Original Article
Expression of RecA and cell-penetrating peptide (CPP) fusion protein in bacteria and in mammalian cells

Xiubao Chang, Yuexian Hou

Department of Biochemistry & Molecular Biology, College of Medicine, Mayo Clinic in Arizona, Scottsdale, AZ, USA

Abstract: Genome editing is a powerful tool to modify a specific gene and to correct a disease-causing mutation. Recently developed new techniques, such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALEN) and clustered regularly interspaced short palindromic repeats/Cas9 (CRISPR/Cas9), significantly facilitate the progression in this field. However, mutations associated with the double strand DNA breaks (DSBs) introduced by these systems hampered their direct usage in clinic. In order to prevent the mutations caused by DSBs, we have designed a novel mean to induce homology-directed recombination (HDR) without DSBs, i.e., the fusion protein of RecA with cell-penetrating peptide (CPP). The involvement of RecA in these fusion proteins will play important roles in formation of the nucleoprotein filament with single strand DNA (ssDNA) *in vitro* and promoting HDR *in vivo*; whereas the involvement of CPP in these fusion proteins will mainly play a role in facilitating cellular intake/uptake of the nucleoprotein filaments. Our results indicated that certain amount of the fusion proteins expressed in bacteria is in soluble fraction, whereas majority of the fusion proteins expressed in baby hamster kidney (BHK) cells is in soluble fraction. Interestingly, expression of these fusion proteins in bacteria completely blocked cell growth, whereas expression of them in BHK cells significantly inhibited cell growth, implying that these fusion proteins may bind to ssDNA regions, such as ssDNA regions in DNA replication forks, and inhibit cell growth. These results suggest that we have functional RecA.CPP fusion proteins ready to test our novel idea of inducing HDR without DSB.

Keywords: RecA, cell-penetrating-peptide (CPP), RecA.CPP fusion protein, single strand DNA (ssDNA), double strand DNA breaks (DSBs), homology-directed recombination (HDR)

Introduction

Many disorders, such as color blindness [1-4], cystic fibrosis (CF) [5, 6], haemochromatosis [7, 8], haemophilia [9-15], phenylketonuria [16, 17], polycystic kidney disease [18, 19], sickle-cell disease [http://ghr.nim.nih.gov/condition/sickle-cell-disease], duchenne muscular dystrophy [20], and some of the cancers are caused by small deletions/insertions as-well-as simple point mutations. These diseases, as well as others, can potentially be corrected by homology-directed recombination (HDR).

HDR is a complex processing of orchestrated reactions involving multiple factors. In this complex processing, presynaptic single strand DNA (ssDNA) invasion plays a crucial role for initiation of the HDR. The greatest challenge in HDR-mediated gene correction is the creation of recombinogenic DNA ends near the mutation site. Development of the new techniques, such as zinc-finger nucleases (ZFNs) [21-23], transcription activator-like effector nucleases (TALEN) [24-26] or clustered regularly interspaced short palindromic repeats/Cas9 (CRISPR/Cas9) [27-34], provide means to cut the DNA near the mutation sites. Unfortunately, non-homologous end-joining (NHEJ), albeit without ensuring restoration of the DNA sequence around the break site, plays a dominant role over HDR for any double stranded DNA (dsDNA) break (DSB) repair in mammalian cells [35, 36]. In addition, the modifications at the break site, including a few nucleotides insertion [37] and/or deletion [38], may cause deleterious mutations, suggesting that safety is another very important issue in ZFNs, TALENs and CRISPR/Cas9 mediated gene correction. In fact, we have found that the frequency of mutations introduced by guideRNA complementary to the target DNA is significantly higher than the gene-correction

Table 1. List of Oligonucleotides

#	Name	Sequence
1	NutRecAfwasu	GCCCGGGACCATGGCTATTGATGAGAATAAAC
2	ET32RecAfwasu	CTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCTATTGAT GAGAATAAAC
3	RecA324rvasu	CAATTTCTTGGCATAGATTGG
4	RecA763fwasu	CCATTCAAACAAGCTGAATTC
5	RecAlinkGfpfwasu	GAAACCAACGAAGAATTTAGTGGCCTACGATCGCGAGCAGCTGCGAACACGATG AGTATTCAACATTTC
6	CDHGFP6658rv	CGGGATAATACCGCGCCAC
7	RecAlinkGfprvasu	GAAATGTTGAATACTCATCGTGTTCGCAGCTGCTCGCGATCGTAGGCCACTAAA TTCTTCGTTGGTTTC
8	Gfp6302rv	GCTTCCCGGCAACAATTAATAG
9	Gfp6019fw	GAGTAAACTTGGTCTGACAG
10	1st.CPPrv	GTGAAGTTGACATCCAAAAAGGATGTTTTCTCGTGCTGCAGCCCAATGCTTAAT CAGTGA
11	2nd.CPPrv	CTTCTTCCTGCCGTAATGGTGATGGTGATGGTGATGGTGATGGTGAAGTTGACA TCCAAA
12	3rd.CPPrv	GCGGCCGCCTATCTTCGTCGCTGTCTCCGCTTCTTCCTGCCGTAATG
13	ET32RecAfw1step	GGTGGCGGCCGCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT ATGGCTATTGATGAGAATAAAC
14	ET32RecArv1step	GTTTATTCTCATCAATAGCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATT ATTTCTAGAGCGGCCGCCACC
15	rmgfpbamh1fw	CGAGCAGCTGCGAACACGGGATCCGCTGCAGCACGAGAAAC
16	rmgfpbamh1rv	GTTTTCTCGTGCTGCAGCGGATCCCGTGTTCGCAGCTGCTCG

mediated by HDR [39]. In addition, the random dsDNA break insertions, such as CRISPR/Cas9 DNA or donor DNA insertion into chromosomes, and/or off-target modifications may also cause mutations that affect normal cell functions. Furthermore, it has been reported that unexpected mutations occurred after CRISPR/Cas9-mediated genome editing *in vivo* [40], suggesting that safety is a very important issue in DSB mediated gene correction. Thus, a more safe technology is critically needed in the design of strategies to correct disease-causing mutations.

In order to prevent the mutations caused by DSB, we have designed a novel mean to induce HDR without DSB, i.e., the fusion protein of RecA with cell-penetrating-peptide (CPP). Involvement of RecA in this fusion protein will play important roles in: 1) formation of the nucleoprotein filament with ssDNA *in vitro*; 2) protecting the ssDNAs wrapped within the nucleoprotein filaments from nucleases-mediated degradation when they are introduced into cells; 3) holding ssDNA and dsDNA molecules together; 4) searching for sequence similarity along the dsDNA; 5) catalyzing a DNA synapsis reaction between a DNA double helix and a complementary region of ssDNA; and 6) promoting HDR [41-43]. Involvement of CPP in this fusion protein will mainly play a role in facilitating cellular intake/uptake of the nucleoprotein filaments [44-47]. Interestingly, expression of these fusion proteins in bacteria completely blocked cell growth, whereas expression of them in mammalian cells significantly inhibited cell growth, implying that these fusion proteins may be functional in binding to ssDNA regions, such as ssDNA regions in DNA replication forks, and block the DNA replication.

Materials and methods

Materials

Most of the chemicals were purchased from Sigma; DMEM/F-12 medium and fetal bovine serum were derived from Thermo Scientific; Restriction endonucleases, from New England Biolabs; QuikChange site-directed mutagenesis kit, from Stratagene; Anti-mouse Ig conjugated with horseradish peroxidase, from Amersham Biosciences; Chemiluminescent substrates for western blotting, from Pierce; RecA DNA (pDONR221.RecA), from DNASU.

RecA.CPP fusion gene construction

In order to express the RecA.CPP fusion protein in mammalian cells, the 5' part of the RecA DNA (pDONR221.RecA was used as template) was amplified by using the primers NutRecAfwasu and RecA324rvasu (**Table 1**); the fusion part between RecA and green fluorescent protein (GFP) was performed by two steps PCR, i.e., the 1st piece (pDONR221.RecA was used as template) was amplified by using RecA763fwasu and RecAlinkgfprvasu (**Table 1**), whereas the 2nd part (pCDH-CMV-MCS-EF1-copGFP was

used as template) was amplified by using RecAlinkgfpfwasu and CDHGFP6658rv (**Table 1**); upon amplification of these two pieces DNA, they were used as templates to put them together by using RecA763fwasu and CDH-GFP6658rv (**Table 1**) as primers; the 3' part of the fusion gene was amplified by three steps, i.e., the 1st piece (pCDH-CMV-MCS-EF1-copGFP was used as template) was amplified by using Gfp6302rv and 1st.CPPrv (**Table 1**) as primers; the 2nd part (the 1st piece of the PCR product was used as template) was amplified by using Gfp6302rv and 2nd.CPPrv (**Table 1**) as primers; whereas the 3rd part (the 2nd part of the PCR product was used as template) was amplified by using Gfp6302rv and 3rd.CPPrv (**Table 1**) as primers. All these pieces of PCR products were cloned into pBluescript and sequenced completely to make sure that there is no mutation occurred in the clones. Two bigger pieces, i.e., the N-terminal half (cloned by combining the XmaI-DraIII fragment from the 1st PCR clone, the DraIII-AseI fragment from pDONR221.RecA and the AseI-HindIII fragment from the RecA. GFP fusion clone) and C-terminal half (cloned by combining the HindIII-ApaL1 fragment from the RecA.GFP fusion clone, the ApaL1-BglI fragment from pCDH-CMV-MCS-EF1-copGFP and the Bgl1-HindIII fragment from the 3rd part of the clone), were cloned into pBluescript and sequenced completely. The N-terminal half and C-terminal half clones were used to make full length fusion gene in pNUT vector [48]. In order to make a shorter version of the fusion protein, the two primers, rmgfpbamh1fw and rmgfp-bamh1rv (**Table 1**), were used to delete the GFP gene from the full length fusion gene by employing the QuikChange Site-directed Mutagenesis kit (Stratagene). The longer version of the fusion gene (named as pNUT.RecA.GFP.CPP) and the shorter version of the fusion gene (named as pNUT.RecA.CPP) were sequenced completely to make sure that there is no mutation occurred in the final clones.

In order to express the RecA.CPP fusion proteins in bacteria, the two primers, ET32RecA-fw1step and ET32RecArv1step (**Table 1**), were used to modify the 5' part of the N-terminal half clone by employing the QuikChange Site-directed Mutagenesis kit. The modified N-terminal half clone and the original C-terminal half clone were used to make full length fusion gene in pET32a expression vector. In order to make shorter version of the fusion protein, the

two primers, rmgfpbamh1fw and rmgfpbam-h1rv (**Table 1**), were used to delete the GFP gene from the full length fusion gene. The longer version of the fusion gene (named as pET32a.RecA.GFP.CPP) and the shorter version of the fusion gene (named as pET32a.RecA. CPP) were sequenced completely to make sure that there is no mutation occurred in the final clones.

Cell culture and transfection

Baby hamster kidney (BHK) cells were grown in DMEM/F-12 medium containing 5% fetal bovine serum at 37°C in 5% CO_2. Subconfluent cells were transfected with plasmid DNAs containing either longer version of the fusion gene (pNUT.RecA.GFP.CPP) or shorter version of the fusion gene (pNUT.RecA.CPP) in the presence of 20 mM HEPES (pH 7.05), 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4, 6 mM dextrose and 125 mM $CaCl_2$ [49]. Whole mixture of the methotrexate-resistant cells was used to determine the expression of the fusion proteins with our multidrug resistance-associated protein 1 (MRP1) monoclonal antibody (mAb) 42.4 [50].

Expression of the RecA.CPP fusion proteins in prokaryotic DL21 cells

The DL21 competent cells were transformed with either pET32a.RecA.GFP.CPP or pET32a. RecA.CPP. The freshly received ampicillin-resistant colonies were used to inoculate 1 mL of 50% Luria-Bertani Broth (LB) and 50% super LB (with 100 µg/mL ampicillin) and cells were grown at 37°C for~6 hours. 10-100 µL (depending on the cell density) of these bacteria were used to inoculate 100 mL of 50% LB and 50% super LB (with 100 µg/mL ampicillin) and the cells were grown overnight at 16°C until the OD600 reaching 0.6-1.0. After adjusting temperature to 4°C, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to 1 mM (final concentration) and cells were grown at this temperature for 16 hours. The cells were harvested by centrifugation at 5,000 × g for 5 minutes at 4°C and the pellets and supernatants were used to determine the expression of these fusion proteins.

Identification of RecA.CPP fusion proteins

Western blot was performed according to the routine protocol. For RecA.CPP fusion proteins expressed in BL21 cells, the following four sam-
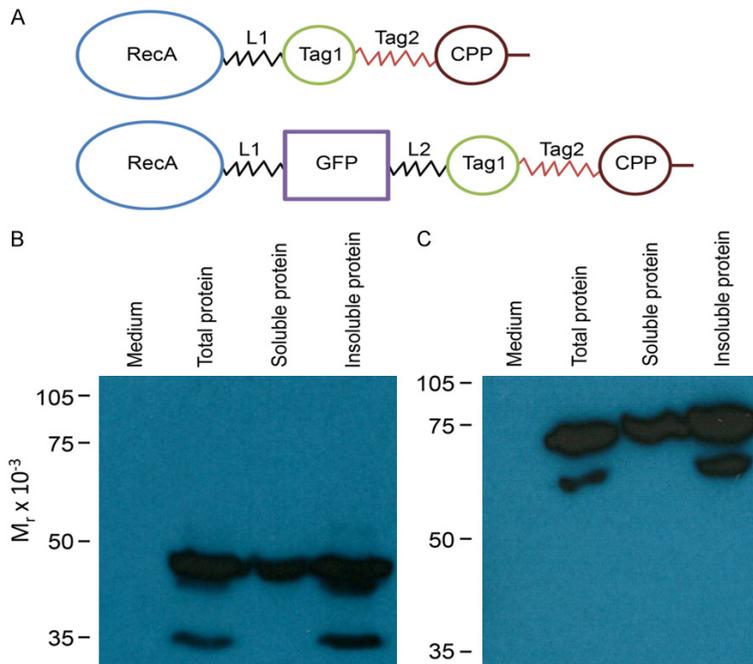
Figure 1. Expression of RecA.CPP fusion proteins in bacteria. A: Design of the RecA.CPP fusion proteins. Two versions of the RecA.CPP fusion proteins, i.e., shorter version (RecA.CPP) and longer version (RecA.GFP.CPP), were designed. The shorter version contains: 1) RecA; 2) L1, a linker that has been used in the past [51, 52]; 3) Tag1, the epitope of our MRP1 monoclonal antibody 42.4 [50]; 4) Tag2, a ten histidine tag; 5) CPP, a transactivator of transcription (TAT) peptide [46, 47, 53-55]. The longer version contains: 1) RecA; 2) L1; 3) GFP, green fluorescent protein; 4) L2, two-alanine short linker; 5) Tag1; 6) Tag2; and 7) CPP. B: A representative western blot (100 μg protein per lane), probed with MRP1 mAb 42.4, showed that certain amount of the shorter RecA.CPP fusion protein expressed in bacteria is in soluble fraction. C: A representative western blot (100 μg protein per lane) showed that certain amount of the longer RecA.GFP.CPP fusion protein expressed in bacteria is also in soluble fraction.

20 bursts to break the DNA. The soluble fraction was collected after centrifugation at 14,000 RPM for 10 minutes); 4) proteins in insoluble fraction (the pellets derived from previous step were dissolved in 1 × nickel bead binding buffer containing 10% glycerol, 1 × protease inhibitors and 2% SDS and then sonicated for 20 bursts to break the DNA).

For RecA.CPP fusion proteins expressed in BHK cells, the following three samples were prepared: 1) Cells lysed with SDS and sonication [Cells were lysed with phosphate buffered saline (PBS) containing 1 × protease inhibitors and 2% SDS and then sonicated for 20 bursts to break the DNA]; 2) Cells lysed with sonication (Cells re-suspended in PBS containing 1 × protease inhibitors were sonicated for 20 bursts to break the DNA); 3) Cells lysed with NP40 buffer [Cells were lysed with NP40 cell lysis buffer (0.1% NP40, 150 mM NaCl, 50 mM Tris, 10 mM Sodium Molybdate, pH 7.6) containing 1 × protease inhibitors by shaking the plates in cold room for 30 minutes. The supernatants were collected after centrifugation at 14,000 RPM].

ples were prepared: 1) the proteins in medium (the proteins in medium were precipitated with trichloroacetic acid and the pellets were dissolved in 1 × sample buffer containing 1 × protease inhibitor cocktail, i.e., Aprotonin, 2 μg/mL; Benzamide, 121 μg/mL; E64, 3.5 μg/mL; Leupeptin, 1 μg/mL; and Pefabloc, 50 μg/mL); 2) total proteins in bacteria [Cell pellets were re-suspended in 1 × nickel bead binding buffer (20 mM Tris/HCl, pH7.9; 500 mM NaCl) containing 10% glycerol, 1 × protease inhibitors and 20,000 units/mL of lysozyme, incubated at 37°C for 15 minutes, added sodium dodecyl sulfate (SDS) to 2% (final concentration) and then sonicated for 20 bursts to break the DNA]; 3) proteins in soluble fraction (Cell pellets were re-suspended in 1 × nickel bead binding buffer containing 10% glycerol, 1 × protease inhibitors and 20,000 units/mL of lysozyme, incubated at 37°C for 15 minutes, and then sonicated for

Samples were subjected to SDS-PAGE, followed by transferring the proteins to nitrocellulose membranes, probed with our MRP1 primary antibody 42.4 [50] overnight at 4°C, washed with PBS containing 0.1% Tween-20 and then incubated with anti-mouse Ig conjugated with horse radish peroxidase. Chemiluminescent film detection was performed according to the manufacturer's recommendations (Pierce).

Statistical analysis

The results in **Figure 4** were presented as means ± SD from the triplicate experiments. The two-tailed *P* values were calculated based on the unpaired t test from GraphPad Software
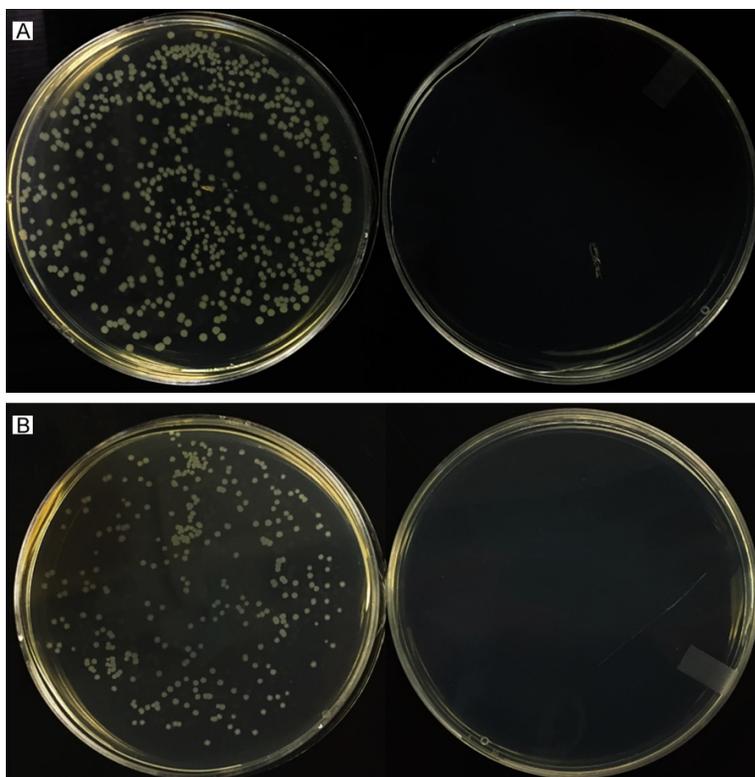
Figure 2. Bacterial growth is completely inhibited by the addition of IPTG at 37°C. A: Upon transformation of the DL21 competent cells with the shorter version of the RecA.CPP fusion construct in pET32a vector, the cells were plated out on plates with 100 μg/mL ampicillin (the plate on the left) or with 100 μg/mL ampicillin and 0.25 mM IPTG (the plate on the right). B: Upon transformation of the DL21 competent cells with the longer version of the RecA.GFP.CPP fusion construct in pET32a vector, the cells were plated out on plates with 100 μg/mL ampicillin (the plate on the left) or with 100 μg/mL ampicillin and 0.25 mM IPTG (the plate on the right).
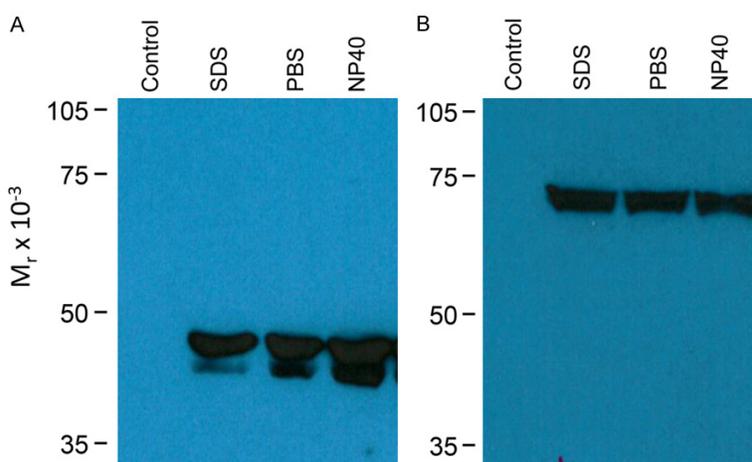

Figure 3. Expression of RecA.CPP fusion proteins in BHK cells. A: A representative western blot (100 μg protein per lane) showed that majority of the shorter RecA.CPP fusion protein expressed in BHK cells are in soluble fraction. B: A representative western blot (100 μg protein per lane) showed that majority of the longer RecA.GFP.CPP fusion protein expressed in BHK cells are also in soluble fraction.

Quick Calcs. By conventional criteria, if *P* value is less than 0.05, the difference between two samples is considered to be statistically significant.

Results

Expression of RecA.CPP fusion proteins in bacteria

Two versions of the RecA.CPP fusion proteins, i.e., shorter version (RecA.CPP) and longer version (RecA.GFP.CPP), were designed (**Figure 1A**). RecA.CPP contains: 1) RecA; 2) L1, a linker that has been used in the past [51, 52]; 3) Tag1, the epitope of our MRP1 mAb 42.4 [50]; 4) Tag2, a ten histidine residue tag; 5) CPP, a cell-penetrating-peptide, i.e., transactivator of transcription (TAT) peptide [46, 47, 53-55]. The longer version, i.e., RecA.GFP.CPP, contains: 1) RecA; 2) L1; 3) GFP, green fluorescent protein; 4) L2, a two-alanine residue short linker; 5) Tag1; 6) Tag2; and 7) CPP.

The results in **Figure 1B** indicated that RecA.CPP fusion protein is clearly expressed in DL21 cells. The protein expressed in DL21 cells is not leaked out to the medium and it also clearly indicated that certain amount the fusion protein is in soluble fraction. The expression of the longer version, i.e., RecA.GFP.CPP, in DL21 cells is similar to the shorter version (**Figure 1C**).

Expression of the fusion proteins completely blocked DL21 cell growth

In order to test whether the expression of the fusion proteins has effect on cell growth or not, the DL21 competent cells transformed with either pET32a.RecA.CPP or pET32a.
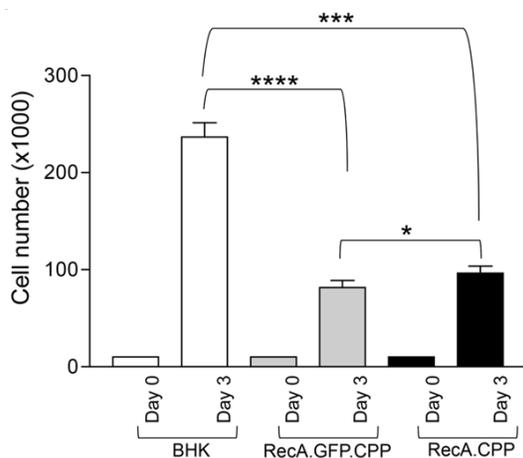
Figure 4. The RecA.CPP fusion protein expression in BHK cells significantly inhibited cell growth. 10,000 cells were plated out on day 0 and counted after 3 days incubation at 37°C. The numbers of cells, after 3 days incubation, are: 236,667 ± 25,403 (BHK); 81,500 ± 12,817 (RecA.GFP.CPP); and 96,300 ± 12,817 (RecA.CPP). *indicates that the *P* value is 0.2302; ***, 0.0010; ****, 0.0007.

RecA.GFP.CPP were plated out on the plates containing either 100 μg/mL ampicillin or 100 μg/mL ampicillin and 0.25 mM IPTG. Interestingly, regardless whether the shorter version or the longer version of the fusion constructs were used, the cells plated out on the plates containing only 100 μg/mL ampicillin grow very well, whereas the cells plated out on the plates containing 100 μg/mL ampicillin and 0.25 mM IPTG did not form visible colonies (**Figure 2A** and **2B**), implying that IPTG induction of the fusion proteins significantly inhibited prokaryotic cell growth.

Expression of RecA.CPP fusion proteins in BHK cells

In order to express these fusion proteins in mammalian cells, the two fusion genes diagramed in **Figure 1A** were inserted into a mammalian expression vector, i.e., pNUT [48]. Upon transformation of BHK cells with these two constructs, i.e., pNUT.RecA.CPP and pNUT. RecA.GFP.CPP, the methotrexate resistant cells were used to determine the expression of these fusion proteins. The results in **Figure 3A** clearly indicated that RecA.CPP fusion protein is expressed in BHK cells. In addition, the amount of the fusion protein in cells lysed with SDS is similar to the cells lysed with NP40 or lysed in PBS, suggesting that majority of the fusion protein expressed in BHK cells is in soluble fraction.

The expression of the longer version, i.e., RecA. GFP.CPP, in BHK cells is similar to the shorter version (**Figure 3B**).

Expression of the fusion proteins significantly inhibited BHK cell growth

In order to test whether the expression of the fusion proteins has effect on mammalian cell growth or not, 10,000 BHK cells expressing either RecA.CPP or RecA.GFP.CPP were plated out on day 0 and counted on day 3. Interestingly, the number of BHK cells expressing RecA.CPP is similar to the cells expressing RecA.GFP.CPP, whereas the number of parental BHK cells is significantly higher than either cells expressing RecA.CPP or RecA.GFP.CPP (**Figure 4**), suggesting that expression of these fusion proteins significantly inhibited mammalian cell growth.

Discussion

The development of customizable sequence-specific nucleases, such as ZFNs, TALENs and CRISPR/Cas9 systems, provides an opportunity to generate recombinogenic DNA ends near the interested sites and has revolutionized genetics by greatly enhancing the efficacy of genome editing. These systems are based on the generation of DSBs near the interested sites. We have found that any DSBs, including matched DNA ends and mismatched DNA ends, can be efficiently ligated, via NHEJ, in eukaryotic cells [56] and the modifications at the break sites [37] could cause deleterious mutations. Thus, despite the versatilities of the ZFNs, TALENs and CRISPR/Cas9 systems, the DSBs generated by these nucleases produced unpredictable modifications, such as indels introduced at the DSB sites [57-62]. Although the mutations associated with the DSBs introduced by these nucleases provide powerful tools to modify genes at the expired site, the mutations introduced by these systems hampered their direct usage in correcting disease-causing mutations in patients. In order to prevent DSBs, a novel technology, i.e., base editor (BE), was developed [63, 64]. This system is based on the inactivated CRISPR/Cas9-mediated programmable conversion of target C:G base pairs (for example) to T:A (mediated by a fusion protein of inactivated CRISPR/Cas9 and cytidine deaminase) without introducing DSBs, meaning that this system will make a point mutation to either mutate a gene or correct a point mutation, but will not make either a dele-

tion or an insertion. Thus, this system may be unable to correct either a disease-causing insertion mutation or a disease-causing deletion mutation, such as the CF-causing deletion of three nucleotides coding for phenylalanine in cystic fibrosis transmembrane conductance regulator gene [5]. Interestingly, we have found that the DSB made by CRISPR/Cas9 in the target site harboring three nucleotide deletion mutation in MRP1 induced HDR and indels, whereas the DSB made by CRISPR/Cas9 in the donor DNA induced only HDR at the three nucleotide deletion target site [39]. Based on these results, we speculated that if ssDNAs complementary to the target sites are introduced into the cell, these ssDNAs may induce HDR without causing DSBs, thereby preventing mutations associated with DSBs. Thus, we have designed the fusion proteins shown in **Figure 1A** in which the RecA portion will play important roles in formation of the nucleoprotein filament with ssDNA and promoting HDR [41-43], whereas the CPP portion will play a role in facilitating cellular intake/uptake of the nucleoprotein filaments [44, 45].

Interestingly, when the two fusion proteins were expressed in prokaryotic cells (the IPTG induction was performed after cells having the OD600 reaching 0.6-1.0), only certain fraction of the fusion proteins in prokaryotic cells, regardless whether it is longer version or shorter version, is in soluble fraction (**Figure 1B** and **1C**), whereas majority of the fusion proteins in eukaryotic cells, regardless whether it is longer version or shorter version, is in soluble fraction (**Figure 3A** and **3B**). In addition, without IPTG induction (on the plates with only ampicillin), the transformed DL21 cells grow very well; in contrast, with IPTG induction (on the plates with ampicillin and IPTG, meaning the IPTG induction started at the beginning of incubation), no colonies were found, suggesting that expression of the RecA.CPP fusion protein significantly inhibited prokaryotic cell growth. Furthermore, continues expression of these fusion proteins in eukaryotic cells also significantly inhibited cell growth (**Figure 4**). Why does expression of these RecA.CPP fusion proteins significantly inhibit prokaryotic cells and eukaryotic cells? RecA protein has been studied for decades and its major function is to bind ssDNA (forming nucleoprotein complex with ssDNA). Where and when the genomic DNA will have ssDNA regions? One region is the DNA replica-

tion forks and another is the transcription region. It has been reported that during DNA replication, RecA filaments are assembled on ssDNA gaps and breaks that form when replication forks stall [41] or collapse with RNA polymerases during transcription [65]. In our case, expression of these fusion proteins in either prokaryotic cells or eukaryotic cells significantly inhibited cell growth. These results suggest that: 1) our RecA.CPP fusion proteins have ssDNA binding function; 2) we have functional fusion proteins ready to test our novel idea of inducing HDR without DSB.

In order to test our novel idea of inducing HDR without DSB, we need to get pure functional proteins. The results in **Figure 1B** and **1C** indicated that certain percentage of the fusion proteins is in soluble fraction. We expected that these soluble fusion proteins should be functional, whereas the fusion proteins in insoluble fraction should not. Therefore we can prepare the soluble fraction and purify the fusion proteins from this fraction. The results in **Figure 3A** and **3B** indicated that majority of the fusion proteins in BHK cells is in soluble fraction and we expected that these soluble fusion proteins should be functional. We can also purify the fusion proteins from these eukaryotic cells. However, in considering the amount of the fusion proteins in prokaryotic cells and in eukaryotic cells (Figure S1) and the cost to take care of prokaryotic cells or eukaryotic cells, we prefer to purify the fusion proteins from prokaryotic cells.

Purification of the functional RecA.CPP fusion proteins will provide a powerful tool to test our novel idea of inducing HDR without DSB. Unfortunately, due to limited support, we are unable to do so at the moment. However, this situation does not prevent us from expecting future applications of the fusion proteins. It has been reported that RecA protein forms filaments with various nucleotides [66], meaning that RecA will non-specifically bind the ssDNA and we can design any kind of ssDNA for RecA-mediated formation of nucleoprotein filaments. In addition, a wide variety of CPPs presents in nature and synthetic CPP analogues with more effective protein transduction properties have been generated [44, 67]. Although the molecular mechanism of CPP-mediated cellular entry is not well studied, CPPs can bring any kind of particle into cells, including human cells, ani-

mal cells, plant cells (protoplasts), fungal cells (protoplasts) and bacterial cells (protoplasts). The RecA.CPP fusion proteins designed in our research bring these two properties together so that we can introduce any kind of ssDNA into any kind of cells to: 1) correct gene mutation in *vitro* or *in vivo* (for example, we can deliver the fusion protein-ssDNA nucleoprotein complex to CF patients' airway to correct the three nucleotide deletion mutation coding for phenylalanine at the position of 508 which counts for ~90% CF mutations); 2) mutate a gene to create animal model of genetic disease; 3) modify a gene to generate new species, etc. Thus, the strategy to transduce cells with RecA.CPP-ssDNA nucleoprotein complex will be a powerful tool for genome editing to generate predictable modifications without introducing DSB, thereby preventing DSB associated mutations, meaning that there are numerous exciting applications in research, medical fields, healthcare, pharmaceutical industry, agriculture and food industry, etc.

Acknowledgements

Disclosure of conflict of interest

Mayo Clinic Arizona, on behalf of the 1st author, has filed a provisional patent related to this work. The other author declares that no competing interests exist.

Address correspondence to: Xiubao Chang, Department of Biochemistry & Molecular Biology, College of Medicine, Mayo Clinic in Arizona, Scottsdale, AZ, USA. E-mail: xbchang@mayo.edu

References

[1] Nathans J, Davenport CM, Maumenee IH, Lewis RA, Hejtmancik JF, Litt M, Lovrien E, Weleber R, Bachynski B, Zwas F and et al. Molecular genetics of human blue cone monochromacy. Science 1989; 245: 831-838.

[2] Weitz CJ, Miyake Y, Shinzato K, Montag E, Zrenner E, Went LN and Nathans J. Human tritanopia associated with two amino acid substitutions in the blue-sensitive opsin. Am J Hum Genet 1992; 50: 498-507.

[3] Winderickx J, Sanocki E, Lindsey DT, Teller DY, Motulsky AG and Deeb SS. Defective colour vision associated with a missense mutation in the human green visual pigment gene. Nat Genet 1992; 1: 251-256.

[4] Mackey DA. Three subgroups of patients from the United Kingdom with Leber hereditary optic neuropathy. Eye 1994; 8: 431-436.

[5] Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M and Tsui LC. Identification of the cystic fibrosis gene: genetic analysis. Science 1989; 245: 1073-1080.

[6] Bobadilla JL, Macek M Jr, Fine JP and Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. Hum Mutat 2002; 19: 575-606.

[7] Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, Domingo R Jr, Ellis MC, Fullan A, Hinton LM, Jones NL, Kimmel BE, Kronmal GS, Lauer P, Lee VK, Loeb DB, Mapa FA, McClelland E, Meyer NC, Mintier GA, Moeller N, Moore T, Morikang E, Prass CE, Quintana L, Starnes SM, Schatzman RC, Brunke KJ, Drayna DT, Risch NJ, Bacon BR and Wolff RK. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 1996; 13: 399-408.

[8] Pietrangelo A, Montosi G, Totaro A, Garuti C, Conte D, Cassanelli S, Fraquelli M, Sardini C, Vasta F and Gasparini P. Hereditary hemochromatosis in adults without pathogenic mutations in the hemochromatosis gene. N Engl Med 1999; 341: 725-732.

[9] Gitschier J, Wood WI, Tuddenham EG, Shuman MA, Goralka TM, Chen EY and Lawn RM. Detection and sequence of mutations in the factor VIII gene of haemophiliacs. Nature 1985; 315: 427-430.

[10] Rees DJ, Rizza CR and Brownlee GG. Haemophilia B caused by a point mutation in a donor splice junction of the human factor IX gene. Nature 1985; 316: 643-645.

[11] Bentley AK, Rees DJ, Rizza C and Brownlee GG. Defective propeptide processing of blood clotting factor IX caused by mutation of arginine to glutamine at position -4. Cell 1986; 45: 343-348.

[12] Davis LM, McGraw RA, Ware JL, Roberts HR and Stafford DW. Factor IXAlabama: a point mutation in a clotting protein results in hemophilia B. Blood 1987; 69: 140-143.

[13] Youssoufian H, Kazazian HH Jr, Phillips DG, Aronis S, Tsiftis G, Brown VA and Antonarakis SE. Recurrent mutations in haemophilia a give evidence for CpG mutation hotspots. Nature 1986; 324: 380-382.

[14] Diuguid DL, Rabiet MJ, Furie BC, Liebman HA and Furie B. Molecular basis of hemophilia B: a defective enzyme due to an unprocessed propeptide is caused by a point mutation in the factor IX precursor. Proc Natl Acad Sci U S A 1986; 83: 5803-5807.

[15] Gitschier J, Wood WI, Shuman MA and Lawn RM. Identification of a missense mutation in

the factor VIII gene of a mild hemophiliac. Science 1986; 232: 1415-1416.

[16] DiLella AG, Marvit J, Brayton K and Woo SL. An amino-acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype 2. Nature 1987; 327: 333-336.

[17] Lyonnet S, Caillaud C, Rey F, Berthelon M, Frezal J, Rey J and Munnich A. Molecular genetics of phenylketonuria in Mediterranean countries: a mutation associated with partial phenylalanine hydroxylase deficiency. Am J Hum Genet 1989; 44: 511-517.

[18] Bisceglia M, Galliani CA, Senger C, Stallone C and Sessa A. Renal cystic diseases: a review. Adv Anat Pathol 2006; 13: 26-56.

[19] Audrezet MP, Cornec-Le Gall E, Chen JM, Redon S, Quere I, Creff J, Benech C, Maestri S, Le Meur Y and Ferec C. Autosomal dominant polycystic kidney disease: comprehensive mutation analysis of PKD1 and PKD2 in 700 unrelated patients. Hum Mutat 2012; 33: 1239-1250.

[20] Aartsma-Rus A, Van Deutekom JC, Fokkema IF, Van Ommen GJ and Den Dunnen JT. Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. Muscle Nerve 2006; 34: 135-144.

[21] Chandrasegaran S and Smith J. Chimeric restriction enzymes: what is next? Biol Chem 1999; 380: 841-848.

[22] Huang B, Schaeffer CJ, Li Q and Tsai MD. Splase: a new class IIS zinc-finger restriction endonuclease with specificity for Sp1 binding sites. J Protein Chem 1996; 15: 481-489.

[23] Kim YG, Cha J and Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A 1996; 93: 1156-1160.

[24] Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ and Voytas DF. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 2010; 186: 757-761.

[25] Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP and Yang B. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res 2011; 39: 359-372.

[26] Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD and Rebar EJ. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 2011; 29: 143-148.

[27] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA and Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337: 816-821.

[28] Pennisi E. The CRISPR craze. Science 2013; 341: 833-836.

[29] Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA and Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 2013; 8: 2281-2308.

[30] Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y and Zhang F. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 2013; 154: 1380-1389.

[31] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA and Zhang F. Multiplex genome engineering using CRISPR/Cas systems. Science 2013; 339: 819-823.

[32] Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK and Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 2013; 31: 822-826.

[33] Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L and Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 2013; 31: 833-838.

[34] Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G and Zhang F. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 2013; 31: 827-832.

[35] Smithies O, Gregg RG, Boggs SS, Koralewski MA and Kucherlapati RS. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature 1985; 317: 230-234.

[36] Thomas KR, Folger KR and Capecchi MR. High frequency targeting of genes to specific sites in the mammalian genome. Cell 1986; 44: 419-428.

[37] Roth DB, Chang XB and Wilson JH. Comparison of filler DNA at immune, nonimmune, and oncogenic rearrangements suggests multiple mechanisms of formation. Mol Cell Biol 1989; 9: 3049-3057.

[38] Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Ann Rev Biochem 2010; 79: 181-211.

[39] Xu Q, Hou YX and Chang XB. CRISPR/Cas9-mediated three nucleotide insertion corrects a deletion mutation in MRP1/ABCC1 and restores its proper folding and function. Mol Ther Nucleic Acids 2017; 7: 429-438.

[40] Schaefer KA, Wu WH, Colgan DF, Tsang SH, Bassuk AG and Mahajan VB. Unexpected mu-

tations after CRISPR-Cas9 editing in vivo. Nat Methods 2017; 14: 547-548.

[41] Bell JC and Kowalczykowski SC. RecA: regulation and mechanism of a molecular search engine. Trends Biochem Sci 2016; 41: 491-507.

[42] Cox MM. Motoring along with the bacterial RecA protein. Nat Rev Mol Cell Biol 2007; 8: 127-138.

[43] Cox MM. The bacterial RecA protein as a motor protein. Ann Rev Microbiol 2003; 57: 551-577.

[44] Okuyama M, Laman H, Kingsbury SR, Visintin C, Leo E, Eward KL, Stoeber K, Boshoff C, Williams GH and Selwood DL. Small-molecule mimics of an alpha-helix for efficient transport of proteins into cells. Nat Methods 2007; 4: 153-159.

[45] Wagstaff KM and Jans DA. Protein transduction: cell penetrating peptides and their therapeutic applications. Curr Med Chem 2006; 13: 1371-1387.

[46] Frankel AD and Pabo CO. Cellular uptake of the tat protein from human immunodeficiency virus. Cell 1988; 55: 1189-1193.

[47] Green M and Loewenstein PM. Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. Cell 1988; 55: 1179-1188.

[48] Palmiter RD, Behringer RR, Quaife CJ, Maxwell F, Maxwell IH and Brinster RL. Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene. Cell 1987; 50: 435-443.

[49] Chang XB, Hou YX and Riordan JR. ATPase activity of purified multidrug resistance-associated protein. J Biol Chem 1997; 272: 30962-30968.

[50] Hou Y, Cui L, Riordan JR and Chang XB. Allosteric interactions between the two non-equivalent nucleotide binding domains of multidrug resistance protein MRP1. J Biol Chem 2000; 275: 20280-20287.

[51] Orban TI, Seres L, Ozvegy-Laczka C, Elkind NB, Sarkadi B and Homolya L. Combined localization and real-time functional studies using a GFP-tagged ABCG2 multidrug transporter. Biochem Biophys Res Commun 2008; 367: 667-673.

[52] Hou YX, Li CZ, Palaniyandi K, Magtibay PM, Homolya L, Sarkadi B and Chang XB. Effects of putative catalytic base mutation E211Q on ABCG2-mediated methotrexate transport. Biochemistry 2009; 48: 9122-9131.

[53] Debaisieux S, Rayne F, Yezid H and Beaumelle B. The ins and outs of HIV-1 Tat. Traffic 2012; 13: 355-363.

[54] Schwarze SR, Hruska KA and Dowdy SF. Protein transduction: unrestricted delivery into all cells? Trends Cell Biol 2000; 10: 290-295.

[55] Dietz GP and Bahr M. Delivery of bioactive molecules into the cell: the Trojan horse approach. Mol Cell Neurosci 2004; 27: 85-131.

[56] Chang XB and Wilson JH. Modification of DNA ends can decrease end joining relative to homologous recombination in mammalian cells. Proc Natl Acad Sci U S A 1987; 84: 4959-4963.

[57] Hwang WY, Fu Y, Reyon D, Maeder ML, Kaini P, Sander JD, Joung JK, Peterson RT and Yeh JR. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PLoS One 2013; 8: e68708.

[58] Lo TW, Pickle CS, Lin S, Ralston EJ, Gurling M, Schartner CM, Bian Q, Doudna JA and Meyer BJ. Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. Genetics 2013; 195: 331-348.

[59] Blitz IL, Biesinger J, Xie X and Cho KW. Biallelic genome modification in F(0) Xenopus tropicalis embryos using the CRISPR/Cas system. Genesis 2013; 51: 827-834.

[60] Mianne J, Codner GF, Caulder A, Fell R, Hutchison M, King R, Stewart ME, Wells S and Teboul L. Analysing the outcome of CRISPR-aided genome editing in embryos: screening, genotyping and quality control. Methods 2017; 121-122: 68-76.

[61] Zischewski J, Fischer R and Bortesi L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. Biotechnol Adv 2017; 35: 95-104.

[62] Kosicki M, Rajan SS, Lorenzetti FC, Wandall HH, Narimatsu Y, Metzakopian E and Bennett EP. Dynamics of indel profiles induced by various CRISPR/Cas9 delivery methods. Prog Mol Biol Transl Sci 2017; 152: 49-67.

[63] Komor AC, Kim YB, Packer MS, Zuris JA and Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 2016; 533: 420-424.

[64] Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH and Liu DR. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T: a base editors with higher efficiency and product purity. Sci Adv 2017; 3: eaao4774.

[65] Kogoma T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiol Mol Biol Rev 1997; 61: 212-238.

[66] Heuser J and Griffith J. Visualization of RecA protein and its complexes with DNA by quick-freeze/deep-etch electron microscopy. J Mol Biol 1989; 210: 473-484.

[67] Guidotti G, Brambilla L and Rossi D. Cell-penetrating peptides: from basic research to clinics. Trends Pharmacol Sci 2017; 38: 406-424.
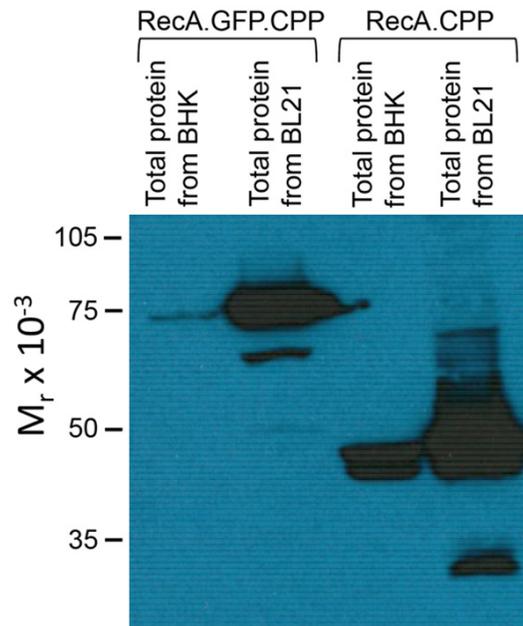
Figure S1. Comparison of the fusion proteins expressed in bacteria and in BHK cells. A representative western blot (100 µg protein per lane), probed with MRP1 mAb 42.4, showed that RecA.GFP.CPP or RecA.CPP expressed in BHK cells is significantly less than in DL21 bacteria cells.