

GENOMICS

TALEs for the masses

An optimized transcription activator–like effector (TALE) and an improved assembly method promise efficient genome editing and transcriptome modulation.

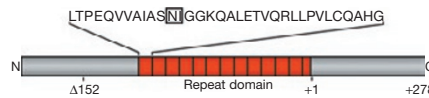
Easy genome manipulation is many a geneticist's dream. The specific targeting of endogenous sequences to introduce mutations, tags or entire new coding sequences would allow precise disease modeling among many other applications.

Zinc-finger nucleases (ZFNs) have partly delivered on this promise, but the design of a ZFN for a genomic sequence of interest remains a challenge for many researchers. The recent discovery of transcription activator–like effector (TALE) proteins, transcription factors from plant pathogens, rekindled the hope that an easier to use DNA targeting molecule might be found.

Like ZFN, TALEs are modular, but unlike ZFN, each module only recognizes one, not three, DNA base pairs. Each TALE is made up of 12 or more modular repeats, flanked by N- and C-terminal sequences, and each repeat consists of 34 amino acids; residues 12 and 13 determine target recognition and are known as repeat variable di-residues (RVDs). Each of the four DNA bases binds a unique RVD, so to target, for example, a 12-base stretch of DNA, one combines 12 modules bearing the appropriate RVDs.

Though shown to work in plants and yeast, TALEs had not been demonstrated to target endogenous genes in mammalian cells until recently, with the work of two independent research groups—one from Sangamo BioSciences led by Edward Rebar (Miller *et al.*, 2011) and the other from Harvard University, led by Feng Zhang and Paola Arlotta in collaboration with George Church (Zhang *et al.*, 2011).

“One of the features that attracted us to TALEs,” says Philip Gregory, chief scientific officer at Sangamo, “was the modularity and simplicity of their recognition code.” In what they refer to as the ‘first shot on goal’ Rebar’s team designed TALEs, fused



Structure and specificity of a TALE. Amino acid sequence for one module is shown with the two amino acids that mediate binding of a DNA base highlighted by a box. Reprinted from *Nature Biotechnology*.

to the activation domain of a transcription factor, to target the promoter of the human nerve growth factor NTF3. They saw strong induction of the endogenous transcript. To make a TALE nuclease, they fused a TALE to the catalytic domain of the nuclease FokI and found that two different C-terminal TALE truncations, retaining only 28 or 63 amino acids of the C-terminal tail, performed optimally. They also demonstrated that a module expressing the RVD asparagine and lysine works better at recognizing the DNA base guanine than the previously described asparagine and asparagine RVD.

The teams at Harvard University wanted to test the potential of TALEs for activating endogenous genes but ran into a roadblock when attempting to commercially synthesize a TALE to a sequence of interest. Zhang recalls that the company he contacted was unsure that they could make the very repetitive sequence of TALEs. “At that point,” Zhang says, “we decided that this was not the way to go. We needed a faster way of making TALEs.” His team then developed a hierarchical ligation-based strategy to engineer TALEs with 12 repeat modules. They altered the DNA sequence of the individual modules to preserve the RVDs but minimized repetitiveness in the rest of the sequence. The process involved 12 PCRs for each of the four modules to create monomers with specific linker sequences, each suitable for a defined position in the final 12-mer, which they then cloned into the backbone vector encoding the N and C termini of the TALE.

To create a TALE transcription factor, the researchers fused the TALE to a transcription activation domain and targeted promoters of four endogenous genes. Two transcripts were upregulated and two were not, probably because the local chromatin structure did not allow access by the TALEs. Arlotta was happy with the results; “we wanted to provide a proof of principle that we can activate an endogenous gene,” she says. “One of our major efforts is to use different transcription factors to map out the regulators that allow cells to differentiate, not only in culture but in the intact organism.” Zhang adds: “our goal is also to make the technique very accessible to people.” They have created a website (<http://www.taleffectors.com/>) with details of the protocol for how to make better TALEs. Zhang wants to make the entire system available as a kit: four plasmids encoding each of the DNA recognition modules and a backbone vector.

Zhang sees a very bright future for TALEs and envisages them as tools for testing disease models in whole animals.

Sangamo’s Gregory is more cautiously optimistic and points out that a lot remains to be discovered about TALEs, such as their off-target activity and their basic biological characterization, including structure and affinity. Also, it is currently not clear how to optimize a TALE that does not perform as expected. With respect to TALEs’ advantage over ZFNs, Gregory says, “we are looking at 15–20 years worth of work designing ZFNs and about 18 months of work designing TALEs. It is a little bit early to say what the true advantages are going to be.”

Nicole Rusk

RESEARCH PAPERS

Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).

Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* **29**, 149–153 (2011).