Modeling of tandem dCas9 complexes bound to DNA for nucleic acids detection

Roman Novikov[1], Julia Kacher[2], Anna Gribkova[2], Petr Zaytsev[2], Grigoriy Armeev[2], Grigoriy Gluhov[1] and Alexey Shaytan[2]

[1]Sirius University of Science and Technology, Sochi, Moskva, Russia, [2]Sirius University of Science and Technology, Sochi, Russia

Detection of pathogens is an acute concern for the diagnosis of a broad spectrum of human, animal, and plant diseases, environmental monitoring, and food safety management. Highly sensitive, specific, fast, and easy-to-use diagnostic methods are required. As the "gold standard" PCR method requires expensive laboratory equipment and qualified personnel, other approaches should be developed. A promising approach is the use of the natural ability of bacterial CRISPR/Cas9 systems to recognize DNA sequences with high specificity under isothermal conditions. r work was dedicated to the design and study of the applicability of a biosensor system based on dCas9 proteins fused to the domains of a split-enzyme reporter system. Two dCas9 proteins can bind to the target DNA sequence at spatially close sites without cutting them. The split reporter enzyme is responsible for the detection of spatial colocalization of these protein complexes. We determined possible mutual orientations of two dCas9 proteins at the target loci of genomic DNA, which are favorable for the interaction of the attached domains of the reporter protein. The optimal distances between the DNA binding sites of dCas9 proteins in different orientations ("PAM-direct", "PAM-in", and "PAM-out") were calculated. The modeling was performed to evaluate the dependence of the system structure on the distance along the DNA between the binding sites of dCas9 proteins.

To model dCas9/sgRNA/DNA complexes in different orientations, the SpdCas9/sgRNA/DNA (PDB ID 5Y36) structure was chosen; and for modeling additional DNA between two dCas9 complexes, the structure of ideal B-DNA was used [1]. For attaching the DNA to the complexes the structural alignment of the two terminal DNA nucleotides with the two terminal nucleotides of the duplex DNA in the dCas9 complexes was performed. These models were created using Python scripts for the UCSF Chimera [2]. To measure distances and angles in the resulting models, we used the MDAnalysis and NumPy packages [3,4]. To evaluate the optimal rotational positioning of dCas9 proteins with respect to the DNA axis the dihedral angle α formed by the following points was calculated: the geometric centers of the N- and C-terminal amino acid residues of the two dCas9 proteins and the geometric centers of the two central C1' atoms of DNA separating dCas9 proteins. Matplotlib package was used for plotting [5].

Using molecular modeling, we have calculated the dependence of the distance between the N- and C-terminal residues of a pair of dCas9 proteins (to which parts of the split reporter system can be attached) on the distance between the PAM sequences along the DNA (Figure 1A). The angle α was calculated as well. On the basis of these data, the optimal variants of the attachment of the split enzyme domains were determined: for PAM-in orientation — the C-termini of dCas9 proteins; for PAM-out orientation — the N-termini of dCas9 proteins; for PAM-direct orientation — the C-terminus of the dCas9 protein located at the 5'-end of the recognized locus and the N-terminus of the dCas9-protein located at the 3'-end of the recognized locus (locus orientation is determined by the DNA strand forming the heteroduplex). Also, for each variant, models with the smallest values of the distances between the attachment points of split reporter system domains and optimal values of α angles without steric overlap were determined. The most optimal options for the distances between the PAM sequences are the following: for PAM-direct

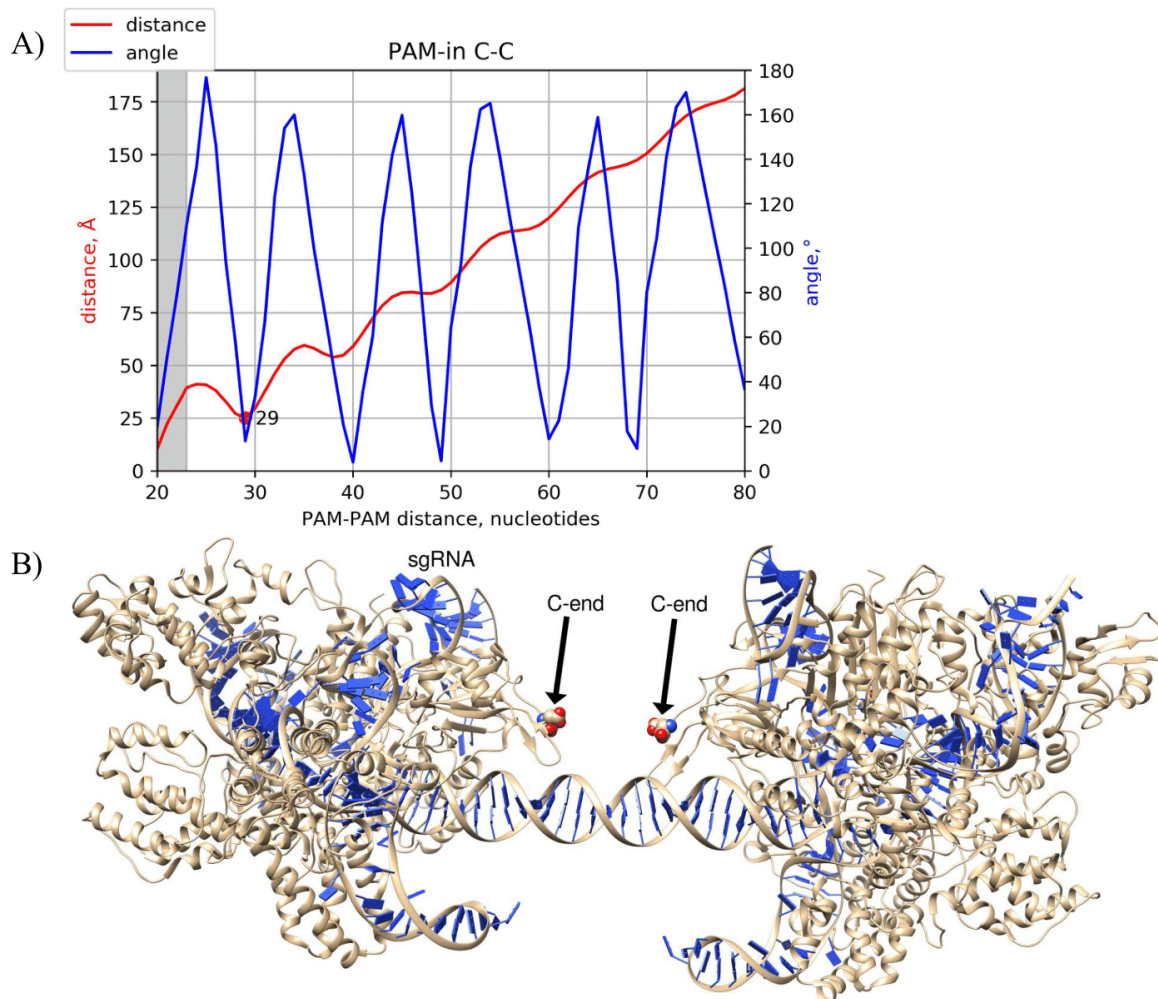orientation — 41 nucleotides, for PAM-in — 29 nucleotides, and for PAM-out — 52 nucleotides (Figure 1B).

Figure 1. A. The dependence of the distance between the C-ends of the two SpdCas9 proteins and α-angle in the PAM-in orientation on the distance between the PAM sequences; a sterically inaccessible area is shaded in gray; the optimal configuration is marked with a dot. B. The structure of the two SpdCas9 proteins forming a nucleo-protein complex in the PAM-in orientation with the distance between PAM sequences of 29 nucleotides; the spheres mark the C-ends of SpdCas9 proteins.

References
1. *Huai C., Li G., Yao R., Zhang Y., Cao M., Kong L., Jia C., Yuan H., Chen H., Lu D., Huang Q.* Structural insights into DNA cleavage activation of CRISPR-Cas9 system: 1 // Nat. Commun. 2017. Vol. 8. N 1. P. 1375.

2. *Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E.* UCSF Chimera—A visualization system for exploratory research and analysis // J. Comput. Chem. 2004. Vol. 25. N 13. P. 1605–1612.

3. *Gowers R., Linke M., Barnoud J., Reddy T., Melo M., Seyler S., Domański J., Dotson D., Buchoux S., Kenney I., Beckstein O.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. 2016. P. 98–105.

4. *Harris C.R. et al.* Array programming with NumPy: 7825 // Nature. 2020. Vol. 585. N 7825. P. 357–362.

5. *Hunter J.D.* Matplotlib: A 2D Graphics Environment // Comput. Sci. Eng. 2007. Vol. 9. N 3. P. 90–95.