Literature Learning Module

Cut Site Selection by the Two Nuclease Domains of the Cas9 RNA-guided Endonuclease

Assessment of Biochemistry/Molecular Biology (BMB) Foundational Concepts

[**Henry Jakubowski**](http://employees.csbsju.edu/hjakubowski/), Ph.D., Professor, [**Chemistry Department**](http://www.csbsju.edu/chemistry), [**College of Saint Benedict/Saint John's University**](http://www.csbsju.edu/)

The questions below are based on data, graphs, and figures from the following article:

Cut Site Selection by the Two Nuclease Domains of the Cas9 RNA-guided Endonuclease.   Hongfan Chen, Jihoon Choi and Scott Bailey.The Journal of Biological Chemistry, **289**, 13284-13294 (2014)

[**These guidelines**](http://www.jbc.org/site/misc/Copyright_Permission.xhtml) apply to the reuse of articles, figures, charts and photos in the Journal of Biological Chemistry, Molecular & Cellular Proteomics and the Journal of Lipid Research.

These questions may be used to help assess student mastery of key foundational concepts in BMB developed by the American Society for Biochemistry and Molecular Biology (ASBMB) and tested in the new MCAT2015 developed by the American Association of Medical Colleges.  The particular foundational concepts and learning objectives relevant for this particular assessment are listed below.

|  |
| --- |
| ASBMB Biochemistry and Molecular Biology Foundational Concepts  -   [**Detailed Learning Objectives**](http://employees.csbsju.edu/hjakubowski/BiochemAssessQuestions/ASBMBCoreConceptsObjectives050314.xlsx) (Excel)Foundational Concept 2:  Macromolecular structure determines function and regulation.* 2. Given a list of macromolecules, students should be able to devise an experiment on how they interact or interpret results of experiments on their interactions
* 3.  Given structural changes of a macromolecule (or ligand), students should be able to predict the impact of structural substitution would have on macromolecule structure and function
* 4. Given experimental data, students should be able to assess how enzymes facilitate biochemical reactions.

Foundational Concept 3:  Information storage and flow are dynamic and interactive * 2. Given an understanding of common mechanisms of gene regulation, students will be able to explain or predict changes in transcription in response to biologic variables.
* 3. Given an understanding of genetic information transfer, students should be able to explain the role of RNA in the flow of genetic information

Foundational Concept 4: . Discovery requires objective measurement, quantitative analysis, and clear communication.  * 2. Given a fundamental understanding of BMB concepts, students should be able to formulate experiments and assess the quality of experiments addressing molecular structure, assays of biological function, and isolation / separation of biomolecules.
* 3. Given a data set, students should be able to assess the reliability of the data and draw appropriate conclusions.
* 4. Given a set of data, students should be able to appropriately present and interpret the data.

MCAT 2015:  [**Foundational Concepts, Detailed Content and Topics**](http://employees.csbsju.edu/hjakubowski/BiochemAssessQuestions/MCAT2015_Biochem_Chem_ConceptsContent.xlsx) (Excel)*Biological and Biochemical Foundations of Living Systems* Foundational Concept 1 : Biomolecules have unique properties that determine how they contribute to the structure and function of cells, and how they participate in the processes necessary to maintain life * 1A. Structure and function of proteins and their constituent amino acids
* 1B. Transmission of genetic information from the gene to the protein

*Chemical and Physical Foundations of Biological Systems* Foundational Concept 5: The principles that govern chemical interactions and reactions form the basis for a broader understanding of the molecular dynamics of living systems. * 5B. Nature of molecules and intermolecular interactions
* 5D. Structure, function, and reactivity of biologically - relevant molecules

Scientific Inquiry and Reasoning Skills* 1.  Knowledge of Scientific Concepts and Principles;
* 2.  Scientific Reasoning and Problem;
* 3.  Reasoning about the Design and Execution of Research;
* 4.  Data-based and Statistical Reasoning
 |

Background Reading:

* [**Binding and the Control of Gene Transcription**](http://employees.csbsju.edu/hjakubowski/classes/ch331/bind/olbindtransciption.html) in Biochemistry Online by Henry Jakubowski.  .

Background Videos:

* [**Genome Repair: CRISPR/Cas 9 Genomic Editing**](http://youtu.be/I6jPLBqzNH0) (20 minutes slowly narrated by a protein chemist; wait about 15 sec for high definition to start)

**Background and Review**

A promising tool for genome manipulation and regulation  in a wide variety of organisms has recently been identified in the RNA-guided DNA endonuclease activity of the CRISPR-Cas (clustered regularly interspaced short palindromic repeat–CRISPR-associated) system. CRISPR-Cas, an inheritable prokaryotic immune system, protects bacteria and archaea against mobile genetic elements via RNA-guided target silencing. CRISPR-Cas systems consist of an array of short direct repeats interspersed by variable invader-derived sequences (spacers)  and a cas operon. During invasion, small fragments of the invading DNA from phage or plasmids (protospacers) are incorporated into host CRISPR loci, transcribed, and processed to generate small CRISPR RNAs (crRNA). The invading nucleic acid is then recognized and silenced by Cas proteins guided by the crRNAs. There are three types of CRISPR-Cas system each characterized by the presence of a signature gene.

Programmed DNA cleavage requires the fewest components in the type II CRISPR-Cas system, requiring only crRNA, a trans-activating crRNA (tracrRNA), and the Cas9 endonuclease, the signature gene of the type II system. The system can be further simplified by fusing the mature crRNA and tracrRNA into a single guide RNA (sgRNA). In addition to its role in target cleavage, tracrRNA also mediates crRNA maturation by forming RNA hybrids with primary crRNA transcripts, leading to co-processing of both RNAs by endogenous RNase III. Cas9 contains two nuclease domains that together generate a double-strand (ds) break in target DNA. The HNH nuclease domain cleaves the complementary strand, and the RuvC-like nuclease domain cleaves the noncomplementary strand.

A short signature sequence, named the protospacer adjacent motif (PAM), is characteristic of the invading DNA targeted by the type I and type II CRISPR-Cas systems. The PAM serves two functions. It has been linked to the acquisition of new spacer sequences, and it is necessary for the subsequent recognition and silencing of target DNA. The sequence, length, and position of the PAM vary depending on the CRISPR-Cas type and organism. PAMs from type II systems are located downstream of the protospacer and contain 2–5 bp of conserved sequence. A variable sequence, of up to 4 bp, separates the conserved sequence of the PAM from the protospacer. This variable region is often included in the definition of the PAM sequence, but for simplicity, we refer to this variable region as the linker and the conserved sequence as the PAM. To date, Cas9 from Streptococcus pyogenes, Cas9 from Streptococcus thermophilus DGCC7710, and Cas9 from Neisseria meningitidis have been employed as tools for genome editing or regulation. For these Cas9 orthologs, the PAMs are GG, GGNG, and GATT, and the linkers are 1, 1, and 4 bp, respectively.

The simplicity of sgRNA design and sequence-specific targeting means the RNA-guided Cas9 machinery has great potential for programmable genome engineering. Cas9 can be employed to generate mutations in cells by introducing dsDNA breaks. The capabilities of Cas9 can be expanded to various genome engineering purposes, such as transcription repression or activation, with its nickase (generated by inactivating one of its two nuclease domains) or nuclease null variants. Another appealing possibility for the Cas9 system is to target different Cas9-mediated activities to multiple target sites, for example transcriptional repression of one gene but activation of another. To achieve this, multiple Cas9 orthologs will need to be employed as a single ortholog cannot concurrently mediate different activities at multiple sites. Therefore to broaden our understanding of Cas9 proteins, we have characterized the Cas9 ortholog from S. thermophilus LMG18311, which we refer to as LMG18311 Cas9.

Questions:

1. Investigators studied the type II CRISPR-Cas system of S. thermophilus LMG18311.

a.  Figure A shows a Coomassie Blue-stained SDS-polyacrylamide gel of Cas9, Cas9 D9A, Cas9 D599A, and Cas9 D9A,D599A where D is aspartic acid and A is alanine and D#A reflects the site of the specific mutation where D was changed to A.  Why did they perform this experiment?



b.  To confirm the PAM sequence for LMG18311 Cas9, the authors performed BLAST searches to identify potential protospacers in viral and plasmid genomes that matched any of the 33 spacer sequences from CRISPR-1. This search generated 41 unique target sequences, from the genomes of bacteriophage known to infect S. thermophilus. We then aligned 50-nucleotide segments from the identified target genomes, inclusive of the 30-nucleotide protospacer and 10-nucleotide flanking regions.  Figure b shows B, logo plot revealing the PAM for LMG18311 Cas9. The positions of the protospacer, PAM, and linker are indicated.



What is the most likely consensus sequence of the PAM sequence?  How far is it away from the protospacer:

c. Figure C shows a schematic representation of the CRISPR nucleic acid complexes.  Label the target DNA, crRNA, tracrRNA,, protospacer, PAM, and linker.  Draw in a connecting piece of RNA if you wish to create a single guide RNA containing both the crRNA and tracRNA.  What is the specific PAM sequence in this figure.



2. To confirm that the identified PAM was functional, investigators  needed a selection system.  They used a  transformation assay in which E. coli cells containing an exogenous type II CRISPR-Cas system were resistant to plasmid transformation, whereas cells lacking the system are competent for transformation.  Explain how this system could be used to determine if specific PAM are functional?

a.  Figure A shows a schematic representation of transformation assay.

  

The target plasmid contained protospacer-1 (whose sequence was identical to the first spacer of CRISPR-1), a 2-bp linker, and the identified PAM. The first control plasmid contained only protospacer-1, whereas the second control plasmid lacked both protospacer-1 and PAM. The target and control plasmids were then used to transformed cells.  in the presence of IPTG and the appropriate antibiotics.

Why did the investigators add IPTG?  What difference would you expect to find in the cells in the left panel compared to the right in their assay?

b.  Figure 2B shows plasmid transformation by LMG18311 Cas9 and sgRNA in E. coli cells. Transformation efficiency is expressed as cfu per 5 ng of plasmid DNA. Average values from at least three biological replicates are shown, with error bars representing 1 S.D. Interpret and explain results.



c. For Figure  C, the investigators made single nucleotide changes in the PAM sequence of the CRISPR+ plasmid and studied the effect on plasmid transformation efficiency. Interpret he results.



d. For Figure D, the effect of the length of the linker length on plasmid transformation efficiency was studied. PS denotes the protospacer sequence. Interpret the results.



3.  The investigators then studied their system *in vitro* by reconstituting the components.  They monitored DNA cleavage by LMG18311 Cas9 as a measure of activity.

a.  Figure 3A shows an electrophoresis gel of reaction mixtures containing 5 nm target plasmid, 25 nm Cas9, 25 nm crRNA (42-nucleotide crRNA containing the sequence derived from first spacer of CRISPR-1), 25 nm (42-nucleotide) tracrRNA, and 10 mm Mg2+ that were incubated for 30 min at 37 °C.   The agarose gels were stained with ethidium bromide.   Interpret the results.  Why does the linear control migrate less far than the uncut plasmid?



b. In Figure B, a cognate sgRNA was substituted for separate crRNA and tracrRNA molecules.  Interpret the gel results..



c.  In Figure C, they studied cleavage in the presence of different sgRNA sequences.  Explain the results.



d. For figure D, the used different active site mutants of Cas  9.  The (D9A) mutant was in the RuvC-like domain while the H599A mutant was in the HNH domain  Explain the results.  What effect does the double mutant have on the activity?



e.  In Figure E, the authors studied which domain was involved in the cleavage of each strand (complementary to the guide DNA or the noncomplementary strand) of the target DNA.  The dsDNA was radiolabeled at the 5′ end of the complementary strand (left hand gel) or the noncomplementary strand (right hand gel). Reactions were performed as in A, and products were separated by 10% denaturing PAGE. The cleavage sites are indicated with arrows in the schematic diagram (bottom). 50 nt, 50 nucleotides; 37 nt, 37 nucleotides. Interpret the results.



f.  Do mutations in the PAM or changes in linker length have the same effect on DNA interference *in vitro* as they did *in vivo*?   The investigators monitored cleavage of these variant plasmids by recombinant LMG18311 Cas9.   Interpret the results in Figure F.



g.  In Figure G, cleavage of plasmid targets containing the indicated linker lengths was studied.  In A–C and E–F, the positions of negatively supercoiled (nSC), linear (L) and nicked or open circle (OC) plasmid are indicated. The linear control is a digestion of the plasmid target with the restriction enzyme AgeI



h.  In the mutational analysis, they substituted asparatic acid for alanine.  Draw the structures of the side chains of D and A.  Why did they choose A to substitute for D?  What role might D have in catalysis in the wild-type protein?



4. To evaluate whether other divalent cations besides Mg2+ can activate DNA cleavage by Cas9, the authors performed plasmid cleavage assays in the presence of one of the following divalent cations: Ca2+, Mn2+, Co2+, Ni2+, and Cu2+.

a.  Figure A shows cleavage of a target plasmid by Cas9 with either no metal or 1 mm of the indicated metal ions. All reactions were treated with 0.5 mm EDTA prior to metal addition.  Interpret the results.



b. To identify which domain was activated, we assayed the single active site mutants of Cas9 (D9A or H599A) in a reaction buffer containing Ca2+.  In both panels, the positions of negatively supercoiled (nSC), linear (L), and nicked or open circle (OC) plasmid are indicated. The linear control is a digestion of the plasmid target with the restriction enzyme AgeI.



5. To determine the effect of PAM sequence and linker length on binding of LMG18311 Cas9 to DNA targets, the authors determined the binding affinity (Kd) of the Cas9-sgRNA complex to 5′ end-labeled dsDNA targets using native gel electrophoresis.

a.  Why did they not include SDS in the gel?

b.  Binding experiments were conducted with the nuclease-deficient mutant of Cas9 (D9A,H599A) in the presence of Mg2+. Fixed concentrations of the dsDNA targets were incubated with increasing concentrations of the Cas9-sgRNA complex.  Why did they use the double mutant of Cas9?

c.  Figure A below shows a representative gel shift assay for Cas9-sgRNA and the binding curve measured from the assay.   From inspection of the graph, what is the approximate Kd for the interaction? The top wedge indicates increasing concentration of the Cas9-sgRNA.



b. A bar graph plotting Kd values for DNA targets with PAM mutations (labeled red) are shown below (B) . Average values from at least three replicates are shown, with error bars representing 1 S.D. Targets where binding was not observed are shown with Kd values at the lower limit (> 1000 nm). PS denotes the protospacer sequence.  Interpret the results.



c.  Next they investigated linker length and binding.  Figure C shows a bar graph plotting Kd values for DNA targets with different linker lengths (labeled red) (C). Average values from at least three replicates are shown, with error bars representing 1 S.D. Targets where binding was not observed are shown with Kd values at the lower limit (> 1000 nm). PS denotes the protospacer sequence. Interpret the results.

