

What history tells us XXXVII. CRISPR-Cas: The discovery of an immune system in prokaryotes

MICHEL MORANGE

Centre Cavallès, République des Savoires: Lettres, Sciences, Philosophie USR 3608, Ecole Normale Supérieure, 29 Rue d'Ulm, 75230, Paris Cedex 05, France

(Fax, 33-144-323941; Email, morange@biologie.ens.fr)

1. Introduction

An increasing number of articles and reviews are being published that demonstrate the efficiency of the CRISPR-Cas9 system in genome editing of very different types of organisms and describe the resulting experimental and therapeutic perspectives. A canonical historical record of the major steps that led to the present situation has already been reported (see, for instance, Horvath and Barrangou 2010 and Doudna and Charpentier 2014). In 1987, a new type of repeated sequence was discovered in prokaryotes (Ishino *et al.* 1987). Between 2000 and 2002 the major characteristics of these short, regularly spaced repeats were described: they consist of clusters of short repeated palindromic sequences of 20–40 bp separated by unique intervening 20–60 bp sequences (Mojica *et al.* 2000). They are present in most archaea and a large percentage of bacteria. In 2002, Ruud Jansen and colleagues (Jansen *et al.* 2002) introduced a new term to designate these sequences – clustered regularly interspaced short palindromic repeats (CRISPRs), and showed that they were often found close to a family of genes already described as a ‘DNA repair system’ (Makarova *et al.* 2002) and now called ‘Cas’ for ‘CRISPR-associated’. However, the origin of the spacer sequences remained unknown and it was only three years later that three groups independently showed that some of these spacer sequences have a bacteriophage origin (Mojica *et al.* 2005; Pourcel *et al.* 2005; Bolotin *et al.* 2005). In 2006, Eugene V Koonin and his collaborators suggested that the CRISPR-Cas system was a prokaryotic RNA interference-based immune system (Makarova *et al.* 2006). In 2007, Rodolphe Barrangou, Philippe Horvath and their collaborators (Barrangou *et al.* 2007)

clearly demonstrated that the CRISPR-Cas system provides acquired immunity against bacteriophages: infection leads to the formation in a polarized way of new spacer sequences and to protection against further infection, which is abolished by the deletion of the spacer sequences or of some of the *Cas* genes. In the following years, the molecular mechanisms were partially unveiled and the structural and functional characteristics of the proteins encoded by the *Cas* genes described, opening the way to the design of the CRISPR-Cas9 editing system.

In this article, I will focus on the early steps and particularly on the multiplicity of explanations that were initially provided for the existence of the CRISPR-Cas system. I will devote the major part of this contribution to the three 2005 publications. In contrast with the previous historical description, these three articles did not simply demonstrate the bacteriophage origin of some of the spacer sequences, but they collectively reached the conclusion that the CRISPR-Cas system was a memory of infection and a defence system against further infections. They provided information on the mechanism of insertion of the new spacers, and proposed hypotheses on the way these spacers prevented further infection. Therefore, the issue is what the 2006 and 2007 publications really provided and why the 2005 contributions were not better recognized?

2. The early functional interpretations of the CRISPR-Cas system

The evidence for the existence of CRISPRs resulted from the rapid accumulation of archaeal and bacterial genome

Keywords. *Cas* genes; CRISPR; genome editing; immune system; RNA interference

sequences at the end of the 1990s. Their higher incidence in archaea and their presence in thermophilic bacteria suggested that they might be associated with particular growth conditions. Without any connection with the previous observations, it was shown that CRISPRs were present a short distance from the origin of replication, an observation correlated with a possible role of these repeated sequences in chromosome partition deduced from the inhibitory effect on this process of added CRISPR sequences (Mojica *et al.* 1995). As short inverted repeats are often binding sites for proteins, CRISPRs might permit the binding of proteins involved in this partitioning. Up to 2002, the functions of the subsequently named *Cas* genes were looked for independently of those of CRISPRs. Sequence comparisons suggested DNA unwinding and nuclease activity for the products of the *Cas* genes. Their role in a fully new repair system was a bold hypothesis proposed by Koonin's group (Makarova *et al.* 2002). However, the established physical link between CRISPRs and *Cas* genes suggested a potential role of the latter in the structural organization of CRISPRs, and in particular in that of the spacer sequences, which were shown to be different from one cluster to another.

3. Three papers, and already the full story

I will present the three 2005 papers in their order of acceptance. The first was a contribution by a group in Alicante that had already widely contributed to the characterization of CRISPRs and to the demonstration that they were the most widely distributed family of repeats among prokaryotic genomes. By characterizing the spacer sequences of CRISPRs in different species, they observed that 65% of them had a bacteriophage and conjugative plasmid origin, and 35% were similar to chromosomal sequences. Referring to previous observations scattered through the scientific literature, they showed that strains with phage-derived spacer sequences are immune to infection by these phages, whereas strains devoid of these spacer sequences are fully sensitive to them: the presence of these spacer sequences makes the prokaryotes immune to an infection by phages bearing the same sequences. They suggested that this action could be mediated by CRISPR RNA molecules, 'similarly to the eukaryotic interference RNA' (Mojica *et al.* 2005, p 181) – arguments in favour of CRISPR transcription having been obtained before by different groups.

The two other contributions came from French groups that had not so far been directly involved in the study of CRISPRs. Working on *Yersinia pestis*, the idea of Christine Pourcel and colleagues (Pourcel *et al.* 2005) was to use the highly variable CRISPRs as a new and robust identification tool for strains of this pathogenic organism, a project already developed on *Mycobacterium tuberculosis* by another group (Kamerbeek *et al.* 1997). Previous studies had wrongly

shown that in *Yersinia pestis* CRISPRs only evolved by loss of spacer sequences.

The diversity of samples of *Yersinia pestis*, easily explainable by the epidemiological surveillance of these organisms, permitted 'old' shared spacer motifs to be distinguished from new ones and the elaboration of a model of CRISPR evolution. The addition of new spacer motifs was shown to be 'polarized', taking place near a leader sequence that had been described in previous studies of CRISPRs (Jansen *et al.* 2002). This simple rule was used to establish phylogenies. The fact that the most active CRISPRs (in terms of new spacer insertions) were those that were the closest to the *Cas* genes supported a role for the products of these genes in spacer insertions. Two-thirds of the new spacers were related by their sequences to a prophage. The authors suggested that CRISPRs are able to take up pieces of foreign DNA as part of a defence mechanism, constituting a memory of past genetic aggressions.

The third article (Bolotin *et al.* 2005) benefitted from the results described in the previous study. Working on two species of *Streptococcus*, one of which had been recently sequenced, the authors confirmed that a large proportion of spacers had a phage and, more generally, an extrachromosomal origin, also confirmed the close association between CRISPRs and *Cas* genes, and described new members of this gene family. More original was the establishment of a quantitative relation between the number of spacers of phage origin and the degree of resistance to phage infection, and the mechanism – synthesis of antisense RNA from the CRISPR – that was proposed to explain this protective effect.

4. Why are these contributions underappreciated in the scientific literature?

I do not argue that the 2006 (Makarova *et al.* 2006) and 2007 (Barrangou *et al.* 2007) papers contributed nothing. The precise structural and functional analysis of the *Cas* gene products reported in the first article provided strong arguments in favour of their role in the structural organization of CRISPRs. This article also drew a striking parallel between the CRISPR-Cas system and the recently described RNA silencing phenomenon in eukaryotes. The 2007 article was a beautifully designed set of experiments demonstrating (by addition and deletion) the role of the spacer sequences in the protection against phage infection, whereas the 2005 contributions only suggested this role. But the 2007 paper would not have been possible without those of 2005. Both types of papers, although very different, are necessary for the construction of well-founded scientific knowledge: the first introduce a new hypothesis and the second test it in a highly demonstrative way.

The fact that the 2006 and 2007 articles provided new information does not justify the improper presentation of the three 2005 papers in the literature, and the underestimation of their scientific value. Such a process of disinformation started early: the 2006 and 2007 papers already gave a too brief (and inappropriate) description of the content of the 2005 papers. Even worse was the short presentation of the 2007 article by the journal *Science* (Marx 2007).

There are two additional reasons for this lack of recognition of the 2005 papers. The first is the existence of a cultural difference between evolutionary biologists (and epidemiologists) and molecular biologists. Statistically significant correlations and associations have a high value of evidence for the former, but not for the latter. The precise order of spacer addition had a demonstrative value for microbiologists and epidemiologists that it did not have for molecular biologists.

The second reason is more sociological. Two of the three groups that published in 2005 had no ambition to discover the function of CRISPRs. They were working on other projects – classification and comparison of microorganisms – which remained their major objectives in the following years, thus maybe preventing them from immediately pushing on in the direction that they had opened up and from fighting for full recognition of their contributions. In addition, lack of recognition in the field made acceptance of their papers and fund-raising more difficult. And both were engaged in application-oriented research projects, which prevented them from rapidly reorienting their work.

5. Conclusion

The story that I have told is not exceptional and many readers have probably experienced similar situations in their careers. What is unusual is the importance of the discovery and its final application – the design of a simple tool for genome editing.

It is highly probable that the Nobel Committee will rapidly acknowledge the importance of these discoveries. Unfortunately, it is obvious that many of the participants in these discoveries will be sidelined and that the Nobel Committee's choice will be strongly influenced by the active process of rewriting history that has already been initiated.

Recently, a strong emphasis was placed on translational research and the development of 'big' programmes such as genome sequencing. In the case that I have described, such programmes were clearly a burden and acted as a brake on the rapid reorientation of the groups involved in the early observations.

There is also a risk that the attention now paid to the use of a modified form of this system for genome editing will prevent further efforts to determine its full physiological significance in prokaryotes and to focus on some of its characteristics described early on – such as the presence of chromosome sequences in spacers – that have yet to be satisfactorily explained.

Acknowledgements

I am indebted to David Marsh for his critical reading of the manuscript.

References

- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, *et al.* 2007 CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315** 1709–1712
- Bolotin A, Quinquis B, Sorokin A and Ehrlich SD 2005 Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. **151** 2551–2561
- Doudna JA and Charpentier E 2014 The new frontier of genome engineering with CRISPR-Cas9. *Science* **346** 6213
- Horvath P and Barrangou R 2010 CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327** 167–170
- Ishino Y, Shinagawa H, Makino K, Amemura M and Nakata A 1987 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
- Jansen R, van Emben JDA, Gaastra W and Schouls LM 2002 Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43** 1565–1575
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, *et al.* 1997 Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35** 907–914
- Makarova KS, Aravind L, Grishin NV, Rogozin IB and Koonin EV 2002 A DNA repair system specific for thermophilic archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* **30** 482–496
- Makarova KS, Grishin NV, Shabalina SA, Wolf YI and Koonin EV 2006 A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.* **1** 7
- Marx J 2007 New bacterial defense against phage invaders identified. *Science* **315** 1650–1651
- Mojica FJM, Ferrer C, Juez G and Rodriguez-Valera F 1995 Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* **17** 85–93
- Mojica FJM, Diez-Villasenor C, Soria E and Juez G 2000 MicroCorrespondence. *Mol. Microbiol.* **36** 244–246
- Mojica FJM, Diez-Villasenor C, Garcia-Martinez J and Soria E 2005 Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60** 174–182
- Pourcel C, Salvignol G and Vergnaud G 2005 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