Efficient and specific gene editing with Cas12a (Cpf1) nuclease variants

Tóth Eszter, PhD

The discovery and the innovative implementation of the bacterial immune system CRISPR (clustered regularly-interspaced short palindromic repeats) as a tool for genome editing during the past few years has triggered an explosion of developments in molecular biology and related fields. At present, developing the palette and versatility of the CRISPR-based genome modification tools is a competitive field with spectacularly intense developments aiming to expand the areas of their applicability by either perfecting or searching for yet new solutions.

The nucleases of the CRISPR system work together with short associated RNA molecules (crRNA or gRNA) that bind the nuclease and guide it to specific target sites on the DNA molecule. A ~20 nt long part of the RNA molecule determines in part the specificity of the nuclease and by altering this sequence (spacer) the nuclease can be redirected to different complementary sites (protospacers) on the DNA. For DNA binding and cleavage by the nuclease, the spacer-complementary protospacer sequence needs to be adjacent to a few nt-long motif, called PAM (protospacer adjacent motif). The two sequences, the protospacer and PAM, together form the target site of the nuclease. Since the PAM sequence is specific to the type of nuclease and it is recognized by one of its domains, this interaction cannot be as easily modified, like the spacer-protospacer interaction. Utilization of nucleases as genome editing tools also relies on the existing repair mechanisms of the cells to eliminate the double strand DNA breaks inflicted.

One of the major limitations of the CRISPR-based method is the appearance of off-target hits. This is not only due to the chance for multiple occurrence of the target sequence of choice in the genome, which may be avoided by using prediction-software, but it is also attributable to its non-perfect fidelity, the mismatch tolerance of the nuclease for up to even 5 mismatches in the complementary target sequence, that may pose serious disadvantage in some of the applications. Our aim is to increase further the productivity and broaden the applicability of CRISPR-based tools. We achieve this by implementing the newly discovered and more precise Cas12a (or Cpf1) nucleases.

We managed to characterize different Cas12a orthologues (from *Acidaminococcus sp.*, from *Lachnospiraceae* bacterium, from *Moraxella bovoculi* and from *Francisella novicida*) and generated their PAM mutant variants (called RVR and RR variants) to broaden the DNA targets available for Cas12a nucleases. Our experiments demonstrate that Fn- and MbCas12a can be used to perform genome engineering tasks with a considerable efficiency in mammalian cells. However, neither Fn-, nor MbCas12a demonstrate a general higher activity on targets with VTTV PAM sequences compared to As- or LbCas12a, suggesting that the sequence requirements in mammalian cells for As- and

LbCas12a are more relaxed than those identified *in vitro* whereas those for Fn- and MbCas12a are stricter. Nevertheless, since their target specificities are not identical and because there is a target dependency of their PAM preferences, they seem destined to become useful complements to As- and LbCas12a. We also generated the so called RVR and RR PAM mutant variants of both Fn- and MbCas12a and characterized them along with the analogous As- and LbCas12a variants and the corresponding wild type proteins. Our results revealed that the mutations altered the PAM preferences of all four Cas12as in a similar manner and redefined the known PAM preferences of the As- and LbCas12a variants: RVR mutants recognize targets with TWTV PAMs, while RR mutants prefer targets with TYYV or TTYV/TCCV PAMs. We also showed here that the mutant variants retained their activity on the canonical TTTV PAM, suggesting that they are superior alternatives with a more relaxed PAM recognition compared to the wild type proteins in most applications (Tóth et al. 2018, Nucleic Acids Research).

We continued the evolution of LbCas12a toward the acceptance of new PAM sequences. We combined the already existing RVR and RR mutants (termed RVRR mutant) and found that the PAM preference of the RVRR mutants of LbCas12a nuclease unite the PAM motifs on which the WT, RVR and RR variants demonstrate activities, and also involve new PAM motifs (TGTN) on which no LbCas12a variants have been demonstrated to exert substantial activities so far. Their target space is considerably broadened, thus the theoretical probability of finding a suitable PAM for the RVRR variants in DNA sequences increases to about three-fold compared to the RR variant, which, until now, has possessed the broadest PAM space among the LbCas12a nucleases. Our results suggest that the RVRR mutants offer superior alternatives to the wild type, RVR or RR variants of LbCas12a in many applications, such as multiplex gene editing, transcriptome modulation, epigenetic editing and C to T base editing. Moreover, introduction of a single point mutation at the N-terminal part of LbCas12a resulted in even more relaxed PAM constraints and slightly enhanced activity. This variant is termed improved or impLbCas12a variant. We have shown that impLbCas12a is a useful alternative for the recently described enAsCas12a (Kleinstiver et al 2019, Nature Biotechnology), owing to their different activities on various PAM sequences. Further advantages of the LbCas12a variants include their less pronounced temperature dependence compared to the AsCas12a nucleases, and high activity at lower temperatures. Thus, our findings suggest that the application of impLbCas12a is more advantageous than that of enAsCas12a for ectotherm animals and plants (Tóth et al 2020, Nucleic Acids Research).