Methods 69 (2014) 188–197

Contents lists available at ScienceDirect

Methods

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

Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression

Ami M. Kabadi [a], Charles A. Gersbach [a,b,c,*]

[a] Department of Biomedical Engineering, Duke University, Durham, NC 27708, United States
[b] Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708, United States
[c] Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC 27710, United States

ARTICLE INFO

Article history:
Received 25 January 2014
Revised 26 June 2014
Accepted 27 June 2014
Available online 8 July 2014

Keywords:
Gene regulation
Transcription factor
Protein engineering
TAL effector
TALE
CRISPR
Cas9
Gene therapy
Genetic reprogramming
Synthetic biology

ABSTRACT

Engineered DNA-binding proteins that can be targeted to specific sites in the genome to manipulate gene expression have enabled many advances in biomedical research. This includes generating tools to study fundamental aspects of gene regulation and the development of a new class of gene therapies that alter the expression of endogenous genes. Designed transcription factors have entered clinical trials for the treatment of human diseases and others are in preclinical development. High-throughput and user-friendly platforms for designing synthetic DNA-binding proteins present innovative methods for deciphering cell biology and designing custom synthetic gene circuits. We review two platforms for designing synthetic transcription factors for manipulating gene expression: Transcription activator-like effectors (TALEs) and the RNA-guided clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system. We present an overview of each technology and a guide for designing and assembling custom TALE- and CRISPR/Cas9-based transcription factors. We also discuss characteristics of each platform that are best suited for different applications.

1. Introduction

The inspiration for many of the technological advances that have transformed biomedical research, such as polymerase chain reaction or RNA interference, has come from natural biological processes. Similarly, recent discoveries of the principles of protein-DNA interactions in various species and systems has guided the development of methods for engineering designer proteins that can be targeted to any DNA target site. These proteins can serve as a scaffold for building enzymes that can modify DNA sequence, transcriptional regulation, or the epigenetic status at any site in the genome. Three main classes of natural biomolecules have been engineered to target new DNA sequences and manipulate gene expression: zinc finger proteins (ZFPs), Transcription activator-like effectors (TALEs), and the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas) system. Each of these programmable DNA-binding proteins can be genetically fused to an effector domain to create custom enzymes that localize the effector function to the DNA target site. Various effectors domains have been widely used to create targeted changes to genome sequence, including nucleases, integrases, and recombinases [1–5]. Alternatively, the fusion of these DNA-binding proteins to transcriptional activation and repression domains enables the control of gene regulation at targeted promoter or enhancer elements [6–8]. The applications of these gene regulation technologies are diverse, including stimulating expression of endogenous therapeutic factors [9–16], controlling cell differentiation [17–20], and regulating synthetic gene circuits [21–23]. This article discusses the origins of these technologies for targeted gene regulation, the methods for engineering programmable transcriptional activators and repressors, and design considerations for implementing the various approaches.

The first engineered proteins designed to target new DNA sequences were based on the Cys_2His_2 zinc finger domain [5,24], the most common DNA-binding motif in the human proteome [25]. Despite many notable successes with this technology [1,3,5,26], engineering new ZFPs with high activity and specificity remains technically challenging for most researchers. Since predictions of ZFP DNA-binding specificity and affinity are complex [27,28], it is typically necessary to build and screen many rationally designed proteins or use high-throughput selections to find

* Corresponding author at: Department of Biomedical Engineering, Room 136 Hudson Hall, Box 90281, Duke University, Durham, NC 27708-0281, United States. Fax: +1 919 668 0795.
E-mail address: charles.gersbach@duke.edu (C.A. Gersbach).

http://dx.doi.org/10.1016/j.ymeth.2014.06.014
1046-2023/

functional proteins within large libraries [24,29,30]. Consequently, academic laboratories are adopting the newer TALE and CRISPR/Cas9 platforms that have straightforward DNA-recognition properties [31,32]. Nevertheless, two decades of studies on targeted gene regulation with engineered ZFPs have provided much of the foundation on which these new platforms are being built [5].

The DNA recognition code for TALE proteins was first reported in 2009 [33,34] and presented a new modular DNA-binding domain that is more easily reprogrammed to target new sequences compared to ZFPs [35–38]. Because of their simple DNA recognition code, the frequency of engineering active TALE-based enzymes is very high [4,39]. Consequently, there have been many recent successes modulating mammalian gene expression with synthetic TALE enzymes [14,18,19,21,22,35,36,40–42]. In 2012, an engineered version of the CRISPR/Cas9 system was developed as the first platform for targeting proteins to DNA target sites through RNA:DNA interactions rather than direct protein-DNA interactions [43]. Since simply changing the expressed RNA molecule alters CRISPR/Cas9 binding, novel binding proteins do not need to be engineered for each new target sequence. Therefore the CRISPR/Cas9 system is typically easier, faster, and more economical to implement in comparison to the ZFP and TALE technologies. After the CRISPR/Cas9 system was shown to be effective as a nuclease platform for genome editing [44–48], it was quickly reengineered for transcriptional regulation [20,49–56]. The early studies on both of these systems for gene regulation have demonstrated that there are advantages and limitations of each approach, and further work is necessary to elucidate properties of each of these systems. Here we present an overview and user's guide for applying the TALE and CRISPR/Cas9 systems to control gene expression.

1.1. Transcription activator like effectors

Transcription Activator Like Effectors (TALEs) are modular DNA-binding proteins derived from the plant pathogenic bacteria *Xanthomonas* [33,34] and *Ralstonia* [57]. As a defense mechanism, these organisms produce TALEs to modulate host gene expression. The TALE DNA-binding domain consists of multiple repeats of 34 amino acids where variability in positions 12 and 13, referred to as the repeat-variable di-residues (RVDs), confer binding specificity for one specific DNA base [33,34,58–60] (Fig. 1A). Multiple TALE monomers can be linked in tandem to recognize the desired DNA sequence [35–38] (Fig. 1B). The array of TALE domains is then fused to an effector domain to induce a specific action at a user-determined genomic locus [61].

1.2. The CRISPR/Cas system

Bacteria and archaea have evolved the CRISPR/Cas system as an RNA-guided defense mechanism against viral parasites that detects and silences foreign nucleic acids [62]. In the naturally-occurring system, bacteria and archaea integrate short fragments of foreign nucleic acids (termed protospacers) into the CRISPR genomic loci. Functioning as molecular memory of previous invaders, the CRISPR locus is transcribed and processed into short CRISPR-derived RNAs (crRNAs). Thus each crRNA contains sequence complementarity to a prior nucleic acid invader. In the type II system, crRNAs associate with transactivating crRNAs (tracrRNAs) and the Cas9 endonuclease. Through complementary base pairing, the crRNA localizes the Cas9 complex to the foreign DNA sequence to induce a double strand break. Characterization of CRISPR cleavage sites identified a short sequence directly downstream from the protospacer required for Cas9-mediated cleavage. This sequence is called the protospacer adjacent motif (PAM) and the identity of the sequence is highly variable between CRISPR systems from different species. For ease of use, a single transcript chimeric guide RNA (gRNA) has been engineered to recapitulate the function of both the crRNA and tracrRNA [43–45]. The chimeric gRNA consists of three regions: a 20 bp protospacer which confers targeting specificity through complementary base pairing with the desired DNA target, a nucleotide hairpin which mimics the crRNA:tracrRNA structure required for Cas9 protein binding, and a transcriptional termination sequence [63]. It was recently demonstrated that the crRNA, tracrRNA, and Cas9 nuclease were all necessary and sufficient to program the RNA-guided Cas9 nuclease activity to new sequences outside of the native host [43]. The gRNA could also substitute for the crRNA and tracrRNA in these experiments. Subsequently, this engineered type II CRISPR system was shown to function effectively for RNA-guided programmable nuclease activity in numerous other hosts, including human cells [44–47], mouse cells [64–66], *Caenorhabditis elegans* [67,68], zebrafish [48,69], and bacteria [70].

Cas9 catalyzes DNA double-stranded breaks via RuvC and HNH endonuclease domains, each of which cleaves one strand of the target DNA. Both of these enzymatic domains can be inactivated by a single amino acid substitution (D10A and H840A), generating a Cas9 protein that has no endonuclease activity but maintains its RNA-guided DNA-binding capacity [43]. This deactivated Cas9 (dCas9), in conjunction with the gRNA, functions as a modular DNA-binding scaffold similar to ZFPs and TALEs. Therefore the discovery of CRISPR-based immunity has led to the development of a new class of modular synthetic enzymes where DNA-binding is directed by an RNA:DNA interaction rather than a protein:DNA interaction (Fig. 2). This dCas9 scaffold has been used to create both RNA-guided repressors [20,49,50] and transcriptional activators [51–56,71].

2. Synthetic repressors

Transcriptional repression typically occurs through one of three main mechanisms: inhibiting formation of the pre-initiation

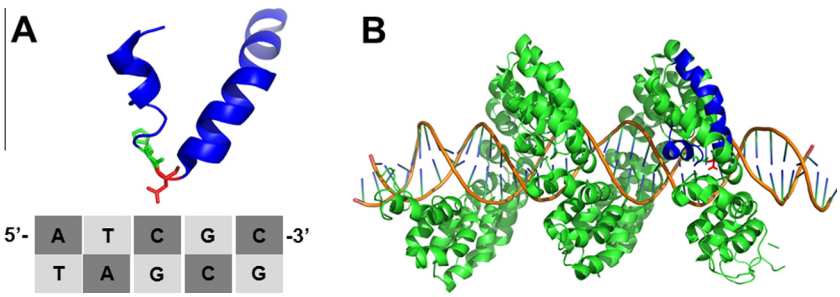

Fig. 1. The TAL effector DNA-binding domain. (A) Through a DNA–protein interaction, each TALE repeat binds one bp of DNA. The TALE repeat is shown in blue, and the repeat variable diresidue (RVD) at the 12th and 13th position are shown in green and red, respectively. (B) TALEs can be linked in tandem to recognize virtually any DNA sequence. The desired string of TALEs is then fused to an effector domain to induce a specific action at a predetermined DNA sequence. Crystal structure adapted from [60].