**Recognition mechanism of Cpf1-crRNA complex in CRISPR-Cpf1 from crystal structure**

Author:
Major:
Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA

**Key Words:**

CRISPR-Cpf1,crRNA recognition, mechanism, crystal structure

**After discovery of the first CRISPR system, they have been widely used as a highly efficient genome editing tool for their steerable process and great specificity as well. But with more discoveries of many new CRISPR system springing out, we also need to better understand their structures in detail in order to utilize the whole system more specifically, which also has great potential contributions to treatment and prevention for abnormal issues when using CRISPR system. Recently, many researchers are devoting them into structure determination work and have made some impressive progress. In this microreview, we talk about the structure of one novel CRIPSR system, different from well-known CRISPR-Cas9 system, named CRISPR-Cpf1, but also belongs to class 2 CRISPR-Cas system. Some researchers recently obtained the crystal structure of Cpf1-crRNA complex, which is tremendously useful to propose mechanism for the significant RNA recognition process. Based on these work, we can make further steps in engineering CRISPR systems but regardless of current discovered diverse sub-systems like CRISPR-Cpf1, and more and more clear understanding for them, there still exist many open unanswered question about detailed structural information and approaches to improve system’s accuracy or flexibly adapt system’s endurance. In all, more brilliant work are needed in this appreciably promising field.**

**Introduction**

The CRISPR-Cas system, known as CRISPR-Cas9, are RNA-guided adaptive immune systems, which are originally discovered in Bacteria and Archaea, where they are used to defend against viral infection. Here the full name of CRISPR is “Clustered regularly interspaced short palindromic repeats”, and as what the full name shows, CRISPR actually contains two parts: spaced and palindromic gene segments. They are arranged alternatively and regularly in relevant regions of genome. Now we have known the general functional pathway of CRISPR-Cas system. We will take CRISPR-Cas9 as the example: After the invader passes their DNAs into the host, the Cas protein will combine those DNAs to create a novel spacer if that is the first time for that kind of DNAs to invade that host. The spacer will be inserted into the CRISPR array within the host genome. After that, expression and processing of the precursor crRNA—CRISPR RNA, produces mature crRNAs. The mature crRNA then guide an effector protein to target and cleave those foreign DNAs (or RNAs in some cases) bearing complementary sequences. Here according to effector protein, we can classify the general CRISPR system into two classes: class 1 with a Cas protein complex and class 2 with a large single Cas protein. What’s fascinating for CRISPR system is after the first immune response to one certain foreign DNA, the system will quickly respond when the same infection happens again because the CRISPR array has recorded the information and bestows adaptive immunity to the host by continuing adding more spacers into the CRISPR array.

As to the effector protein, many researches are interested in the class 2 CRISPR system which has a largely single Cas protein. Typical class 2 CRISPR system include well-characterized CRISPR-Cas9. The combination of Cas9 from *Streptococcus pyogenes* (SpyCas9) and a synthetic single-guide RNA (sgRNA) that contains a guide region and duplex of crRNA and *trans*-activating crRNA (tracrRNA) has been engineered as a two-component programmable system for genetic manipulation of various organisms[1].

**Recent Progress**

Some researchers now identify the structure of the other class 2 CRISPR system—CRISPR-Cpf1[1]. Different from CRISPR-Cas, in CRISPR-Cpf1, Cpf1 serves as effector protein. Although functionally conserved, Cpf1 and Cas9 are different in many aspects like guide RNA and substrate specificity. According to De Dong’s work, we have an insight of this new class 2 CRISPR system based on the 2.38Å crystal structure of crRNA-boundCpf1.

Cpf1, different from Cas9 system, comprises the N-terminal helical domain, an RNA-recognizing domain (OBD), the C-terminal RuvC domain (nuclease domain). Interestingly, the three portions form a triangle-shaped architecture with a large positively charged channel at the center. What’s more, a looped-out helical domain (LHD) from OBD region is positioned nearly perpendicular to the planar triangle. The OBD and LHD may be the sites to recognize and bind crRNA and target dsDNA because of their structural similarity to one RNA-bind protein (PDB,1WJX) and λintegrase (PDB,2OXO), respectively.

The crystal structures clearly shows that the Cpf1-bound crRNA is highly distorted, containing a short stem-loop-like structure. This twisted conformation is markedly different from that of the sgRNA bound by SpyCas9 or SaCas9 but stabilized through extensive intramolecular interactions. Those intramolecular interactions are very complicated to some extent due to the crowded space but abundant atoms there, but can be coarsely divided into following categories: (1) hydrogen bonds and stacking interaction within nucleobases of crRNA.(2)electrostatic interaction. (3) interactions between crRNA and Cpf1, including van der Waals.

The Mg ion plays a very important role in stabilizing structure, providing the positive charges to neutralize crRNA backbone’s negative charges and removal of Mg ion by EDTA shows the binding affinity between crRNA and Cpf1 was reduced about 50-fold as compared to that in the absence of EDTA[1].

Compared to intramolecular interactions in crRNA, intermolecular interactions between crRNA and Cpf1 seem more diverse. These interactions not only help distorted crRNA to stay stable, but contribute to crRNA recognition too. Most of these intermolecular interactions are mediated by the sugar-phosphate backbone of the bound-crRNA. For example, one uracil nucleobase around the stem-looped region splayed out, pointing to protein residues, forming non-conventional pairing like U-U, stacking force and hydrogen bonds. At the same time, the three U-interacting residues are well conserved among Cpf1 orthologues, suggesting a consistent function in crRNA recognition. Additionally, several 2’-OH groups on crRNA are involved in interactions with Cpf1, explaining the recognition. More importantly, like those three U-interacting residues, structure-based sequence alignment revealed that crRNA-interacting residues are largely conserved, indicating a conserved crRNA recognition mechanism.

The crRNA in CRISPR-Cpf1 system can be divided into two parts: direct repeat sequence and guide sequence. The guide sequence is used to bind target dsDNA but direct repeat sequence is used for structural stabilization. De Dong[1] also finds the conformational change before and after crRNA binding Cpf1 protein. They observe the reduced frictional coefficient of Cpf1, indicating the crRNA binding resulted in a more compact conformation of Cpf1 protein. Besides, this conformational change seems not responsive to the presence of guide sequence because they still observe the contraction even after removing the guide sequence (crRNA\*). This can be explained by less interaction between guide sequence and the complex This is markedly different from Cas9-bound sgRNA that requires both the direct repeat and guide sequences to alter Cas protein conformation.

As the dsDNA binding site, which is proposed as LHD. As described above, its position is out of the planar triangle and its conformation is very similar toλintegrase. Compared to Cas9 system, where dsDNA binding sites have two factors: PAM and Wedge, interestingly, LHD is located at the equivalent position, considering structural similarity between Cas9 and Cpf1. What’s more, the removal of Cpf1 has no effect to triangle structure due to little contact.

**Discussion**

 Thanks to those efforts, especially the achievement of crystal structure, we can have an insight for the structure and properties. So we can analyze possible recognition mechanism based on the interactions in structure. But there still remains some problems which need further efforts.

Firstly, the relevant publication[1] has much discussion based on structure analysis, such as recognition mechanism, but still lacks more variable-controlled experiments to verify them. So more experiments like mutation comparison are needed to find out more details inside. For example, we have known that various interactions have contributions to structural stabilization and specific recognition. But we have no idea which interaction is dominant and which one is not indispensable.

Secondly, recent work does not implement enough stability analysis. Only structure-based interaction enumeration is not enough for stability analysis. As to protein stability, the most convincing data is the parameter like melting temperature, represent ΔG between folded and unfolded state. These data can better show what the conformation favors resulting from crRNA binding.

 At last, the final goal is to apply our understanding for CRISPR system to practice like gene editing. So there are many open opportunities of engineering Cpf1 to make it more specific and therapeutic applications.

**References**

[1] De Dong, Kuan Ren, Xiaolin Qiu, etc. “The crystal structure of Cpf1 in complex with CRISPR RNA”. Nature.532(2016):522-526