Unravelling the Mysteries of CRISPR Memory Generation

Yoganand Knr
Indian Institute of Technology, Guwahati
Email: k.yoganand@iitg.ac.in

Adaptive immune response plays a vital role in the survival and evolution of a living being. This mechanism helps to fight against various pathogenic infections and ensures their abolishment upon future occurrences. Our micro co-inhabitants, like unicellular bacteria and archaea, do face such life-threatening challenges by phages (viruse made up of proteins and nucleic acids like DNA or RNA). These viruses make fatal use of bacterial cellular machinery for their propagation. To protect themselves from extinction, bacteria have acquired and developed an adaptive immune system called as CRISPR-Cas (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats - **C**RISPR **as**sociated genes).

This immune system contains a CRISPR locus, which is present within the genome of bacteria. It harbours numerous short repetitive DNA sequences termed 'repeats'. Upon phage infection, the Cas protein machinery (Adaptation complex) uptakes small fragments of nucleic acids that are specifically derived from the infecting phages and incorporates them at the first repeat of the CRISPR locus. Such incorporated sequences (called 'spacers') partition the array of repeats (refer Figure 1). This process instils the molecular memory of infection within the bacteria. The leader sequence in proximity to the first repeat encompasses a signal for expressing the spacer and repeat information in the form of regulatory CRISPR RNA. Another set of Cas proteins (Maturation complex) processes this CRISPR RNA to generate functional guide RNA (gRNA). This active form of gRNA contains the sequence of a CRISPR repeat and a single spacer. The repeat region of the gRNA signals the assembly of various Cas proteins on to it, thus forming an RNA-

---

* Mr. Yoganand Knr, Ph.D. Scholar from Indian Institute of Technology, Guwahati, is pursuing his research on "Functional Characterization of Adaptation Stage of CRISPR-Cas System." His popular science story entitled "Unravelling the Mysteries of Crispr Memory Generation." has been selected for AWSAR Award.

protein surveillance complex. Upon recurring phage infection, CRISPR-Cas surveillance complex can identify the phage genetic material by sequence similarity to the spacer. This detection signals the interfering Cas nuclease to silence the infection by rapid degradation of phage genetic material. This, in turn, protects the bacteria from fatal phage encounters (refer Figure 1).

By incorporating new spacers, CRISPR memory expands during each phage invasion. This repository of infection memory passes onto the next generation of bacteria and ensures their fitness against evolutionary pressures such as phage attacks. The spacer located in the proximity of the leader is known to achieve immediate response against the infections. Manoeuvring this bias, CRISPR machinery incorporates the spacers derived from fresh phage invasions at the leader proximal repeat (amidst presence of numerous repeats in CRISPR array). This process maintains the chronology of spacer insertion events such that the spacers corresponding to newest infections are located at the leader proximity. Despite being such a vital step in the CRISPR mediated immune response, the molecular events guiding the directionality of spacer insertion remains elusive.

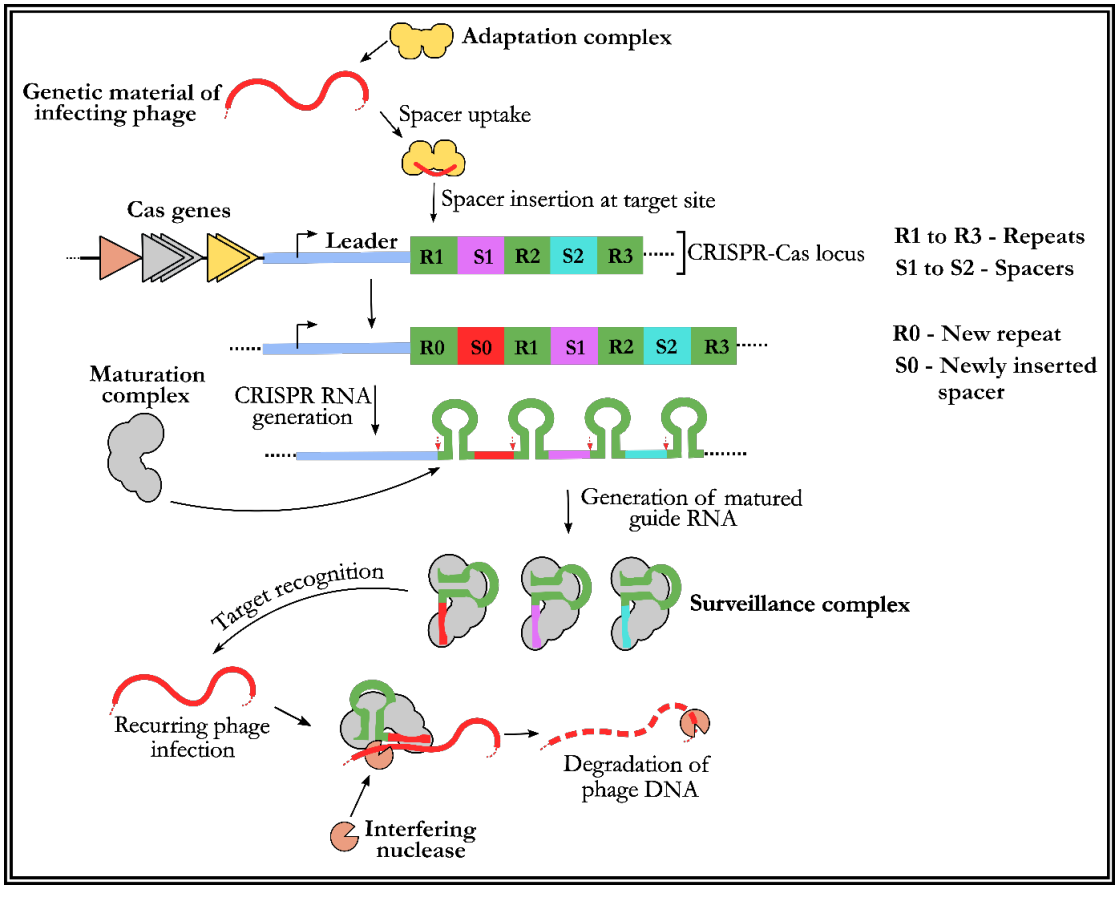

Figure 1:*Pictorial depiction of CRISPR-Cas mediated adaptive immune response against phage infections*

A research team at Indian Institute of Technology, Guwahati (IIT-G) sought to understand the mechanism by which CRISPR-Cas machinery identifies the first repeat as target site of spacer insertion in a bacterium named *Escherichia coli*. A complex formed by the two proteins namely, Cas1 and Cas2 is known to select spacers and insert them at the target region (i.e., first repeat). Various genetic and *in vitro* experiments performed by Dr B Anand's research group have resulted in identifying the involvement of a protein called **I**ntegration **H**ost **F**actor (IHF) in the expansion of CRISPR memory. IHF is a DNA architectural protein that sharply bends the linear DNA (in the shape of 'U') upon recognising a specific sequence (**I**HF **B**inding **S**ite – IBS). Employing various biochemical assays, the research team could identify IBS in the CRISPR leader and monitor the bending upon IHF binding. While investigating the indispensable requirement of this structural

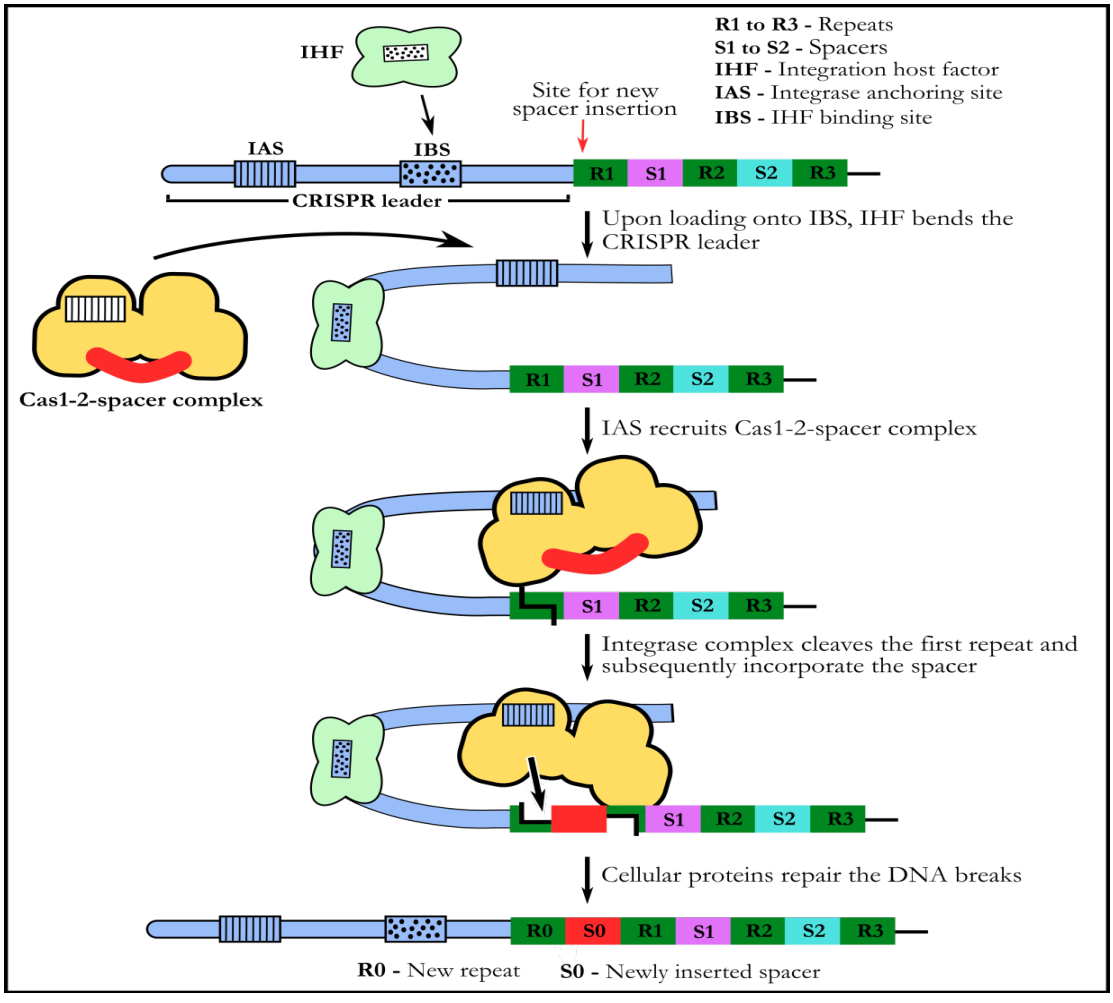

Figure 2: *Schematic representation of molecular events guiding the directional incorporation of spacers during CRISPR memory generation*

architecture for spacer insertion, the research team also discovered another key sequence in the leader (IAS - **I**ntegrase**A**nchoring **S**ite). In addition, the research group demonstrated that IAS is crucial for recruiting the Cas1-2-spacer complex and is distantly located from the spacer insertion site (refer Figure 2). Observations made by the team have proven that IAS is brought in close proximity of spacer insertion site by IHF mediated bending of CRISPR leader. This structural change recruits the Cas1-2-spacer complex around the first repeat and promotes the directional incorporation of the spacer. The molecular interplay of key regions in the leader, Cas1-2 and IHF ensures spacer insertion at the first repeat alone. In turn, these mechanistic features empower the bacteria with a quick immune response against recent viral attacks and ensures a productive fight to evade infections.

The current research exploration in the area of CRISPR-Cas immunity opens up the possibility of developing novel applications in a plethora of fields ranging from therapeutics to digital memory storing devices. An upcoming area of therapeutics, called phage therapy, employs virulent phages to specifically target and kill disease-causing bacteria in humans. Many of these bacteria possess CRISPR-Cas immune response and can successfully evade phage invasions, thus leading to the failure of medical treatments. Understanding the molecular events leading to CRISPR-Cas immunity paves the way for designing drug inhibitors to silence CRISPR-Cas response and helps to promote the efficacy of phage therapy. Recent research generated from George Church's lab at Harvard University, harnesses the potential of CRISPR-Cas system to integrate spacers. The scientists utilise the spacer integration ability of the immune system to transform bacterial cells into data storage devices. Usually, the electronic devices such as, hard disks store the data in combination of binary codes '1' and '0' which corresponds to the positive and negative polarity of a magnet, respectively. Combination of these binary codes can be assigned to a character of a data, pixel of a picture or a frame of a video. The sequential arrangement of these binary codes stores this digital information within the electronic memory device. Scientists have developed an analogous system utilising various combinations of nucleotides (building blocks of a DNA polymer) i.e., Adenosine triphosphate (A), Guanosine triphosphate (G), Cytidine triphosphate (C) and Thymidine triphosphate (T). In the proposed concept of a DNA digital data storage device, various nucleotide combinations are assigned to a data. As the CRISPR-Cas system has the ability to collect and store short spacer DNA information, scientists have repurposed this mechanism to store synthetic spacers that are encoded with the desired data module in a sequential fashion. Using advanced sequencing techniques, the spacer information encoded within the CRISPR locus was read in a serial order and the output was obtained in the form of images and videos. In this context, the research performed at Dr B Anand's lab in IIT-G helps to shed light on the molecular mechanism by which spacer information can be stored in a sequential order within a CRISPR locus. Empowered with these mechanistic details of CRISPR memory generation, the scientific community could potentially fine-tune the DNA storage devices to achieve utmost precision in data capture and storage.