DOI: 10.19884/j.1672-5220.202506004

Analysis of *Cne* Prp8 Intein Substrate Preference Based on mCherry Reporter System

CHEN Tianqi[1#], ZHANG Xiaomeng[1#], GAO Xianling[2], LIN Ying[1*]
1. College of Biological Science and Medical Engineering, Donghua University, Shanghai 201620, China
2. Shandong Guoli Biotechnology Co., Ltd., Jinan 250101, China

Abstract: Intein-mediated protein splicing provides a valuable tool for protein ligation. The amino acid sequences of the flanking exteins serve as the intein's natural substrates. Alterations to these extein sequences may cause significant changes in splicing efficiency. To date, how extein sequences at different positions (such as −1 and +2 positions) influence intein splicing activity remains to be systematically investigated. To address this issue, an mCherry-based fluorescence reporter system that links fluorescence intensity to splicing activity was established for high-throughput screening. A saturation mutagenesis library targeting the −1 and +2 positions of the *Cne* Prp8 intein was constructed via homologous recombination. After induction, 38 high-fluorescence clones were identified from 336 clones. Western blot analysis revealed that three substrate combinations (A/S, S/V, and R/S) achieved splicing efficiencies above 80%, significantly higher than that of the wild-type sequence. These findings not only reveal amino acid preferences at key extein positions but also provide valuable insights for the rational design of efficient splicing systems, thereby advancing the application of inteins in protein engineering.

Keywords: intein; extein; protein splicing; fluorescent protein
CLC number: Q51 **Document code**: A
Article ID: 1672-5220(2026)03-0085-07

Open Science Identity (OSID)

0 Introduction

Inteins are polypeptide sequences embedded within precursor proteins that can excise themselves with high precision through an autocatalytic process during protein maturation. This reaction simultaneously ligates the flanking extein sequences via a peptide bond, a process known as protein splicing[1]. Protein splicing proceeds through a four-step cascade of reactions that do not require any cofactors. The central mechanism involves acyl transfer[2]. Due to their versatility as protein ligation tools, inteins have been widely applied in various fields of protein engineering, including protein purification[3-4], peptide backbone cyclization[5-7], site-specific labeling[8-9], and biosensor construction[10-11].

Notably, the amino acid residues at the junction between the intein and its adjacent exteins significantly influence splicing activity. These adjacent exteins can thus be regarded as substrates of inteins. Previous studies have shown that native substrates are not necessarily optimal for splicing[12-14]. Intein substrate specificity and the relationship between different types of adjacent exteins and splicing efficiency are not well understood, and the substrate preferences of most natural inteins have yet to be systematically characterized. Investigating intein substrate preferences not only broadens the application of inteins as protein ligation tools but also enhances our understanding of the structural and functional relationships within intein-extein complexes, which holds both theoretical significance and practical value.

The β-barrel structure of the red fluorescent protein mCherry contains multiple loop regions that are relatively tolerant to structural perturbations. Ren et al.[15] employed mCherry as a reporter protein to monitor intein splicing activity. They inserted an intein in the loop region between β8 and β9, enabling successful conversion of intein splicing activity into a fluorescent signal. In this study, the 142−143 site of mCherry (approaching the β8 – β9 loop region) was selected as the insertion point. Inserting an intein mutant with no splicing activity resulted in no fluorescence, whereas insertion of a highly active intein restored fluorescence. This successfully coupled intein splicing activity with a fluorescent signal, establishing a reporter system based on mCherry. Saturation mutagenesis was performed at the −1 and +2 extein positions flanking the *Cne* Prp8 mini-intein (CP-M), resulting in a mutant library comprising 400 variants. High-fluorescence clones were identified by using the mCherry reporter system, and protein splicing was further confirmed by Western blot analysis. Sanger sequencing enabled rapid determination of the −1 and +2 extein

Received date: 2025-06-08
Foundation item: National Natural Science Foundation of China (No. 31470836)
* Correspondence should be addressed to LIN Ying, email: liny@dhu.edu.cn
# These two authors contribute equally to this work
Citation: Chen T Q, Zhang X M, Gao X L, et al. Analysis of *Cne* Prp8 intein substrate preference based on mCherry reporter system[J]. *Journal of Donghua University (English Edition)*, 2026, 43(3): 85-91.

residues. These data revealed the relationship between intein splicing activity and its substrates. Statistical analysis of the substrate sequences allowed the identification of extein preferences of the intein.

1 Materials and Methods

1.1 Construction of recombinant plasmids

The original sequence of the CP-M was obtained from the InBase database (https://inbase.ligsciss.com/iwai/). The plasmids pMCP-M and pET-mCherry were derived from previous work conducted in our laboratory[16]. The recombinant plasmids pETmCherry-CP-M (expressing mCherry with CP-M insertion) and pETmCherry-CP-M-AA (expressing mCherry with CP-M splicing-deficient intein mutant (C1A/N178A) insertion), both carrying an N-terminal 6 × His tag, were constructed using the NovoRec ® Plus One-Step PCR Cloning Kit following the manufacturer's standard protocol. A saturation mutagenesis library targeting the −1 and +2 extein positions of CP-M was constructed using NNK degenerate primers[17].

1.2 Protein splicing analysis

Recombinant plasmids were chemically transformed into *Escherichia coli* (*E. coli*) BL21(DE3) competent cells. A single colony was inoculated into 1 mL Luria-Bertani (LB) broth containing 100 μg/mL ampicillin and grown to mid-log phase (with an optical density at 600 nm of about 0.6). Protein expression was induced by adding isopropyl *β-D*-1- thiogalactopyranoside (IPTG) to a final concentration of 0.3 mmol/L, followed by incubation at 25 ℃ for 16 h. Cells were harvested by centrifugation and lysed by boiling in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer at 100 ℃ for 10 min. The lysates were then analyzed by electrophoresis on a 15% SDS-PAGE gel. Western blot analysis was performed using an anti-His antibody and the WesternBreeze™ detection kit, following the manufacturer's standard protocol.

1.3 Measurement of fluorescence intensity

A 200 μL aliquot of the induced bacterial culture was subjected to red fluorescence detection, which was measured using a SpectraMax iD3 microplate reader with excitation at 577 nm and emission at 617 nm. Each sample was measured in triplicate. To calculate the normalized fluorescence intensity, background fluorescence (Mock control) was subtracted from each sample's raw fluorescence. The normalized fluorescence intensity was then calculated as

$$F_{\text{norm}} = \frac{F_{\text{sample}} - F_{\text{LB}}}{D_{600}} - \frac{F_{\text{BL21}} - F_{\text{LB}}}{D_{600}},$$

where F_{norm} represents the normalized fluorescence intensity, which refers to the fluorescence intensity per *E. coli* cell; F_{sample} represents the fluorescence intensity of the induced bacterial culture (experimental sample); F_{BL21} represents the fluorescence intensity of the BL21 strain (Mock control); F_{LB} represents the fluorescence intensity of the LB broth; D_{600} represents the optical density at 600 nm.

1.4 Prediction and simulation of protein structure

The three-dimensional structure of the target protein was predicted using AlphaFold3, a deep learning-based algorithm developed by DeepMind. The full-length amino acid sequence was input into the AlphaFold3 pipeline, and the predicted protein structure was generated using default parameters. The model with the highest predicted local distance difference test (pLDDT) score was downloaded in protein data bank (PDB) format and simulated using PyMOL software. Structural alignment was conducted using the "align" command to compare the model with homologous proteins from the PDB.

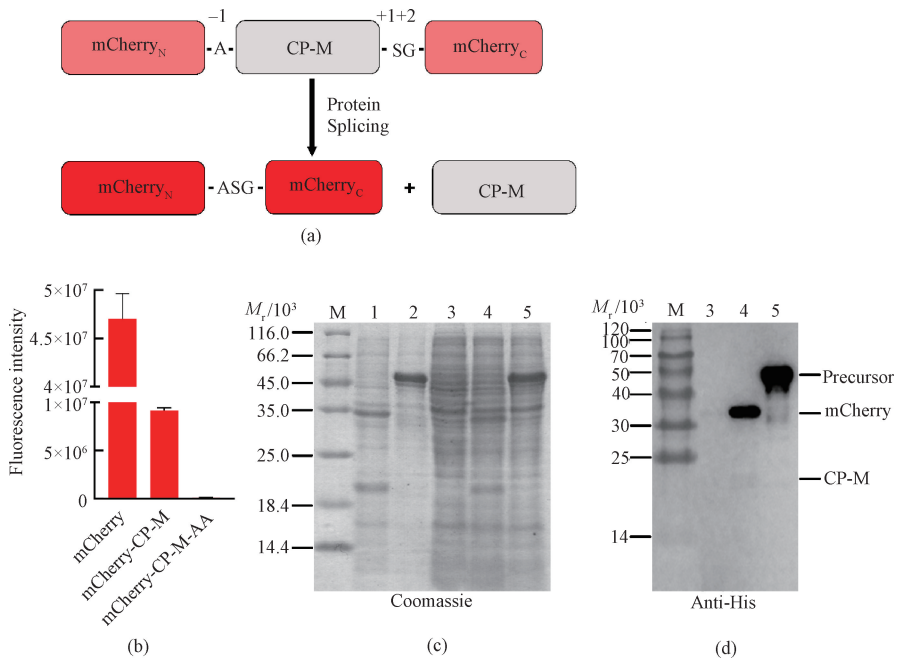
2 Results and Discussion

2.1 Construction of fluorescent protein reporter system

To evaluate the substrate specificity of intein, we inserted intein into a fluorescent protein to construct a fluorescence-based reporter system. The selected insertion site was required to meet two criteria: 1) the addition of three exogenous amino acids must not impair fluorescence; 2) the insertion of a splicing-inactive intein must abolish fluorescence. The red fluorescent protein mCherry features a unique β-barrel structure and flexible loop regions, making it an ideal scaffold for foreign protein insertion without disrupting its overall conformation[18-20]. We constructed the fluorescent reporter system by inserting the CP-M intein, along with its native −1, +1, and +2 extein residues, between the residues 142 and 143 of mCherry (approaching the $\beta 8 - \beta 9$ loop region). In this system, successful protein splicing results in the excision of the intein, restoring the intact mCherry protein and generating detectable red fluorescence, and the −1, +1, and +2 extein residues remain within the interior of the mCherry protein. Conversely, if splicing is inhibited, the intein remains embedded in mCherry, causing steric hindrance that suppresses fluorescence (Fig. 1(a)). As shown in Fig. 1(a), the mCherry protein is split by the CP-M into an N-terminal fragment ($\text{mCherry}_{\text{N}}$) and a C-terminal fragment ($\text{mCherry}_{\text{C}}$), where the flanking extein residues A, S, and G correspond to positions −1, +1, and +2, respectively. This fluorescence-based reporter system links intein splicing activity to a fluorescence signal, offering a rapid and visual strategy for evaluating splicing efficiency.

To verify the direct correlation between mCherry fluorescence and intein splicing activity, we constructed a splicing-deficient intein mutant (C1A/N178A) as a negative control. This mutant is incapable of catalyzing

the protein splicing reaction, and the retained intein disrupts the mCherry conformation, resulting in a non-fluorescent precursor protein. The fluorescence intensity of wild-type mCherry protein reaches 4.7×10^7, the splicing mutant (mCherry-CP-M) exhibits strong red fluorescence as well (9.1×10^6), whereas the non-splicing AA mutant (mCherry-CP-M-AA) displays no detectable fluorescence (Fig. 1(b)). SDS-PAGE and Western blot analysis reveal the spliced mCherry product (relative molecular mass $M_r = 30\,200$) and excised intein ($M_r = 20\,900$) in mCherry-CP-M, whereas mCherry-CP-M-AA only exhibits the unspliced precursor ($M_r = 51\,100$) (Figs. 1(c) and 1(d)). These results demonstrate that intein splicing activity is directly coupled to fluorescence output, confirming the reliability of the fluorescent protein-based reporter system.



M—marker; 1—non-induced mCherry-CP-M; 2—non-induced mCherry-CP-M-AA; 3—wild type mCherry; 4—mCherry-CP-M; 5—mCherry-CP-M-AA.

Fig. 1 Establishment of fluorescent reporter protein system: (a) schematic diagram of fluorescent reporter system; (b) fluorescence intensity comparison; (c) protein splicing in *E. coli* by SDS-PAGE; (d) Western blot analysis (Anti-His)

2.2 Saturation mutation analysis of adjacent exteins

Given the high conservation of the +1 extein position and its critical role in intein splicing activity, we selected the flanking −1 and +2 extein positions as mutagenesis sites. Using NNK degenerate primers, we performed saturation mutagenesis at the −1 and +2 extein positions of the CP-M intein, generating a mutant library covering all 400 possible amino acid combinations. We evaluated the quality and completeness of the mutant library using next-generation sequencing (NGS). The NGS analysis showed an average sequencing depth of 45 049 at the mutated sites, with a 100% mutation rate and 99.68% library coverage. The codon mutation frequencies at the −1 extein position were as follows: first base—T (38.5%), G (32.9%), C (16.7%), A (11.9%); second base—T (43.8%), G (26.0%), A (15.6%), C (14.6%); third base—T (50.2%), G (49.8%). At the +2 extein position, codon mutation frequencies were as follows: first base—A (28.2%), T (27.1%), C (24.1%), G (20.6%); second base—T (33.8%), A (31.1%), C (21.0%), G (14.1%); third base—T (53.4%), G (46.6%). These results confirm the successful construction of a comprehensive saturation mutagenesis library of the CP-M flanking exteins, encompassing all 400 possible amino acid combinations at the −1 and +2 extein positions.

2.3 Screening intein splicing mutants using mCherry reporter system

Library plasmids were transformed into *E. coli* BL21 (DE3) competent cells, and protein expression was induced for subsequent screening. By using the mCherry reporter system, the red fluorescence intensity of each clone was calculated. We observed significant variations in fluorescence intensity among individual clones, ranging from a maximum of 5.7×10^6 to a minimum of 1.3×10^4. Native extein clones (with −1 and +2 residues of A and G) exhibited a red fluorescence intensity of 1.5×10^6. Clones with fluorescence intensity exceeding 1×10^6 were defined as high-fluorescence clones. From a total of 336 clones, 38 high-fluorescence and six low-fluorescence clones were selected for SDS-PAGE (Fig. 2(a)) and Western blot analysis (Fig. 2(b)). The −1 and +2 amino acid combinations corresponding to each mutant clone were labeled above the lanes.

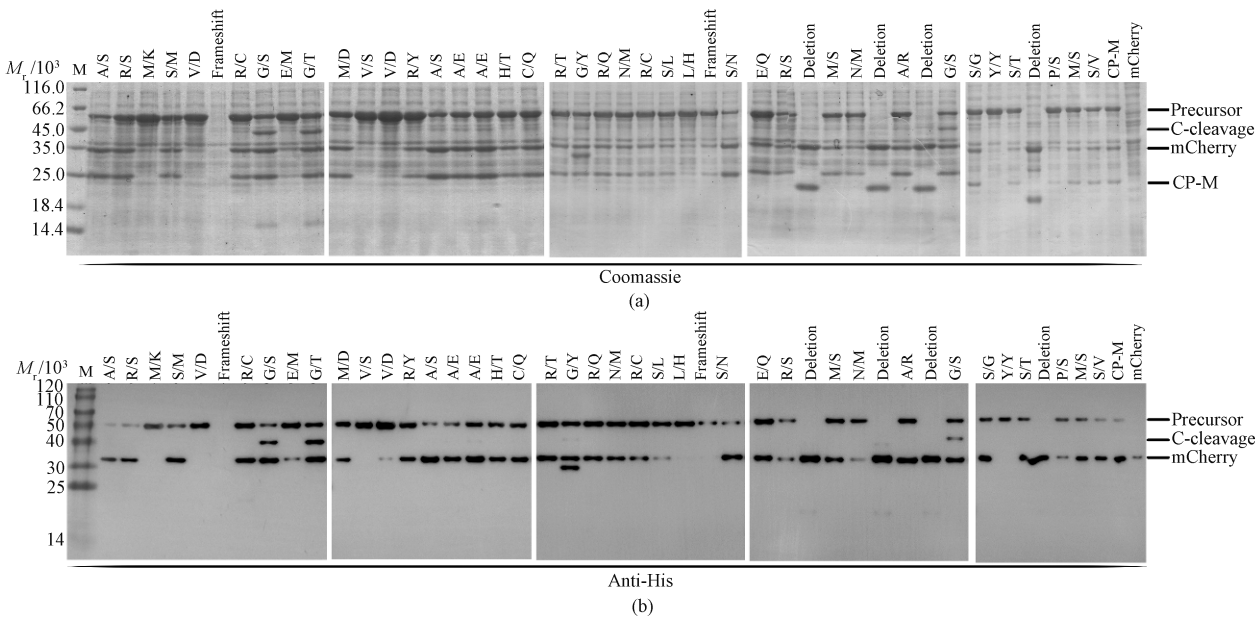
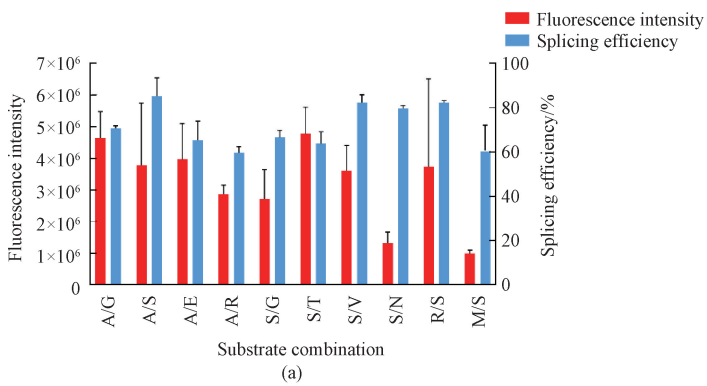
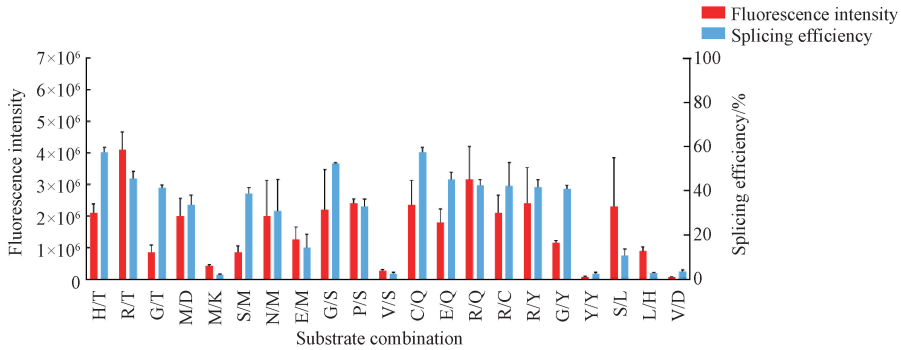


Fig. 2 Identification of mutant clones: (a) SDS-PAGE; (b) Western blot analysis (Anti-His)

SDS-PAGE and Western blot analysis reveal that most high-fluorescence clones contain both the precursor protein (M_r = 51 200) and the mCherry protein (M_r = 30 200) produced by protein splicing. Some clones exhibit false positives due to frameshifts or deletions, while low-fluorescence clones primarily retain the precursor form. The native −1 and +2 substrate combination for CP-M is A and G, yielding a splicing efficiency of 71.9%. Compared to the native substrate clone, some mutant clones exhibit significantly enhanced splicing efficiency, while the majority show reduced efficiency. Additionally, a few mutant clones display C-terminal cleavage (M_r = 40 600) as a side reaction. Clones corresponding to −1 and +2 substrate combinations with splicing efficiency exceeding 60% are categorized as high-splicing efficiency clones; the remaining are designated as low-splicing efficiency clones. Figure 3 illustrates the correlation between fluorescence intensity and splicing efficiency for each clone. Nine substrate combinations display splicing efficiency exceeding 60%, with three combinations (A/S, S/V, R/S) exhibiting efficiencies over 80%, significantly higher than the native substrate (Fig. 3(a)). All nine mutant clones with high splicing efficiency exhibit fluorescence intensities exceeding 1×10^6, belonging to the category of high fluorescence clones. This result indicates a positive correlation between splicing activity and fluorescence intensity, highlighting the coupling between the two processes. While some low-splicing efficiency substrate combinations reach up to 50% splicing efficiency, most other combinations exhibit efficiencies below 40%, representing a significant reduction compared to the native substrate (Fig. 3(b)). Notably, some low-splicing substrate combinations result in increased fluorescence intensity, which is likely attributable to higher protein expression levels in the corresponding mutant clones.



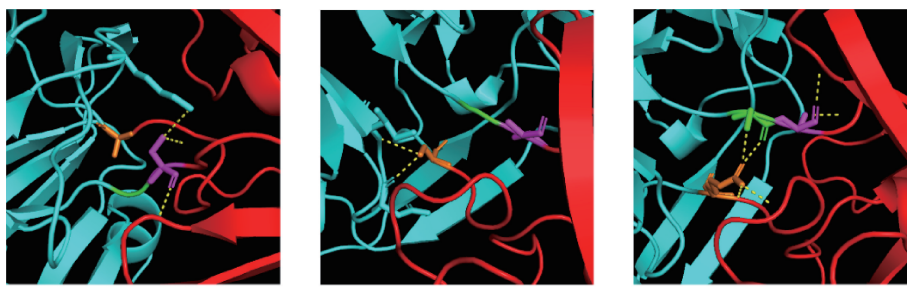


(b)

Fig. 3 Comparison of fluorescence intensity and splicing efficiency of mutant clones: (a) high splicing efficiency mutant clones; (b) low splicing efficiency mutant clones

2.4 *Cne* Prp8 substrate preference analysis

This study partially screened the −1 and +2 substrates of CP-M using a fluorescent protein reporter system, yielding 32 substrate combinations. Among these, the A/S, S/V, and R/S substrate combinations significantly enhanced protein splicing efficiency to over 80% (Fig. 3(a)). To further investigate their splicing mechanisms, these three substrates were subjected to structural prediction and simulation using AlphaFold and PyMOL software. The results are presented in Fig. 4.



(a) (b) (c)

Red—mCherry; cyan—CP-M; orange—−1 substrate; green—+1 substrate; magenta—+2 substrate; yellow—hydrogen bond.

Fig. 4 Prediction and simulation results of substrate structure: (a) A/S substrate combination; (b) S/V substrate combination; (c) R/S substrate combination

Based on the high splicing efficiencies and structural predictions for the A/S, S/V, and R/S substrate combinations, it is hypothesized that CP-M exhibits substrate preferences for serine, arginine, alanine, and valine. As a nonpolar amino acid with a small side chain, A is less prone to steric hindrance. The electrostatic interference caused by A is also negligible on intein. Structural predictions indicate that A at the −1 position does not form hydrogen bonds with other amino acids, thereby having minimal impact on intein splicing efficiency (Fig. 4(a)). In addition to the native substrate combination A/G, the combinations A/S, A/E, and A/R also support efficient protein splicing (Fig. 3(a)), suggesting that when A is present at the −1 position, the +2 position can be occupied by G, S, E, or R. The −1 position S may form hydrogen bonds with an internal intein histidine, thereby facilitating acyl rearrangement during protein splicing and enhancing the stability of the linear ester intermediate (PolB16 OarG intein)[21]. Additionally, AlphaFold structural predictions of CP-M suggest that the −1 position S may also form hydrogen bonds with asparagine and isoleucine, which are located on both sides of histidine (Fig. 4(b)). R, carrying a positive charge and possessing a long side chain, is predicted by structural analysis to form hydrogen bonds with the +1 position S when at the −1 position (Fig. 4(c)). This interaction may affect the stability of the intein active site and introduce significant steric hindrance[16]. However, when R is present at the −1 position, CP-M still exhibits relatively high splicing efficiency, suggesting a degree of tolerance to −1 position substrate. When at the +2 position, S, a nucleophilic residue, may undergo its hydroxyl group attacking the ester bond between the intein C-terminus and the downstream extein, leading to intein cleavage (PolB16 OarG intein)[21], which is also detected in the Western blot analysis of CP-M intein with G/S combination but not in the other +2 S combinations. Structural predictions of CP-M indicate that the +2 position S forms hydrogen bonds with conserved lysine within the intein, which is hypothesized to enhance the stability of the splicing reaction intermediate (Fig. 4(a)). Structural predictions for V indicate that it does not form hydrogen bonds with conserved amino acids within the

intein (Fig. 4(b)). Although V is uncharged, its relatively large side chain may introduce significant steric hindrance, potentially reducing splicing activity[5]. However, when V is located at the +2 position, CP-M still exhibits efficient protein splicing, suggesting a certain level of substrate tolerance at this position.

In this study, three −1 and +2 substrate combinations, A/S, S/V, and R/S, were identified to elevate CP-M splicing efficiency above 80%. Under these substrate conditions, protein splicing occurred with high efficiency. These findings provide theoretical guidance for selecting insertion sites when applying CP-M to heterologous proteins, thereby enhancing its splicing activity, reducing constraints in heterologous systems, and expanding its applicability as a protein ligation tool.

3 Conclusions

In this study, we developed a fluorescent protein reporter system that successfully coupled intein splicing activity with a fluorescent signal. This system offers a rapid and visual tool for investigating intein splicing. Using degenerate primers, we constructed a saturation mutagenesis library of the −1 and +2 extein positions for CP-M intein. After induction, protein splicing for each substrate combination was validated by fluorescence intensity data and Western blot analysis. From this, we identified 32 substrate combinations, including nine with high splicing efficiency. Notably, the A/S, S/V, and R/S substrate combinations exhibited splicing efficiency above 80%, significantly outperforming the native substrate. These results further support the theory that an intein's native substrate may not always be its optimal splicing substrate, providing valuable guidance for selecting insertion sites when applying CP-M intein in heterologous proteins.

References

[1] Nanda A, Nasker S S, Mehra A, et al. Inteins in science: evolution to application [J]. *Microorganisms*, 2020, 8(12): 2004.

[2] Tori K, Perler F B. Expanding the definition of class 3 inteins and their proposed phage origin [J]. *Journal of Bacteriology*, 2011, 193(8): 2035-2041.

[3] Zhu H F, Kuang Z Y, Wang Y Y, et al. An industrial available platform for high-yield, plasmid-free recombinant protein production in *E. coli* based on advanced cSAT scheme [J]. *International Journal of Biological Macromolecules*, 2025, 307: 142208.

[4] Gharios R, Li A, Kopyeva I, et al. One-step purification and N-terminal functionalization of bioactive proteins via atypically split inteins [J]. *Bioconjugate Chemistry*, 2024, 35(6): 750-757.

[5] Yayci A, Huang Y H, Harvey P J, et al. Recombinant production of the cyclotide kalata B1 by conditional split inteins [J]. *ChemBioChem*, 2025, 26(4): e202400591.

[6] Lin T T, Zhang S C, Zhang D, et al. Use of the redox-dependent intein system for enhancing production of the cyclic green fluorescent protein [J]. *Protein Expression and Purification*, 2023, 207: 106272.

[7] Nandy S, Maranholkar V M, Crum M, et al. Expression and characterization of intein-cyclized trimer of staphylococcus aureus protein a domain Z [J]. *International Journal of Molecular Sciences*, 2023, 24(2): 1281.

[8] Heikkinen H A, Aranko A S, Iwaï H. The NMR structure of the engineered halophilic DnaE intein for segmental isotopic labeling using conditional protein splicing [J]. *Journal of Magnetic Resonance*, 2022, 338: 107195.

[9] Li X, Zhang X L, Cai Y M, et al. Site specific labeling of two proteins in one system by atypical split inteins [J]. *International Journal of Biological Macromolecules*, 2018, 109: 921-931.

[10] Yao Z, Kim J, Geng B, et al. A split intein and split luciferase-coupled system for detecting protein-protein interactions [J]. *Molecular Systems Biology*, 2025, 21(2): 107-125.

[11] Son A, Smetana J S, Horowitz S, et al. An intein-based biosensor to measure protein stability in vivo [J]. *Protein Science*, 2024, 33(3): e4925.

[12] Mills K V, Johnson M A, Perler F B. Protein splicing: how inteins escape from precursor proteins [J]. *Journal of Biological Chemistry*, 2014, 289(21): 14498-14505.

[13] Shah N H, Eryilmaz E, Cowburn D, et al. Extein residues play an intimate role in the rate-limiting step of protein trans-splicing [J]. *Journal of the American Chemical Society*, 2013, 135(15): 5839-5847.

[14] Cheriyan M, Pedamallu C S, Tori K, et al. Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues [J]. *Journal of Biological Chemistry*, 2013, 288(9): 6202-6211.

[15] Ren W, Ji A, Ai H W. Light activation of protein splicing with a photocaged fast intein [J]. *Journal of the American Chemical Society*, 2015, 137(6): 2155-2158.

[16] Lin Y, Li M M, Song H L, et al. Protein trans-splicing of multiple atypical split inteins engineered from natural inteins [J]. *PLoS One*, 2013, 8(4): e59516.

[17] Kille S, Acevedo-Rocha C G, Parra L P, et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis [J]. *ACS Synthetic*

Biology, 2013, 2(2): 83-92.
[18] Li Y K, Sierra A M, Ai H W, et al. Identification of sites within a monomeric red fluorescent protein that tolerate peptide insertion and testing of corresponding circular permutations [J]. *Photochemistry and Photobiology*, 2008, 84(1): 111-119.
[19] Fan J Y, Cui Z Q, Wei H P, et al. Split mCherry as a new red bimolecular fluorescence complementation system for visualizing protein-protein interactions in living cells [J]. *Biochemical and Biophysical Research Communications*, 2008, 367(1): 47-53.
[20] Carlson H J, Campbell R E. Circular permutated red fluorescent proteins and calcium ion indicators based on mCherry [J]. *Protein Engineering, Design and Selection*, 2013, 26(12): 763-772.
[21] Pasch T, Schröder A, Kattelmann S, et al. Structural and biochemical analysis of a novel atypically split intein reveals a conserved histidine specific to cysteine-less inteins [J]. *Chemical Science*, 2023, 14(19): 5204-5213.

基于 mCherry 荧光蛋白报告系统的 *Cne* Prp8 内含肽底物偏好性分析

陈天琦[1#]，张晓萌[1#]，高先岭[2]，林　瑛[1*]
1. 东华大学 生物与医学工程学院，上海 201620
2. 山东国力生物科技有限公司，山东 济南 250101

摘　要：内含肽介导的蛋白质剪接技术为蛋白质连接提供了重要的工具。内含肽侧翼外显蛋白的氨基酸序列可作为内含肽的天然底物，这些外显蛋白序列的改变可能导致内含肽剪接效率的大幅变化。迄今为止，关于不同位点（如−1 和+2 位点）外显蛋白序列如何影响内含肽的剪接活性仍缺乏系统性的研究。为解决这一问题，该研究构建了基于 mCherry 的荧光蛋白报告系统，将荧光信号与剪接活性偶联，建立高通量筛选方法。利用同源重组技术构建了 *Cne* Prp8 内含肽−1 和+2 位点的饱和突变库。经诱导表达后，从 336 个克隆中筛选出 38 个高荧光克隆。Western blot 分析表明，3 种底物组合 A/S、S/V、R/S 的剪接效率超过 80%，显著高于野生型序列。这些结果不仅揭示了外显肽关键位点的氨基酸偏好性，也为理性设计高效剪接系统提供了重要参考，有助于推动内含肽在蛋白质工程中的应用。

关键词：内含肽；外显肽；蛋白质剪接；荧光蛋白