






A split protease-*E. coli* ClpXP system quantifies protein-protein interactions in *Escherichia coli* cells

Shengchen Wang ¹, Faying Zhang¹, Meng Mei¹, Ting Wang ¹, Yueli Yun¹, Shihui Yang ¹, Guimin Zhang ¹✉ & Li Yi ¹✉

Characterizing protein-protein interactions (PPIs) is an effective method to help explore protein function. Here, through integrating a newly identified split human Rhinovirus 3 C (HRV 3 C) protease, super-folder GFP (sfGFP), and ClpXP-SsrA protein degradation machinery, we developed a fluorescence-assisted single-cell methodology (split protease-*E. coli* ClpXP (SPEC)) to explore protein-protein interactions for both eukaryotic and prokaryotic species in *E. coli* cells. We firstly identified a highly efficient split HRV 3 C protease with high re-assembly ability and then incorporated it into the SPEC method. The SPEC method could convert the cellular protein-protein interaction to quantitative fluorescence signals through a split HRV 3 C protease-mediated proteolytic reaction with high efficiency and broad temperature adaptability. Using SPEC method, we explored the interactions among effectors of representative type I-E and I-F CRISPR/Cas complexes, which combining with subsequent studies of Cas3 mutations conferred further understanding of the functions and structures of CRISPR/Cas complexes.

¹State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Hubei, China. ✉email: zhangguimin6@hotmail.com; liyihubu@hubu.edu.cn

Proteins serve essential functions in nearly all biological processes in living organisms. It has been revealed that over 80% of proteins exist in complexes¹, suggesting that protein is normally not carrying out its function alone. Therefore, characterizing the protein–protein interactions (PPIs) could help reveal the functions of the proteins, and also prompt the understanding of biological processes which they are involved in. Many methods have been developed for characterizing the PPIs, including bacteria/yeast two-hybrid (B2H/Y2H)^{2,3}, affinity-based techniques^{4,5}, quantitative proteomics⁶, and split protein complementation assays⁷. For example, Y2H has developed 30 years ago, which has been used to confirm suspected interactions, identify protein–drug interaction, and discover inhibitors of the PPIs⁸. In addition, quantitative proteomics is a newly developed method, which was recently used to map receptor–ligand interactions in the activation mechanism of class B G-protein-coupled receptors (GPCRs) upon interaction with peptide hormones in signal-transduction pathways of mammalian cells and elucidate the protein interactome in chromatin biology⁹.

Besides these methods, split-protein assays, also termed protein fragment complementation assays (PCAs), are another set of methods for characterizing PPIs, including split-GFP¹⁰, split-luciferase¹¹, split-ubiquitin¹², and split-TEV protease¹³ et al. PCAs have been widely applied in both prokaryotic and eukaryotic species to characterize transient and irreversible PPIs, and manipulate signal pathways, which consist of modular elements to prompt the system flexibility and functionality under in vivo environment. For example, split TEV protease assay was used to monitor the neuregulin-induced ErbB2/ErbB4 receptor tyrosine kinase heterodimerization¹³, and GPCR activities caused by drug-induced changes¹⁴. Although PCAs has been proven to be a robust method for explore PPIs, designing functional split proteins is a very challenging task, which needs to maintain the stability as well as the functionality of the split proteins to ensure the recovered biological activity of the reassembled proteins¹⁵. Many split proteins have been identified so far, among which split protease has the advantages of signal amplification and clean background because of its high catalytic activity and substrate specificity. Small amount of reassembled proteases could cleave large amount of its substrates, thus generating enhanced signals to increase the systematic sensitivity.

E. coli is the dominant model species in biological research due to its easy manipulation and clear genetic background. Recently, engineered *E. coli* was also used for eukaryotic protein studies^{16,17}, and the good adaption to low temperature further favors its use in academia and industry. Therefore, it is attractive to develop a sensitive PCA for characterizing PPIs in *E. coli* within a broad temperature range. Through combining the ClpXP-SsrA protein degradation machinery¹⁸, sfGFP fluorescence reporter, we developed a split human Rhinovirus (HRV) 3 C protease mediated PCA in *E. coli* cells, termed as Split Protease-*E. coli* ClpXP (SPEC) system. Our studies showed that splitting at K82 position of HRV 3 C protease could generate a high effective split HRV 3 C protease, and its incorporation to the SPEC system can make it characterize PPIs in *E. coli* quantitatively with high sensitivity, easy manipulation, and broad temperature adaptability from 18 to 37 °C. Using SPEC method, we further analyzed the interactions of effects of type I-E and I-F CRISPR/Cas to understand its possible complex structure and functions.

Results

Development of SPEC system. The SPEC system consists of two sub-parts, the ClpXP-SsrA protein degradation machinery mediated sfGFP fluorescence reporting sub-system and the split-HRV 3 C protease-mediated protein interacting sub-system (Fig. 1a). In

E. coli, ClpXP protease can recognize the SsrA sequence (AAN-DENYALAA), thus causing the degradation of SsrA-tagged protein in cells¹⁸. When the prey protein and bait protein do not interact with each other, sfGFP-SsrA is degraded through the ClpXP-SsrA mediated protein degradation pathway accompanied with a low cellular sfGFP fluorescence intensity. However, when the prey protein and bait protein interact with each other, their fused N-terminal and C-terminal split HRV 3 C proteases can re-assemble to a functional HRV 3 C protease, cleaving the SsrA peptide off sfGFP-SsrA to cause an increased sfGFP fluorescence intensity.

We first validated the principle of SPEC system in *E. coli* BL21 (DE3) by the designed SPEC-VA and SPEC-VB constructs (Supplementary Fig. 1a), which express the sfGFP-LEVLFQGP-SsrA polypeptide and intact HRV 3 C protease under the control of T7 and lac promoters, respectively. Our results showed that cells presented almost no sfGFP fluorescence when bearing the SPEC-VA vector only, while exhibited strong recovered sfGFP fluorescence when carrying both SPEC-VA and SPEC-VB vectors (Fig. 1b). Subsequently, we compared the functionality of Tobacco Etch Virus (TEV) and human Rhinovirus 3 C (HRV 3 C) proteases in SPEC system. Both are virus-derived proteases possessing high proteolytic activity and substrate specificity, with HRV 3 C protease exhibiting better proteolytic efficiency and toleration to low temperature¹⁹. Four common temperatures (37 °C, 30 °C, 25 °C, and 18 °C) were firstly used for system evaluation, because they are normally used for heterologous protein expression in *E. coli*. Our results showed that HRV 3 C protease presented up to 4 folds stronger sfGFP fluorescence intensity than using TEV protease, especially at low temperatures (Fig. 1c). With HRV 3 C protease, the highest levels of sfGFP fluorescence intensity in *E. coli* at 37 °C, 30 °C, 25 °C, and 18 °C were $2.0 \pm 0.1 \times 10E4$, $1.8 \pm 0.1 \times 10E4$, $1.6 \pm 0.1 \times 10E4$ and $1.7 \pm 0.1 \times 10E4$, which was obtained after 4 h, 6 h, 10 h and 20 h IPTG induction, respectively.

Identification of an optimized split HRV 3 C protease in SPEC system.

Next, we identified the split HRV 3 C protease modules for SPEC system (Fig. 2a, b). The solved HRV 3 C protease structure indicated that it was constituted of two characteristic six-stranded β -barrel subdomains, connected by a random coil²⁰. This structure is similar to the structure of TEV protease, which can also be dissected into two characteristic N-terminal and C-terminal subdomains. The reported split TEV protease modules have the split sites located on the random coil connected N-terminal and C-terminal subdomains¹³. Therefore, K82, L94, and N107, which were located on the random coil and the following β -sheet of the structure of HRV 3 C protease, were selected for generating the split HRV 3 C protease modules. At the same time, two protein pairs, Cas1/Cas2-3 from Type I-F CRISPR/Cas complex of prokaryotic *Zymomonas mobilis*²¹ and Yae1/Lto1 from Fe-S protein assembly (CIA) machinery of eukaryotic *Saccharomyces cerevisiae*²², were used to evaluate the effectiveness of split HRV 3 C protease and generality of the SPEC system (Fig. 2b, Supplementary Fig. 1b, Supplementary Fig. 1c). All three split-sites could prompt a re-assembled functional HRV 3 C protease, with K82 being the best. A total of 85–97% cells expressing split HRV 3 C (K82) protease showed sfGFP fluorescence, which is close to the 98% cells expressing the HRV 3 C protease (Fig. 2b). In comparison, only ~40% activity of the TEV protease could be recovered with the reported split TEV protease in a previous report¹³. Moreover, split HRV 3 C (K82) protease outperformed split TEV protease in SPEC system under different temperatures (Fig. 2c, d). It is also worth noting that the highest level of cellular sfGFP fluorescence intensity with re-assembled split HRV 3 C (K82) protease at 30 °C, 25 °C, or 18 °C, was more than 4 folds higher than that of the re-assembled split TEV protease (Fig. 2c, d).

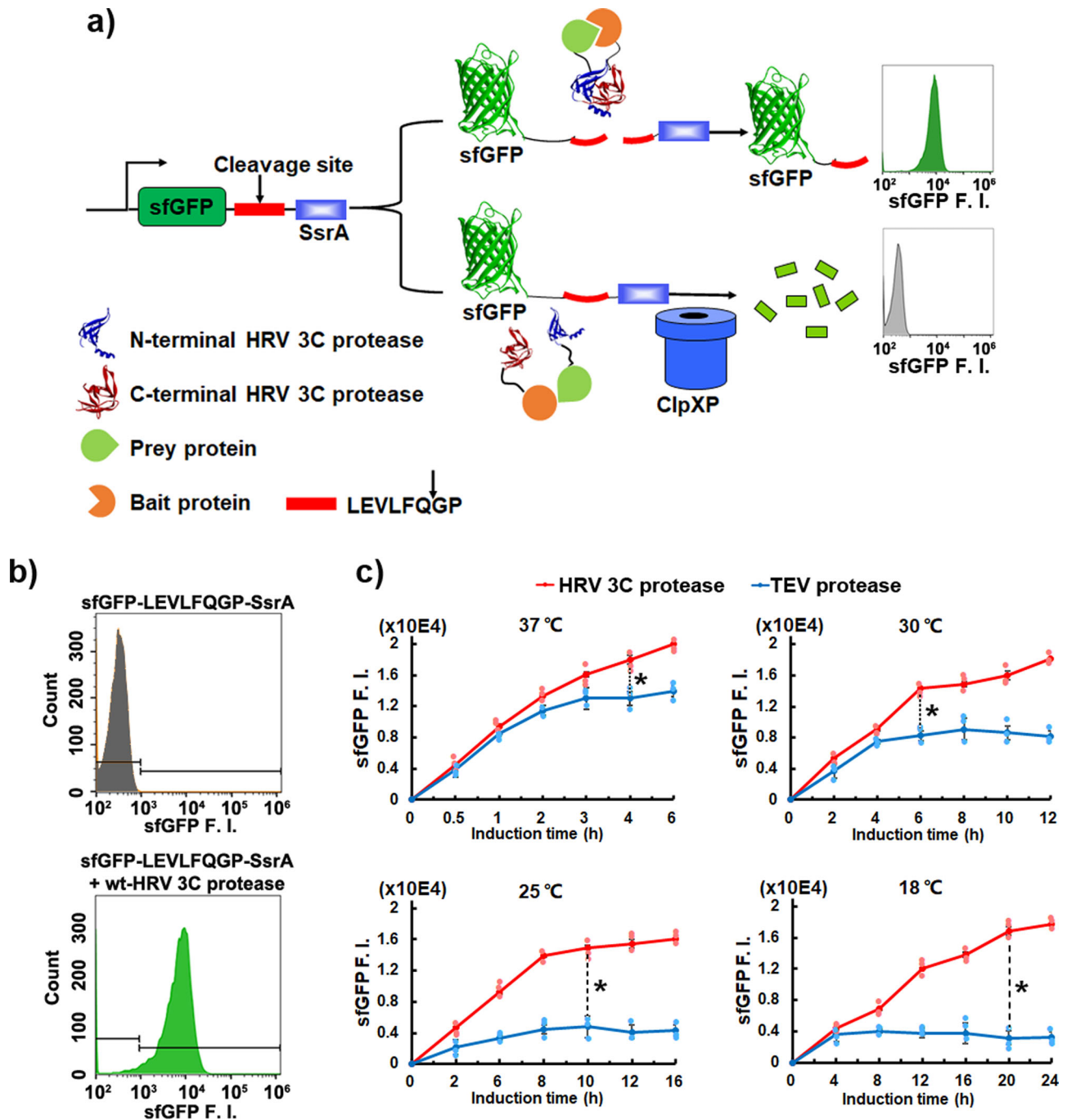


Fig. 1 Development of the concept of SPEC system. **a** Schematic diagram of Split Protease-*E. coli* ClpXP (SPEC) system. sfGFP is fused with an HRV 3C protease cleavage sequence (LEVLFQ↓GP) at its C-terminus followed by a SsrA sequence (AANDENYALAA), forming a sfGFP-LEVLFQGP-SsrA cassette. The prey protein and bait protein are fused to the N-terminal and C-terminal split-HRV 3C proteases, respectively. Through ClpXP-SsrA degradation pathway, highly expressed sfGFP-LEVLFQGP-SsrA cassette is fast degraded in *E. coli* cells, thus very low green fluorescence is detected. Nevertheless, when prey protein interacts with bait protein, the N-terminal split and C-terminal split-HRV 3C proteases will re-assemble to a functional HRV 3C protease, which will cleave off the SsrA peptide from the sfGFP-LEVLFQ↓GP-SsrA cassette, leading to an accumulation of sfGFP for increased fluorescence intensity. **b** Validation of the ClpXP-SsrA protein degradation machinery mediated sfGFP degradation in *E. coli*. The total cellular sfGFP fluorescence intensity (sfGFP F. I.) of cells bearing different plasmids were quantitated by flow cytometry (detected with FITC channel, 525/40 nm BP). The experiments were performed under 25 °C with 8 h induction. Upper panel: cells bearing pSPEC-VA vector, Lower panel: cells bearing both pSPEC-VA and pSPEC-VB vectors (Supplementary Table 1, Supplementary Fig. 1). **c** Comparing TEV protease (blue line, pSPEC-T, Supplementary Table 1) with HRV 3C protease (red line, pSPEC-VB, Supplementary Table 1) in SPEC system at different temperatures of 37 °C, 30 °C, 25 °C, and 18 °C, respectively. All the vector information can be found in Supplementary Table 1. Data are presented as mean ± SEM ($n = 3$ independent experiments) with Student's t test being performed, # $P > 0.05$, * $P \leq 0.05$.

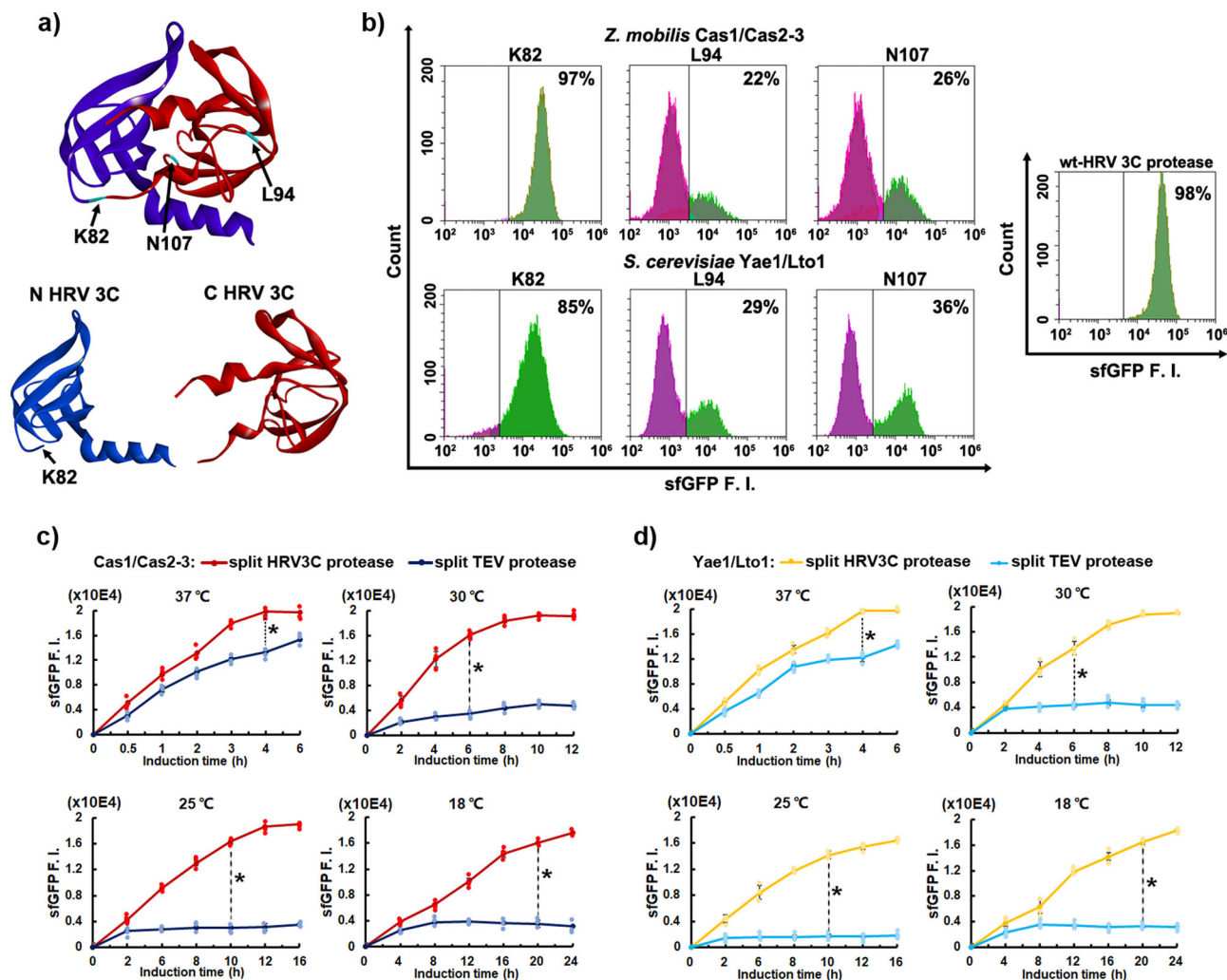


Fig. 2 Identification of a highly efficient split HRV 3C protease and its comparison with split TEV protease in the SPEC system. **a** Designing split HRV 3C protease based on its structure (PDB: 2B0F) with three split positions (K82, L94, N107). Blue: N-terminal HRV 3C protease, Red: C-terminal HRV 3C protease, Purple: the random coil between N-terminal and C-terminal HRV 3C protease, Cyan: the designed split sites. **b** Characterizing different split HRV 3C proteases in SPEC system using pSPEC-N107/L94/K82-PP and pSPEC-N107/L94/K82-EP vectors, respectively. Cells bearing pSPEC-VB vector containing full length HRV 3C protease was used as a positive control (right panel), and cells bearing only pSPEC-VA vector was used as a negative control (Fig. 1b). The total cellular sfGFP F. I. of cells bearing different vectors (pSPEC-K82-EP to pSPEC-N107-PP, Supplementary Table 1) were quantitated through flow cytometry. Upper-left panel, representative flow cytometry histograms of the total cellular sfGFP F. I. of cells expressing different split HRV 3C proteases fused with *Z. mobilis* Cas1/Cas2-3 protein pair. Lower-left panel, representative flow cytometry histograms of the total cellular sfGFP F. I. of cells expressing different split HRV 3C proteases fused with *S. cerevisiae* Yae1/Lto1 protein pair. The experiments were performed at 25 °C with 8 h induction. **c** Comparison of split HRV 3C (K82) protease (Wine line, pSPEC-VC1) with split TEV protease (Dark blue line, pSPEC-VC5) using Cas1/Cas2-3 protein pair in SPEC system at 37 °C, 30 °C, 25 °C, and 18 °C, respectively. **d** Comparison of split HRV 3C (K82) protease (Yellow line, pSPEC-VC2) with split TEV protease (Cyan line, pSPEC-VC6) using Yae1/Lto1 protein pair in SPEC system at 37 °C, 30 °C, 25 °C, and 18 °C, respectively. All the vector information can be found in Supplementary Table 1. Data are presented as mean \pm SEM ($n = 3$ independent experiments) with Student's *t* test being performed, $\#P > 0.05$, $*P \leq 0.05$.

Construction of the SPEC system in the recombinant *E. coli* BL21 (DE3)-SPEC cells. To simplify the manipulation of SPEC system and increase its stability in *E. coli*, we used the modified no-SCAR (Scarless Cas9 Assisted Recombineering) technology²³ to integrate the *sfGFP-LEVLFGQP-SsrA* into *E. coli* chromosome followed by temperature sensitive plasmid curing²⁴, generating the recombinant *E. coli* BL21(DE3)-SPEC strain (Fig. 3a). Additionally, the prey protein fused N-terminal split HRV 3C (K82) protease and bait protein fused C-terminal split HRV 3C (K82) protease were separated into SPEV-VD (Kan^R) and SPEC-VE (Amp^R) vectors, respectively (Supplementary Fig. 1d). Our results showed a very low sfGFP background fluorescence in

BL21(DE3)-SPEC cells, while cells bearing the SPEC-VB vector exhibited a high sfGFP fluorescence intensity (Fig. 3b). Similar high sfGFP fluorescence intensity was also obtained in BL21 (DE3)-SPEC bearing both SPEC-VD and SPEC-VE vectors, validating the effectiveness of this system constituting of BL21 (DE3)-SPEC strain and SPEC-VD/SPEC-VE vectors. Subsequently, we evaluated this final SPEC system using Cas1/Cas2-3 and Yae1/Lto1 protein pairs in BL21(DE3)-SPEC at different temperatures (Fig. 3c). Similar tendencies were obtained that split HRV 3C (K82) protease conferred a higher sensitivity than that of split TEV protease, especially under low temperatures (Fig. 2c, d).

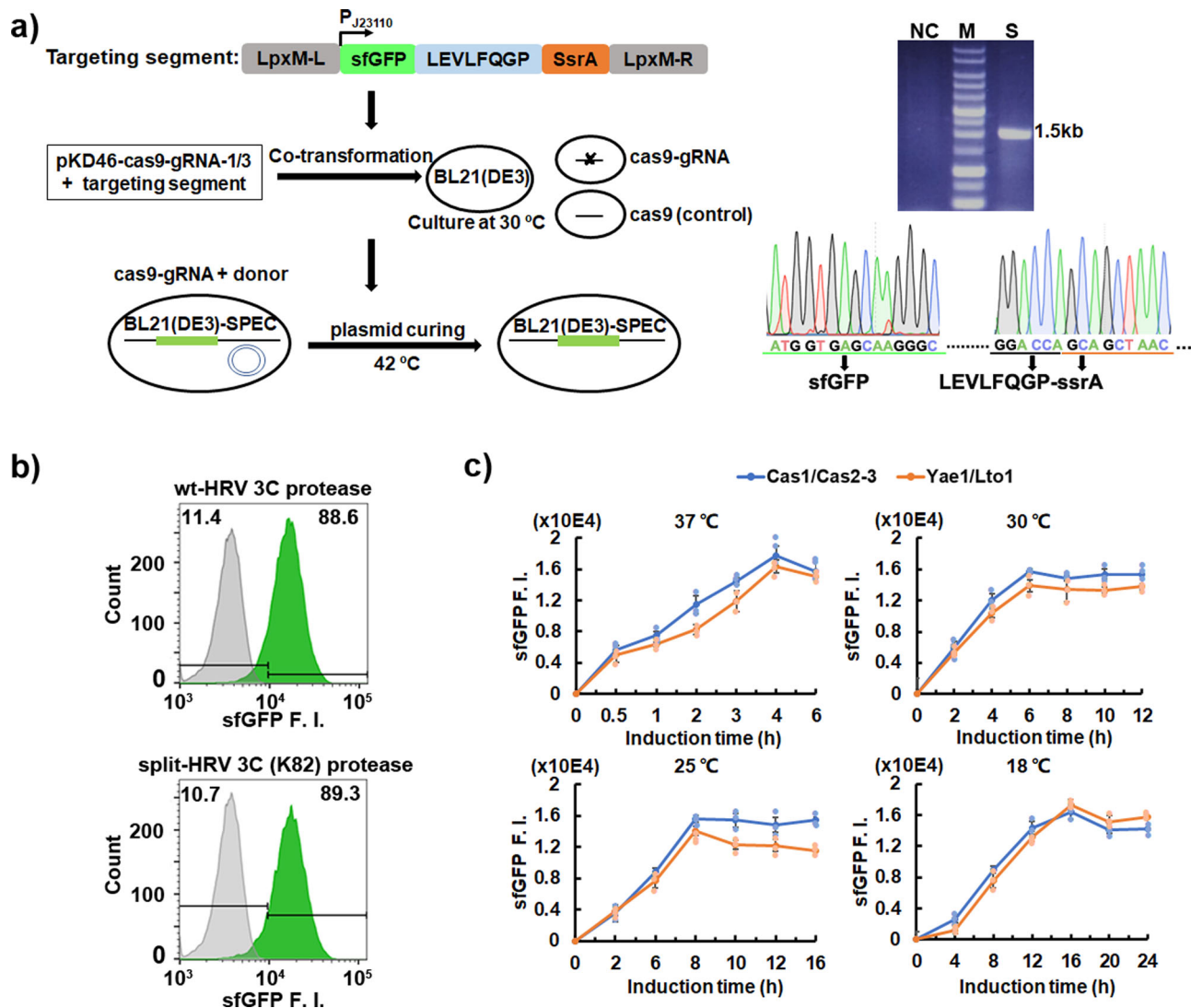


Fig. 3 Development of SPEC system in the engineered BL21(DE3)-SPEC cells. **a** Generation of the BL21(DE3)-SPEC cells using no-SCAR incorporated with temperature sensitive plasmid curing strategy. Left panel, scheme of experimental procedure. The *lpxM* gene fragment in the genome of BL21(DE3) strain is replaced by sfGFP-LEVLQGP-SsrA fragment using CRISPR/Cas9 assisted λ -Red recombinase technology. The pKD46 based vector is used to prompt the temperature sensitive plasmid curing. Upper-right panel, PCR validation of the inserted sfGFP-LEVLQGP-SsrA fragment in BL21(DE3)-SPEC strain. Negative control (NC): wild-type BL21(DE3) strain. S: BL21(DE3)-SPEC strain. M: 1Kb DNA ladder. Lower-right panel, Sanger sequencing of BL21 (DE3) genome to verify the insertion of the sfGFP-LEVLQGP-SsrA fragment. **b** Validation of the SPEC system in the BL21(DE3)-SPEC strain by comparing the HRV 3 C protease (pSPEC-VB) and split HRV 3 C (K82) protease (pSPEC-VC1) with Cas1/Cas2-3 protein pair. The experiments were performed at 25 °C with 8 h induction. **c** Characterizing split HRV 3 C (K82) protease in BL21(DE3)-SPEC strain with *Z. mobilis* Cas1/Cas2-3 (pSPEC-VC1) and *S. cerevisiae* Yae1/Lto1 (pSPEC-VC2) protein pairs at 37 °C, 30 °C, 25 °C, and 18 °C, respectively. Data are presented as mean \pm SEM ($n = 3$ independent experiments). BL21(DE3)-SPEC cells bearing pSPEC-VB plasmid containing HRV 3 C protease was used as a positive control, and the total cellular sfGFP F. I. were quantitated through flow cytometry. All the vector information can be found in Supplementary Table 1.

Characterization of interaction of different effectors in the type I-E and I-F CRISPR/Cas complex. CRISPR/Cas system is a newly emerged gene-editing tool, in which different effectors work in a correlative manner²⁵. We characterized the effector interactions of the type I-E CRISPR/Cas complex from *E. coli* K12 using SPEC system (Fig. 4a, Supplementary Fig. 2). Cas1 and Cas2 exhibited very strong interactions (92%), which is consistent with the reported structure of Cas1-Cas2 complex²⁶. Cas1/Cas2, Cas1/Cas5, Cas1/Cas11, Cas3/Cas5, and Cas6/Cas7 presented strong interactions above 80%, and Cas2/Cas5, Cas2/Cas6, Cas3/Cas6, and Cas5/Cas8 exhibited weak interactions. Interestingly, interaction between Cas3 and Cas2 enhanced 32% and 28% with co-overexpression of Cas1 and Cas8, respectively (Supplementary

Fig. 2), confirming that the assembly of Cas3 could be prompted by Cas1-Cas2 complex and Cas8 mediated Cascade/R-loop/Cas3 complex²⁷. For a comparison, the effector interactions of the type I-F CRISPR/Cas system from *Z. mobilis* were also profiled using the SPEC system (Fig. 4b, Supplementary Fig. 3). Type I-F CRISPR/Cas system is the smallest in Class 1 CRISPR system, containing a characteristic fused Cas2-3 subunit²⁸. Similar to type I-E CRISPR/Cas system, a very strong interaction also exists between Cas1 and Cas2-3 (99%), consistent with the reported structure of Cas1₄-Cas2-3₂ complex²⁹. Differently, most effectors had interaction intensities higher than 60%, among which the interaction intensity of Cas1/Cas2-3, Cas1/Csy2, Cas1/Csy4, Cas2-3/Csy2, Cas2-3/Csy3, Cas2-3/Csy4, Csy1/Csy2, and

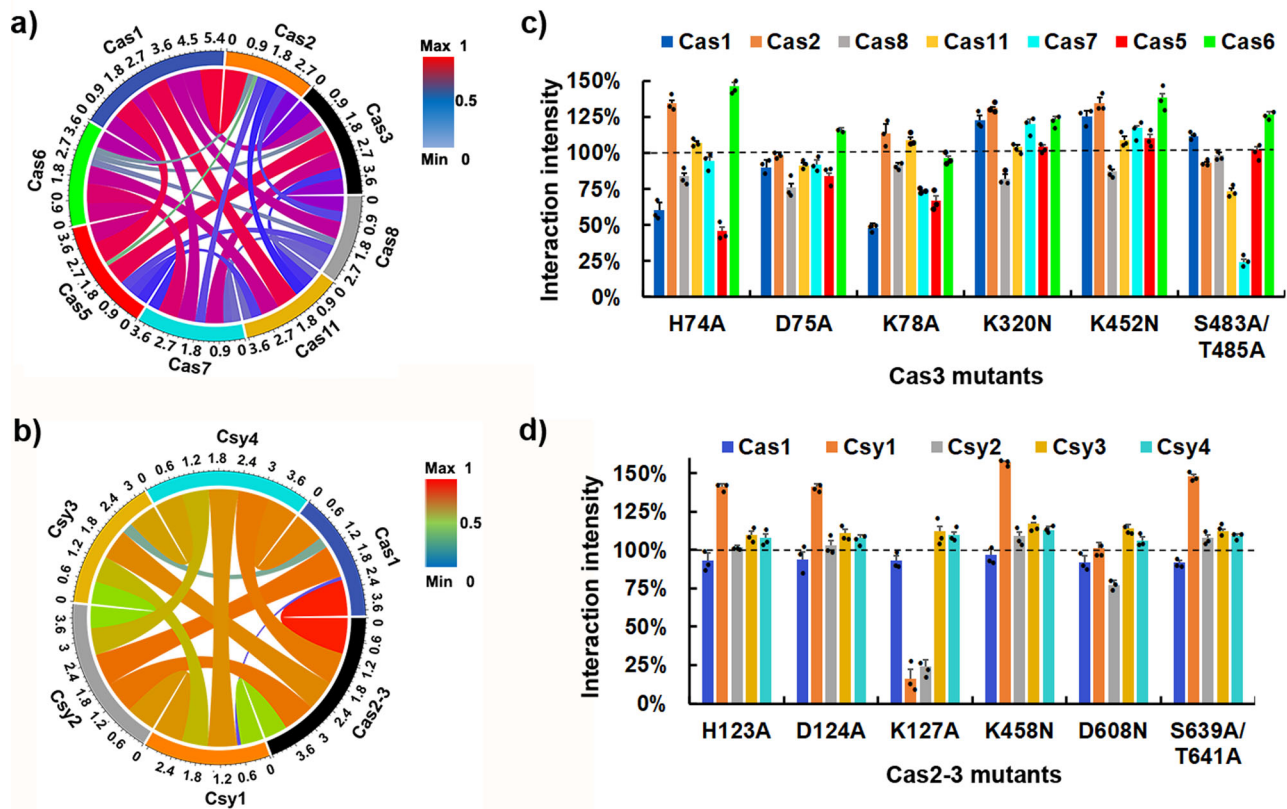


Fig. 4 Characterization of protein-protein interactions of effectors in the Type I-E and I-F CRISPR/Cas complex. **a** Protein-protein interactions among eight effectors (Cas1, Cas2, Cas3, Cas8, Cas11, Cas7, Cas5, and Cas6) of *E. coli* K12 Type I-E CRISPR/Cas complex was interpreted using the circular visualization R program⁴⁴. **b** Protein-protein interactions among six effectors (Cas1, Cas2-3, Csy1, Csy2, Csy3, and Csy4) of *Z. mobilis* Type I-F CRISPR/Cas complex was interpreted using the circular visualization R program⁴⁴. **c** The histogram of protein-protein interaction between Cas3 mutants (H74A, D75A, K78A, K320N, K452N, and S483A/T485A, respectively) and seven other effectors (Cas1, Cas2, Cas5, Cas6, Cas7, Cas8, and Cas11) of *E. coli* K12 Type I-E CRISPR/Cas complex. Cas1/GST was used as a negative control (Supplementary Fig. 2b), and Cas1/Cas2 was used as a positive control. Percentage numbers presented correspondent interaction intensities of protein pairs, which were normalized by the interaction between Cas3 and correspondent effectors (Supplementary Fig. 2b). **d** The histogram of protein-protein interaction between Cas2-3 mutants (H123A, D124A, K127A, K458N, D608N, and S639A/T641A, respectively) and five other effectors (Cas1, Csy1, Csy2, Csy3, and Csy4) of *Z. mobilis* Type I-F CRISPR/Cas complex. Cas1/GST was used as a negative control (Supplementary Fig. 3b), and Cas1/Cas2-3 was used as a positive control. Percentage numbers presented correspondent interaction intensities of protein pairs, which were normalized by the interaction between Cas2-3 and correspondent effectors (Supplementary Fig. 3b). All the vector information can be found in Supplementary Table 1. Data are presented as mean \pm SEM ($n = 3$ independent experiments).

Csy1/Csy4 were above 80%. However, Cas1 presented very weak interaction with Csy3 (14%) and no interaction with Csy1 (0%), indicating a characteristic type I-F CRISPR/Cas complex structure.

To further understand the difference between type I-E and I-F CRISPR/Cas complexes, we profiled the interactions of Cas3 mutants³⁰ with all other effectors (Fig. 4c, d). Surprisingly, the Cas3 mutants presented different effects in these two complexes. In type I-E CRISPR/Cas complex, most Cas3 mutants increased its interaction with Cas2 and Cas6, while mutations altering Mg²⁺-dependent HD-nuclease activity (H74A, D75A, K78A) and ATP-dependent helicase activity (K320N, K452N, S483A/T485A) either decreased or increased its interaction with Cas1, respectively. In comparison, most Cas3 mutants did not affect the interaction of Cas2-3 with other effectors in type I-F CRISPR/Cas complex. However, K127A significantly decreased the interaction of Cas2-3 with Csy1 (85%) and Csy2 (75%), while other mutants largely increased its interaction with Csy1 (>40%). These results indicated different roles of Cas3 in type I-E and I-F CRISPR/Cas systems, in which Cas3 might possess a more stable structure due to its fusion with Cas2 effector. Besides, this difference on effector interaction might lead to the different functional mechanisms between these two CRISPR/Cas systems.

Discussion

In our studies here, a designed split HRV 3 C protease was utilized in the SPEC system for quantitating protein-protein interaction from both prokaryotic and eukaryotic species. Compared to the widely used split TEV protease^{13,31}, HRV 3 C protease is catalytically faster, thus producing further enhanced sfGFP signals in SPEC system (Fig. 1c). More importantly, HRV 3 C protease exhibit constantly high proteolytic activity in a broad temperature range from 18 °C to 37 °C (Fig. 1c). This property is very important because many heterologous proteins could only be expressed and folded well at low temperature in *E. coli* cells. Besides the common temperatures for *E. coli* cell growth, we also evaluated the SPEC system at 42 °C and 12 °C. Similar results were obtained that split HRV 3 C (K82) protease presented higher sensitivity than that of split TEV protease at 12 °C, while their difference is not obvious at 42 °C (Supplementary Fig. 4). These results further confirmed that split HRV 3 C (K82) protease possesses high sensitivity in a broad temperature range, which is better than the split TEV protease, especially at low temperature. Besides the advantageous adaption to low temperature, it is also noticed that the reassembled split HRV 3 C proteases presented almost the identical proteolytic activity to the intact HRV 3 C protease (Fig. 2b), while it was reported that reassembled split

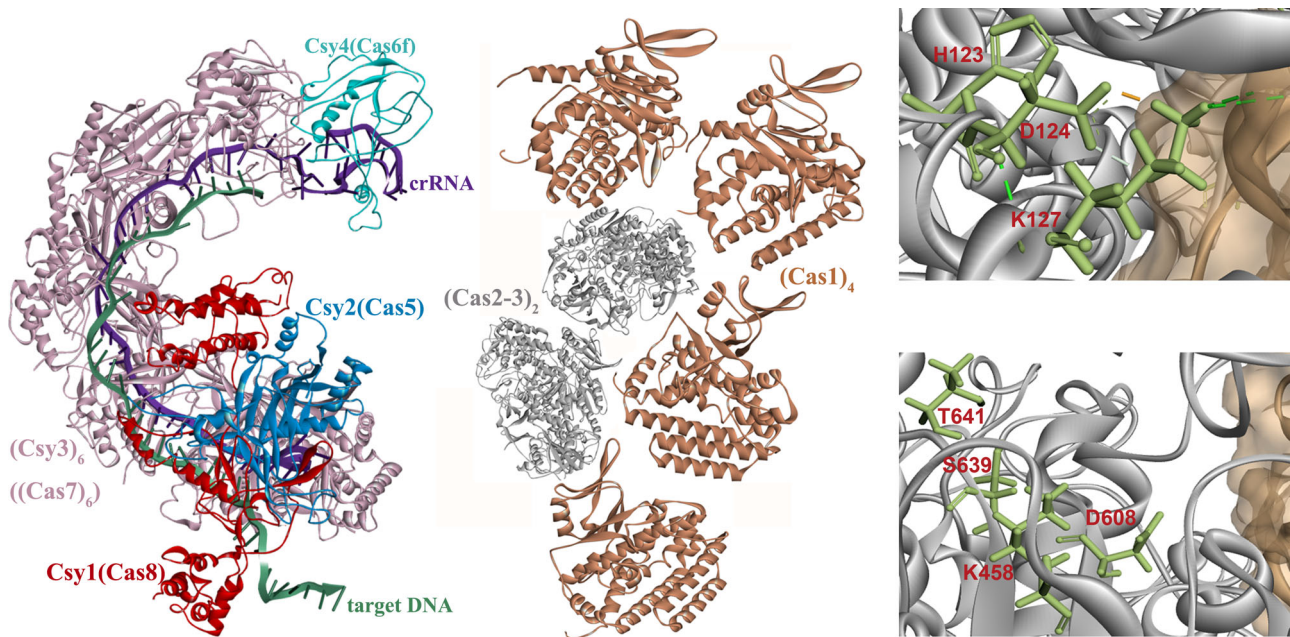


Fig. 5 Simulated structure diagram of Type I-F CRISPR/Cas complex. Left panel: Model protein structures of six proteins (Cas1, Cas2-3, Csy1, Csy2, Csy3 and Csy4) in CRISPR/Cas complex. Right panel: Location of mutation sites (H123, D124, K127, K458, D608, S639, and T641; red font) in Cas2-3.

TEV protease only recovers its 40% proteolytic activity¹³. The high activity and low-temperature adaptability of split HRV 3 C (K82) protease make it more efficient in SPEC system.

SPEC systems have some unique advantages. Comparing to split GFP method, the split HRV 3 C protease can amplify the signals of PPIs, thus detecting the weak interactions. Some other PCAs can also amplify the signals of PPIs, such as split luciferase assay³². However, SPEC system has another advantage that it quantitates the PPIs through fluorescent intensity of cellular sfGFP, not requiring additional exogenous substrates. It also needs to point out that SPEC system also has limitations. Since SPEC system is an *E. coli* cell-based assay, it requires that both the prey and bait proteins can be expressed and functionally folded in *E. coli* cells. Therefore, it may not be suitable for complex eukaryotic proteins that cannot be folded in *E. coli* cells.

To stabilize the production of sfGFP in *E. coli* cells, we integrated the sfGFP gene into the *E. coli* genome in SPEC, which could minimize the heterogeneity caused by episomal plasmid (Fig. 3). The modular elements make the SPEC system highly flexible. Using other reporters, SPEC system may be performed with preferred signal readout. Additionally, owing to the sensitivity and flexibility of split HRV3C protease, we believe that it could be used in the future to design more useful biosensors, in particular for high-throughput applications requiring signal enhancement. In addition, SPEC system uses a two-plasmid strategy in the recombinant *E. coli* BL21(DE3)-SPEC strain. The separation of prey and bait proteins in different plasmids not only significantly simplifies the manipulation, but also facilitates the analysis of large-scale PPIs. Combined with the flow cytometry and high-throughput sequencing technology, it could be believed that SPEC system can be extended to characterize PPIs in a library against library set.

Combining our PPIs results with the previous reported structures of Cas1₄-Cas2-3₂ complex²⁹ and resembled type I-E CRISPR/Cas complex²⁶, we simulated the model structure scheme of type I-F CRISPR/Cas complex (Fig. 5). Our results showed that Cas2-3 presented strong interactions with all other effectors, and Cas1 presented a very strong interaction with Cas2-3, Csy2, and Csy4, but no interaction of Csy1 and very low

interaction with Cys3 (Fig. 4b, Supplementary Fig. 3). We speculated that Cas1 effector is located outside of the whole complex, wrapping up the Cas2-3 effectors, with Csy1 and Csy3 effectors being located at the other side of the whole complex. This unique structure might be caused by the requirement of the “flexible” PAM recognition function of Csy1³³, and the binding of the spacer region by Csy3 in type I-F CRISPR systems³⁴. Additionally, our results also confirmed that Cas2 might function as adaption protein and the helicase and nuclease functions of the Cas3 domain were involved in interference process³⁵.

The simulated type I-F CRISPR/Cas complex structure (Fig. 5) exhibits high similarity with that of the type I-E cascade complex²⁶. Interestingly, Cas2 and Cas3 only presented an interaction of 56%, while this interaction was significantly enhanced to 90% and 86% when co-expression of Cas1 and Cas8 (Fig. 4a, Supplementary Fig. 2), respectively. This might explain the characteristically fused Cas2-Cas3 domain in type I-F CRISPR/Cas complex, because it lacks Cas8 effector. Cas3 is the catalytic effector in type I CRISPR/Cas systems, and the fused Cas2-Cas3 domain is believed to maintain a more stable function of Cas3 in type I-F CRISPR/Cas complex, which is further supported in our studies of Cas3 mutants in both type I-E and I-F CRISPR/Cas complexes (Fig. 4c, d). Comparing to the Cas3 mutants in type I-E CRISPR/Cas complex, which exhibited promiscuous effect, the correspondent mutations in type I-F CRISPR/Cas complex all presented similar interactions with all other effectors, except the nuclease activity related Cas2-3 (K127A) mutant presented obviously decreased interaction with Csy1/Csy2 (from 63% to 10%, 88% to 21%, respectively). Besides, Cas2-3(H123A, D124A, K458N, and S639A/T641A) presented a significantly enhanced interaction with Csy1. All these Cas3 related mutations affect the interactions of Cas2-Cas3 with Csy1/Csy2 effectors, which were located at N-terminus of crRNA-Cascade complex, and speculated to function as blocking nuclease recruitment, crRNA and target binding³⁶.

In principle, any protein pairs that are able to be functionally expressed in *E. coli* cells at the temperature range from 12 °C and 42 °C. can be characterized in SPEC system. Considering the exist of similar ClpXP-SsrA mediated protein degradation machinery

in other bacteria, such as the Gram-positive model species *Bacillus subtilis*³⁷, the concept of SPEC system can be easily implanted in other bacteria.

Methods

Bacterial strains and reagents. *E. coli* XL10-Gold strain (Stratagene, San Diego, USA) was used for plasmid construction, and *E. coli* BL21(DE3) strain was used for protein expression and flow cytometry analysis. Culture media, chemicals, and restriction enzymes were purchased from BD Biosciences (San Jose, CA, USA), Sigma-Aldrich (St. Louis, MO, USA), and New England Biolabs (Ipswich, MA, USA), respectively.

Plasmid construction. pET28a and pUC19 plasmids were used as backbone to construct pSPEC plasmids in this study (Supplementary Fig. 1). Total of 64 constructs were generated and their vector information can be found in Supplementary Table 1. The split HRV 3 C protease domain (N HRV 3 C or C HRV 3 C) was fused with different protein pairs (prey and bait proteins), forming a prey-N HRV 3 C protease or bait-C HRV 3 C protease cassette. HRV 3C protease substrate sequence was flanked with sfGFP and SsrA tag (AANDENYALAA) sequences, forming the sfGFP-LEVLQGP-AANDENYALAA cassette. A well-characterized Tobacco Etch Virus (TEV) protease variant containing a S219P mutation with enhanced stability³⁸ was used and annotated as TEV Protease in our research. Similarly, TEV protease substrate sequence (ENLYFQS) was also flanked with sfGFP and SsrA tag sequences, forming the sfGFP-ENLYFQS-AANDENYALAA cassette. These cassettes were cloned into different pSPEC plasmids.

Plasmids for this study was constructed using T5 exonuclease DNA assembly (TEDA) assay with minor modification³⁹. Firstly, 1 mL 5 × TEDA solution contained 0.5 M Tris-HCl pH7.5, 50 mM dithiothreitol, 0.25 g of PEG-8000, and 1 μL of 10 U/μL T5 exonuclease (New England Biolabs, USA) was prepared for cloning. For the experiment, 4 μL assembly mixture aliquot was thawed on ice, followed by adding 16 μL of DNA solution containing the linear vector and target inserts. The molar ratio of linearized vector to target insert was set as 1:3, with vector amount approximately 50 to 200 ng. Reactions were then performed at 30 °C for 40 min, followed by immediate transformation into *E. coli* competent cells.

Knock-in of sfGFP-LEVLQGP-SsrA fragment into the genome of *E. coli*. *E. coli* BL21(DE3) (*lpxM*:: sfGFP-LEVLQGP-SsrA) strain, named as BL21 (DE3)-SPEC, was generated using CRISPR/Cas9 assisted λ-Red recombinase system⁴⁰ followed by temperature sensitive plasmid curing. Cas9 and gRNA were incorporated into pKD46, a temperature-sensitive vector⁴¹ carrying λ-Red genes (γ, β, exo), which promotes recombination, to construct pKD46-cas9-gRNA plasmid (Supplementary Fig. 1). Subsequently, pKD46-cas9-gRNA vector DNAs targeting the *lpxM* gene in *E. coli* genome were transformed into *E. coli* BL21(DE3) followed by cultivation and induction with 0.2% L-arabinose in LB media containing 100 μg/mL ampicillin at 30 °C to an OD₆₀₀ of approximately 0.6. The electrocompetent cells were then made by centrifugation, followed by washing twice with ice-cold deionized water. Meanwhile, the donor fragment (sfGFP-LEVLQGP-SsrA) was gel-purified, and 10 ng of donor DNA and 50 μL of electrocompetent cells were then mixed on ice for electroporation. One mL LB media were then added into the post-shocked cells immediately following by recovering at 37 °C for 1 h, and then spreading on LB plates with chloramphenicol. *E. coli* transformants carrying sfGFP-LEVLQGP-SsrA fragments in the genome were screened and selected by colony PCR and Sanger sequencing. The validated transformants were then grown in LB media at 42 °C for 12 h to eliminate the temperature-sensitive pKD46-cas9-gRNA-donor plasmid, generating BL21(DE3)-SPEC strain for further experiments.

Protein-Protein Interaction analysis with CytoFLEX by SPEC assay. The pSPEC constructs were transformed into *E. coli* BL21(DE3) or BL21(DE3)-SPEC cells followed by cultivating at 37 °C until OD₆₀₀ reached 0.8. The cells were then induced by 0.5 mM IPTG (final concentrations) at different temperatures ranging from 18 °C to 37 °C.

Two hundred μL induced cells were collected by centrifugation at 3,000 × g at 4 °C for 5 min and then resuspended into 400 μL 1 × PBS buffer (200 mM NaCl, 66 mM Na₂HPO₄, 34 mM NaH₂PO₄, 2H₂O, pH 7.4), followed by sfGFP fluorescent analysis using Beckman Coulter CytoFLEX Flow Cytometer (Beckman Coulter, USA) equipped with a 488 nm laser and a 525/40 nm band-pass filter. sfGFP fluorescent intensity (sfGFP F. I.) of the cells was measured by the mean fluorescence intensity of the fluorescent cells.

The protein-protein interaction (PPI) intensity of cells expressing different targeting protein pairs were normalized using the mean fluorescence intensity of the cells expressing wt HRV 3 C protease as a control. The normalized interaction intensity was calculated as: [Normalized interaction intensity] = ([Mean fluorescence intensity percentage of the cells] / [Percentage of whole cells]) × 100%.

Structure modeling and circular visualization analysis for PPIs data. The structure modeling of six proteins (Cas1, Cas2-3, Csy1, Csy2, Csy3, and Csy4) from *Z. mobilis* Type I-F CRISPR/Cas complex were performed by using the I-TASSER on-line server⁴² respectively. The docking analyses of type I-F CRISPR/Cas

complex were performed by using the ZDOCK program in BIOVIA Discovery Studio client 2020⁴³. The circular visualization map of protein-protein interactions among six proteins (Cas1, Cas2-3, Csy1, Csy2, Csy3, and Csy4) from *Z. mobilis* Type I-F CRISPR/Cas complex was generated by using *circIize* package in R⁴⁴.

Statistics and reproducibility. sfGFP fluorescent analysis and the protein-protein interaction (PPI) intensity of cells assays were independently replicated three times and individual data points are reported for each experiment.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The source data for the graphs in the main figures is available as Supplementary Data 1. The vectors in Supplementary Figure 1 have been deposited in to Addgene with ID numbers 171974-171989. The datasets generated and/or analyzed during the current study are available from the authors on reasonable request.

Received: 27 December 2020; Accepted: 10 June 2021;

Published online: 06 July 2021

References

- Berggard, T., Linse, S. & James, P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **7**, 2833–2842 (2007).
- Lin, J.-S. & Lai, E.-M. Protein-protein interactions: yeast two-hybrid system. *Methods Mol. Biol.* **1615**, 177–187 (2017).
- Karimova, G., Gauliard, E., Davi, M., Ouellette, S. P. & Ladant, D. Protein-protein interaction: bacterial two-hybrid. *Methods Mol. Biol.* **1615**, 159–176 (2017).
- Burckstummer, T. et al. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat. Methods* **3**, 1013–1019 (2006).
- Rigaut, G. et al. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030–1032 (1999).
- Blagoev, B. & Mann, M. Quantitative proteomics to study mitogen-activated protein kinases. *Methods* **40**, 243–250 (2006).
- Wehr, M. C. & Rossner, M. J. Split protein biosensor assays in molecular pharmacological studies. *Drug Discov. Today* **21**, 415–429 (2016).
- Moosavi, B., Mousavi, B., Yang, W. C. & Yang, G. F. Yeast-based assays for detecting protein-protein/drug interactions and their inhibitors. *Eur. J. Cell Biol.* **96**, 529–541 (2017).
- Nguyen, T. A., Cigler, M. & Lang, K. Expanding the genetic code to study protein-protein interactions. *Angew. Chem. Int. Ed. Engl.* **57**, 14350–14361 (2018).
- Ghosh, I., Hamilton, A. D. & Regan, L. Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* **122**, 5658–5659 (2000).
- Kim, S. B., Ozawa, T., Watanabe, S. & Umezawa, Y. High-throughput sensing and noninvasive imaging of protein nuclear transport by using reconstitution of split *Renilla* luciferase. *Proc. Natl Acad. Sci. USA* **101**, 11542–11547 (2004).
- Kittanakom, S. et al. Analysis of membrane protein complexes using the split-ubiquitin membrane yeast two-hybrid (MYTH) system. *Methods Mol. Biol.* **548**, 247–271 (2009).
- Wehr, M. C. et al. Monitoring regulated protein-protein interactions using split TEV. *Nat. Methods* **3**, 985–993 (2006).
- Wehr, M. C., Galinski, S. & Rossner, M. J. Monitoring G protein-coupled receptor activation using the protein fragment complementation technique split TEV. *Methods Mol. Biol.* **1272**, 107–118 (2015).
- Ray, P. et al. Noninvasive quantitative imaging of protein-protein interactions in living subjects. *Proc. Natl Acad. Sci. USA* **99**, 3105–3110 (2002).
- Rodrigues, A. L., Becker, J., de Souza Lima, A. O., Porto, L. M. & Wittmann, C. Systems metabolic engineering of *Escherichia coli* for gram scale production of the antitumor drug deoxyviolacein from glycerol. *Biotechnol. Bioeng.* **111**, 2280–2289 (2014).
- Rodrigues, A. L. et al. Systems metabolic engineering of *Escherichia coli* for production of the antitumor drugs violacein and deoxyviolacein. *Metab. Eng.* **20**, 29–41 (2013).
- Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12**, 1338–1347 (1998).
- Raran-Kurussi, S., Tözsér, J., Cherry, S., Tropea, J. E. & Waugh, D. S. Differential temperature dependence of tobacco etch virus and rhinovirus 3C proteases. *Anal. Biochem.* **436**, 142–144 (2013).

20. Bjorn Dahl, T. C., Andrew, L. C., Semenchenko, V. & Wishart, D. S. NMR solution structures of the apo and peptide-inhibited human rhinovirus 3C protease (Serotype 14): structural and dynamic comparison. *Biochemistry* **46**, 12945–12958 (2007).
21. Zheng, Y. et al. Characterization and repurposing of the endogenous Type I-F CRISPR-Cas system of *Zymomonas mobilis* for genome engineering. *Nucleic Acids Res.* **47**, 11461–11475 (2019).
22. Paul, V. D. et al. The deca-GX3 proteins Yae1-Lto1 function as adaptors recruiting the ABC protein Rli1 for iron-sulfur cluster insertion. *eLife* **4**, e08231 (2015).
23. CR, R. & KL, P. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* **5**, 15096 (2015).
24. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Appl. Environ. Microbiol.* **81**, 5103–5114 (2015).
25. Mitsunobu, H., Teramoto, J., Nishida, K. & Kondo, A. Beyond native Cas9: manipulating genomic information and function. *Trends Biotechnol.* **35**, 983–996 (2017).
26. Nuñez, J. K. et al. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 (2014).
27. Xiao, Y., Luo, M., Dolan, A. E., Liao, M. & Ke, A. Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science* **361**, 6397 (2018).
28. Pickar-Oliver, A. et al. Targeted transcriptional modulation with type I CRISPR-Cas systems in human cells. *Nat. Biotechnol.* **37**, 1493–1501 (2019).
29. Rollins, M. F. et al. Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc. Natl Acad. Sci. USA* **114**, E5113–E5121 (2017).
30. Westra, E. R. et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46**, 595–605 (2012).
31. Wintgens, J. P., Rossner, M. J. & Wehr, M. C. Characterizing dynamic protein-protein interactions using the genetically encoded split biosensor assay technique split TEV. *Methods Mol. Biol.* **1596**, 219–238 (2017).
32. Stefan, E. et al. Quantification of dynamic protein complexes using *Renilla* luciferase fragment complementation applied to protein kinase A activities in vivo. *Proc. Natl Acad. Sci. USA* **104**, 16916–16921 (2007).
33. Cass, S. D. et al. The role of Cas8 in type I CRISPR interference. *Biosci. Rep.* **35**, e00197 (2015).
34. Makarova, K. S., Zhang, F. & Koonin, E. V. SnapShot: Class 1 CRISPR-Cas Systems. *Cell* **168**, 946–946 e941 (2017).
35. Fagerlund, R. D. et al. Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proc. Natl Acad. Sci. USA* **114**, E5122–E5128 (2017).
36. Rollins, M. F. et al. Structure reveals a mechanism of CRISPR-RNA-guided nuclease recruitment and anti-CRISPR viral mimicry. *Mol. Cell* **74**, 132–142 e135 (2019).
37. Sauer, R. T. & Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80**, 587–612 (2011).
38. Kapust, R. B. et al. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**, 993–1000 (2001).
39. Xia, Y. et al. T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Res.* **47**, e15 (2019).
40. Thomason, L. C., Costantino, N. & Court, D. L. Examining a DNA replication requirement for bacteriophage λ red- and rac prophage recET-promoted recombination in *Escherichia coli*. *mBio* **7**, e01443-16 (2016).
41. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
42. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
43. Chen, R. & Weng, Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins* **47**, 281–294 (2002).
44. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. *circIzize* Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

Acknowledgements

This research was supported by the National Key Technology R & D Program of China (2018YFA0901100).

Author contributions

L. Y., G. Z., and S. W. designed the research; S. W., F. Z., M. M., T. W., and Y. Y. performed the experiments; S. W., S. Y., and L. Y. performed data analysis and the figures drafting; S. W., L. Y., and G. Z. wrote and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02374-w>.

Correspondence and requests for materials should be addressed to G.Z. or L.Y.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Audrone Lapinaite and Anam Akhtar.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021