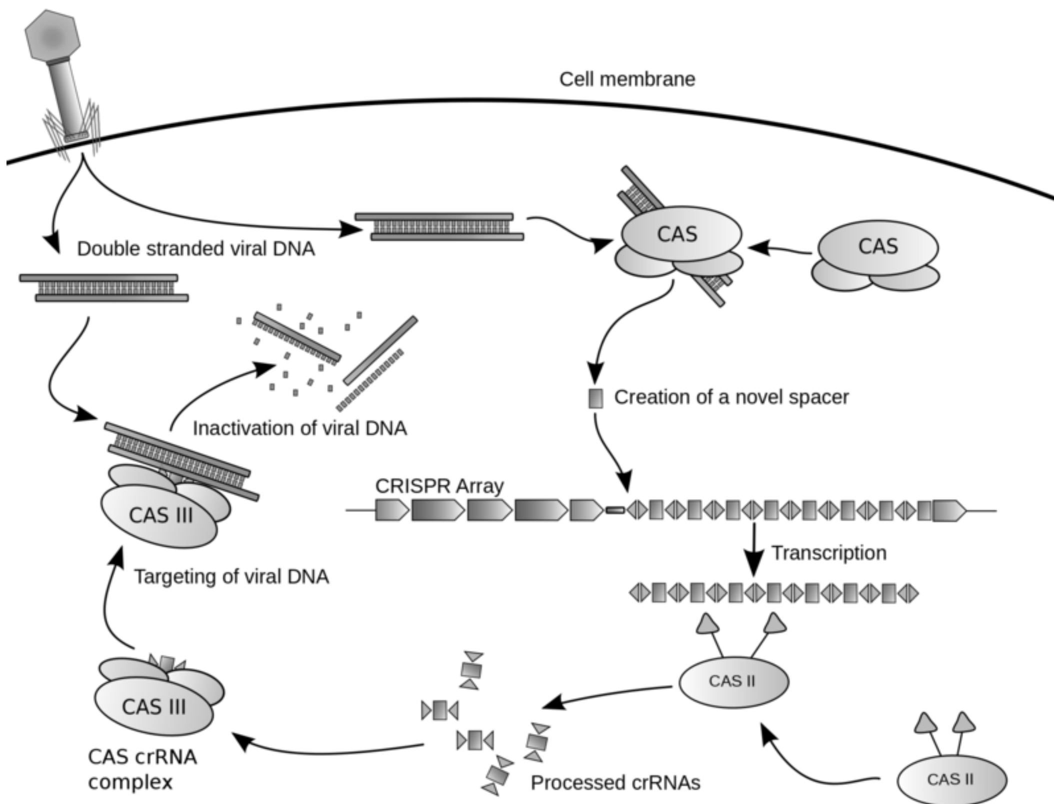
CRISPR

From Wikipedia, the free encyclopedia

Clustered regularly-interspaced short palindromic repeats (abbreviated as **CRISPR**, pronounced *crisper*[2]) are segments of prokaryotic DNA containing short repetitions of base sequences. Each repetition is followed by short segments of "spacer DNA" from previous exposures to a bacterial virus or plasmid.[3]

The **CRISPR/Cas system** is a prokaryotic immune system that confers resistance to foreign genetic elements such as plasmids and phages,[4][5] and provides a form of acquired immunity. CRISPR spacers recognize and cut these exogenous genetic elements in a manner analogous to RNA interference in eukaryotic organisms.[3] CRISPRs are found in approximately 40% of sequenced bacteria genomes and 90% of sequenced archaea.[6][note 1]

Diagram of the CRISPR prokaryotic viral defense mechanism.[1]



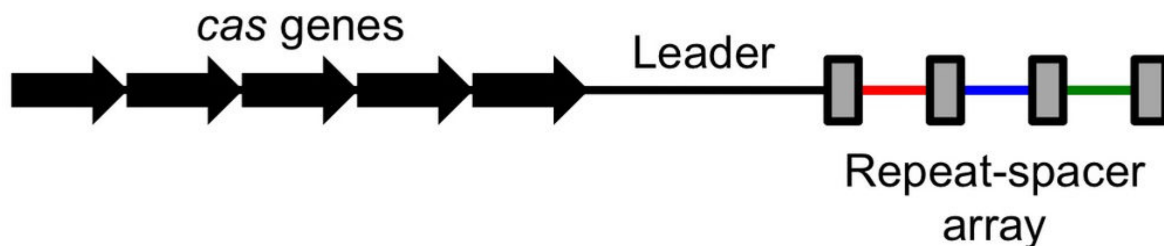
CRISPR has enormous potential application, including altering the germline of humans, animals and other organisms, and manipulating the genes of food crops. By delivering the Cas9 protein and appropriate guide RNAs into a cell, the organism's genome can be cut at

any desired location.[7][8][9] CRISPRs have been used in concert with specific endonuclease enzymes for gene editing (adding, disrupting or changing the sequence of specific genes) and gene regulation in species throughout the tree of life.[10] Ethical concerns have been expressed about this nascent biotechnology and the prospect of intentional gene editing by humans.[11]

History

CRISPR is part of a normally occurring bacterial process, though it has only recently been studied, as bacteria may incorporate foreign DNA in other circumstances and even scavenge damaged DNA from their environment.[12]

Clustered repeats were first described in 1987 for the bacterium *Escherichia coli* by Yoshizumi Ishino, but at that time their function was not known.[13] In 2000, similar repeats were identified in other bacteria and archaea, and were termed Short Regularly Spaced Repeats (SRSR).[14] SRSR were renamed CRISPR in 2002.[15] A set of genes was found to be associated with CRISPR repeats, and was named the *cas*, or *CRISPR-associated*, genes. The *cas* genes encode putative nuclease or helicase proteins, which are enzymes that can cut or unwind DNA.[15]



Simplified diagram of a CRISPR locus. The three major components of a CRISPR locus are shown: *cas* genes, a leader sequence, and a repeat-spacer array. Repeats are shown as grey boxes and spacers are colored bars. The arrangement of the three components is not always as shown.[1][3] In addition, several CRISPRs with a similar sequences can be present in a single genome, only one of which is associated with cas genes.[6]

In 2005, three independent research groups showed that some CRISPR spacers are derived from phage DNA and extrachromosomal DNA such as plasmids.[16][17][18] In effect, the spacers are fragments of DNA gathered from viruses that have previously tried to attack the cell. The source of the spacers was a sign that the CRISPR/cas system could have a role in adaptive immunity in bacteria.[1][19] Koonin and colleagues proposed that spacers serve as a template for RNA molecules, analogous to a system called RNA interference used by eukaryotic cells.[20]

In 2007, Barrangou, Horvath (food industry scientists at Danisco) and Moineau's group at Université Laval (Canada) showed that they could use spacer DNA to alter the resistance of *Streptococcus thermophilus* to phage attack.[20]

Doudna and Charpentier had independently been exploring CRISPR-associated proteins to learn how bacteria use spacers in their immune defenses.[21] They jointly studied a simpler CRISPR system that relies on a protein called Cas9. They found that bacteria respond to an invading phage by transcribing spacers and palindromic DNA into a long RNA molecule. The cell then uses tracrRNA and Cas9 to cut this long RNA molecule into pieces called crRNAs.[20]

Cas9 is a nuclease, an enzyme specialized for cutting DNA. It has two active cutting sites (HNH and RuvC), one for each strand of the DNA's double helix. The team demonstrated that they could disable one or both sites while preserving Cas9's ability to home in on its target DNA. Jinek combined tracrRNA and spacer RNA into a "single-guide RNA" molecule that, mixed with Cas9, could find and cut the correct DNA targets. Jinek *et al* proposed that such synthetic guide RNAs could be used for gene editing.[22]

CRISPR was first shown to work as a genome engineering/editing tool in human cell culture in 2012.[22][23] It has since been used in a wide range of organisms including baker's yeast (*S. cerevisiae*),[24] zebrafish (*D. rerio*),[25] fruit flies (*D. melanogaster*),[26] axolotl (*A. mexicanum*),[27] nematodes (*C. elegans*),[28] plants,[29] mice,[30] monkeys and[31] human embryos.[32]

CRISPR has been modified to make programmable transcription factors that allow scientists to target and activate or silence specific genes.[33]

Libraries of tens of thousands of guide RNAs are available.[20]

Predecessors

In the early 2000s, researchers developed zinc finger nucleases, synthetic proteins whose DNA-binding domains enable them to create double-stranded breaks in DNA at specific spots. In 2010, synthetic nucleases called TALENs provided an easier way to target a double-strand break to a specific location on the DNA strand. Both zinc-finger nucleases and TALENs require researchers to make a custom protein for each targeted DNA sequence, which is a more difficult and time-consuming process than that for guide RNAs. CRISPRs are much easier to design because they make a short RNA sequence that is paired to the targeted DNA sequence, rather than engineering a custom protein.[34]

Locus structure

Repeats and spacers

CRISPR repeats range in size from 24 to 48 base pairs.[35] They usually show some dyad symmetry, implying the formation of a secondary structure such as a hairpin, but are not truly palindromic.[36] Repeats are separated by spacers of similar length.[35] Some CRISPR spacer sequences exactly match sequences from plasmids and phages,[16][17][18] although some spacers match the prokaryote's genome (self-targeting spacers).[16][37] New spacers can be added rapidly as part of the immune response to phage infection.[38]

Cas genes and CRISPR subtypes

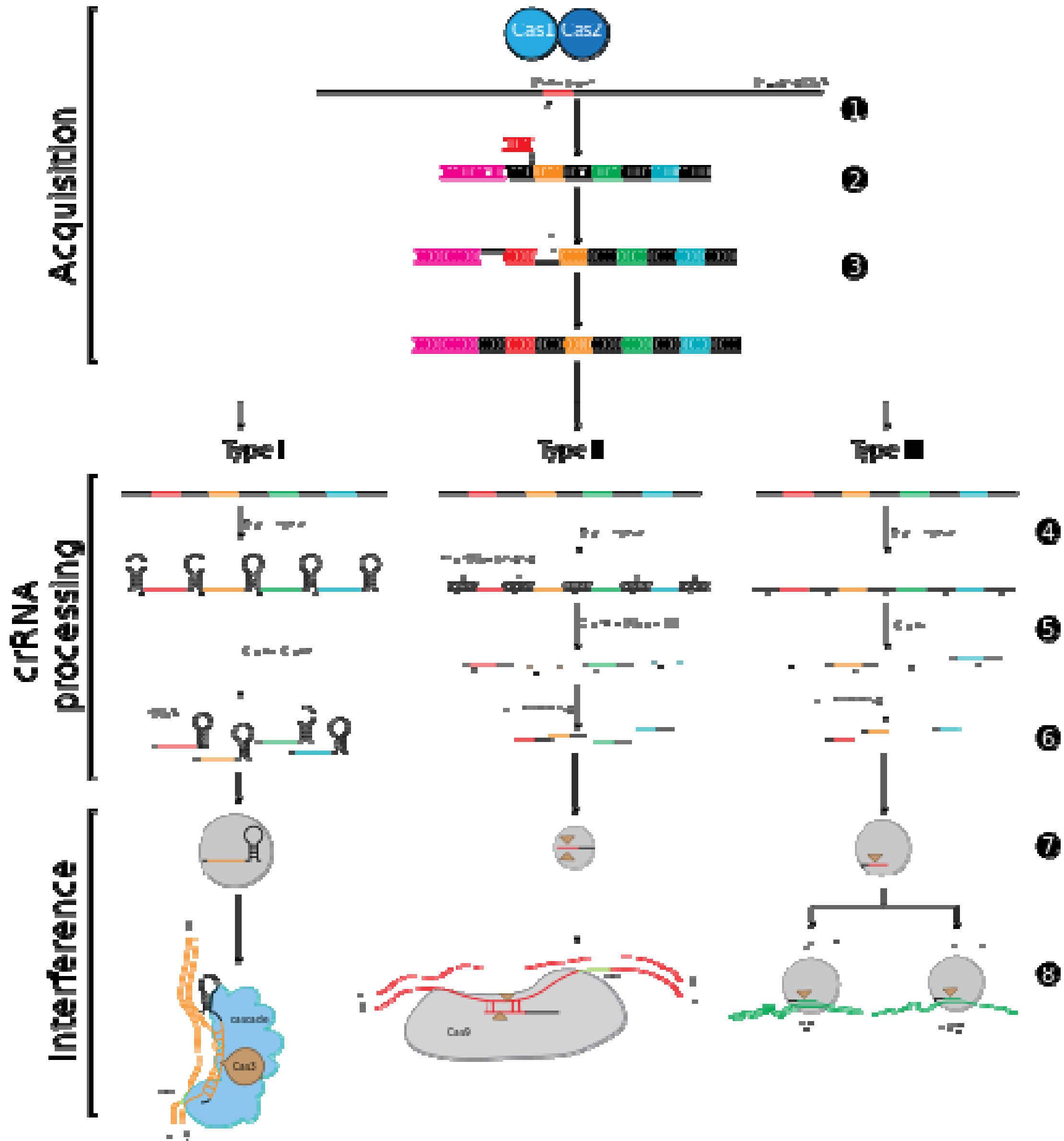
Cas genes are often associated with CRISPR repeat-spacer arrays. Comparative genomics identified multiple cas genes; an initial analysis of 200 bacterial and archaeal genomes suggested as many as 45 cas gene families. Only cas1 and cas2 genes are present in all 45 families.[35] The current CRISPR classification groups cas operons into three major divisions, each with multiple subdivisions based on cas1 phylogeny and cas operon gene complement.[39] Aside from cas1 and cas2, each major division's operons have a common set of constituent genes. Each subdivision is characterised by a 'signature gene' found exclusively in that subdivision. Many organisms contain multiple CRISPR-Cas systems suggesting that they are compatible and may share components.[40][41] The sporadic distribution of the CRISPR/Cas subtypes suggests that the CRISPR/Cas system is subject to horizontal gene transfer during microbial evolution.

Signature genes and their putative functions for the major and minor CRISPR-cas types.

Cas type	Signature gene	Function	Reference
I	Cas3	Single-stranded DNA nuclease (HD domain) and ATP-dependent helicase	[42]
IA	Cas8a	Subunit of the interference module. Important in targeting of invading DNA by recognizing the PAM sequence	[43][44]
IB	Cas8b		
IC	Cas8c		
ID	Cas10d	contains a domain homologous to the palm domain of nucleic acid polymerases and nucleotide cyclases	[39][45]
IE	Cse1		
IF	Csy1	Not determined	
II	Cas9	Nucleases RuvC and HNH together produce DSBs, and separately can produce single-strand breaks. Ensures the acquisition of functional spacers during adaptation.	[46][47]
IIA	Csn2	Not Determined	
IIB	Cas4	Not Determined	
IIC		Characterized by the absence of either Csn2 or Cas4	[48]
III	Cas10	Homolog of Cas10d and Cse1	[45]

IIIA	Csm2	Not Determined
IIIB	Cmr5	Not Determined

Mechanism



The stages of CRISPR immunity for each of the three major types of adaptive immunity. (1) Acquisition begins by recognition of invading DNA by Cas1 and Cas2 and cleavage of a protospacer. (2) The protospacer is ligated to the direct repeat adjacent to the leader sequence and (3) single strand extension repairs the CRISPR and duplicates the direct repeat. The crRNA processing and interference stages occur differently in each of the three major CRISPR systems. (4) The primary CRISPR transcript is cleaved by cas genes to produce crRNAs. (5) In type I systems Cas6e/Cas6f cleave at the junction of ssRNA and dsRNA formed by hairpin loops in the direct repeat. Type II systems use a trans-activating (tracr) RNA to form dsRNA, which is cleaved by Cas9 and RNaseIII. Type III systems use a Cas6 homolog that does not require hairpin loops in the direct repeat for cleavage. (6) In type II and type III systems secondary trimming is performed at either the 5' or 3' end to produce mature crRNAs. (7) Mature crRNAs associate with Cas proteins to form interference complexes. (8) In type I and type II systems, interactions between the protein and PAM sequence are required for degradation of invading DNA. Type III systems do not require a PAM for successful degradation and in type III-A systems basepairing occurs between the crRNA and mRNA rather than the DNA, targeted by type III-B systems.

Spacer acquisition

When a microbe is invaded by a virus, the first stage of the immune response is to capture viral DNA and insert it into a CRISPR locus in the form of a spacer. Cas1 and Cas2 are found in all three types of CRISPR-Cas immune systems, which indicates that they are involved in spacer acquisition. Mutation studies confirmed this hypothesis, showing that removal of cas1 or cas2 stopped spacer acquisition, without affecting CRISPR immune response.[43][49][50][51][52]

Multiple Cas1 proteins have been characterised and their structures resolved.[53][54][55] Cas1 proteins have diverse amino acid sequences. However, their crystal structures are similar and all purified Cas1 proteins are metal-dependent nucleases/integrases that bind to DNA in a sequence-independent manner.[40] Representative Cas2 proteins have been characterised and possess either ssRNA-[56] or dsDNA-[57][58] specific endoribonuclease activity.

In the I-E system of *E. coli* Cas1 and Cas2 form a complex where a Cas2 dimer bridges two Cas1 dimers.[59] In this complex Cas2 performs a non-enzymatic scaffolding role,[59] binding double-stranded fragments of invading DNA, while Cas1 binds the single-stranded flanks of the DNA and catalyses their integration into CRISPR arrays.[60][61][62]

Protospacer adjacent motifs

Bioinformatic analysis of regions of phage genomes that were excised as spacers (termed protospacers) revealed that they were not randomly selected but instead were found adjacent to short (3 – 5 bp) DNA sequences termed protospacer adjacent motifs (PAM). Analysis of CRISPR-Cas systems from the three major divisions showed PAMs to be important for type I and type II, but not type III systems during

acquisition.[18][63][64][65][66][67] In type I and type II systems, protospacers are excised at positions adjacent to a PAM sequence, with the other end of the spacer cut using a ruler mechanism, thus maintaining the regularity of the spacer size in the CRISPR array.[68][69] The conservation of the PAM sequence differs between CRISPR-Cas systems and appears to be evolutionarily linked to Cas1 and the leader sequence.[67][70]

New spacers are added to a CRISPR array in a directional manner,[16] occurring preferentially,[38][63][64][71][72] but not exclusively, adjacent[66][69] to the leader sequence. Analysis of the type I-E system from *E. coli* demonstrated that the first direct repeat, adjacent to the leader sequence is copied, with the newly acquired spacer inserted between the first and second direct repeats.[51][68]

The PAM sequence appears to be important during spacer insertion in type I-E systems. That sequence contains a strongly conserved final nucleotide (nt) adjacent to the first nucleotide of the protospacer). This nucleotide becomes the final base in the first direct repeat.[52][73][74] This suggests that the spacer acquisition machinery generates single-stranded overhangs in the second-to-last position of the direct repeat and in the PAM during spacer insertion. However, not all CRISPR-Cas systems appear to share this mechanism as PAMs characterised in other organisms do not show the same level of conservation in the final position.[70] It is likely that in those systems, a blunt end is generated at the very end of the direct repeat and the protospacer during acquisition.

Insertion variants

Recent analysis of *Sulfolobus solfataricus* CRISPRs revealed further complexities to the canonical model of spacer insertion as one of its six CRISPR loci inserted new spacers randomly throughout its CRISPR array, as opposed to inserting closest to the leader sequence.[69]

Multiple CRISPRs contain many spacers to the same phage. The mechanism that causes this phenomenon was elucidated in the type I-E system of *E. coli*. A significant enhancement in spacer acquisition was detected where spacers already target the phage, even mismatches to the protospacer. This ‘priming’ requires the Cas proteins involved in both acquisition and interference to interact with each other. Newly acquired spacers that result from the priming mechanism are always found on the same strand as the spacer that caused the priming.[52][73][74] This observation led to the hypothesis that the acquisition machinery slides along the foreign DNA after priming to find a new protospacer.[74]

Interference stage

The CRISPR immune response occurs through two steps: CRISPR-RNA (crRNA) biogenesis and crRNA-guided interference.

Biogenesis

A CRISPR array is transcribed from a promoter in the leader into a single long transcript.[43][75][76] This transcript is processed by cleavage inside the repeat sequence to form crRNAs. The mechanisms to produce mature crRNAs differ greatly between the three main CRISPR-Cas systems. In both type I-E and type I-F systems, the proteins Cas6e and Cas6f respectively, recognise stem-loops[77][78][79] created by the palindromic nature of the direct repeats.[36] These proteins cleave the primary transcript at the junction between double-stranded and single-stranded RNA, leaving an 8 nt 5 -handle[*clarification needed*] originating from the repeat on mature crRNAs along with a single spacer sequence.

Type III systems also use Cas6, however their repeats do not produce stem-loops. Cleavage instead occurs by the primary transcript wrapping around the Cas6 to allow cleavage 8 nt upstream of the repeat spacer junction.[80][81][82]

Type II systems lack the Cas6 gene and instead utilize RNaseIII for cleavage. Functional type II systems encode an extra small RNA that is complementary to the repeat sequence, known as a trans-activating RNA (tracrRNA).[49] Transcription of the tracrRNA and the primary CRISPR transcript results in base pairing and the formation of dsRNA at the repeat sequence, which is subsequently targeted by RNaseIII to produce crRNAs. Unlike the other two systems the crRNA does not contain the full spacer but instead is truncated at one end by 10 nt.[46]

CrRNAs associate with Cas proteins to form ribonucleotide complexes that recognize foreign nucleic acids. CrRNAs show no preference between the coding and non-coding strands, which is indicative of an RNA-guided DNA-targeting system.[5][43][52][83][84][85][86] The type I-E complex (commonly referred to as Cascade) requires five Cas proteins arranged in a ‘seahorse’ conformation, bound to a single crRNA that runs down the spine.[87][88]

Interference

During the interference stage in type I systems the PAM sequence is recognized on the crRNA-complementary strand and is required along with crRNA annealing. In type I systems correct base pairing between the crRNA and the protospacer signals a conformational change in Cascade that recruits Cas3 for DNA degradation.

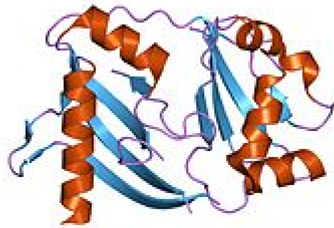
Type II systems rely on a single multifunctional protein, Cas9, for the interference step.[46] Cas9 requires both the crRNA and the tracrRNA to function and cleaves DNA using its dual HNH and RuvC/RNaseH-like endonuclease domains. Basepairing between the PAM and the phage genome is also required in type II systems, however the PAM is recognized on the same strand as the crRNA (the opposite strand to type I systems).

Type III systems, like type I require a multi-protein complex to associate with the crRNA. Biochemical and structural analyses of complexes from *S. solfataricus* and *Pyrococcus furiosus* have elucidated that six or seven cas proteins bind to crRNAs, respectively.[89][90] Surprisingly, the type III systems analysed from *S. solfataricus* and *P. furiosus* both

target the mRNA of phage/plasmids,[41][90] which may make these systems uniquely capable of targeting RNA-based phage genomes.[40]

The mechanism for distinguishing self from foreign DNA during interference is built into the crRNAs and is therefore likely common to all three systems. Throughout the distinctive maturation process of each major type, all crRNAs contain a spacer sequence and some portion of the repeat at one or both ends. It is the partial repeat sequence that prevents the CRISPR-Cas system from targeting the chromosome as base pairing beyond the spacer sequence signals self and prevents DNA cleavage.[91] RNA-guided CRISPR enzymes are classified as type V restriction enzymes.

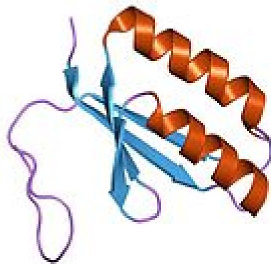
CRISPR associated protein



crystal structure of a crispr-associated protein from thermus thermophilus

Identifiers	
Symbol	CRISPR_assoc
Pfam	PF08798
Pfam clan	CL0362
InterPro	IPR010179
CDD	cd09727
[show]Available protein structures:	

CRISPR associated protein Cas2



crystal structure of a hypothetical protein tt1823 from thermus thermophilus

Identifiers

Symbol	CRISPR_Cas2
Pfam	PF09827
InterPro	IPR019199
CDD	cd09638
[show]Available protein structures:	

CRISPR-associated protein Cse1

Identifiers	
Symbol	CRISPR_Cse1
Pfam	PF09481
InterPro	IPR013381
CDD	cd09729
[show]Available protein structures:	

CRISPR-associated protein Cse2

Identifiers	
Symbol	CRISPR_Cse2
Pfam	PF09485
InterPro	IPR013382
CDD	cd09670
[show]Available protein structures:	

Evolution and diversity

The basic model of CRISPR evolution is one where newly incorporated spacers drive phages to mutate their genomes to avoid the bacteria immune response, creating diversity in both the phage and host populations. To fight off a phage infection, the sequence of the CRISPR spacer must correspond perfectly to the sequence of the target phage gene. Phages can continue to infect their hosts given point mutations in the spacer.[91] Similar stringency is required in PAM or the bacterial strain will remain phage sensitive.[64][91]

A study of 124 *S. thermophilus* strains showed that 26% of all spacers were unique and that different CRISPR loci showed different rates of new spacer acquisition.[63] Particular CRISPR loci evolve more rapidly than others, which allowed the strains' phylogenetic relationships to be determined. A comparative genomic analysis showed that *E. coli* and *S. enterica* evolve much more slowly than *S. thermophilus*. The latter's strains that had diverged 250 thousand years ago still contained the same spacer complement.[92]

Metagenomic analysis of two acid mine drainage biofilms showed that one of the analyzed CRISPRs contained extensive deletions and spacer additions in comparison to the other biofilm, suggesting a higher phage activity/prevalence in one community compared to the other.[38] In the oral cavity, a temporal study determined that 7-22% of

spacers were shared over 17 months within an individual while less than 2% of spacers were shared between individuals.[72]

From the same environment a single strain was tracked using PCR primers specific to its CRISPR system. Unlike the broad-level results of spacer presence/absence, which showed significant diversity, this CRISPR added 3 spacers over 17 months,[72] suggesting that even in an environment with significant CRISPR diversity some loci evolve slowly. CRISPRs have also been analysed from the metagenomes produced for the human microbiome project.[93] Although most CRISPRs were body-site specific, some CRISPRs within a body site are widely shared among individuals. One of these CRISPR loci originated from streptococcal species and contained ~15,000 spacers, 50% of which were unique. Similar to the targeted studies of the oral cavity, some of the CRISPRs showed little evolution over time.[93]

CRISPR evolution has been studied in chemostats using *S. thermophilus* to explicitly examine spacer acquisition rates. In one week, *S. thermophilus* strains acquired up to three spacers when challenged with a single phage.[94] During the same interval the phage developed single nucleotide polymorphisms that became fixed in the population, suggesting that CRISPR targeting had prevented phage replication absent these mutations.[94] Other *S. thermophilus* experiments showed that phages can still infect and replicate in hosts that have only one targeting spacer and that sensitive hosts can exist in environments with high phage titres.[95] The chemostat and observational studies suggest many nuances to CRISPR and phage (co)evolution.

Identification

CRISPRs are widely distributed among bacteria and archaea[39] and show some sequence similarities.[36] However their most notable characteristic is their repeating spacers and direct repeats. This characteristic makes CRISPRs easily identifiable in long sequences of DNA, since the number of repeat copies decreases the likelihood of a false positive match. Three programs are used for CRISPR repeat identification that search for regularly interspaced repeats in long sequences: CRT,[96] PILER-CR[97] and CRISPRfinder.[98]

Analysis of CRISPRs in metagenomic data is more challenging, as CRISPR loci do not typically assemble due to their repetitive nature or through strain variation, which confuses assembly algorithms. Where many reference genomes are available, polymerase chain reaction (PCR) can be used to amplify CRISPR arrays and analyse spacer content.[63][72][99][100][101] However, this approach yields information only for specifically targeted CRISPRs and for organisms with sufficient representation in public databases to design reliable PCR primers.

The alternative is to extract and reconstruct CRISPR arrays from shotgun metagenomic data. This is computationally more difficult, particularly with second generation sequencing technologies (e.g. 454, Illumina), as the short read lengths prevent more than two or three repeat units appearing in a single read. CRISPR identification in raw reads has been achieved using purely denovo identification[102] or by using direct repeat

sequences in partially assembled CRISPR arrays from contigs (overlapping DNA segments that together represent a consensus region of DNA)[93] and direct repeat sequences from published genomes[103] as a hook for identifying direct repeats in individual reads.

Evolutionary significance

A bioinformatic study showed that CRISPRs are evolutionarily conserved and cluster into related types. Many show signs of a conserved secondary structure.[36]

Through the CRISPR/Cas mechanism, bacteria can acquire immunity to certain phages and thus halt further transmission of targeted phages. For this reason, Eugene Koonin has described CRISPR/Cas as a Lamarckian inheritance mechanism.[104] However, this has been disputed by a recent critic noting "We should remember [Lamarck] for the good he contributed to science, not for things that resemble his theory only superficially. Indeed, thinking of CRISPR and other phenomena as Lamarckian only obscures the simple and elegant way evolution really works.[105]

Analysis of CRISPR sequences revealed coevolution of host and viral genomes.[106] Cas9 proteins are highly enriched in pathogenic and commensal bacteria. CRISPR/Cas-mediated gene regulation may contribute to the regulation of endogenous bacterial genes, particularly during interaction with eukaryotic hosts. For example, *Francisella novicida* uses a unique, small, CRISPR/Cas-associated RNA (scaRNA) to repress an endogenous transcript encoding a bacterial lipoprotein that is critical for *F. novicida* to dampen host response and promote virulence.[107]

Use by phages

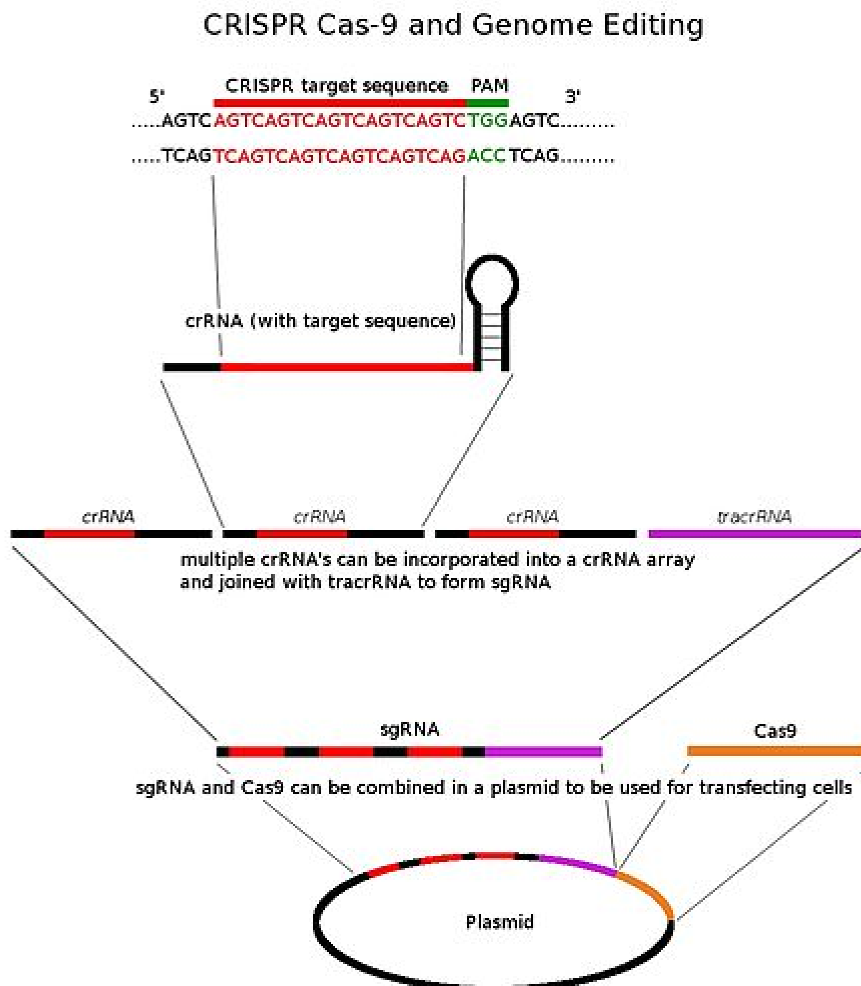
Another way for bacteria to defend against phage infection is by having chromosomal islands. A subtype of chromosomal islands called phage-inducible chromosomal island (PICI) is excised from a bacterial chromosome upon phage infection and can inhibit phage replication.[108] The mechanisms that induce PICI excision and how PICI inhibits phage replication are not well understood. One study showed that lytic ICP1 phage, which specifically targets *Vibrio cholerae* serogroup O1, has acquired a CRISPR/Cas system that targets a *V. cholera* PICI-like element. The system has 2 CRISPR loci and 9 Cas genes. It seems to be homologous to the 1-F system found in *Yersinia pestis*. Moreover, like the bacterial CRISPR/Cas system, ICP1 CRISPR/Cas can acquire new sequences, which allows phage and host to co-evolve.[109]

Applications

By the end of 2014 some 600 research papers had been published that mentioned CRISPR.[110] The technology has been used to functionally inactivate genes in human cell lines and cells, to study *Candida albicans*, to modify yeasts used to make biofuels and to genetically modify crop strains.[110]

Genome engineering

CRISPR/Cas9 genome editing is carried out with a Type II CRISPR system. When utilized for genome editing, this system includes Cas9, CRISPR RNA (crRNA), trans-activating crRNA (tracrRNA) along with an optional section of DNA repair template that is utilized in either Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR).



graphical overview of CRISPR Cas9 plasmid construction[111][112]

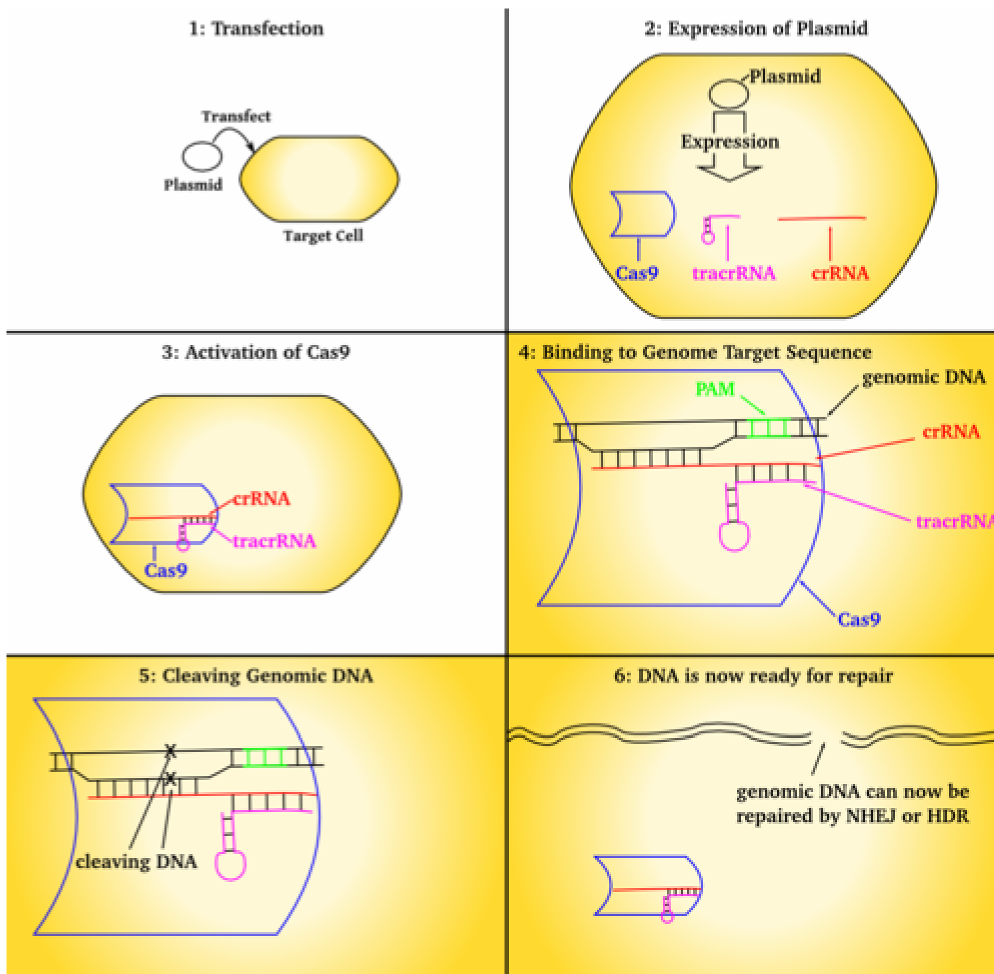
Major components

Component	Function
crRNA	Contains the RNA used by Cas9 to guide it to the correct section of host DNA along with a region that binds to tracrRNA (generally in a hairpin loop form) forming an active complex with Cas9
tracrRNA	Binds to crRNA and forms an active complex with Cas9

sgRN	Single guide RNAs are a combined RNA consisting of a tracrRNA and at least one crRNA
Cas9	Protein that in its active form is able to modify DNA utilizing crRNA as its guide. Many variants exist with differing functions (i.e. single strand nicking, double strand break, DNA binding) due to Cas9's DNA site recognition function that is independent of its two DNA cleaving domains (one for each strand).
Repair template	DNA that guides the cellular repair process allowing insertion of a specific DNA sequence

CRISPR/Cas9 often employs a plasmid to transfect the target cells. The main components of this plasmid are displayed in the image and listed in the table The crRNA needs to be designed for each application as this is the sequence that Cas9 uses to identify and directly bind to the cell's DNA. The crRNA must bind only where editing is desired. The repair template must also be designed for each application, as it must overlap with the sequences on either side of the cut and code for the insertion sequence.

Multiple crRNA's and the tracrRNA can be packaged together to form a single-guide RNA (sgRNA). This sgRNA can be joined together with the Cas9 gene and made into a plasmid in order to be transfected into cells (see image for overview).



overview of the transfection and DNA cleaving by CRISPR Cas9 (crRNA and tracrRNA are often joined as one strand of RNA when designing a plasmid)[113]

Structure

CRISPR/Cas9 is a widely used system for genome editing due to its high degree of fidelity and relatively simple construction. CRISPR/Cas9 depends on two factors for its specificity – the CRISPR target sequence and the PAM. The CRISPR target sequence is 20 bases long as part of each CRISPR locus in the crRNA array.[113] A typical crRNA array has multiple unique target sequences. Cas9 proteins select the correct location on the host's genome by utilizing the sequence for base pair bonding with the host DNA. The sequence is not part of the Cas9 protein and as a result is customizable and can be independently synthesized.[114][115]

The PAM sequence on the host genome is recognized by the protein structure of Cas9 and generally cannot be easily modified to recognize a difference sequence. However this is not limiting as it is a short sequence and not very specific (e.g. the SpCas9 PAM sequence is 5'-NGG-3' and in the human genome occurs roughly every 8 to 12 base pairs).[113]

Once these have been assembled into a plasmid and transfected into cells the Cas9 protein with help of the crRNA finds the correct sequence in the host cell's DNA and – depending on the Cas9 variant – creates a single or double strand break in the DNA. Properly spaced single strand breaks in the host DNA can trigger homology directed repair, which is less error prone than non-homologous end joining that typically follows a double strand break. Providing a section of DNA repair template allows for the insertion of a specific DNA sequence at an exact location within the genome. The repair template should extend 40 to 90 base pairs beyond the Cas9 induced DNA break.[113] The goal is for the cell's HDR process to utilize the provided repair template and thereby incorporate the new sequence into the genome. Once incorporated, this new sequence is now part of the cell's genetic material and passes into its daughter cells.

Many online tools are available to aid in designing effective sgRNA sequences.[116]

Editing

CRISPRs have been used to cut five[20] to 62 genes at once: pig cells have been engineered to inactivate all 62 Porcine Endogenous Retrovirus in the pig genome, which eliminated infection from the pig to human cells in culture.[117] CRISPR's low cost compared to alternatives is widely seen as revolutionary.[7][8]

Selective engineered redirection of the CRISPR/Cas system was first demonstrated in 2012 in:[118][119]

- Immunization of industrially important bacteria, including some used in food production and large-scale fermentation
- Cellular or organism RNA-guided genome engineering. Proof of concept studies demonstrated examples both *in vitro*[9][22][46] and *in vivo*[30][120][121]
- Bacterial strain discrimination by comparison of spacer sequences

Reversible knockdown

Main article: CRISPR interference

Like RNAi, CRISPR interference (CRISPRi) turns off genes in a reversible fashion by targeting, but not cutting a site. The targeted site is methylated so the gene is epigenetically modified. This modification inhibits transcription. Cas9 is an effective way of targeting and silencing specific genes at the DNA level.[122] In bacteria, the presence of Cas9 alone is enough to block transcription. For mammalian applications, a section of protein is added. Its guide RNA targets regulatory DNA, called promoters that immediately precede the gene target.[20]

Activation

Main article: CRISPR interference

Cas9 was used to carry synthetic transcription factors (protein fragments that turn on genes) that activated specific human genes. The technique achieved a strong effect by targeting multiple CRISPR constructs to slightly different spots on the gene's promoter.[20]

Some of the affected genes tied to human diseases, including those involved in muscle differentiation, cancer, inflammation and fetal hemoglobin.[20]

Disease models

CRISPR simplifies creation of animals for research that mimic disease or show what happens when a gene is knocked down or mutated. CRISPR may be used at the germline level to create animals where the gene is changed everywhere, or it may be locally targeted.[123][124][125]

CRISPR can also be utilized to create human cellular models of disease. For instance, CRISPR was applied to human pluripotent stem cells to introduce targeted mutations in genes relevant to two different kidney diseases, polycystic kidney disease and focal segmental glomerulosclerosis.[126] These CRISPR-modified pluripotent stem cells were subsequently grown into human kidney organoids, which exhibited disease-specific phenotypes. Kidney organoids from stem cells with polycystic kidney disease mutations formed large, translucent cyst structures from kidney tubules. Kidney organoids with mutations in a gene linked to focal segmental glomerulosclerosis developed junctional defects between podocytes, the filtering cells affected in that disease. Importantly, these disease phenotypes were absent in control organoids of identical genetic background, but lacking the CRISPR mutations.[126] A similar approach has been taken to model long QT syndrome in cardiomyocytes derived from pluripotent stem cells.[127] These CRISPR-generated cellular models, with isogenic controls, provide a new way to study human diseases and test drugs that might work in human patients.

Gene drive

In 2003 evolutionary biologist Austin Burt envisioned attaching a gene that coded for a desired trait to “selfish” DNA elements that could copy themselves from one chromosome spot to another. That would bias daughter cells to inherit it, quickly spreading it throughout a population. In 2015 a U.S. team used CRISPR to create a “mutagenic chain reaction,” that drove a pigmentation trait in lab-grown Drosophila to the next generation with 97% efficiency. With another research group they created a gene drive in mosquitoes that spread genes that prevented the insects from harboring malaria parasites. Only weeks later, the team reported the a second drive with genes that rendered female mosquitoes infertile and could quickly wipe out a population. The work was done in the lab, leading to debates over the desirability of field testing.[128]

Biomedicine

Using “dead” versions of Cas9, eliminates CRISPR’s DNA-cutting ability, while preserving its ability to target desirable sequences. Multiple groups added various regulatory factors to dead Cas9s, enabling them to turn almost any gene on or off or subtly adjust its level of activity.[128]

In another 2015 experiment the 20,000 or so known human genes were separately targeted turning them on one by one in groups of cells to identify those involved in resistance to a melanoma drug. Each such gene manipulation is itself a separate "drug", potentially opening the entire genome to CRISPR-based regulation.[128]

Clinical researchers are applying it to develop tissue-based treatments for cancer and other diseases.[128]

CRISPR may revive the concept of transplanting animal organs into people. Retroviruses present in animal genomes could harm transplant recipients. In 2015 a team eliminated 62 copies of a retrovirus’s DNA from the pig genome.[128]

Patents and commercialization

As of December 2014, patent rights to CRISPR were contested. Several companies had been formed to develop related drugs and research tools.[129]

As of November 2013 SAGE Labs had exclusive rights from one of those companies to produce and sell genetically engineered rats and nonexclusive rights for mouse and rabbit models.[130]

Society and culture

Human germline modification

In light of plans or ongoing research to apply CRISPR to human embryos in at least four labs in the US, labs in China and the UK, and by a US biotechnology company called Ovascience,[131] scientists including an inventor of CRISPR, urged a worldwide moratorium on applying CRISPR to the human germline, especially for clinical use. They said "scientists should avoid even attempting, in lax jurisdictions, germline genome modification for clinical application in humans" until the full implications "are discussed among scientific and governmental organizations".[32][132] These scientists support basic research on CRISPR and do not see CRISPR as developed enough for any clinical use in making inheritable changes to people.[133]

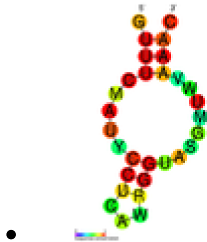
In April 2015, scientists from China published a paper in the journal *Protein & Cell* reporting results of an attempt to alter the DNA of non-viable human embryos using CRISPR to correct a mutation that causes beta thalassemia, a lethal heritable disorder.[134][135] According to the paper's lead author, the study had previously been rejected by both *Nature* and *Science* in part because of ethical concerns; the journals did

not comment to reporters.[136] The experiments resulted in changing only some of the genes, and had off-target effects on other genes. The scientists who conducted the research stated that CRISPR is not ready for clinical application in reproductive medicine. One said to a reporter at *Nature*: "If you want to do it in normal embryos, you need to be close to 100%.... That's why we stopped. We still think it's too immature."[136]

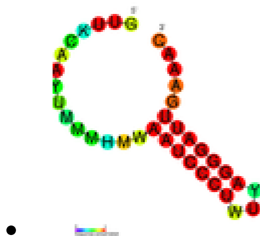
In December 2015, the International Summit on Human Gene Editing took place in Washington under the guidance of David Baltimore. Members of national scientific academies of America, Britain and China discussed the ethics of germline modification. In conclusion, they agreed to proceed further with basic and clinical research under appropriate legal and ethical guidelines. A specific distinction was made between clinical use in somatic cells, where the effects of edits are limited to a single individual, versus germline cells, where genome changes would be inherited by future generations. This could have unintended and far-reaching consequences for human evolution, genetically (e.g. gene/environment interactions) and culturally (e.g. Social Darwinism), hence altering of gametocytes and embryos to generate inheritable changes in humans was claimed irresponsible. In addition, they agreed to initiate an international forum where these concerns will be continuously addressed, and regulations in research harmonised across countries.[137]

Recognition

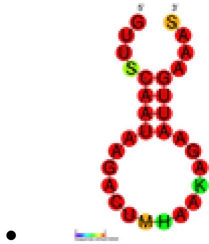
American Jennifer Doudna and French-born Emmanuelle Charpentier co-authored a key study published in August 2012 that demonstrated the technical power of Crispr-Cas9 to cut and splice genes with extreme efficiency at the highest resolution possible. In 2012 and 2013, CRISPR was a runner-up in *Science Magazine*'s Breakthrough of the Year award. In 2015, it was the winner of that award.[128]

- 

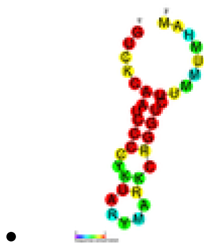
  CRISPR-DR2: Secondary structure taken from the Rfam database. Family RF01315.

- 

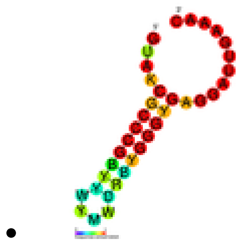
CRISPR-DR5: Secondary structure taken from the Rfam database. Family RF011318.

- 

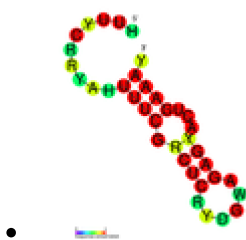
  CRISPR-DR6: Secondary structure taken from the Rfam database. Family RF01319.

- 

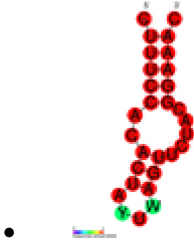
  CRISPR-DR8: Secondary structure taken from the Rfam database. Family RF01321.

- 

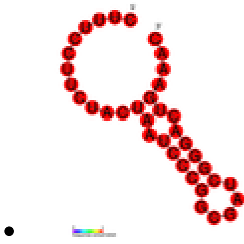
  CRISPR-DR9: Secondary structure taken from the Rfam database. Family RF01322.

- 

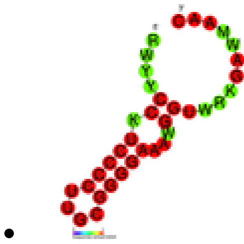
  CRISPR-DR19: Secondary structure taken from the Rfam database. Family RF01332.

- 

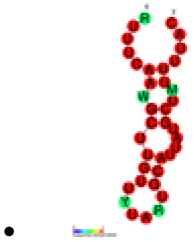
  CRISPR-DR41: Secondary structure taken from the Rfam database. Family RF01350.

- 

  CRISPR-DR52: Secondary structure taken from the Rfam database. Family RF01365.

- 

  CRISPR-DR57: Secondary structure taken from the Rfam database. Family RF01370.

- 

  CRISPR-DR65: Secondary structure taken from the Rfam database. Family RF01378.