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
Received December 19, 2017; Accepted December 25, 2017; Epub February 8, 2018; Published February 15, 2018

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.nlm.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
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 means 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...
RecA and cell-penetrating-peptide fusion protein

used as template) was amplified by using RecAlinkgfpwasu and CDHGFP6658rv (Table 1); upon amplification of these two pieces DNA, they were used as templates to put them together by using Rec763fwasu 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 Xmal-Dralll fragment from the 1st PCR clone, the DraII-AseI fragment from pDONR221.RecA and the Asel-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-BglII fragment from pCDH-CMV-MCS-EF1-copGFP and the BglI-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 rmgfpbamh1rv (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% CO₂. 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 Na₂HPO₄, 6 mM dextrose and 125 mM CaCl₂ [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-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 x 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-
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].

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.
RecA and cell-penetrating-peptide fusion protein

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.

Quick Calcs. By conventional criteria, if \( P \) value is less than 0.05, the difference between two samples is considered to be statistically significant.
RecA and cell-penetrating-peptide fusion protein

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-
RecA and cell-penetrating-peptide fusion protein

...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 replication 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-
RecA and cell-penetrating-peptide fusion protein

normal 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

The authors would like to thank: 1) Dr. Maoqing Dong for preparation of the figures; 2) Leah W. Moore for participating in manuscript preparation.

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


[40] Schaefer KA, Wu WH, Colgan DF, Tsang SH, Bassuk AG and Mahajan VB. Unexpected mu-
RecA and cell-penetrating-peptide fusion protein

10

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.