

CRISPR'S MYSTERIES

While the world rallies round the gene-editing tool that is revolutionizing biotechnology, basic questions about how it works and where it came from still loom large.

BY HEIDI LEDFORD

Francisco Mojica was not the first to see CRISPR, but he was probably the first to be smitten by it. He remembers the day in 1992 when he got his first glimpse of the microbial immune system that would launch a biotechnology revolution. He was reviewing genome-sequence data from the salt-loving microbe *Haloflex mediterranei* and noticed 14 unusual DNA sequences, each 30 bases long. They read roughly the same backwards and forwards, and they repeated every 35 bases or so. Soon, he saw more of them. Mojica was entranced, and made the repeats a focus of his research at the University of Alicante in Spain.

It wasn't a popular decision. His lab went years without funding. At meetings, Mojica would grab the biggest bigwigs he could find and ask what they thought of the strange little

repeats. "Don't care about repeats so much," he says that they would warn him. "There are many repeats in many organisms — we've known about them for years and still don't know how many of them work."

Today, much more is known about the clustered, regularly interspaced short palindromic repeats that give CRISPR its name and help the CRISPR-Cas microbial immune system to destroy invading viruses. But although most in biomedicine have come to revere the mechanics of the system — particularly of a version called CRISPR-Cas9 — for the ways in which it can be harnessed to edit genes, Mojica and other microbiologists are still puzzling over some basic questions about the system and how it works. How did it evolve, and how did it shape microbial evolution? Why do some microbes use it, whereas others don't? And might it have

TESSA QUAY/DAVID PRANGSHVILI/GERRARD FEHAU-ARNAUDET/JEAN-MARC PANAUD

Prokaryotes use CRISPR–Cas to fight off viruses such as the one that formed this structure on a cell.

other, yet-to-be-appreciated roles in their basic biology?

“A lot of the attention paid to CRISPR systems in the media has really been around its use as a technology — and with good reason. That’s where we’re seeing incredible impact and opportunities,” says Jennifer Doudna, a molecular biologist at the University of California, Berkeley, and one of the first scientists to reveal CRISPR–Cas’s agility as a gene-editing tool. “At the same time, there’s a lot of interesting fundamental biology research to be done.”

WHERE DID IT COME FROM?

The biological advantages of something like CRISPR–Cas are clear. Prokaryotes — bacteria and less-well-known single-celled organisms called archaea, many of which live in extreme environments — face a constant onslaught of genetic invaders. Viruses outnumber prokaryotes by ten to one and are said to kill half of the world’s bacteria every two days. Prokaryotes also swap scraps of DNA called plasmids, which can be parasitic — draining resources from their host and forcing it to self-destruct if it tries to expel its molecular hitch-hiker. It seems as if nowhere is safe: from soil to sea to the most inhospitable places on the planet, genetic invaders are present.

Prokaryotes have evolved a slew of weapons to cope with these threats. Restriction enzymes, for example, are proteins that cut DNA at or near a specific sequence. But these defences are blunt. Each enzyme is programmed to recognize certain sequences, and a microbe is protected only if it has a copy of the right gene. CRISPR–Cas is more dynamic. It adapts to and remembers specific genetic invaders in a similar way to how human antibodies provide long-term immunity after an infection. “When we first heard about this hypothesis, we thought that would be way too sophisticated for simple prokaryotes,” says microbiologist John van der Oost of Wageningen University in the Netherlands.

Mojica and others deduced the function of CRISPR–Cas when they saw that DNA in the spaces between CRISPR’s palindromic repeats sometimes matches sequences in viral genomes. Since then, researchers have worked out that certain CRISPR-associated (Cas) proteins add these spacer sequences to the genome after bacteria and archaea are exposed to specific viruses or plasmids. RNA made from those spacers directs other Cas proteins to chew up any invading DNA or RNA that matches the sequence.

How did bacteria and archaea come to possess such sophisticated immune systems? That question has yet to be answered, but the leading theory is that the systems are derived from transposons — ‘jumping genes’ that can hop from one position to another in the genome. Evolutionary biologist Eugene Koonin of the US National Institutes of Health in Bethesda,

Maryland, and his colleagues have found¹ a class of these mobile genetic elements that encodes the protein Cas1, which is involved in inserting spacers into the genome. These ‘casposons’, he reasons, could have been the origin of CRISPR–Cas immunity. Researchers are now working to understand how these bits of DNA hop from one place to another — and then to track how that mechanism may have led to the sophistication of CRISPR–Cas.

HOW DOES IT WORK?

Many of the molecular details of how Cas proteins add spacers have been worked out in fine detail² in recent years. But viral DNA is chemically nearly identical to host DNA. How, in a cell packed with DNA, do the proteins know which DNA to add to the CRISPR–Cas memory?

The stakes are high: if a bacterium adds a piece of its own DNA, it risks suicide by auto-immune attack, says Virginijus Siksnys, a

a spacer becomes obsolete. And it can be a burden for microbes to retain extra DNA. “A bacterium cannot inflate its genome forever,” says Rotem Sorek, a geneticist at the Weizmann Institute of Science in Rehovot, Israel.

WHAT ELSE MIGHT IT BE DOING?

The origin of some spacers presents another mystery. Less than 3% of spacers observed so far match any known sequences in DNA databases.

It could be a reflection of how little is known about viruses. Most sequencing efforts have concentrated on those that infect people, livestock or crops. “We know very little about the enemies of bacteria, and especially the enemies of crazy archaea,” says Michael Terns, an RNA biologist at the University of Georgia in Athens.

It is also possible that some spacers are the ghosts of viruses no longer around or mutated beyond recognition. But a third possibility has the field buzzing. Researchers have found exam-

“A MAJOR QUESTION IS HOW MUCH BIOLOGY IS THERE THAT GOES BEYOND DEFENCE.”

biochemist at Vilnius University in Lithuania. “These enzymes are a double-edged sword.”

It may be that populations of bacteria and archaea can absorb some error, says Rodolphe Barrangou, a microbiologist at North Carolina State University in Raleigh. A few cellular suicides may not matter if other cells can thrive after a viral attack.

In fact, when viruses infiltrate a bacterial ecosystem, often only about one bacterium in 10 million will gain a spacer that lets it defend itself. Those odds make it hard to study what drives spacer acquisition, and to learn why a cell succeeded where others failed. “It’s difficult to catch that bacterium when it actually is happening,” says Luciano Marraffini, a microbiologist at the Rockefeller University in New York City.

Sorting out how suitable spacers are recognized — and boosting the rate at which they are incorporated — could be useful. Some work has shown that cells containing CRISPR–Cas machinery could serve as a recording device of sorts, cataloguing DNA and RNA sequences that they have encountered³. This might allow researchers to track a cell’s gene expression or exposure to environmental chemicals over time.

Researchers would also like to learn how old memories are pruned from the collection. Most microbes with CRISPR–Cas systems contain a few dozen spacers; some have only one. The archaeon *Sulfolobus tokodaii*, by contrast, dedicates 1% of its genome to its 5 CRISPR–Cas systems, including 458 spacers.

There may be little incentive to hang on to old spacers: if a virus mutates to avoid CRISPR–Cas,

ples of CRISPR–Cas systems doing more than warding off genetic intruders. In some bacteria, CRISPR–Cas components control DNA repair, gene expression and the formation of biofilms. They can also determine a bacterium’s ability to infect others: *Legionella pneumophila*, which causes Legionnaires’ disease, must have the Cas protein Cas2 in order to infect the amoeba that is its natural host. “A major question is how much biology is there that goes beyond defence,” says Erik Sontheimer, a molecular biologist at the University of Massachusetts Medical School in Worcester. “That is something where there’s still quite a few shoes to drop in the coming years.”

Sontheimer adds that it creates an enticing parallel with the discovery of RNA interference, a system that silences gene expression in plants, animals and other non-prokaryotic organisms. RNA interference was also primarily thought of as a defence mechanism early on, and it was only later that researchers noticed its role in regulating host gene expression.

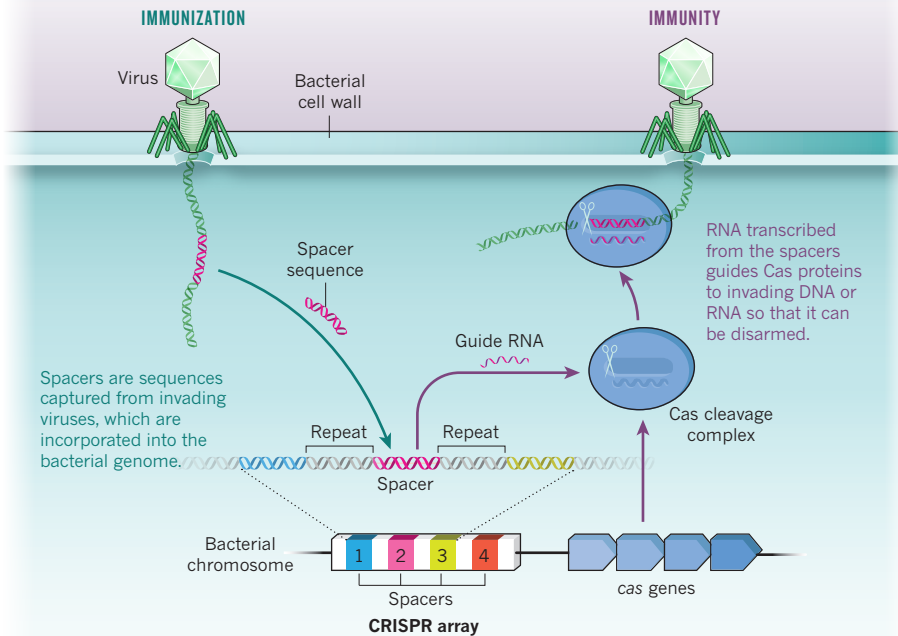
This could also explain why some spacers do not match known viruses or plasmids, says Stan Brouns, a microbiologist at Delft University of Technology in the Netherlands. “The systems are not tuned to be perfect: they grab the viral DNA as well as their own,” he says. “As soon as they start pulling in new pieces of DNA, they can gain new functions — if they don’t die.”

WHY DO ONLY SOME MICROBES USE IT?

Whatever other functions CRISPR–Cas has, it is clear that some microbes use it more than others. More than 90% of archaea have

LASTING PROTECTION

About 90% of known archaea and one-third of bacteria have some form of CRISPR–Cas immunity. This is controlled by a cluster of short DNA repeats separated by ‘spacer’ sequences and a series of nearby genes that encode CRISPR-associated (Cas) proteins.



CRISPR-based immunity, whereas only about one-third of sequenced bacteria bother with it, says Koonin. And no non-prokaryotic organisms, even single-celled ones, have been caught troubling with CRISPR–Cas at all.

One archaeon, called *Nanoarchaeum equitans*, lives as a parasite on another archaeon in near-boiling waters and has dispensed with many of its genes related to energy production and general cellular house-keeping. Yet in its minuscule, 490,000-letter DNA instruction manual, *N. equitans* has held on to a CRISPR–Cas system with about 30 spacers. “A big chunk of its genome is still dedicated to CRISPR,” says Malcolm White, a molecular biologist at the University of St Andrews, UK. “CRISPR must be so important, yet we don’t really know why.”

Such differences suggest that there are key ecological factors that favour CRISPR–Cas systems, prizing viral defence — or other benefits — over the risks of cellular suicide, says Edze Westra, a microbiologist at the Penryn campus of the University of Exeter, UK. Extreme environments seem to favour CRISPR–Cas systems, but Westra notes that the frequency of such systems also varies among bacteria in more-hospitable habitats. The bird pathogen *Mycoplasma gallisepticum*, for example, tossed out its CRISPR–Cas equipment when it switched hosts from chickens to wild finches. Why the system was useful in a chicken but not a finch is anyone’s guess, says Westra.

Mathematical models and some early laboratory experiments suggest that CRISPR–Cas may be more of an advantage when there are

only a few types of virus to contend with^{4,5}. CRISPR–Cas spacers can record a limited number of viral sequences before the added DNA becomes a genomic burden. If the diversity of viruses in the environment greatly outweighs the number of possible spacers, CRISPR–Cas systems may be of little use, says Koonin. Another possibility is that archaea in extreme environments cannot rely as heavily on other means of defence. One common way

“CRISPR MUST BE SO IMPORTANT, YET WE DON’T REALLY KNOW WHY.”

for bacteria to thwart invaders is to mutate the proteins found in their own outer casing, called an envelope. Some archaea, however, may have less freedom to tinker with these envelopes because the envelopes’ structure is so crucial to the organism’s survival in harsh conditions. “This makes alternative systems such as CRISPR more relevant,” says Mojica.

HOW MANY FLAVOURS OF CRISPR–CAS EXIST?

Humans tend to focus on the CRISPR–Cas9 system, which is prized for its simplicity and versatility in genome editing, but microbes

don’t play favourites. Instead, they tend to mix and match different systems, quickly picking up new ones from other bacteria and discarding the old.

Researchers have officially recognized 6 different types of CRISPR system, with 19 subtypes. “And we really only know how a fraction of them actually work,” says Marraffini.

Unravelling those mechanisms could hold the key to finding new biotechnological applications for CRISPR–Cas systems. The beloved CRISPR–Cas9, for example, is a type II system, which uses RNA molecules transcribed from spacer sequences to direct an enzyme to cut invading viral or plasmid DNA (see ‘Lasting protection’). But enzymes in type VI systems — discovered last year⁶ — cut up RNA rather than DNA. And type IV systems contain some genes associated with CRISPR–Cas, but lack the repeats and the machinery to insert spacers.

Type III systems are among the most commonly found CRISPR–Cas systems in nature — and among the least understood. Evidence so far suggests that they respond not to the invading DNA or RNA itself, but to the process of transcribing DNA into RNA. If that proves to be the case, it would be a new form of regulation that could expand the CRISPR–Cas toolbox for genome editing, says Doudna.

Other systems may yet crop up, particularly as researchers extend their search beyond microbes that have been grown in culture, to include genetic sequences from environmental DNA samples. “We have already said a couple of times that we reached the end,” says van der Oost — only to be surprised when a new CRISPR–Cas system surfaced.

For Mojica, exploring that diversity and answering basic questions about CRISPR systems hold more allure than the revolution they sparked. This puzzles many of his colleagues, he says. He has immersed himself in CRISPR–Cas biology for a quarter of a century, and although there’s a lot of funding available for those who wish to edit genomes, there is considerably less for the kind of work he does.

“I know that it’s a great tool. It’s fantastic. It could be used to cure diseases,” says Mojica. “But it’s not my business. I want to know how the system works from the very beginning to the end.” ■

Heidi Ledford is a senior reporter for Nature in London.

1. Krupovic, M., Makarova, K. S., Forterre, P., Prangishvili, D. & Koonin, E. V. *BMC Biol.* **12**, 36 (2014).
2. Nuñez, J. K., Lee, A. S. Y., Engelman, A. & Doudna, J. A. *Nature* **519**, 193–198 (2015).
3. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. *Science* **353**, aaf1175 (2016).
4. Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S. & Koonin, E. V. *mBio* **3**, e00456-12 (2012).
5. Westra, E. R. et al. *Curr. Biol.* **25**, 1043–1049 (2015).
6. Abudayyeh, O. O. et al. *Science* **353**, aaf5573 (2016).

CORRECTION

An earlier version of the graphic in this article misrepresented the nuclease activity of the Cas cleavage complex. It has been corrected.