Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Applications of CRISPR-Cas for synthetic biology and genetic recording

Florian Schmidt[1] and Randall J. Platt[1,2]

Abstract

The epic arms race between microbes and their predators was the driving force behind the evolution and diversification of the truly remarkable microbial adaptive immune system CRISPR-Cas. CRISPR-Cas systems mediate defense through three stages: recording of nucleic acid species, multiplexed RNA expression and processing, and eventually RNA-guided cleavage of hostile genetic elements. The entire process is orchestrated by a plethora of effector proteins endowed with specialized functions for manipulating genetic material. Investigating this treasure trove substantially fostered the development of the RNA-guided DNA endonuclease Cas9 into a versatile molecular tool for synthetic biology and biomedicine. Here, we review the developments of Cas9 and other CRISPR-Cas components for applications in synthetic biology as well as highlight emerging CRISPR-Cas-based genetic recorders and memory devices.

Addresses

[1] Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel, Switzerland

[2] Department of Chemistry, University of Basel, Petersplatz 1, 4003 Basel, Switzerland

Corresponding author: Platt, Randall J (rplatt@ethz.ch)

Current Opinion in Systems Biology 2017, **5**:9–15

Edited by **Danielle Tullman-Ercek** and **Martin Fussenegger**

This review comes from a themed issue on **Synthetic biology (2017)**

For a complete overview see the Issue and the Editorial

Available online 31 May 2017

http://dx.doi.org/10.1016/j.coisb.2017.05.008

Keywords

CRISPR-Cas, Cas9, Synthetic biology, Genome editing, Genetic recording.

Introduction

Genome editing technologies enabling the manipulation of genetic sequences are fueling major advancements in both basic and applied research. With an expanding molecular toolkit, our capacity to manipulate DNA, elucidate causal relationships between genotypes and phenotypes, and engineer biological systems is ever expanding, holding immense promise for the future of biomedical research. The DNA revolution began with recombinant DNA and restriction enzyme technology, enabling the manipulation of DNA in a test tube. Next came a suite of discoveries enabling the manipulation of genomic DNA in living cells, which includes integrases, recombinases, transposases, and site-specific DNA endonucleases (e.g. meganucleases, zinc finger nucleases, and transcriptional activator-like effector (TALE) nucleases). Now, with the microbial defense system CRISPR-Cas, we are undergoing yet another revolution in genome editing and biomedical science. Since this transformative discovery, sculpted by dozens of scientists over three decades, we now have a vast CRISPR-Cas-based genome editing toolkit [1–5]. With this expanded toolkit, new possibilities emerge, including DNA mutagenesis, transcriptional activation or repression, epigenetic engineering, and molecular barcoding. Here, we provide a brief introduction of the CRISPR-Cas system and review emerging genome editing applications in synthetic biology. In particular, we highlight how genome editing technologies can be utilized to write information into the genome, and discuss current approaches that repurpose CRISPR-Cas for lineage tracing and genetic recording of biological activity.

The CRISPR-Cas defense system

The microbial adaptive immune system CRISPR-Cas evolved as a defense system against exogenous genetic material (extensively reviewed elsewhere [1–5]). In brief, CRISPR-Cas orchestrates defense in three steps, adaptation, biogenesis, and interference. During the adaptation step, short stretches of DNA or RNA derived from invading genetic elements (i.e. protospacers) are recorded into a CRISPR array with the help of associated Cas proteins. Acquired foreign sequences within the CRISPR array (i.e. spacers) are flanked by short palindromic direct repeat sequences (i.e. DRs). During the biogenesis step, the entire CRISPR array is transcribed as a single RNA molecule and subsequently processed to yield mature guide RNAs. During the interference step, the guide RNAs direct CRISPR-associated nucleases to hostile nucleic acids and mediate their destruction. The protospacer adjacent motif (PAM), a sequence present in the invading element but not the CRISPR array or guide RNA, facilitates self versus non-self discrimination.

The most widely investigated protein of the CRISPR-Cas system is the RNA-guided DNA endonuclease Cas9 from *Streptococcus pyogenes* (SpCas9, referred to as Cas9 herein). Cas9 recognizes it's target site via Watson-Crick base-pairing between guide RNA and target DNA. Intriguingly, Cas9 is easily targeted to seemingly any DNA sequence through modification of the 20 nucleotide targeting sequence of the guide RNA. Upon target

recognition, Cas9 induces a DNA double strand break (DSB), which is repaired by either of two alternate DNA repair pathways. The first, termed non-homologous end joining (NHEJ) is an error-prone process leading to insertions and deletions (indels) at the target site, thereby frequently disrupting the coding sequence if targeted to an exon. The second, named homology-directed repair (HDR), is a high fidelity process that utilizes a homologous recombination template matching the target site to repair the break. Thus, Cas9 allows both gene disruption and precision editing, and can be applied to the engineering of mammalian genomes [6,7].

The development and applications of Cas9-based technologies for mammalian genome editing are diverse, immense, and extensively reviewed elsewhere [1–5]. Many of these technologies have been widely adopted and adapted, resulting in a plethora of new tools for synthetic biology. These include Cas9 orthologs (e.g. StCas9 [8], SaCas9 [9], NmCas9 [10], and FnCas9 [11] as well as other Class 2 CRISPR-Cas nucleases (e.g. Cpf1/Cas12a) [12], which enable orthogonal genome editing applications. Furthermore, fusing catalytically inactive Cas9 (dCas9) to effector proteins enables transcriptional activation (CRISPRa) [13–19], transcriptional repression (CRISPRi) [19–22], epigenetic modifications [14,19,23,24] and single base editing via a cytidine deaminase [25]. Likewise, inducible systems based on a split Cas9 architecture or nuclear localization have been achieved [26–31]. Many of these tools based on the CRISPR-Cas system have replaced their predecessors in synthetic circuits due to increased orthogonality, modularity, and multiplexability.

CRISPR-Cas-based memory devices for lineage tracing and recording biological activity

Cellular behaviors are complex, dynamic, and often require the integration and orchestration of multiple stimuli and responses, respectively. Currently, we rely on time point and inference experiments that often miss the dynamic and complex single cell characteristics underlying biological phenomena. Genetic recording devices hold great potential for elucidating these diverse cellular behaviors.

Early work on genetic recording relied on synthetic circuits comprised of toggle switches or multiple quasi-stable states based on protein expression [32–34]. Later, DNA recombinases were leveraged to record transient cellular events into genomic DNA [35,36]. These early technologies lacked scalability, orthogonality, and the capacity for capturing dynamic information, and therefore, future developments combined recombinases with inducible retron expression to encode dynamic analog memory into populations of bacteria [37]. More recently, CRISPR-Cas-based systems utilizing either Cas9 or CRISPR-Cas adaptation proteins were applied as genetic recorders (Table 1), facilitating lineage tracing (Box 1), molecular barcoding, and intracellular recording of biological activity.

Scartrace

To induce sequence diversification for tracing of cellular lineages in zebrafish, Junker and colleagues injected GFP transgenic zebrafish embryos at the single-cell stage with Cas9 and GFP-targeted guide RNA ribonucleoprotein complexes (Figure 1A) [39]. They discovered, that Cas9-mediated DSBs resulted in several hundred unique indels at the target site within GFP, which allowed them to trace cellular lineages during zebrafish development and caudal fin regeneration. While this method, termed Scartrace, is easily adoptable and does not require elaborate expression constructs, the information storage capacity is limited by the diversity of indel mutations at a single target site.

GESTALT

An alternative technique, termed genome editing of synthetic target arrays for lineage tracing (GESTALT), expanded the information storage capacity of CRISPR-Cas9-based memory devices by concatemerizing multiple guide RNA target sites (Figure 1B) [40]. This allows multiple genome editing events in the same synthetic array within the same cell. Using GESTALT, McKenna and colleagues reconstructed the lineage of 200,000 cells in the adult zebrafish, and revealed that much of the organism was formed from just a few embryonic cells. One drawback of this approach however, is the fact that simultaneous editing at two target sites of the 10 barcode concatemer can result in depletion of the intervening sequence and thus in a loss of previously stored information.

mSCRIBE and homing CRISPR – evolvable and self-targeting guide RNAs

Expanding the storage capacity of CRISPR-Cas9-based memory devices even further, Perli et al. and Kalhor et al. developed self-targeting guide RNAs [41] and homing guide RNAs (referred to here as Homing CRISPR) [42], respectively (Figure 1C). With these approaches, the expression cassette encoding the guide RNA also contains the PAM sequence, and is therefore self-targeted. Upon cleavage of the guide RNA expression cassette, indels are created within the targeting sequence, and yet the guide RNA remains transcribed. Therefore, the locus undergoes multiple consecutive rounds of self-targeted cleavage and mutagenesis. This process is interrupted only when the PAM sequence is deleted, the guide RNA scaffold is destroyed, or when the guide RNA targeting sequence is shortened below 16 base pairs. Extension of the targeting sequence beyond the 20 base pairs of the canonical guide RNA resulted in genetic recordings of longer durations and sequences of greater diversity.