Chapter 36
Genome Editing and CRISPR/Cas System of Extremophiles and Its Applications

Suneeta Gireesh Panicker
Dr. D.Y. Patil Arts, Commerce, and Science College, Pune, India

ABSTRACT

Extremophiles will be the choice of next generation industrial biotechnology (NGIB) as they are known to be contaminant resistant, but engineering their genomes has always been difficult and time consuming task. CRIPR/Cas (clustered regularly interspaced short palindromic repeat and CRISPR associated proteins) system can be employed for this reason. The genome of an industrially important halophile (i.e., Halomonas) was edited to study a combined effect of four different genes on glucose breakdown and production of poly (3-hydroxybutyrate-co-3-hydroxyvalerate). This editing has resulted in 16-fold increase of 3HV, and the mutants generated by CRIPR/Cas system were significantly effective in synthesizing PHBV. Unfortunately, this system does not always work, specifically in extremophilic microorganisms because Cas9 or Cpf1 are from mesophilic bacteria. Therefore, alternatively, the endogenous CRISPR/Cas system is used for editing the genomes of such organisms. This genome editing of extremophiles will open the doors for developing next generation industrial biotechnology (NGIB).

INTRODUCTION

The Nobel Prize of year 2020 in Chemistry 2020 was awarded to Emmanuelle Charpentier and Jennifer Doudna for discovering the genetic scissors called CRISPR/Cas9: **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeat - CRISPR-associated sequences (CRISPR-Cas).

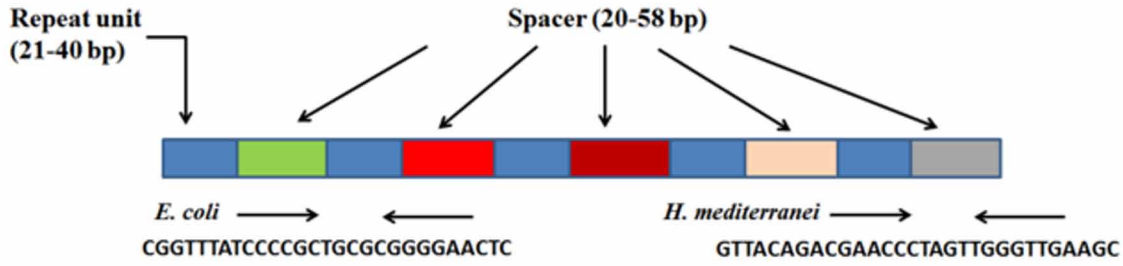
The Discovery of the CRISPR-Cas System in Prokaryotes

In 1993, for the first time, CRISPRs were observed in archaea, named *Haloferax mediterranei* (Mojica et al., 1993) and then subsequently as the life science moved into the genomic era, these were detected in

DOI: 10.4018/979-8-3693-3026-5.ch036

more and more number of bacteria and archaea genomes. The powerful CRISPR-Cas9 system discovery initiated with the identification of repeated genome structures present in bacteria and Archaea. An unusual repeated structure in the *Escherichia coli* genome was reported in 1987. This sequence consisted of five extremely homologous sequences of 29 base pairs (bp) along with a reverse complementary (dyad symmetry) sequence of 14 bp that were interspersed by variable spacer sequences of 32 bp (Ishino et al., 1987). Later, the halophilic Archaea *Haloferax mediterranei*, showed the presence of similar, repeated structures in its genome as shown in figure 1. It had 14 nearly conserved sequences of 30 bp, repeated at regular distances (Mojica et al., 1993). Subsequently, these types of repeats were revealed through bioinformatics analyses of various prokaryotes. It was thus shown that such repeats possessed a characteristic feature of short, partially palindromic sequences in clusters which were separated by exclusive interfering elements of fixed length. It was also observed that these intervening sequences differed from organism to organism, but had an ancestral origin and high biological relevance (Mojica et al., 2000). These arrays of repeated sequences are called **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats, abbreviated as **CRISPR** (Jansen et al., 2002). The analysis of the unique, non-repetitive sequences in CRISPR showed they matched with the genetic code of different viruses, pointing towards the protective immunity against viruses. This led to a hypothesis that if a bacterium overcomes a virus infection and succeeds in surviving then it will add a piece of that virus genetic code to its genome, like a memory of that infection. In the efforts of understanding the function of CRISPR, researchers discovered special genes called CRISPR-associated, abbreviated as *cas*. This group of genes were located only in the CRISPR-containing prokaryotes and were always found adjacent to CRISPRs. The *cas* genes encoded proteins that had helicase and nuclease activity, advocating their role in DNA metabolism or gene expression (Jansen et al., 2002). In the coming years, a number of Cas protein subfamilies were discovered (Haft et al., 2005; Makarova et al., 2006). Cas proteins control the three functional stages of CRISPR adaptive defence. These are adaptation, processing, and interference as shown in figure 2.

Figure 1. CRISPR structure in prokaryotes



The adaptation stage was best studied in a type I-E CRISPR system of *E. coli.* In this system, a stable complex was formed between two strongly conserved nucleases (Cas1 and Cas2) (Nuñez et al., 2014). Two Cas1 dimers were linked to one Cas2 dimer and this is considered as the minimum requirement for the *de novo* spacer acquisition (Yosef et al., 2012). crRNA processing is the second stage (Bhaya et al., 2011), which involves shortening of the long crRNA. When the crRNA is transcribed from its gene, the tracrRNA (trans-activating CRISPR RNA which has complementary sequences for palindromic se-