierają pełną sekwencję łącznikową w środku i połowy sąsiadujących sekwencji DR zlokalizowanych na ich końcach 5' i 3'. Wiadomo, że obecność w sekwencjach DR motywu palindromów umożliwia tworzenie drugorzędowych struktur i wypełnienia sekwencji łącznikowych. Fakt, że transkrypty traktów CRISPR są trawione do krótkich sRNA, a w proces ten zaangażowane są domeny białek CAS, stał się podstawą do sformułowania hipotezy, że bakteryjny system CRISPR jest funkcjonalnym analogiem systemu siRNAs (ang. *small interference RNA*) organizmów eukariotycznych [23], który spełnia ważną rolę jako system obrony komórki przed RNA wirusów i transpozonów [14, 32]. Zgodnie z tą hipotezą, bakteryjne siRNAs, krótkie produkty trawienia transkryptu CRISPR, są analogami eukariotycznych siRNA. Tworzą one bowiem kompleksy z białkami CAS, podobnie jak siRNAs eukariotów z białkiem Dicer, i rozpoznają homologiczne sekwencje w syntetyzowanych przez obce elementy genetyczne transkryptach, prowadząc w końcowym efekcie do ich sukcesywnej degradacji przez inne białka CAS (Rys. 2).

Ostatnio przeprowadzono niezwykle owocne badania aktywności systemu CRISPR na modelu genetycznie rekonstruowanych szczepów pochodnych *E. coli* K12 wrażliwych na infekcję fagiem λ [4]. W genomie tego gatunku znajduje się zarówno trakt powtórzonych sekwencji DR-sekwencja łącznikowa, jak i 8 genów *cas* : *cas3*, *cas1*, *cas2*, kodujących, odpowiednio, HD-nukleazę/helikazę, integrazę, endorybonukleazę, i 5 genów oznaczonych *casABCD*, kodujących kompleks białek systemu CRISPR odpowiedzialnego za ochronę komórek przed infekcją fagową, nazwanych CASCADE (*CRISPR-associated complex for antiviral defense*) (Rys. 1B). Autorzy cytowanej pracy skonstruowali różne rekombinowane szczepy. Jedne z nich nosiły jeden fragment DNA faga λ , komplementarny do DNA czterech genów niezbędnych dla rozwoju faga, wklonowany pomiędzy pierwszy i drugi motyw

DR traktu CRISPR. W inne skonstruowane szczepy wklonowano cztery fragmenty DNA faga. Jeszcze inne, pozbawione traktu DR-sekwencja łącznikowa, nosiły geny *cas*, ale aktywne tylko w określonych kombinacjach (*cascade*, *cas3*, *cascade+cas3*, *cascade-casE+cas3*, *cascade+cas1+cas2*, *cascade+cas3*, *cas1+cas2*). Nie wnikając w metodyczne szczegóły przeprowadzonych badań należy stwierdzić, że wklonowanie fragmentów DNA faga λ w region traktu CRISPR wrażliwego szczepu *E. coli* K12 pozwoliło wyselekcjonować szczepy 10 milionów razy bardziej odporne na infekcję tym fagiem, iż szczep wyjściowy. Zidentyfikowano i scharakteryzowano również kompleks białkowy CASCADE i udowodniono jego bezpośredni udział w degradacji mRNA regionu CRISPR, z wytworzeniem fragmentów o długości 57 nukleotydów CRISPR-RNA. Każdy z tych fragmentów zawierał sekwencję łącznikową otoczoną na końcach 5' i 3' fragmentami przylegających sekwencji DR. Pełna efektywność kompleksu CASCADE była zapewniona w obecności białka Cas3. Ustalono także niezwykle specyficzną specyficznosc kompleksu CASCADE w skonstruowanym szczepie *E. coli* K12 opornym na faga wykazując, że nie jest on aktywny w stosunku do transkryptów innych CRISPR, w którym znajdują się sekwencje DR z innych szczepów *E. coli*. Co więcej, udało się udowodnić, że aktywność systemu CRISPR zapewniają transkrypty zarówno nici sensu jak i antysensu wklonowanych fragmentów DNA faga λ . Wyniki te wskazują, że docelowym elementem dla małych CRISPR-RNA może być także DNA faga. Zatem fundamentalna różnica pomiędzy bakteryjnym i eukariotycznym systemem obrony przed infekcją obcym DNA polega na tym, że w pierwszym przypadku degradacji ulega DNA w drugim zaś komplementarny RNA.

Można obecnie oczekiwać, że w najbliższej przyszłości zostaną wyjaśnione kolejne pytania i wątpliwości dotyczące systemu CRISPR-CAS, a w tym w szczególności regulacji transkrypcji regionu CRISPR, regulacji ekspresji białek CAS i ich składania w komórce w postaci kompleksu CASCADE, a także enzymatycznych mechanizmów degradacji mRNA-CRISPR oraz degradacji docelowych sekwencji DNA fagów. Nie wiadomo jeszcze w jaki sposób sekwencje fagów są *in vivo* selekcjonowane, wycinane i wbudowywane w trakty CRISPR jako sekwencje łącznikowe. Nie wyklucza się, że system CRISPR-CAS może być zaangażowany w wyciszenie także endogennych genów bakterii, bowiem 7–35% sekwencji łącznikowych w traktach CRISPR ma chromosomalne pochodzenie [1, 3, 26].

W świetle omawianych wyżej wyników można lepiej zrozumieć równowagę biologiczną pomiędzy bakteriami i szeroko rozpowszechnionymi w środowiskach naturalnych bakterii fagami. Szacuje się, że co dwa dni około połowa populacji bakterii żyjących w biosferze

jest zabijana w wyniku infekcji różnymi fagami. Bakterie są jednak wyposażone w różne naturalne mechanizmy obrony przed fagami jak: blokowanie ich adsorpcji, trawienie fagowego DNA restryktazami oraz degradacja obcego DNA przy udziale szeroko rozpowszechnionego systemu CRISPR-CAS. Sugeruje się, że być może, zostaną wykryte również fagowe białka specyficznie blokujące system CRISPR-CAS, co w świetle wyników opisanych przez dla białek *Sulfolobus* ma pewne uzasadnienie [29]. Z drugiej strony, wysokie tempo mutacji fagów pozwala na generowanie mutantów zdolnych do przełamania mechanizmów obronnych komórki, zaś selekcyjna presja bakteryjnych systemów obrony, nadają tej równowadze bardzo dynamiczny charakter, zarówno w aspekcie ekologicznym jak i ewolucyjnym.

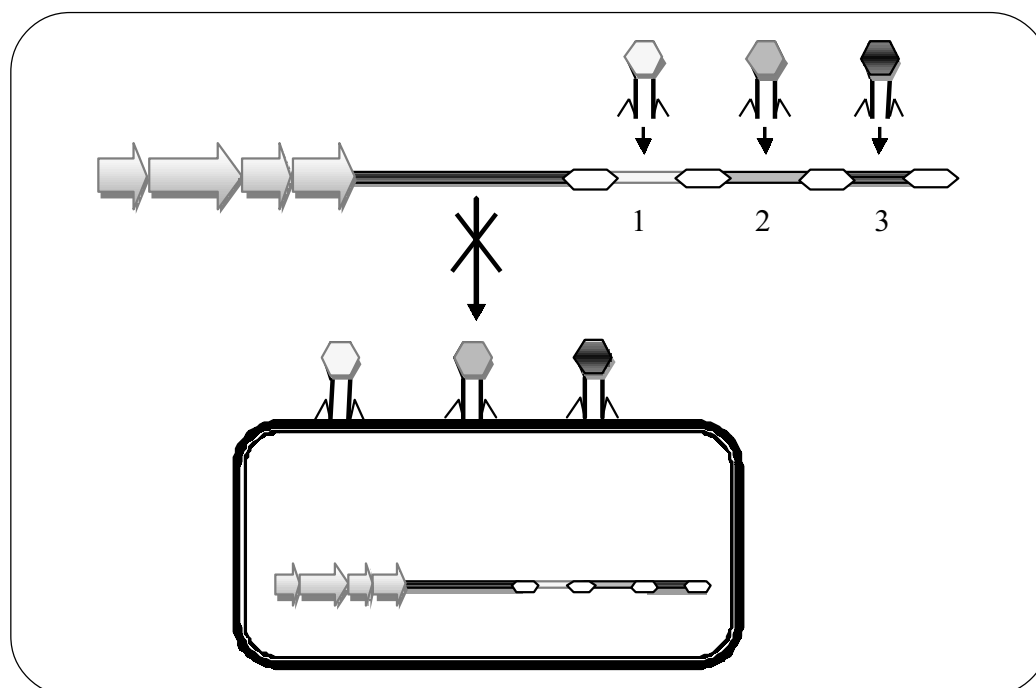
4. Możliwości praktycznych zastosowań systemu CRISPR-CAS

Wiele gałęzi przemysłu spożywczego, farmaceutycznego wykorzystuje bakterie w procesach biotechnologicznych, a stąd ochrona przed infekcjami fagowymi czystych kultur bakterii, wykorzystywanych w różnych procesach biotechnologicznych, ma ogromne znaczenie praktyczne i ekonomiczne. Dla przykładu, infekcja fagami szczepów wykorzystywanych w przemyśle mleczarskim w produkcji napojów mlecznych i serów jest przyczyną ogromnych problemów

i strat finansowych [33]. Fagi replikują się w bakteriach bardzo szybko, a stąd pełna liza hodowli bakteryjnych następuje w czasie kilku godzin. Ustalono, że dodanie do hodowli *E. coli* o gęstości 10^6 komórek/ml 1 cząstki faga T7 prowadzi w krótkim czasie do pełnej lizy 99,9% komórek oraz syntezy 10^9 nowych cząstek fagowych.

Ustalenie organizacji i funkcji biologicznej systemu CRISPR-CAS stworzyło obecnie realną szansę zarówno dla selekcji szczepów przemysłowych opornych na określone fagi jak i konstrukcji takich szczepów dostępnymi metodami genetycznymi. Klonowanie określonych fragmentów DNA różnych fagów w trakty CRISPR szczepów wrażliwych i selekcjonowanie szczepów opornych na infekcje nie następuje obecnie żadnych problemów technicznych (Rys. 3). Jednak regulacje prawne, szczególnie w krajach Europy, dotyczące genetycznie modyfikowanych mikroorganizmów (GMO), praktycznie nie dopuszczają do ich stosowania na skalę przemysłową w przemyśle spożywczym. Jednak w tym przypadku można z powodzeniem selekcjonować takie szczepy także poprzez infekcję szczepów przemysłowych różnymi naturalnymi fagami oraz poszukiwanie szczepów opornych, które nabyły i stabilnie wbudowały w trakty CRISPR określone fragmenty DNA fagów na drodze naturalnej.

Podobieństwo bakteryjnego systemu CRISPR-CAS do eukariotycznego systemu RNAi nasuwa możliwość jeszcze innych zastosowań. Eukariotyczny system interferencyjnego RNA spełnia nie tylko funkcję obrony



Rys. 3. Możliwe sposoby konstrukcji szczepów opornych na infekcję fagową

1, 2, 3, odpowiednio, fragmenty DNA trzech różnych fagów wklonowanych jako sekwencje łącznikowe w trak CRISPR. Alternatywnie, sposób naturalnej selekcji komórek bakterii opornych na infekcję po kolejnych cyklach litycznych fagami 1, 2, 3.

komórki przed obcym DNA (RNAi), ale także reguluje ekspresję endogennych genów chromosomalnych (micro-RNAs). Nie jest wykluczonym, że zgromadzona wiedza na temat systemu CRISPR-CAS może pozwolić na opracowanie sposobu hamowania ekspresji endogennych genów prokariotów, przydatnego nie tylko dla badań poznawczych, ale także praktycznego wykorzystania w wyciszaniu genów odpowiedzialnych za patogenezę.

Wspominano wcześniej, że loci CRISPR w świecie bakterii są bardzo zróżnicowane zarówno pod względem liczby powtórzeń DR oraz ich sekwencji nukleotydowej, jak i obecności różnorodnych sekwencji łącznikowych. U niektórych gatunków trakt CRISPR jest zróżnicowany nawet u bardzo blisko spokrewnionych szczepów pod względem innych wyznaczników genetycznych. Jednym z przykładów są szczepy *Leptospirillum* charakteryzujące się dużym, klonalnym zróżnicowaniem tego traktu. [36]. Innym przykładem jest gatunek *Mycobacterium tuberculosis*, w obrębie którego szczepy są bardzo zróżnicowane pod względem zarówno liczby sekwencji łącznikowych jak i ich sekwencji nukleotydowych. Wiedza ta stała się podstawą do opracowania szeroko już stosowanej metody genotypowania izolatów *M. tuberculosis* dla celów szybkiej i pewnej diagnostyki oraz dochodzeń epidemiologicznych zakażeń gruźliczych [5, 7].

Z racji faktu, że nowe sekwencje łącznikowe są wbudowywane z reguły pomiędzy pierwszym i drugim powtórzeniem traktu CRISPR można sądzić, że sekwencje łącznikowe coraz bardziej oddalone od sekwencji liderowej mają, odpowiednio, coraz starsze pochodzenie ewolucyjne. Stąd, te pierwsze są unikatowe dla poszczególnych szczepów, natomiast sekwencje starsze ewolucyjnie mogą być wspólne dla wielu izolatów w obrębie gatunku a nawet różnych gatunków [30]. Zatem obecność różnych sekwencji łącznikowych, specyficznych dla poszczególnych szczepów w obrębie określonego gatunku oraz specyficznych dla różnych gatunków w obrębie rodzajów, może stanowić bardzo dobry marker dla ustalania filogenetycznych pokrewieństw i przydatny także w taksonomii bakterii.

Piśmiennictwo

- Barrangou R., Fremaux C., Deveau H., Richards M., Moineau S., Romero D.A., Horvath P.: CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712 (2007)
- Bland C., Ramsey T.L., Sabree F., Lowe M., Brown K., Kyrioples N.C., Hugenholtz P.: CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly palindromic repeats. *BMC Bioinformatics*, **8**, 209 (2007)
- Bolotin A., Quinquis B., Sorokin A., Ehrlich S.D.: Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**, 2551–2561 (2005)
- Brouns S.J.J.: Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960 (2008)
- Brudey K., Driscoll J.R. i wsp. (cytowana praca jest dziełem 66 autorów): *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* **6**, 23 (2006)
- Bult C. J., White O. i wsp. (cytowana praca jest dziełem 23 autorów): Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058–1073 (1996)
- Crawford J.T.: Genotyping in contact investigations: a CDC perspective. *Int. J. Tuberc. Lung. Dis.* **7**, S453–S457 (2003)
- Deveau H., Barrangou R., Garneau J.E., Labonte J., Fremux C., Boyaval P., Romero D.A., Horvath P., Moineau S.: Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008)
- Ebihara A., Yao M., Ryoji M., Isao T., Shigeyki Y., Seiki K.: Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci.* **15**, 1494–1499 (2006)
- Edgar R.C.: PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics*, **8**, 18 (2007)
- Godde J.S., Bickerton A.: The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* **62**, 718–729 (2006)
- Grissa I., Vergnaud G., Pourcel C.: The CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, **8**, 172 (2007)
- Haft D.H., Selengut J., Mongodin E.F., Nelson K.E.A.: A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60 (2005)
- Hannon G.J.: RNA interference. *Nature*, **418**, 244–251 (2002)
- Hermans P., van Soolingen D., Bik E.M., de Haas P.E., Dale J.W., van Embaden J.D.: Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect. Immun.* **59**, 2695–2705 (1991)
- Ishino Y., Shinagawa H., Makino K., Amemura M., Nakata A.: Nucleotide sequence of the *iap* gene responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987)
- Jansen R., van Embaden J.D., Gaastra W., Schouls L.M.: Identification of a novel family sequence repeats among prokaryotes. *OMICS*, **6**, 23–33 (2002)
- Jansen R., van Embaden J.D., Gaastra W., Schouls L.M.: Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002)
- Klenk H.P.: The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, **390**, 364–370 (1997)
- Kunin V., Sorek R., Hugenholtz P.: Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**, R61 (2007)
- Lillestol R.K., Redder P., Garret R.A., Brüger K.: A putative viral defense mechanisms in archeal cells. *Archea*, **2**, 59–72 (2006)

22. Makarova K.S., Aravind L., Grishin N.V., Rogozin I.B., Koonin E.V.A.: DNA repair system specific for thermophilic Archea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* **30**, 482–498 (2002)
23. Makarova K.S., Grishin N.V., Shabalina S.A., Wolf Y.I., Koonin E.V.A.: 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. *Biol. Direct.* **1**, 7 (2007)
24. Mojica F.J., Ferrer C., Juez G., Rodriguez-Valera F.: Long stretches of hord tandem repeats are present in the largest replicons of the Archea *Haloferax mediterranei* an *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* **17**, 85–93 (1995)
25. Mojica F.J., Diez-Villasenor C., Soria E., Juez G.: Biological significance of a family of regularly spaced repeats in the genomes of Archea, Bacteria and mitochondria. *Mol. Microbiol.* **36**, 244–246 (2000)
26. Mojica F.J., Diez-Villaasenor C., Garcia-Martinez J., Soria E.: Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005)
27. Nakata A., Amemura M., Makino K.: Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol.* **171**, 3553–3556 (1989)
28. Nelson K., Clayton R.A. i wsp.(cytowana praca jest dziełem 29 autorów): Evidence for lateral gene transfer between Archea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, **399**, 323–329 (1999)
29. Peng X., Brugger K., Shen B., Chen L., She Q., Garrett R.A. Genus-specific protein binding to the large clusters of DNA repeats (short regularly spaced repeats) present in Sulfolobus genomes. *J. Bacteriol.* **185**, 2410–2417 (2003)
30. Pourcel C., Salvignol G., Vergnaud G.: CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, **151**, 653–663 (2005)
31. Smith D.R., Doucette-Stamm L.A. i wsp. (cytowana praca jest dziełem 25 autorów): Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135–7155 (1997)
32. Sontheimer E.J.: Assembly and function of RNA silencing complexes. *Nature Rev. Mol. Cell. Biol.* **6**, 127–138 (2005)
33. Sturino J.M., Klaenhammer T.R.: Engineered bacteriophage-defence system in bioprocessing. *Nature Rev. Microbiol.* **4**, 395–404 (2006)
34. Tang T. H., Bachelierie J.P., Rozhdestvensky T., Bortolin M.L., Huber H., Drungowski M., Elge T., Brosius J., Huttenhofer A.: Identification of 86 candidates for small non-messenger RNAs from the archeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. USA*, **99**, 7536–7541 (2002)
35. Tang T.H., Polacek N., Zywicki M., Huber H., Brugger K., Garrett R., Bachelierie J.P., Hittenhofer A.: Identification of a novel non-coding RNAs as potential antisense regulators in the archeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **55**, 469–481 (2005)
36. Tyson G.W., Banfield J.F.: Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10**, 200–207 (2008)
37. Young R.F. III: Secret Weapon. *Science*, **321**, 922–923 (2008)