Journal of Biotechnology 241 (2017) 50–60

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Review

Applications of genome editing by programmable nucleases to the metabolic engineering of secondary metabolites

Ana Lúcia Leitão [a,b,*], Marina C. Costa [c], Francisco J. Enguita [c,**]

[a] Departamento de Ciências e Tecnologia da Biomassa, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, Campus de Caparica, 2829-516 Caparica, Portugal
[b] MEtRICs, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, Campus de Caparica, 2829-516 Caparica, Portugal
[c] Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 18 April 2016
Received in revised form 6 November 2016
Accepted 10 November 2016
Available online 11 November 2016

Keywords:
Secondary metabolism
Genome editing
TALEN
CRISPR-Cas9
Hybrid metabolites
Epigenome editing

ABSTRACT

Genome engineering is a branch of modern biotechnology composed of a cohort of protocols designed to construct and modify a genotype with the main objective of giving rise to a desired phenotype. Conceptually, genome engineering is based on the so called genome editing technologies, a group of genetic techniques that allow either to delete or to insert genetic information in a particular genomic locus. Ten years ago, genome editing tools were limited to virus-driven integration and homologous DNA recombination. However, nowadays the uprising of programmable nucleases is rapidly changing this paradigm. There are two main families of modern tools for genome editing depending on the molecule that controls the specificity of the system and drives the editor machinery to its place of action. Enzymes such as Zn-finger and TALEN nucleases are protein-driven genome editors; while CRISPR system is a nucleic acid-guided editing system. Genome editing techniques are still not widely applied for the design of new compounds with pharmacological activity, but they are starting to be considered as promising tools for rational genome manipulation in biotechnology applications. In this review we will discuss the potential applications of programmable nucleases for the metabolic engineering of secondary metabolites with biological activity.

Contents

* Corresponding author at: Departamento de Ciências e Tecnologia da Biomassa, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, Campus de Caparica, 2829-516 Caparica, Portugal.
** Corresponding author.
E-mail addresses: aldl@fct.unl.pt (A.L. Leitão), fenguita@fm.ul.pt (F.J. Enguita).

http://dx.doi.org/10.1016/j.jbiotec.2016.11.009
0168-1656/

1. Introduction: synthetic biology meets secondary metabolism

The foundations of synthetic biology rely on the concept that all living systems are constituted by functional and structural modules. The intrinsic "modularity" of biological systems offers unlimited possibilities for the construction of new living entities. At the genomic level, the dissection and systematic combination of these modules by the use of biological engineering methods will allow building organisms that are not easily generated by the evolution. Synthetic biology combines biology and engineering to design and build novel biological entities with new or improved functions, including the construction of new biological devices and also the re-design of existing natural systems for advantageous functions (Schwille, 2015). The successful achievement of the synthetic biology goals is supported by three pillars: the molecular knowledge of living organisms to characterize their functional modules; the development of computational models to integrate and combine these functional blocks in order to design new living entities; and the availability of wet-lab methods to construct and validate the new synthetic organisms (Quin and Schmidt-Dannert, 2014; Seyedsayamdost and Clardy, 2014; Zotchev et al., 2012). The principles of synthetic biology have been elegantly illustrated in recent years by the heterologous production of biologically active compounds by combination of biosynthetic genes from different organisms (Paddon et al., 2013), and pushed to the limits by the creation of whole functional cells from a completely synthetic chromosome by the Craig Venter's research team (Gibson et al., 2010; Suzuki et al., 2015).

Secondary metabolism is constituted by a group of biological processes which are dispensable for cell growth. In microorganisms such as fungi and filamentous bacteria, secondary metabolism is an important source of biologically active compounds with applications in medical chemistry and biotechnology (Leitao and Enguita, 2014). Secondary metabolites derived from microorganisms tend to have significantly greater chemical complexity than the majority of synthetic drugs, harboring a great variety of diversity and biological activities. However, they are typically not available as pure compounds and their isolation from complex mixtures is frequently a tedious and complicated process (Seyedsayamdost and Clardy, 2014). Also due to the technical innovations as robotic high-throughput screening and combinatorial chemical synthesis, secondary metabolites and other natural products have been disregarded from the drug discovery pipelines (Mitchell, 2011). Despite of the availability of new methods, the drug discovery pipelines are not giving the desired results, as demonstrated by the recession in the numbers of new drugs introduced to the market in the last years (Kaitin and DiMasi, 2011). Future solutions for unmet clinical needs must be dependent on a revisited strategy for the discovery of new compounds based on natural sources, and secondary metabolism could be one of the most reliable sources of these compounds. This strategy is supported by the recent convergence of next-generation sequencing methods with the ideas of synthetic biology, allowing rapid genetic screening for new functional modules facilitating their rational combination (Hirai, 2015; Mattern et al., 2015).

The genes encoding for the enzymes involved in secondary metabolism are frequently clustered showing a modular organization and are transcriptionally co-regulated (Enguita et al., 1998). The increasing number of available complete genomes arising from the application of next generation sequencing techniques have allowed to discover an immense variety of biosynthetic clusters for secondary metabolites in actinomycetes and fungi, some of them not expressed under laboratory conditions (Jiang et al., 2015a; Lin et al., 2013). The enzymes encoded by these genes are typically active over a limited number of metabolite precursors arising from the anabolic branch of the primary metabolism such as amino acids, lipids, sugars, nucleotides and other low-molecular weight compounds (Lewis, 2013). The vast majority of secondary metabolism gene clusters contain a central gene encoding for a high-molecular weight condensing enzyme, which is responsible for the assembly of the precursors, and other genes encoding for auxiliary enzymes that will be in charge of the chemical modifications of the secondary metabolite chemical core (Coque et al., 1995a,b; Fisch et al., 2010). The clusters often include additional regulatory genes involved in gene expression and transport (Coque et al., 1993). The intrinsic characteristic of secondary metabolism makes it a perfect target for the application of synthetic biology principles with an almost unlimited potential for the construction of new cell factories producing new bioactive compounds. However, the engineering of secondary metabolic pathways is constrained by the knowledge of the function and working rules of the functional blocks governing the biosynthesis of a particular metabolite and also by the availability of specific techniques of genetic manipulation (Leitao and Enguita, 2014). The advances in automated DNA synthesis together with the uprising of genome editing techniques has dramatically improve the portfolio of available techniques for synthetic biology applications in the last five years (Kim et al., 2015). In this review we will discuss the potential applications of targeted genome editing to the engineering of secondary metabolites following the rules of synthetic biology.

2. Genome editing with programmable nucleases

2.1. Strategies for targeted genome editing

Targeted genome editing is the specific modification of a predetermined locus within a genome with the objective of repurposing the functions of this specific region. Genome editing technologies must have the prerequisite of generating genetically stable organisms, be easy to use against a range of DNA sequences, and ideally must be also easy to perform in a wide range of organisms (Kim and Kim, 2014). Targeted genomic modifications are based on the deletion or insertion of genetic information, and are strongly dependent on genetic recombination. Classical approaches used for genome editing are based on homologous recombination, a process with an extremely low efficiency especially in higher eukaryotic cells which hindered its routine application. However, several solutions have been developed to increase the efficiency of genetic recombination, including the use of tunable and guided nucleases. This family of enzymes is able to produce DNA double-strand breaks at specific locations within the genome, increasing the efficiency of homologous recombination events and also triggering the non-homologous end joining process (NHEJ), an error-prone RNA repair mechanism that leads to targeted mutations (Boettcher and McManus, 2015). The generation of double strand breaks at specific DNA loci can be also used to facilitate homologous recombination to insert DNA fragments and generate recombinant organisms (Ran et al., 2013b). There are two main family of programmable nucleases, a protein-guided nuclease family, composed by Zn-finger and transcription activator-like effector nucleases (TALENs), and a nucleic acid-guided nuclease family, mainly represented by the CRISPR-Cas9 system (Kim and Kim, 2014).

2.2. Protein-guided nucleases

The first programmable nucleases were developed at the end of the 1990's as customizable restriction enzymes designed to cut any DNA in a predetermined sequence. At the time, those enzymes were designated as "hybrid restriction enzymes" since they were constituted by the fusion of a catalytic nuclease domain (FokI) and a