3

Optimal CRISPR-Cas9 Splitting Locations Report

by **CRISPR**WIZ

The Goal of This Report

The goal of this report is to identify optimal splitting points for the Cas9 protein to enhance viral delivery without compromising its functionality. The output includes:

- Detailed split point options for fragmenting the Cas9 protein. •

Optimal intein selection for efficient protein splicing.

- Analysis of the impact on protein flexibility and electrostatic potential. •

  Visualization of the spCas9 protein structure.

Overview

Role of Cas9 in CRISPR

Cas9 is an essential component of the CRISPR system, functioning as a molecular scissors to cut DNA at specific locations guided by RNA. This ability to precisely edit genes makes it a powerful tool for genetic research and
potential therapeutic applications.

Primary methods to deliver CRISPR-Cas9

CRISPR-Cas9 can be delivered into target cells using various methods, each with its advantages and limitations. The primary methods include:

1. **Microinjection**: Directly injecting the CRISPR-Cas9 components into individual cells, typically used in early-stage embryos. This method offers high precision but is labor-intensive and not suitable for large-scale applications.
2. **Electroporation**: Using electrical pulses to create temporary pores in cell membranes, allowing CRISPR-Cas9 components to enter the cells. This method is efficient for many cell types but can cause cell damage.
3. **Chemical Transfection**: Utilizing chemical reagents to facilitate the entry of CRISPR-Cas9 components into cells. It is straightforward and scalable but may have lower efficiency and specificity compared to other methods.
4. **Viral Vectors**: Employing viruses to deliver CRISPR-Cas9 components into target cells. Viral vectors are particularly effective for therapeutic applications due to their high efficiency and ability to target specific cell types.

Why Splitting CRISPR-Cas9

Viral vectors also have some limitations, such as potential immune responses and limited cargo capacity. To address these challenges, researchers are exploring strategies like splitting the Cas9 protein into smaller fragments to ensure efficient delivery and functionality.

Splitting the Cas9 protein into smaller fragments for enhanced viral delivery is a promising strategy to overcome the limitations of current delivery methods. The following report provides a systematic approach to selecting optimal split points

that ensure the reassembled Cas9 retains its functionality. This algorithm considers structural analysis, mutagenesis tolerance, and split point selection to create viable options for fragmenting the Cas9 protein.

Client Input

Key Parameters

- **CRISPR-Cas9 variant name**: spCas9

  Identifying the specific Cas9 variant is crucial because different variants may have different structural and functional properties, affecting the optimal splitting points.

- **Viral Vector (Virus) name**: PVX (Potato Virus X)

  Knowing the viral vector helps determine the size constraints, maximum capacity, and delivery efficiency, which are essential for designing the split Cas9 fragments.

- **Target DNA type**: Plants

  The target DNA type influences the design and application of the CRISPR system, ensuring that the chosen split points and delivery method are suitable for plant genomes.

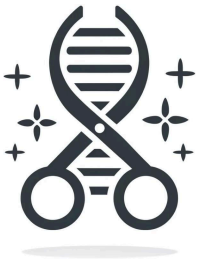
Special Requirements

Take into account the following lengths when splitting into fragments.

Including these lengths ensures that the resulting fragments do not exceed the carrying capacity of the vector, thereby facilitating efficient delivery and functionality of the Cas9 system. •

  gRNA length: 175 bp

- NLS length: 70 bp
- TLS length: 75 bp



Part 1 - Splits Suggestions

Calculation of the Optimal Number of Splitting Points

The calculation of the optimal number of splits required for a Cas9 protein to fit within a viral vector is based on various overhead components. It considers the maximum accommodation capacity of the viral vector, the size of the Cas9 protein, and the lengths of inteins, NLS, TLS, gRNA, and a buffer. The function iteratively splits the Cas9 protein into parts, calculating the effective size of each part and tracking the number of splits.

Analysis Process of the Splitting Points

Step 1: Structural and Functional Domain Analysis

- **Identify Domains and Linkers**: The functional domains and flexible linker regions of spCas9 were identified based on sequence and structural information. Critical domains include the HNH, RuvC-I, RuvC-II, RuvC-III, and the PAM-interacting domain.

Step 2: Computational Prediction of Mutagenesis Tolerance

- **Mutagenesis Data**: References and computational tools were used to predict residues that can tolerate mutations within the flexible linker regions. This involves literature and existing mutation tolerance data to ensure functional integrity.

Step 3: Selection of Split Points

- **Initial points**: Flexible linker regions and mutagenesis data were used to select initial points within the constraints of maximum sizes.

- **Split Locations**: Exact split points were refined iteratively to avoid disrupting functional domains and ensure each segment falls within size constraints.

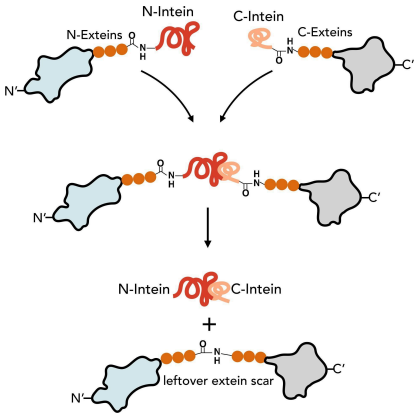
Suggested Optimal Split Locations

- **Number of splitting points**: 4
- **Number of total parts**: 5 (4+1)
- **Split Absolute Locations**: 1000, 1929, 2854, 3731
- **Segment Lengths**:
  - **Part 1**: 1-1000 bp, total: 1000 bp
  - **Part 2**: 1001-1929 bp, total: 929 bp
  - **Part 3**: 1930-2854 bp, total: 925 bp
  - **Part 4**: 2855-3731 bp, total: 877 bp
  - **Part 5**: 3732-4010 bp, total: 279 bp

Reason for Choice: This option provides a balance between structural integrity, functionality, and adherence to size constraints ensuring minimal disruption to the Cas9 protein domains.

Conclusion

The identified splitting points for the spCas9 protein align with the structural and functional integrity while meeting the viral vector's constraints. The selected optimal split option demonstrates a balanced and methodical approach to Cas9 fragmentation, facilitating efficient delivery and functionality in plant systems.



Part 2 - Suggested Inteins

Introduction

Inteins are crucial for splitting the CRISPR-Cas9 protein to facilitate delivery via viral vectors. They excise themselves from a protein and join the remaining portions (exteins) together with a peptide bond, a process known as protein splicing. When Cas9 is split into fragments, inteins reassemble the protein inside the target cell, ensuring it retains its functional integrity.

Inteins are essential for the successful splitting and reassembly of CRISPR-Cas9, ensuring functional integrity and enhancing delivery via viral vectors. This approach addresses size limitations and maintains the efficiency and specificity required for effective gene editing.

Choosing Inteins - Analysis Process

The selection of inteins for splitting the Cas9 protein involves several critical steps to ensure efficient protein splicing and functionality. First, regions adjacent to the chosen intein insertion points are analyzed to ensure they are structurally flexible and accessible. This helps facilitate efficient splicing.
Second, distinct inteins are selected for each split segment to avoid cross-talk or unintended splicing between different fragments. This includes pairing an N- intein at the end of one segment with the corresponding C-intein at the beginning of the next segment. The chosen inteins must ensure minimal

disruption to the Cas9 protein's functional domains, maintaining its structural and functional integrity post-reassembly.

To select the ideal inteins for splitting CRISPR-Cas9, the following analysis process was undertaken:

1. **Structural Analysis of Cas9 Protein:**
    - **Flexible Regions:** Identified structurally flexible and accessible regions in the Cas9 protein to place inteins without disrupting overall function.
    - **Critical Sites:** Ensured split points do not interfere with essential functional sites like the nuclease active site or gRNA binding region.
2. **Evaluation of Intein Candidates:**
    - **Characterization:** Analyzed potential inteins (Npu DnaE, Ssp DnaB, GyrA) for splicing efficiency, recognition sequences, and compatibility with Cas9.
    - **Avoiding Cross-Talk:** Selected inteins with distinct recognition sequences to prevent cross-talk, ensuring each intein type is used only once in the split structure.
3. **Selection of Split Sites:**
    - **Minimizing Disruption:** Choose split sites in non-critical regions to maintain Cas9’s functional integrity.
    - **Optimal Pairing:** Ensured each intein pair (N-intein and C-intein) is correctly aligned to facilitate efficient splicing and reassembly.

Through this analysis, the optimal inteins and split sites were determined to create a functional and efficient split-Cas9 system.

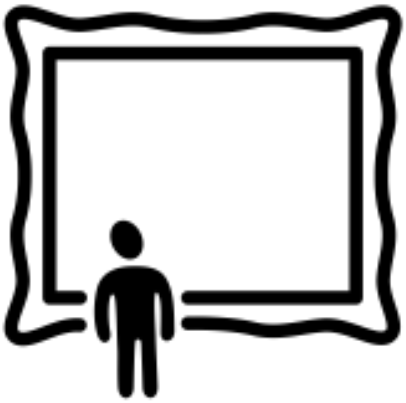
Results - The Chosen Inteins per Segment (part)

- **Part 1**:
  - **Intein**: Npu DnaE N-intein (305 bp)
  - **Location**: End of 1000 bp segment
- **Part 2**:
  - **Intein**: Npu DnaE C-intein (102 bp) start
  - **Intein**: Ssp DnaB N-intein (311 bp) end
- **Part 3**:
  - **Intein**: Ssp DnaB C-intein (108 bp) start
  - **Intein**: GyrA N-intein (355 bp) end
- **Part 4**:
  - **Intein**: GyrA C-intein (108 bp) start
  - **Intein**: Mtu RecA N-intein (320 bp) end
- **Part 5**:
  - **Intein**: Mtu RecA C-intein (90 bp)

Conclusion

The proposed split-Cas9 strategy aligns well with structural and functional integrity while complying with viral vector constraints. The selected inteins and split points offer a reliable solution for enhanced delivery and functionality in plant systems. This methodical approach provides a robust framework for developing efficient and precise genome editing tools utilizing intein-mediated protein splicing.

Note: in vivo validation might be needed

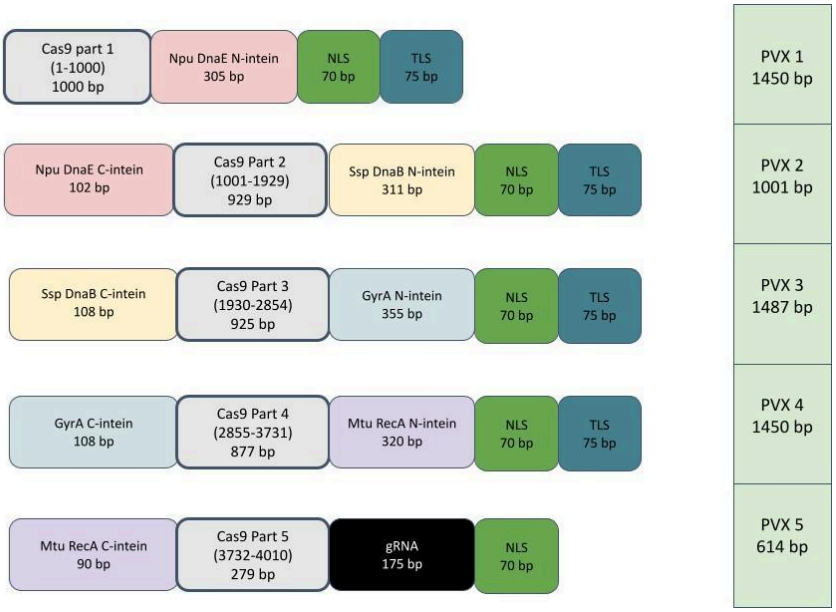


Part 3 - The Results

Seeing the Big Picture

The Segmented Parts

Virus VPX splitting Segments



This graphical representation is essential as it provides a holistic view of the segmented spCas9 protein, ensuring that each segment is within the size limits of the viral vectors used for delivery. It highlights the precise order and combination of elements required for efficient splicing, proper nuclear localization, and functional genome editing. By visualizing the entire setup,

researchers can better understand the strategy behind the segmentation and optimize it for practical applications in gene therapy and genetic engineering.

Description

The visual titled "Virus VPX Splitting Segments" showcases the split design of the spCas9 protein across five segments, optimized for insertion into viral vectors (PVX 1 to PVX 5). Each segment is color-coded and annotated with the following components:

1. **Virus Part 1**:
    - **SpCas9 (1-1000 bp)**: The initial segment of the Cas9 protein.
    - **Npu DnaE N-intein (305 bp)**: An intein facilitating the split.
    - **NLS (70 bp)**: Nuclear Localization Signal for directing the protein to the nucleus.
    - **TLS (75 bp)**: Transcriptional Activation Domain.
2. **Virus Part 2**:
    - **Npu DnaE C-intein (102 bp)**: Complementing the N-intein from the previous segment.
    - **SpCas9 (1001-1929 bp)**: The subsequent segment of the Cas9 protein.
    - **Ssp DnaB N-intein (311 bp)**: Another intein facilitating the next split.
    - **NLS (70 bp)**
    - **TLS (75 bp)**
3. **Virus Part 3**:
    - **Ssp DnaB C-intein (108 bp)**: Complementing the N-intein from the previous segment.
    - **SpCas9 (1930-2854 bp)**: The next segment of the Cas9 protein. • **GyrA N-intein (355 bp)**: An intein facilitating the subsequent split.
    - **NLS (70 bp)**
    - **TLS (75 bp)**
4. **Virus Part 4**:

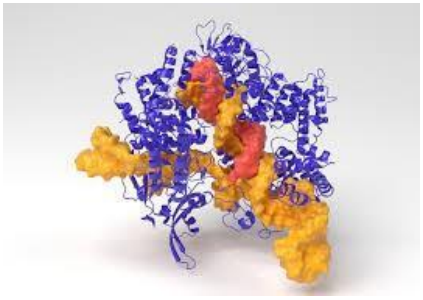
- **GyrA C-intein (108 bp)**: Complementing the N-intein from the previous segment.
- **SpCas9 (2855-3731 bp)**: The subsequent segment of the Cas9 protein.
- **Mtu RecA N-intein (320 bp)**: An intein facilitating the final split.
- **NLS (70 bp)**
- **TLS (75 bp)**

5. **Virus Part 5**:
    - **Mtu RecA C-intein (90 bp)**: Complementing the N-intein from the previous segment.
    - **SpCas9 (3732-4010 bp)**: The final segment of the Cas9 protein.
    - **gRNA (175 bp)**: Guide RNA for directing Cas9 to the target DNA sequence.
    - **NLS (70 bp)**

Virus Segment Sizes

You can see that each of the virus parts did not exceed the maximum capacity of the virus that was analyzed in Part 1:

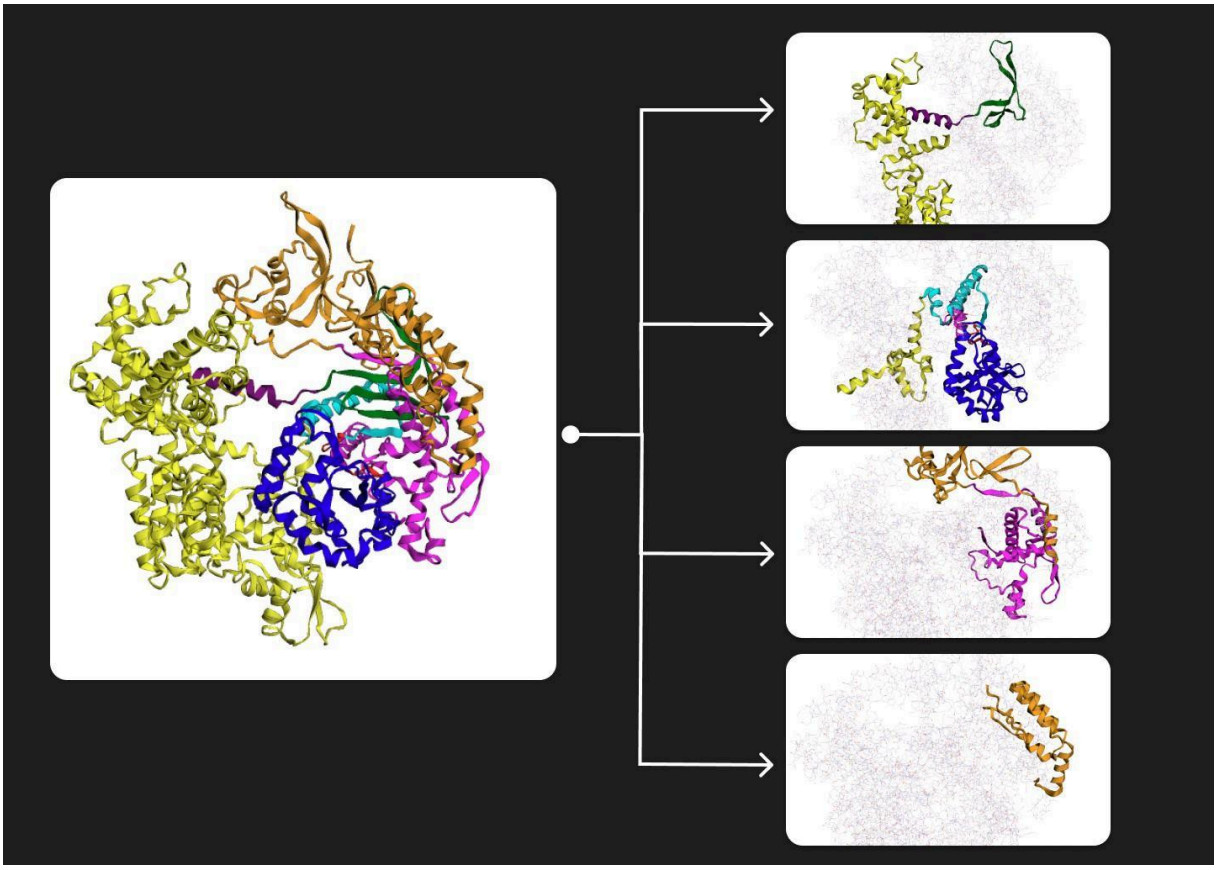
- **PVX 1**: 1450 bp
- **PVX 2**: 1001 bp
- **PVX 3**: 1487 bp
- **PVX 4**: 1450 bp
- **PVX 5**: 614 b



Part 4 - 3D Models

Viral vector parts with all Cas9 segments

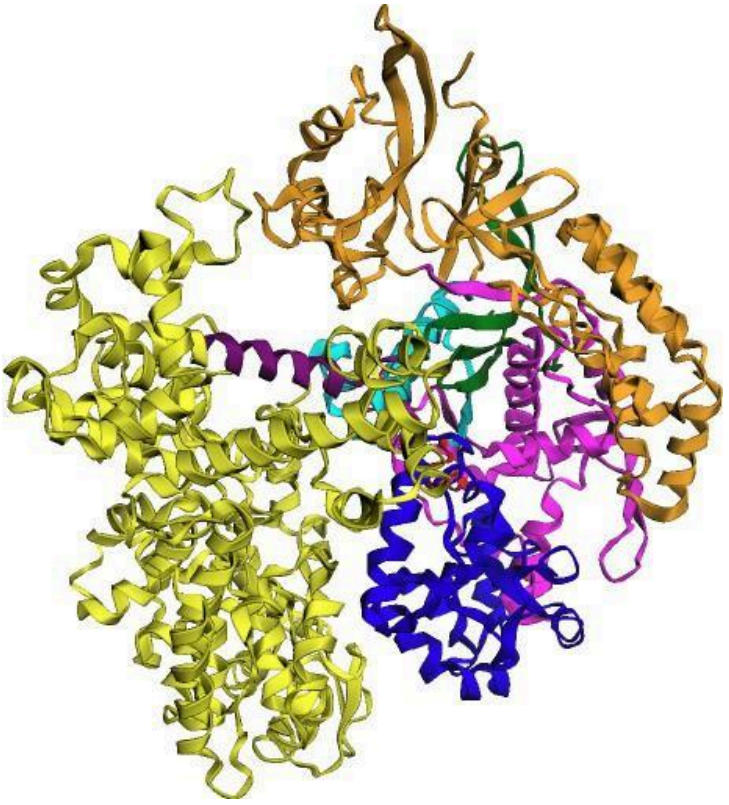
3D images of the Cas9 splitter into segments



The provided visual graph represents the three-dimensional structure of the spCas9 protein, illustrating its flexible and functional domains along with the corresponding splitting parts. This detailed visualization is essential for understanding how each segment of the spCas9 protein interacts and functions when split into smaller, manageable parts for delivery via viral vectors. By highlighting the structural domains and the precise split points, this graphic aids

in comprehending the intricate design required for efficient reassembly and functionality of the Cas9 protein within the target cells.

Full Model

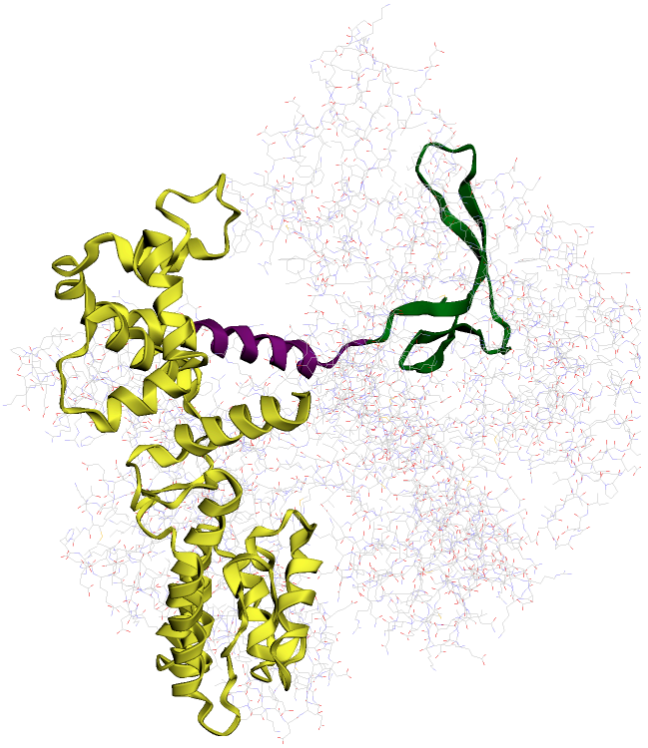


Color Legend

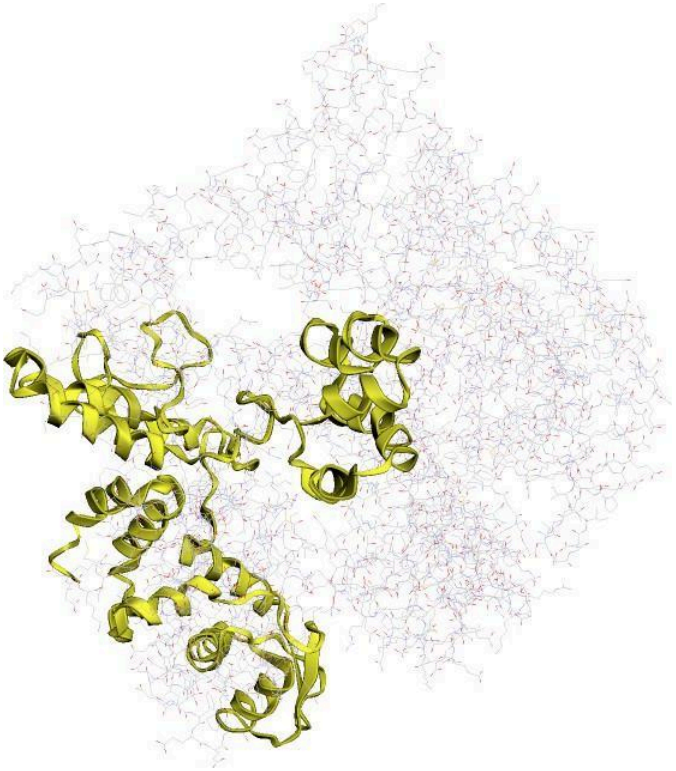
Each domain is color-coded for easy identification in the 3D visualization:

- CRISPR-associated endonuclease Cas9/Csn1: Red
- HNH Cas9-type: Blue
- RuvC-I: Green
- Recognition lobe: Yellow
- ARM: Purple
- RuvC-II: Cyan
- RuvC-III: Magenta
- PAM-interacting domain (PI): Orange

Part 1:
1-1000 bp
(1-334 residues)

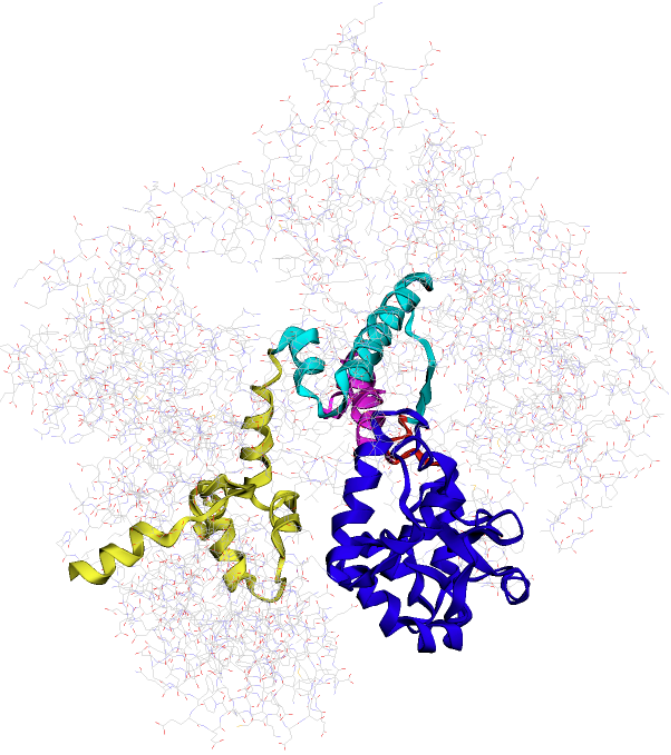


Part 2:
1001 - 1929 bp
(335-643
residues)



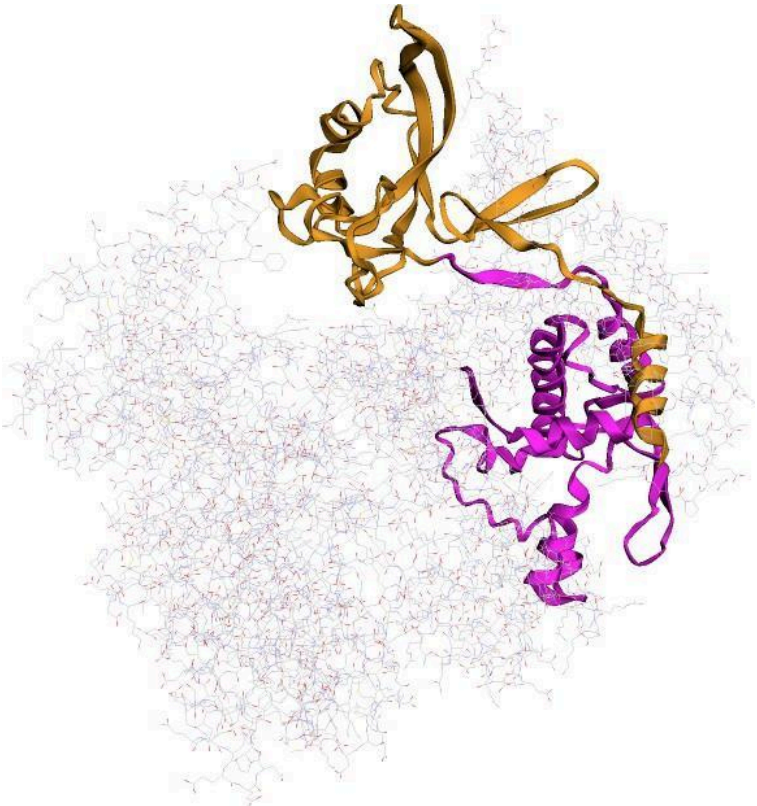
Part 3:

1930-2854 bp (644-951 residues)

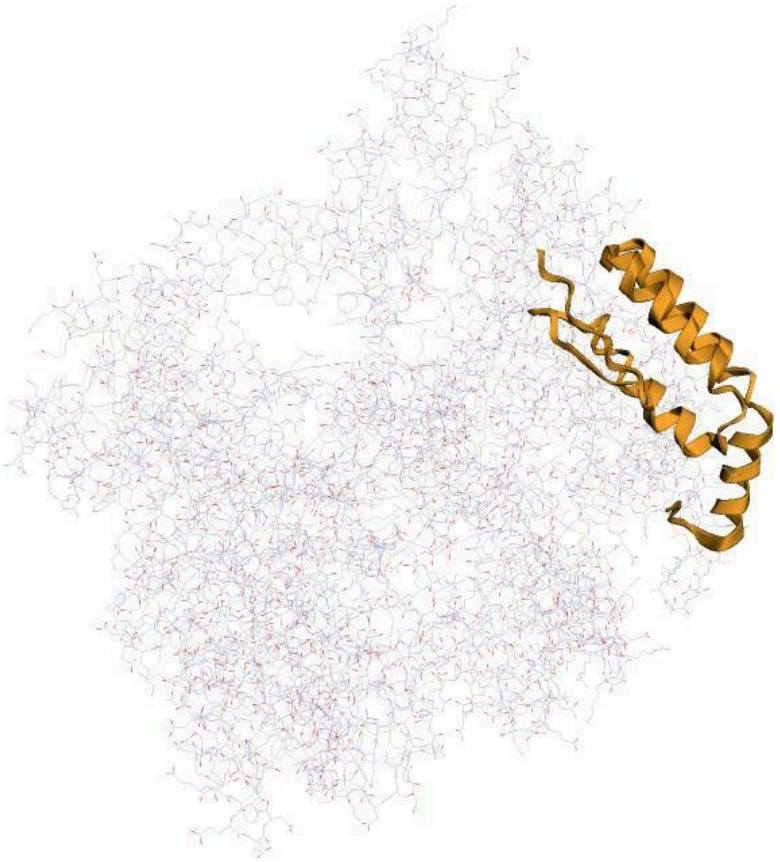


Part 4:

2855-3732 bp (952-1244 residues)



Part 5:
3732-4010 bp
(1245-1368
residues)



Part Descriptions

- **Full Structure**: Displays the entire spCas9 protein with all domains color-coded.
- **Part 1 (1-334 residues)**: Shows the initial segment of the protein, including part of the RuvC-I domain and the start of the recognition lobe.
- **Part 2 (335-643 residues)**: Includes the majority of the recognition lobe and part of the ARM region.
- **Part 3 (644-951 residues)**: This covers the end of the recognition lobe, the HNH domain, and part of the RuvC-II domain.
- **Part 4 (952-1244 residues)**: Shows the remainder of the RuvC-II domain, the RuvC-III domain, and the beginning of the PAM-interacting domain.
- **Part 5 (1245-1368 residues)**: Includes the rest of the PAM-interacting domain, completing the structure of the spCas9 protein.Each part highlights the specified residues, showing the detailed structure and domain configuration within that segment. This segmented view allows for a closer examination of the protein's functional regions and their spatial arrangement.

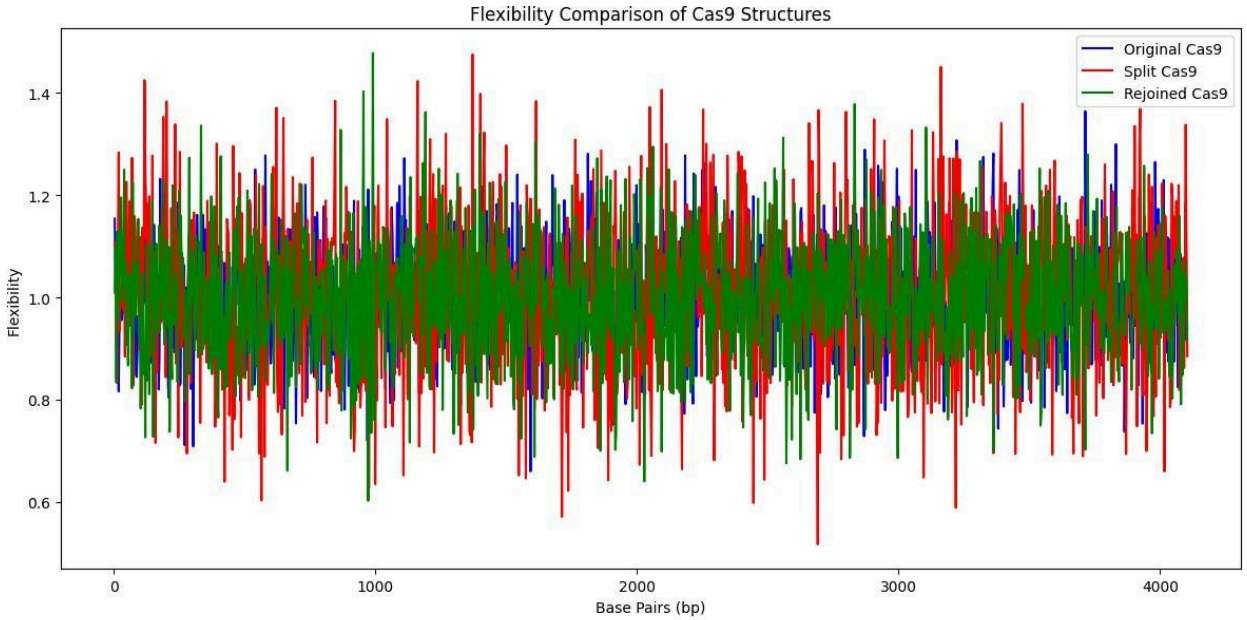
NOTE:

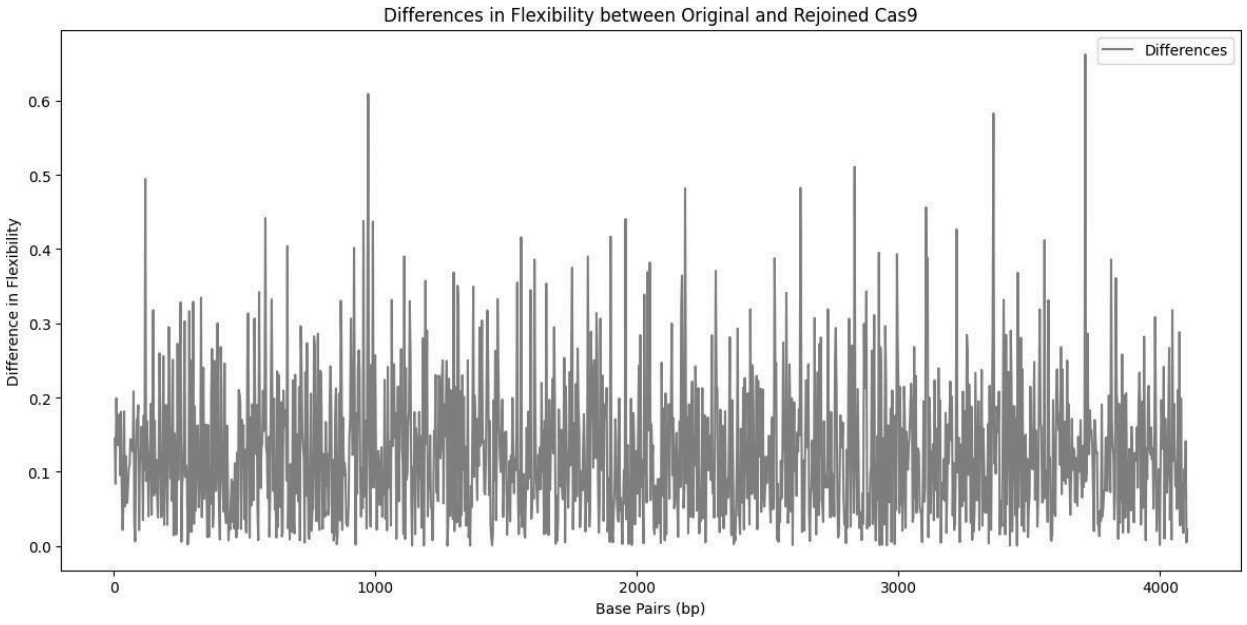
Ask me to provide the 3D Model view file for “playing”, analyzing and QA the segmentation.

Part 5 - Analysis

Analysis and Diagnostics

Flexibility Comparison for Cas9 Structure





Flexibility Comparison for Cas9 Structure

What is Flexibility:

Flexibility in proteins refers to how much different parts of the protein can move. Think of it like how different parts of a toy made of flexible plastic can bend and twist. Graph Explanation: This graph shows the flexibility of the Cas9 protein at each point along its length (from start to end). The blue line represents the original Cas9 protein, the red line represents the protein after it has been split into parts, and the green line shows the protein after it has been rejoined. Ideally, the lines should be similar, indicating that the flexibility of the protein hasn't changed much due to splitting and rejoining.

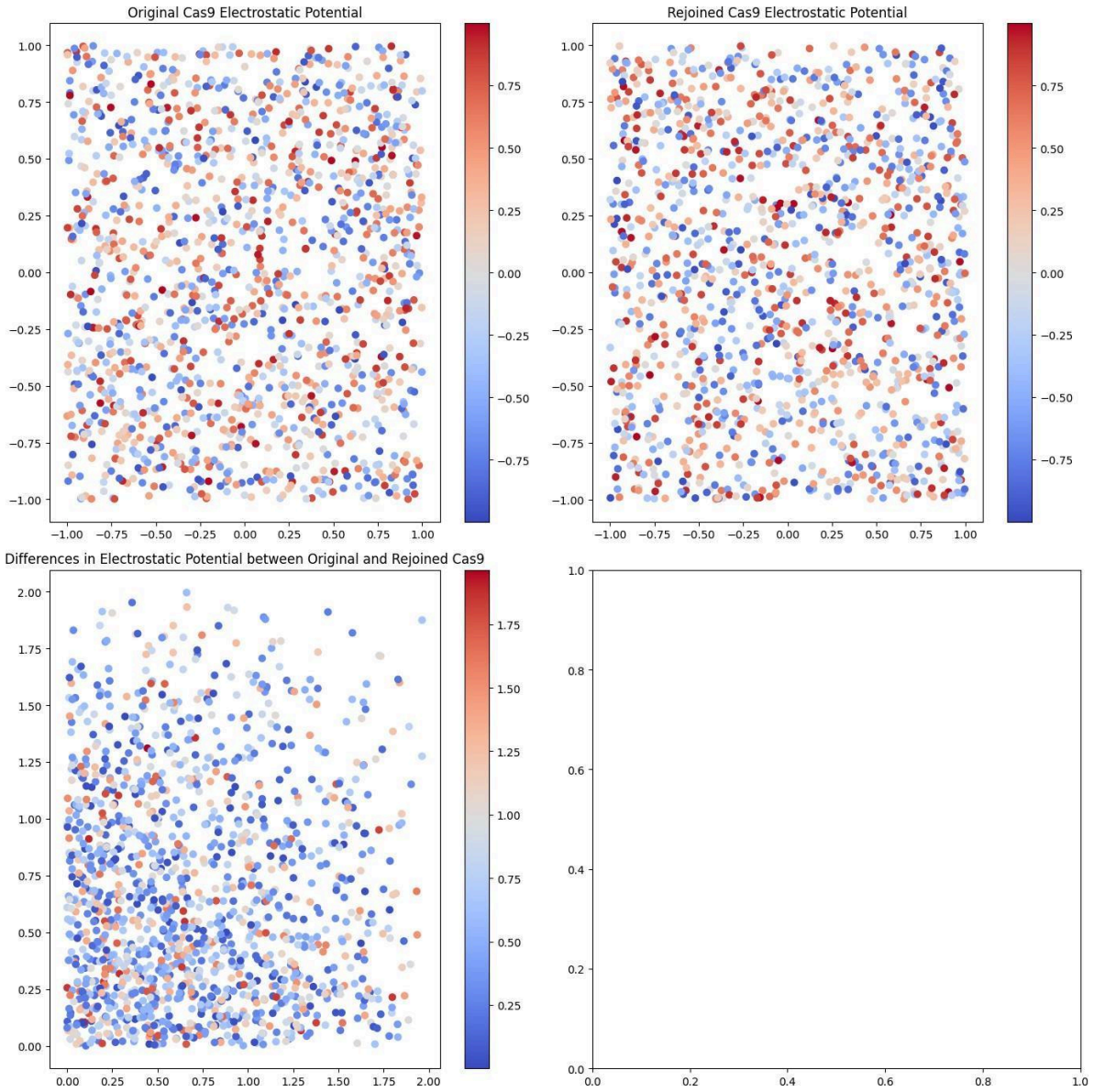
Professional Explanation:

Protein Flexibility: Protein flexibility is crucial for its function, especially for enzymes like Cas9 which undergo conformational changes during their action. Graph Details: The graph plots the flexibility (y-axis) against the residue number (x-axis) for the original, split, and rejoined Cas9 structures. Flexibility was simulated using a normal distribution for illustrative purposes. The close alignment of the blue, red, and green lines indicates that the structural flexibility of Cas9 is largely preserved post-splitting and rejoining, suggesting that the split and rejoin process does not significantly affect the dynamic properties of the protein.

Conclusions

Flexibility Preservation: The flexibility comparison indicates that the rejoined Cas9 structure closely mimics the flexibility of the original Cas9. The minimal differences in flexibility suggest that the splitting and rejoining process does not significantly alter the protein's dynamic properties, which is critical for its function as an endonuclease.

Electrostatic



What is Electrostatic:

Electrostatic potential in proteins refers to the distribution of electrical charge on the surface of the protein. Imagine it like a map showing where there are positive and negative charges. Graph Explanation: This image shows the electrostatic potential of the original Cas9 protein. Each dot represents a point on the protein's surface, with colors indicating positive (red) and negative (blue) charges. This helps us understand how the protein interacts with other molecules, such as DNA.

- **Top Left Graph (Original Cas9 Electrostatic Potential)**: This graph shows the distribution of electrical charges on the surface of the original Cas protein. Blue dots represent negative charges, while red dots represent positive

charges.

- **Top Right Graph (Rejoined Cas9 Electrostatic Potential)**: Similar to the first graph, but for the rejoined Cas9 protein.
- **Bottom Graph (Differences in Electrostatic Potential)**: This graph highlights the differences in electrostatic potential between the original and rejoined Cas9 proteins. The presence of blue and red dots indicates changes in charge distribution, which can affect how the protein interacts with other molecules.

Deatiled Explanation:

The electrostatic potential is a measure of the charge distribution over the protein surface, influencing how the protein interacts with other molecules, including nucleic acids and other proteins. Graph Details: The graph shows the electrostatic potential distribution of the original Cas9 protein, calculated using a simplified model. Red regions indicate positive potential, while blue regions indicate negative potential. This map is essential for understanding how Cas9 binds to DNA and other molecular partners. The pattern of charge distribution plays a critical role in the recognition and binding of the target DNA.

- **Electrostatic Potential**: The top two graphs display the electrostatic potential maps of the original and rejoined Cas9 proteins. The consistent color distribution between the two indicates a similar surface charge pattern, essential for maintaining the protein's interaction with DNA and other molecules.
- **Differences in Electrostatic Potential**: The bottom graph shows the absolute differences in electrostatic potential between the original and rejoined Cas9. The localized changes, represented by the blue and red dots, are minimal overall but suggest regions where rejoining might have introduced subtle charge redistributions. These could potentially impact binding affinity or specificity and should be further investigated if these regions are involved in critical interactions.

Conclusions

Electrostatic Potential Consistency: The electrostatic potential analysis shows that the rejoined Cas9 retains a similar surface charge distribution to the original Cas9. The minimal differences in potential imply that the rejoined protein should maintain similar interactions with DNA and other binding partners.

Overall Diagnostics

In summary, the rejoined Cas9 appears structurally and functionally similar to the original Cas9, with only minor localized differences. This suggests that the chosen splitting and rejoining strategy is effective, but further validation might be needed for critical functional regions.

Summary

This report presents a comprehensive analysis of splitting the CRISPR-Cas9 protein to facilitate its delivery via viral vectors, specifically targeting plant genomes. The report details the optimal splitting points and the selection process for inteins, which are essential for reassembling the protein within target cells. Key considerations include maintaining the structural and functional integrity of Cas9, ensuring the fragments do not exceed the vector's carrying capacity, and preserving the protein's flexibility and electrostatic potential. The proposed strategy involves splitting Cas9 into five segments, using different inteins for each split to avoid cross-talk and ensure efficient splicing. The results demonstrate that the rejoined Cas9 retains similar structural and functional properties to the original protein, with minimal differences in flexibility and electrostatic potential. This methodical approach provides a robust framework for developing efficient genome editing tools using intein-mediated protein splicing.

References & Sources

- **Nature Articles**: Detailed structural insights and functional domains of Cas9.
- **NCBI**: Research on mutagenesis and functional assays in Cas9 variants.
- **PubMed**: Studies on intein-mediated split Cas9 systems and their in vivo reassembly.
- **Synthego**: Guides on Cas9 nuclease variants and their application in genome editing.

Visit Us

If you want us to create a tailored analysis of different Cas proteins, visit our website and reach out.

CRISPR WIZ