Design of TALEN to Edit the Genome of Burkholderia Glumae

[1]Ramya, I.R., [2]Neeta Shivakumar

Department of Biotechnology, R V College of Engineering, Mysore Road, Bangalore, Karnataka, India
Email: ir.ramya@gmail.com

Abstract — Burkholderia glumae (BGR1) is a bacteria that causes panicle blight in rice crops. It produces toxoflavin, a phytotoxin, which is a key virulence factor. ToxR is a LysR-type regulator that activates both the toxABCDE (toxoflavin biosynthesis genes) and toxFGHI (toxoflavin transporter genes) operons in the presence of toxoflavin as a coinducer. Here we genetically engineer a hybrid protein that consists of a high-specificity Transcription Activator Like effector (TAL) that is a DNA binding domain fused with a FokI endonuclease to form TALEN. This hybrid protein functions in pairs to perform the DNA double-strand break in the desired location. Binding specificity is determined by customizable arrays of polymorphic amino acid repeats in the TAL effectors. Once the two monomers of the FokI endonuclease dimerize, the cut is made in both the strands in the spacer region. This technique can be used to eradicate the virulence causing genes of BGR1.

I. INTRODUCTION

Burkholderia glumae (BGR1) is the bacterial pathogen that causes bacterial panicle blight (BPB) in rice. Its growth and pathogenicity is favored by high temperatures and has affected several countries around the globe. Toxoflavin is one among its major virulence factors. The complete genome of BGR1 was first sequenced in Korea in 2009. It consists of two chromosomes and four plasmids. Chromosome 1 is 3,906,529 base pairs in size and has 3290 coding sequences, 144 pseudogenes, 3 rRNA operons and 56 tRNAs. Chromosome 2 is 2,827,355 base pairs in size and has 2079 coding sequences, 192 pseudogenes, 2 rRNA operons and 8 tRNAs. Toxoflavin biosynthesis and transporter genes and other pathogencity related genes involved in the hypersensitive response and the type III protein secretion system are found in Chromosome 2, whose origin of replication is similar to the plasmids [1].

Genome editing has been in use over several years to modify the traits of an organism. One such modification discussed here is gene knock-out. We use custom engineered hybrid protein that composes of a transcription activator like effector (TAL) domain and an endonuclease domain to make a TALEN.

Before the advent of TALENs, custom-designed zinc finger nucleases (ZFN) were used to edit DNA. Hybrid proteins derived from the DNA-binding domains of zinc finger (ZF) proteins were fused with the non-specific DNA-cleavage domain of the FokI endonuclease. Though this technique was successful on many different organisms such as yeast, plants, mammals and even human cells, it was time consuming, labor intensive and prone to a high rate of failure [2].

The TAL effectors found in nature in various strains of Xanthomonas sp. use a Type III secretion system to enter the host cells and thus are called Type III effectors. Once they enter the host cells, they can transcriptionally activate the host genes either for strain virulence or avirulence [3]–[6].

The key component of TALEN is the DNA recognition domain that binds to a precise location. The endonuclease then makes a double-strand break (DSB). In solution, FokI exists as an inactive monomer. In order to become functional, two molecules of FokI each bind to either strands on the DNA. They then dimerized in the space between them. Once activated, the dimer can cleave the DNA strand and cause a double strand break.

The TAL effectors consist of a tandem array of repeat region. In nature this region can have 13–29 repeats. Each repeat is approximately 34 amino acids (aa) long and corresponds to one nucleotide from the target region of the host DNA. For example, if the target DNA is 16 nucleotides long, there will be 16 repeats in the binding domain of the TALEN. The specificity of binding is decided by the 12^{th} and 13^{th} aa of each repeat [7]–[8]. We have designed the array of repeats to which we attached a FokI endonuclease to complete the hybrid protein. This procedure was repeated for both strands of the DNA on opposite sides of the cut site. Once the binding domains bind to the specific binding sites on the host DNA, the FokI dimerize to make the cut. Hence the virulence causing genes in BGR1 can be edited out using this technique.

II. MATERIALS AND METHODS

A. Obtaining sequence

The toxoflavin biosynthesis-related gene cluster was obtained from NCBI. ToxR, the toxoflavin regulator gene for both synthesis and transportation, is coded in the range 174-1175. Hence we perform two cuts at both the

specified locations in order to completely remove the regulator.

B. RVD Selection

DNA sequence was supplied to the TALEN Targeter tool. All four architectures were used. NH was chosen as G substitute and the specific cut sites were provided to filter the options. Upstream base was retained at T as recommended [9]–[10]. The RVDs with the highest percentage of HD or NN/NH among all architectures were selected for each TAL.

C. TAL Plasmid Sequence Assembly

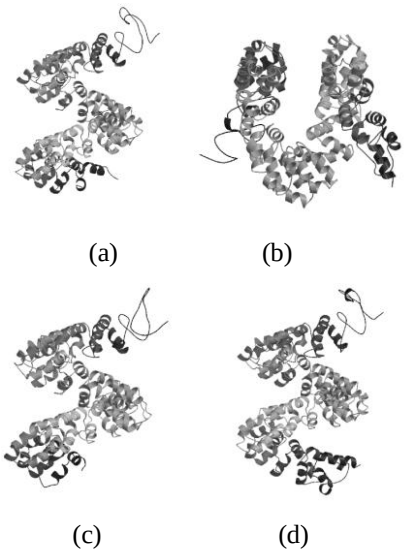
The RVD1 and RVD2 obtained for each cut site were fed to TAL plasmid sequence assembly tool. Cermak et al. (Golden Gate) assembly method was chosen with pTAL3 as the destination vector. NH was chosen as the default Guanosine RVD to maintain consistency. Sequences were confirmed using the TAL Plasmids Assembly Tool (http://bit.ly/assembleTALsequences).

D. Obtaining 3D Structure

3D structure of each of the four TALs was obtained using homology modeling submitted to SWISS-MODEL workspace. In comparison, most of the sequences generated as artificial TALENs here are nearly ninety percent similar to the ones that occur in nature. The SWISS-MODEL template library (SMTL version 18-06-14, PDB release 13-06-2014) was searched with Blast and HHBlits for evolutionary related structures matching the target sequence [11]–[13].

E. Visualizing The 3D Structure

The 3D structures obtained from SWISS MODEL can be visualized using RasMol Molecular Graphics software. Three models were generated for each TAL. Here, we list the first model of each TAL.



(a) (b)

(c) (d)

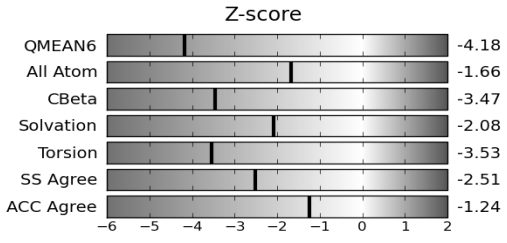
Fig. 5. 3D structures of TALs obtained from SWISS MODEL. a) TAL1 b) TAL2 c) TAL3 d) TAL4

RESULTS

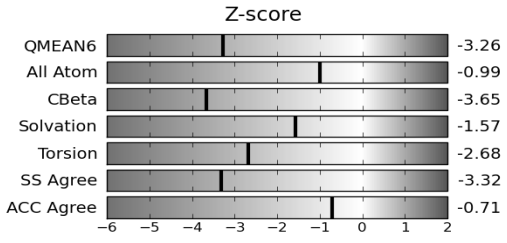
Estimating the quality of protein structure models is a vital step in protein structure prediction. Often best candidate shall be selected from a set of alternative models. QMEAN is a composite scoring function which is able to derive both global (i.e. for the entire structure) and local (i.e. per residue) error estimates on the basis of one single model. The QMEAN Z-score provides an estimate of the absolute quality of a model by relating it to reference structures solved by X-ray crystallography. The QMEAN Z-score is an estimate of the "degree of nativeness" of the structural features observed in a model by describing the likelihood that a model is of comparable quality to high-resolution experimental structures.

The three plots available for download visualize the quality of a given model with respect to these reference structures. The reference structures are a non-redundant subset of the PDB sharing less than 30% pairwise sequence identity among each other and are solved at a resolution below 2 Å.

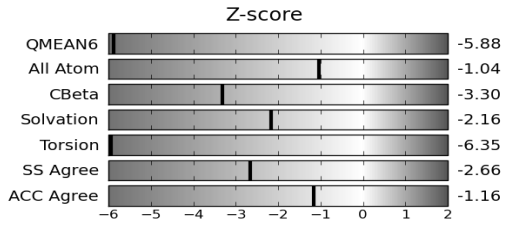
The analysis of these Z-scores of the individual terms can help identifying the geometrical features responsible for an observed large negative QMEAN Z-score. Models of low quality are expected to have strongly negative Z-scores for QMEAN but also for most of the contributing terms. Large negative values correspond to red regions in the color gradient. "Good structures" are expected to have all sliders in the light red to blue region [14].



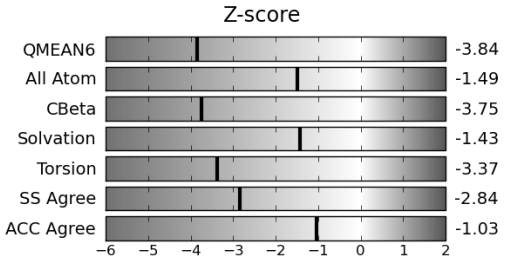
(a)



(b)



(c)



(d)

Fig. 6. Z-scores of TALs obtained from SWISS MODEL. a) TAL1 b) TAL2 c) TAL3 d) TAL4

REFERENCES

[1]. Hari Sharan Karki, "Physiological, biochemical and molecular characteristics associated with virulence of burkholderia glumae: the major causative agent of bacterial panicle blight of rice," M.S. thesis, The Dept. of Plant Pathology and Crop Physiology, Louisiana State University, 2010.

[2]. Kim,Y.G., Cha,J. and Chandrasegaran,S., "Hybrid restriction enzymes: zinc finger fusions to fok I cleavage domain," Proc. Natl Acad. Sci. USA, 93, 1156–1160, 1996

[3]. Gu,K., Yang,B., Tian,D., Wu,L., Wang,D., Sreekala,C., Yang,F., Chu,Z., Wang,G.L., White,F.F. et al., "gene expression induced by a type-III effector triggers disease resistance in rice," Nature, 435, 1122–1125, 2005.

[4]. Yang,B., Sugio,A. and White,F.F., "Os8N3 is a host disease-susceptibility gene for bacterial blight of rice," Proc. Natl Acad. Sci. USA, 103, 10503–10508, 2006.

[5]. Sugio,A., Yang,B., Zhu,T. and White,F.F., "Two type III effector genes of Xanthomonas oryzae pv. oryzae control the induction of the host genes OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice," Proc. Natl Acad. Sci. USA, 104, 10720–10725, 2007.

[6]. Romer,P., Hahn,S., Jordan,T., Strauss,T., Bonas,U. and Lahaye,T., "Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene," Science, 318, 645–648, 2007.

[7]. Moscou,M.J. and Bogdanove,A.J., "A simple cipher governs DNA recognition by TAL effectors," Science, 326, 1501, 2009.

[8]. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U., "Breaking the code of DNA binding specificity of TAL-type III effectors," Science, 326, 1509–1512, 2009.

[9]. Doyle E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., VanDyk, J.K., and Bogdanove, A.J., "TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction," Nucleic Acids Res. doi: 10.1093/nar/gks608, 2012.

[10]. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.J., Bogdanove, A.J., and Voytas, D.F., "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting," Nucleic Acids Res. 39:e82, 2011.

[11]. Marco Biasini, Stefan Bienert, Andrew Waterhouse, Konstantin Arnold, Gabriel Studer, Tobias Schmidt, Florian Kiefer, Tiziano Gallo Cassarino, Martino Bertoni, Lorenza Bordoli, Torsten Schwede, "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information," Nucleic Acids Research; doi: 10.1093/nar/gku340, 2014.

[12]. Arnold K., Bordoli L., Kopp J., and Schwede T., "The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling," Bioinformatics, 22,195-201, 2006.

[13]. Benkert, P., Biasini, M. and Schwede, T. "Toward the estimation of the absolute quality of individual protein structure models," Bioinformatics, 27, 343-350, 2011.

[14]. Benkert P, Biasini M, Schwede T., "Toward the estimation of the absolute quality of individual protein structure models," Bioinformatics, 27(3):343-50, 2011.

◈◈◈