Video Article

Identification of Functional Protein Regions Through Chimeric Protein Construction

Juan M. Adrian-Segarra[1], Holger Lörchner[1,2], Thomas Braun[1,2], Jochen Pöling[1,2]

[1]Department of Cardiac Development and Remodelling, Max Planck Institute for Heart and Lung Research

[2]Partner site Rhein-Main, German Centre for Cardiovascular Research (DZHK)

Correspondence to: Jochen Pöling at Jochen.Poeling@mpi-bn.mpg.de

URL: https://www.jove.com/video/58786
DOI: doi:10.3791/58786

Keywords: Chimeric proteins, structural similarity, molecular biology, overlap PCR, structure-function analysis, protein domains, protein-protein interactions, protein binding properties

Date Published: 12/19/2018

Citation: Adrian-Segarra, J.M., Lörchner, H., Braun, T., Pöling, J. Identification of Functional Protein Regions Through Chimeric Protein Construction. *J. Vis. Exp.* (), e58786, doi:10.3791/58786 (2018).

Abstract

The goal of this protocol encompasses the design of chimeric proteins in which distinct regions of a protein are replaced by their corresponding sequences in a structurally similar protein, in order to determine the functional importance of these regions. Such chimeras are generated by means of a nested PCR protocol using overlapping DNA fragments and adequately designed primers, followed by their expression within a mammalian system to ensure native secondary structure and post-translational modifications.

The functional role of a distinct region is then indicated by a loss of activity of the chimera in an appropriate readout assay. In consequence, regions harboring a set of critical amino acids are identified, which can be further screened by complementary techniques (*e.g.* site-directed mutagenesis) to increase molecular resolution. Although limited to cases in which a structurally related protein with differing functions can be found, chimeric proteins have been successfully employed to identify critical binding regions in proteins such as cytokines and cytokine receptors. This method is particularly suitable in cases in which the protein's functional regions are not well defined, and constitutes a valuable first step in directed evolution approaches to narrow down the regions of interest and reduce the screening effort involved.

Video Link

The video component of this article can be found at https://www.jove.com/video/58786/

Introduction

Several types of proteins, including cytokines and growth factors, are grouped in families whose members share similar three-dimensional structures but often exert distinct biological functions[1,2]. This functional diversity is usually the consequence of small differences in amino acid composition within the molecule's active sites[3]. Identification of such sites and functional determinants do not only offer valuable evolutionary insights but also to design more specific agonists and inhibitors[4]. However, the large number of differences in residue composition frequently found between structurally related proteins complicates this task. Even though constructing large libraries containing hundreds of mutants is nowadays feasible, assessing every single residue variation and combinations of them still remains a challenging and time-consuming effort[5].

Techniques assessing the functional importance of large protein regions are thus of value to reduce the number of possible residues to a manageable number[6]. Truncated proteins have been the most used approach to tackle this issue. Accordingly, regions are considered to be functionally relevant if the protein function under study is affected by the deletion of a particular region[7,8,9]. However, a major limitation of this method is that deletions can affect the protein's secondary structure, leading to misfolding, aggregation and the inability to study the intended region. A good example is a truncated version of the cytokine oncostatin M (OSM), in which an internal deletion larger than 7 residues resulted in a misfolded mutant that could not be further studied[10].

The generation of chimeric proteins constitutes an alternative and innovative approach that permits the analysis of larger protein regions. The goal of this method is to exchange regions of interest in a protein by structurally related sequences in another protein, in order to assess the contribution of the replaced sections to specific biological functions. Widely used in the field of signaling receptors to identify functional domains[11,12], chimeric proteins are particularly useful to study protein families with little amino acid identity but conserved secondary structure. Appropriate examples can be found in the class of interleukin-6 (IL-6) type cytokines, such as interleukin-6 and ciliary neurotrophic factor (6% sequence identity)[13] or leukemia inhibitory factor (LIF) and OSM (20% identity)[6], on which the following protocol is based.

Copyright © 2018 Journal of Visualized Experiments

Protocol

1. Chimeric Protein Design

1. Select a suitable protein (donor) to exchange regions with the protein of interest (recipient) The donor protein should be structurally similar, ideally belonging to the same protein family, but lacking the biological activity to be used as readout. If no structurally related proteins are known, potential candidates can be identified using an automated tool such as the Vector Alignment Search Tool (VAST)[14,15]
    1. Access the Protein Data Bank (PDB)[16] European website (https://www.ebi.ac.uk/pdbe/), introduce the name of the protein of interest in the search box on the upper right corner, and click 'Search'. Provided that a crystal structure is available, not down the PDB identifier (PDB ID; *e.g.* 1evs for OSM).
    NOTE: if structural data is not available at the PDB, a homology model of the protein might be generated by a tool such as SWISS-MODEL[17] instead, using available step-by-step protocols[18].
    2. Access the VAST website (https://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml). In case a PDB ID is available, scroll down to the 'Retrieve pre-computed results' section, imput the PDB ID in the 'Show Similar Structures for' box, and click 'Go'. In the following screen, click on 'Original VAST' and then 'Entire Chain' to see a list of PDB IDs for potential structurally similar candidates.
        1. If a PDB ID is unavailable but a homology model was generated, scroll down to the 'Search with a new structure' section and click on the 'VAST search' link. Upload the PDB file of the model by clicking 'Browse' next to 'Submit PDB file', selecting the file and clicking 'Submit'.
        2. After the PDB file is uploaded, click the 'Start' button to start the VAST calculation. Once the calculation is performed, click on 'Entire Chain' under Domains to see the PDB IDs of structurally similar proteins.

    3. Assess the biological functions of interest (*e.g.* receptor activation, enzymatic activity, transcription factor activity) of the top candidates, either experimentally in the readout system of choice or through a literature search. Select a donor protein with divergent function in comparison to the protein of interest.

2. Obtain the protein amino acid sequences of the recipient and donor proteins from the Reference Sequence (RefSeq) database[19] .
    1. Access the gene section in the RefSeq webpage (https://www.ncbi.nlm.nih.gov/gene), type the name of the protein of interest in the search box and click 'Search'. Click on the gene name for the desired species in the resulting list.
    2. Scroll down to the RefSeq section to see all documented isoforms. Click on the sequence identifier for the isoform of interest (starting with NM), scroll down and click on 'CDS' to highlight the protein-coding region of the gene. On the bottom right of the screen, click on 'FASTA' and copy the gene sequence.
    3. Save the DNA sequence using a suitable DNA editing software. When using the freely available ApE[20], open the program, paste the copied sequence in the blank box, select the sequence name and click 'Save'.
    NOTE: Repeat steps 1.2.1. to 1.2.3. for the one or more donor proteins selected.

3. Choose the protein regions to be substituted in the different chimeric constructs.
    1. Divide the protein sequence of interest in distinct structural regions. Ideally, different domains for the protein in question will have been described in the literature. If this is not the case, the existence of distinct conserved structural features (helices, loops) should be evaluated in steps 1.3.1.1. to 1.3.1.4.
        1. Download the structural data of the protein of interest from the PDB website (see step 1.1.1.). Access the PDB page for the protein, and download the PDB file by clicking 'download' at the right side of the screen.
        2. Open the PDB file in a molecular visualization system like PyMOL (https://pymol.org/). In PyMOL, display the nucleotide sequence (by clicking Display > Sequence On), hide the default structural data (by clicking the H next to the PDB ID, and selecting 'everything') and select the 'cartoon' view to clearly visualize the protein's structural features (clicking the S next to the PDB ID, and selecting 'cartoon').
        3. Click on the nucleotide sequence at the top of the screen to highlight different parts of the molecule, noting down the amino acids corresponding to each distinctive structural feature.

    2. Annotate the distinct structural regions on the DNA sequence in ApE. To do so, open the DNA sequence from step 1.2.3., select the nucleotides coding for the amino acids in a region, right-click on the selection and select 'New Feature' to give it a name and a color. Repeat the process for each structural region identified in the previous step.
    NOTE: The nucleotide selection can be double-checked by clicking ORFs > Translate, then clicking OK, to ensure that they code for the correct amino acid sequence.
    3. Align the residue sequences of the two proteins employing a protein alignment tool (*e.g.* Clustal Omega[21]).
    NOTE: Since the chimeras are to be produced in a mammalian expression system, these sequences should include the proteins' signal peptides.
        1. Obtain the full amino acid sequences of donor and receptor proteins in ApE, by opening the DNA sequences from step 1.2.3., selecting them and clicking ORFs > Translate.
        2. Access the Clustal Omega webpage (https://www.ebi.ac.uk/Tools/msa/clustalo/) and imput the amino acid sequences of the two proteins, then scroll down and click 'Submit'. Each sequence should be proceded by a text line with '>ProteinName' to be properly identified..
        3. Retrieve the alignment file by clicking the 'Download alignment file' tab and save it. This file can be opened by any text editing program.
        4. Using the alignment file as a reference, annotate the corresponding structural regions of the donor protein in their DNA sequence (see step 1.3.2.).

    4. Decide which protein regions to exchange in the chimeric proteins and design the appropriate nucleotide sequences for the chimeras.

NOTE: In the absence of detailed information regarding functional importance of the different regions, it is suggested to select large substitutions such as whole loops or helices to evaluate which of them have an impact on protein function. This first exploratory experiment can then be followed by a second round of chimeric protein design, focused on smaller substitutions within the relevant regions.

1. Create a copy of the annotated DNA sequence of the receptor protein from step 1.3.2. and rename it as a chimeric protein. Open the renamed DNA sequence in ApE, select and delete the nucleotide sequence coding for the region to be exchanged, and replace it by the corresponding region in the donor protein (copied from the annotated sequence created in step 1.3.3.), then save the changes.
   NOTE: Create a new copy and repeat this step for each different chimera designed.

2. Preparation for Molecular Cloning

1. Select a plasmid vector suitable for the expression system of choice. For mammalian expression, a high-expression vector like pCAGGS[22] or the pcDNA vector series are recommended.
   1. For the restriction enzyme-based cloning demonstrated in this protocol, ensure that the unique restriction sites present in the multiple cloning site (MCS) of the vector are compatible with the protein of interest. To do so, open the DNA sequence of the chimeric constructs with ApE, click 'Enzymes > Enzyme selector' and verify that at least two of the restriction sites in the MCS are absent in the sequence (displaying a zero next to their name).

2. Design the terminal primers using a DNA editor such as ApE.
   1. Create a new DNA file (File > New) and initiate the N-terminal primer with a leader sequence (3-9 extra base pairs, *e.g.* AAAGGGAAA), followed by the first restriction site selected in the vector's MCS (6-8 base pairs, *e.g.* TTAATTAA for PacI), an optional spacer (*e.g.* GCTAGCGCATCGCCACC in the pCAGGS vector used in the example) and the initial 18-27 base pairs of the gene of interest (*e.g.* ATGGGGGTACTGCTCACACAGAGGACG for OSM).
   2. In a new DNA file, the C-terminal primer sequence starts with the final 18-27 base pairs of the gene of interest (*e.g.* CTCGAGCACCACCACCACCACCACTGA for a gene with a 6xHistidine C-terminal tag), followed by an optional spacer (*e.g.* TAGCGGCCGC in the pCAGGS vector), the second restriction site chosen (*e.g.* GGCGCGCC for AscI) and a leader sequence (*e.g.* AAAGGGAAA). Highlight the whole sequence, right-click and select 'Reverse-Complement' to obtain the reverse primer.

3. Design primers for each of the border regions in the chimeric constructs.
   1. Open the DNA sequence of the chimera (created in step 1.3.4.1.) and highlight a 30 base pair region in the zone where the original and inserted sequences are in contact, comprising 15 base pairs of each sequence. Copy the region (right-click and select 'Copy') and paste it in a new DNA file; this sequence will be the forward primer.
   2. Make a copy of the forward primer generated in the previous step and rename it as reverse primer. Highlight the primer sequence, right-click and select 'Reverse-Complement' to generate the reverse primer sequence.
   3. Repeat steps 2.3.1. and 2.3.2. for each contact zone in the chimeric DNA sequence. Generally, two sets of forward/reverse primers are required to generate one chimera, unless the replacement occurs at the N-terminal or C-terminal regions.

4. Order the terminal and internal primers from an oligonucleotide synthesis provider.
   NOTES: Using highly purified terminal primers (*e.g.* HPLC-purified) can have a positive impact in the protocol's success rate. Desalted internal primers usually provide good results.
5. Obtain template sequences of the donor and receptor genes.
   NOTE: Employing plasmids containing the open reading frames (ORFs) of these sequences as a template greatly facilitates the procedure and is recommended. Alternatively, complementary DNA (cDNA) generated from a cell line known to express these genes can be used as a template for subsequent steps.

3. Polymerase Chain Reaction (PCR) Amplification of the Individual DNA Fragments Forming the Chimera

1. Prepare an individual PCR reaction mixture for each of the fragments composing the chimeric protein. A typical chimeric protein will require three individual fragments: the N-terminal part, the region to be inserted, and the C-terminal part.
   NOTES: Use a high-fidelity DNA polymerase (*e.g.* Phusion High Fidelity DNA Polymerase) to avoid introducing mutations in the sequence. The PCR reaction can be set up in the evening and run overnight.
   1. Set a 1.5 mL microfuge tube on ice and pipet the different reagents of the PCR mixtures in the order shown in **Table 1**, ensure correct primers and templates are employed for each PCR reaction (see **Table 2**).
   2. Label two thin-walled 0.2 mL PCR tubes for each reaction, and transfer 20 µL of the corresponding PCR mixture in each tube. Transfer the PCR tubes into a PCR thermocycler and initiate the protocol detailed in **Table 3**.
      NOTE: annealing temperature should be at least 5 degrees lower than the melting temperature of the designed primers.

2. While the PCR is running, prepare 100 mL of a 1% agarose gel in Tris-Acetate-EDTA (TAE) buffer. For this purpose, weigh 1 g of agarose, mix with 100 mL of TAE buffer in a glass flask and microwave until the agarose is completely dissolved, swirling the flask every 30-40 seconds. Allow cooling to approximately 50 °C, add 2-3 µL of ethidium bromide (or an equivalent DNA dye) and pour in a gel tray with the desired well combs.
   CAUTION: ethidium bromide is a known mutagen, ensure the use of proper protective equipment.
3. After the PCR reaction is completed, add 4 µL of 6x DNA loading buffer in each tube. Insert the 1% agarose gel in an electrophoresis unit, cover with TAE buffer and carefully load the samples into the gel along with a molecular weight ladder.
   NOTE: at the time of loading, it is preferable to leave empty lanes between the samples to facilitate DNA recovery afterwards.

4. Run the gel at 80-120 V for 20-45 minutes.
   NOTE: bands under 1000 base pairs can usually be electrophoresed in around 20 minutes at 120 V, while larger bands will require longer running times.
5. Turn off the electrophoresis unit, take out the agarose gel and visualize the amplified DNA bands under UV light. Using a razor blade, cut out the individual DNA fragments from the gel, and transfer them to labeled 2 mL microfuge tubes.
   NOTE: Minimize the exposure time to UV light in order to avoid DNA damage.
6. Use a PCR clean-up kit (see **Table of Materials**) to purify the different DNA fragments.
   1. Add 500 µL of the NTI buffer provided by the kit to each tube containing a gel fragment. Transfer to a thermomixer at 50-55 °C and shaking at 1000 rpm until the gel is completely dissolved into the buffer.
   2. Transfer each solution to a labeled kit column, spin in a microcentrifuge (30-60s, 11000xg) and discard the flowthrough. Add 700 µL of the kit's NT3 wash buffer, centrifuge again under the same settings and discard the flowthrough. Centrifuge the columns again for 1-2mins at 11000xg to dry the silica membrane inside the column.
   3. Transfer the column to a labeled 1.5 mL microfuge tube, pipet 30 µL of nuclease-free water into the column, let stand for 1 minute and centrifuge 30-60s at 11000 x g to elute the DNA.

7. Quantify the amount of DNA recovered by measuring the absorbance of the sample at 260 nm and 340 nm in a spectrophotometer (see **Table of Materials**). The DNA concentration is calculated by subtracting the 340 nm reading from the 260 nm figure, then multiplying the result by DNA's extinction coefficient (50 µg/mL)[23].
   NOTES: Additional measurements at 230 nm and 280 nm allow for evaluation of DNA purity: 260/280 and 260/230 ratios above 1.8 are generally regarded as pure for DNA. The protocol can be paused after this step, storing the eluted DNA at 4 °C (short-term) or -20 °C (longer term).

4. PCR Amplification to Generate the Chimeric DNA Sequence

1. Set up 50 µL of a PCR reaction to fuse the separate constituents of the chimera. Follow the same steps detailed in 3.1, employing the N-terminal and C-terminal primers along with 10 ng of each of the DNA fragments obtained in step 3.7.
   NOTE: The PCR reaction can be set in the evening and run overnight.
2. Repeat steps 3.2 to 3.7 to recover and quantify the purified DNA fragment in 30 µL of nuclease-free water. This fragment contains the chimeric DNA sequence, flanked by the restriction sites included in the terminal primers.

5. Insertion of the Chimeric DNA into an Expression Vector

1. Label two 1.5 mL microfuge tubes and pipet the different reagents in the order indicated in **Table 4**, adding 1 µg of the selected expression vector in one tube and 1 µg of the recovered DNA fragment in the other. Incubate for 1-4 h at 37 °C to perform the digestion with the chosen restriction enzymes.
   NOTES: Ensure both restriction enzymes are compatible with the buffer employed. In case they require completely different buffers, perform these steps first with only one of the enzymes and repeat. The time required for the digestion can vary depending on the restriction enzymes selected: refer to the manufacturer's instructions for detailed information.
2. Repeat steps 3.2 to 3.7 to purify and recover the digested DNA fragment and expression vector in 30 µL of nuclease-free water. Quantify the amount of DNA recovered as before.
3. Calculate the amount of insert DNA required for a 3:1 insert/vector molar ratio in the ligation reaction, using the following equation.
   Insert (*ng*) = *Molar ratio* * *Vector* (*ng*) * *Insert size* (*bp*) / *Vector size* (*bp*)
   NOTE: Different insert/vector molar ratios can be tested, although usually a 3:1 ratio is sufficient to obtain adequate results.
4. Set up 20 µL of a ligation reaction in a 1.5 mL microfuge tube with 40 ng of the expression vector the amount of chimeric DNA calculated in step 5.3, buffer, and T4 DNA ligase following the order indicated in **Table 5**, and incubate overnight at 16 °C.
   NOTE: Ligation efficiency is generally increased by performing the reaction overnight at 16 °C, but alternatively the reaction can be incubated for 2 h at room temperature.
5. Transform 5-10 µL of the ligation mixture into chemically competent *Escherichia coli* (*E. coli*) prepared following standard protocols[24] Grow in selection plates and pick single colonies for expansion and plasmid DNA isolation according to established protocols[25].
   NOTE: The *E. coli* XL1-Blue strain was employed for this protocol, but other *E. coli* variants can also be used.
6. Digest 1 µg the isolated plasmids with the appropriate restriction enzymes following the instructions in step 5.1, and assess the presence of a DNA band of a size corresponding to the chimeric sequence by electrophoresis in an agarose gel (see steps 3.2-3.5).
   NOTE: It is recommended that the inserted sequence is verified by means of a DNA sequencing service.
7. Upon successful sequence verification, the plasmids can be produced in larger amounts and employed in a mammalian expression system by following well-established protocols[26,27].

Representative Results

Construction and generation of a chimeric protein (**Figure 1**) will be exemplified with two members of the interleukin-6 cytokine family, OSM and LIF, which were the subject of a recently published study[6]. **Figure 2** shows the three-dimensional structure of these proteins. Both molecules adopt the characteristic secondary structure of class I cytokines, with four helices (termed A to D) packed in a bundle and joined by loops[28]. The aligned amino acid structures of the human proteins can be seen in **Figure 3A**. In this example the BC loop region of OSM was exchanged by the corresponding LIF sequence to create an OSM-LIF chimera with the amino acid sequence as shown in **Figure 3B**.

For this purpose, the DNA sequence of OSM and LIF were obtained and the encoding amino acid region corresponding to the BC loop was identified for both cytokines and replaced (**Figure 4**). A 6-histidine tag was additionally incorporated in the C-terminus to facilitate downstream protein purification. Next, a suitable vector for mammalian expression (pCAGGS) was chosen, and unique restriction sites within its multiple cloning site were selected (*PacI* and *AscI*) after ensuring that they were not present in the chimeric gene sequence (see step 2.1.1.).

Primers were designed as shown in **Table 4.** The N-terminal forward OSM primer included a leading sequence of 9 base pairs, followed by the *PacI* restriction site, a plasmid-specific spacer, and the initial 27 base pairs of OSM. The C-terminal reverse primer incorporated the leading sequence followed by the *AscI* restriction site, a spacer, and the 27 last base pairs of the gene, which in this particular case corresponded to the C-terminal histidine tag. In addition, 30-base pair primers spanning the junction points of the BC loop were required in both forward and reverse orientations.

The first PCR amplification step consisted of three separate reactions. The N-terminal OSM fragment, which required N-terminal OSM forward and BC start reverse primers, used OSM as template. The LIF BC loop was obtained through BC start forward and BC end reverse primers using LIF as template. The C-terminal OSM fragment used BC end forward and C-terminal OSM reverse primers, as well as OSM as template. These three fragments, with expected sizes of 385, 75 and 321 base pairs respectively, can be seen in **Figure 5A** after separation in a 1% agarose gel.

These purified fragments were then used as a template in the second PCR reaction, along with N-terminal OSM forward and C-terminal OSM reverse primers. The result of this amplification, corresponding to the OSM-LIF BC loop gene sequence and is shown in **Figure 5B**. This step was followed by purification, a 4-hour digestion of the gene fragment and the chosen plasmid, gel electrophoresis and purification, overnight ligation at 16 °C, and transformation into *E. coli* XL1-Blue. Individual plasmids were isolated and screened by restriction enzyme digestion for proper insertion of the DNA fragment (**Figure 6**). Finally, positive hits were sent for sequencing to verify that the sequence corresponded to the intended OSM-LIF BC loop chimera before proceeding to protein expression, purification and testing in functional assays[6].

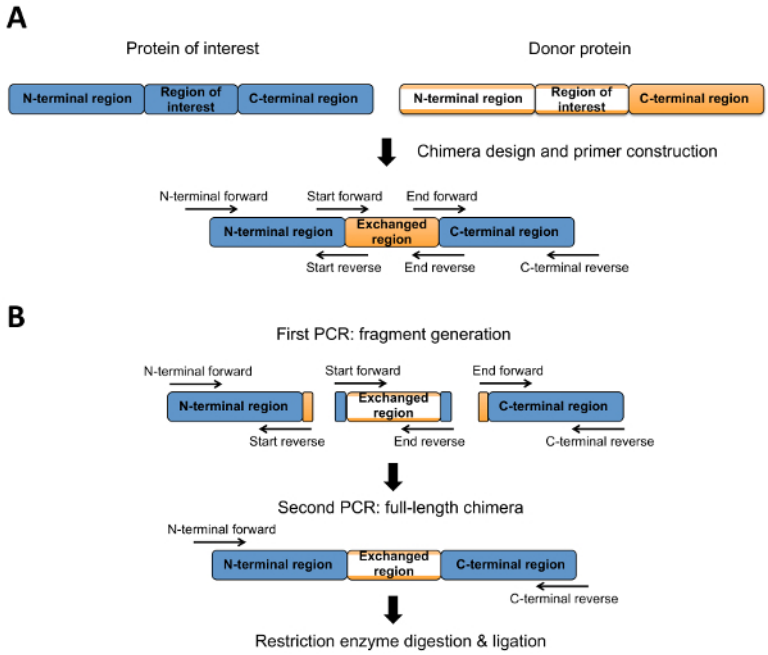

Figure 1: Schematic representation of chimeric protein generation. (A) Chimeric design process: after selection of the regions to be exchanged, the sequence of the desired chimera and the necessary primers are constructed by means of DNA editing software. **(B)** The key steps in the generation of chimeric proteins are depicted. Two steps of PCR amplification produce a chimeric gene sequence, which is then digested with the appropriate restriction enzymes and ligated into an expression vector. Please click here to view a larger version of this figure.

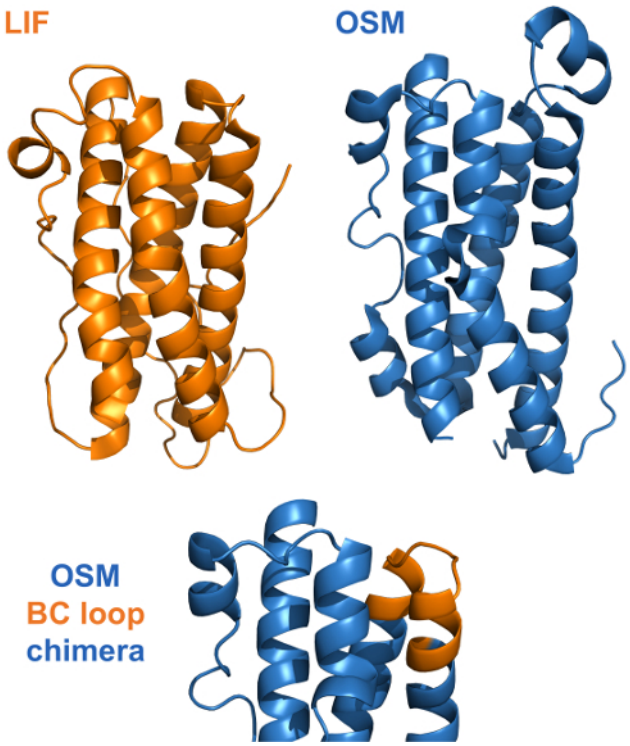

Figure 2: Structural similarities between OSM and LIF. Representation of the crystal structures of OSM[29] (PDB: 1EVS) and LIF[30] (PDB: 2Q7N), along with an approximate representation of the designed chimera. These cytokines adopt a four-helical bundle conformation joined by loops. This research was originally published in the Journal of Biological Chemistry. Adrian-Segarra, J. M., Schindler, N., Gajawada, P., Lörchner, H., Braun, T. & Pöling, J. The AB loop and D-helix in binding site III of human Oncostatin M (OSM) are required for OSM receptor activation. J. Biol Chem 2018; 18:7017-7029. © the Authors[6]. Please click here to view a larger version of this figure.

A

```
LIF      SPLPITPVNATCAIRHPCHNN---LMNQIRSQLAQLNGSANALFI
OSM      ------------AAIGSCSKEYRVLLGQLQKQTD-LMQDTSRLLD
                     *    * ::   *: *::.*   *  .:. *:

LIF      LYYTAQGEPFPNNLDKLCGPNVTDFPPFHAN-GTEKAKLVELYRI
OSM      PYIRIQGLDVPK-LREHCRERPGAFPSEETLRGLGRRGFLQ--TL
          *   **  .*: * : *  .    **  .:  *  :  :::   :
                                BC loop
LIF      VVYLGTSLGNITRD-QK-----ILNPSALSL--HSKLNATADILR
OSM      NATLGCVLHRLADLEQRLPKAQDLERSGLNIEDLEKLQMARPNIL
          . **  * .::   *:      *: *.*.:   .**: :   :

LIF      GLLSNVLCRLCSK--------YHVGHVDVTYGPDTSGKDVFQKKK
OSM      GLRNNIYCMAQLLDNSDTAEPTKAGRGASQPPTPTPASDAFQRKL
         ** .*: *              :.*:        * ..*.**:*

LIF      LGCQLLGKYKQIIAVLAQAF----------
OSM      EGCRFLHGYHRFMHSVGRVFSKWGESPNRSR
          **::*  *::::  :.:.*
```

B

```
OSM      AAIGSCSKEYRVLLGQLQKQTDLMQDTSRLLDPYIRIQGLDVPKLRE
BC loop  HCRERPGAFPSEETLRGLGRRGFLQTLNATLGCVLHRLADQKILNPS
chimera  ALSLHSKLQMARPNILGLRNNIYCMAQLLDNSDTAEPTKAGRGASQP
         PTPTPASDAFQRKLEGCRFLHGYHRFMHSVGRVFSKWGESPNRSR
```

Figure 3: Comparison of OSM and LIF amino acid sequences. (A) Alignment of the full-length amino acid sequences of human OSM and LIF, with the BC loop region highlighted. Asterisks (*) indicate fully conserved residues, colons (:) correspond to amino acids with strongly similar properties and periods (.) denote those with weakly similar features. **(B)** Amino acid sequence of the OSM BC loop chimera, with the BC loop region of OSM replaced by its LIF equivalent. Please click here to view a larger version of this figure.

OSM BC loop chimera DNA sequence

```
ATGGGGGTACTGCTCACACAGAGGACGCTGCTCAGTCTGGTCCTTGCACTCCTGTTTCCAA
GCATGGCGAGCATGGCGGCTATAGGCAGCTGCTCGAAAGAGTACCGCGTGCTCCTTGGCCA
GCTCCAGAAGCAGACAGATCTCATGCAGGACACCAGCAGACTCCTGGACCCCTATATACGT
ATCCAAGGCCTGGATGTTCCTAAACTGAGAGAGCACTGCAGGGAGCGCCCCGGGGCCTTCC
CCAGTGAGGAGACCCTGAGGGGGCTGGGCAGGCGGGGCTTCCTGCAGACCCTCAATGCCAC
ACTGGGCTGCGTCCTGCACAGACTGGCCGACCAGAAGATCCTCAACCCCAGTGCCCTCAGC
CTCCACAGCAAGCTCCAGATGGCGAGGCCGAACATCCTCGGGCTCAGGAACAACATCTACT
GCATGGCCCAGCTGCTGGACAACTCAGACACGGCTGAGCCCACGAAGGCTGGCCGGGGGGC
CTCTCAGCCGCCCACCCCCACCCCTGCCTCGGATGCTTTTCAGCGCAAGCTGGAGGGCTGC
AGGTTCCTGCATGGCTACCATCGCTTCATGCACTCAGTGGGGCGGGTCTTCAGCAAGTGGG
GGGAGAGCCCGAACCGGAGCCGGCTCGAGCACCACCACCACCACCACTGA
```

Figure 4: DNA sequence of the OSM BC loop chimera. Sequence of the chimeric OSM protein. The region inserted from LIF is highlighted in orange. Please click here to view a larger version of this figure.

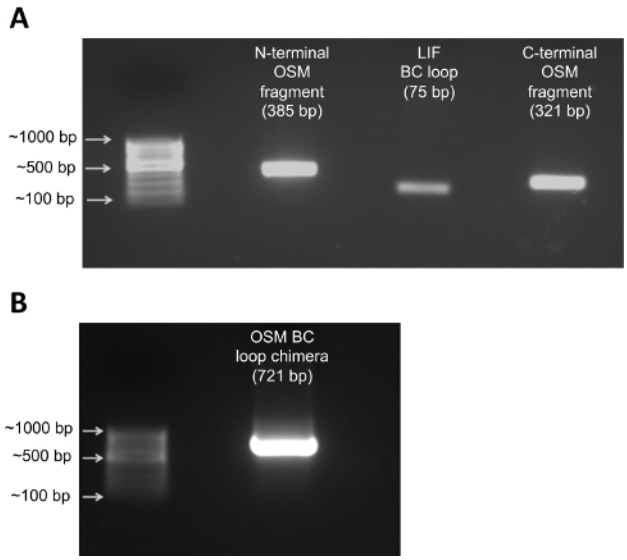

Figure 5: Amplification of the OSM BC loop chimera DNA fragments. (A) Result from the first PCR amplification, with bands corresponding to the N-terminal region (lane 2), BC loop (lane 3) and C-terminal region (lane 4). **(B)** Result from the second PCR amplification, in which the three bands obtained in the first amplification are combined to generate the OSM chimera. Please click here to view a larger version of this figure.

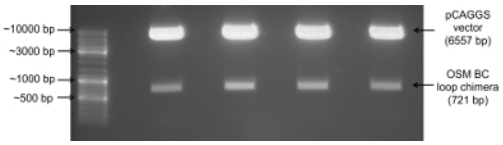

Figure 6: Insertion of the OSM BC loop chimera into plasmid vector. Restriction enzyme digestion of the generated plasmids, with a lower band present at ~700 base pairs indicating the correct insertion of the OSM chimera gene sequence. Please click here to view a larger version of this figure.

Reagent	Stock Concentration	Volume (in 50 µL)
Sterile Water		to 50 µL
PCR Buffer	10X	5 µL
dNTPs	10 mM	1 µL
DMSO	100%	1.5 µL
Forward Primer	10 µM	1 µL
Reverse Primer	10 µM	1 µL
Template DNA	Variable	As required for 2.5-12.5 ng of plasmid DNA
Phusion High Fidelity DNA Polymerase	2 Units/µL	0.5 µL

Table 1: Reagents required for the PCR reaction mixture.

	N-terminal fragment	Chimeric insertion	C-terminal fragment
Forward Primer	N-terminal forward	First junction forward	Second junction forward
Reverse Primer	First junction reverse	Second junction reverse	C-terminal reverse
Template DNA	Recipient	Donor	Recipient

Table 2: Primer and templates needed for the generation of a standard chimeric protein.

Cycle step	Temperature	Time	Number of cycles
Initial Denaturation	98 ºC	30s	1
Denaturation	98 ºC	10s	23-25
Annealing	68 ºC*	25s	23-25
Extension	72 ºC	30s per kilobase	23-25
Final Extension	72 ºC	300s	1
Hold	4 ºC	∞	1
*At least 5 ºC lower than the primer melting temperature			

Table 3: PCR protocol used to amplify the chimeric fragments and the full chimeric protein.

Reagent	Stock Concentration	Volume (in 50 µL)
Sterile Water		to 50 µL
Restriction Enzyme Buffer*	10X	5 µL
Template DNA	Variable	As required for 1 µg of DNA
Enzyme #1	Variable	1 µL
Enzyme #2	Variable	1 µL
*Ensure both enzymes are compatible with the buffer selected		

Table 4: Components of the restriction enzyme digestion reaction.

Reagent	Stock Concentration	Volume (in 20 µL)
Sterile Water		to 20 µL
T4 DNA Ligase Buffer	10X	2 µL
Vector DNA	Variable	As required for 40 ng of DNA
Insert DNA	Variable	As required for a 3:1 molar ratio
T4 DNA Ligase	Variable	1-2 µL

Table 5: Reagents required for the ligation reaction.

Name	Primer sequence
N-terminal OSM forward primer (PacI)	AAAGGGAAA-TTAATTAA-GCTAGCGCATCGCCACC-ATGGGGGTACTGCTCACACAGAGGACG
C-terminal HisTag reverse primer (AscI)	TTTCCCTTT-GGCGCGCC-GCGGCCGCTA-TCAGTGGTGGTGGTGGTGGTGCTCGAG
BC loop start forward	AACATCACCCGGGACTTAGAGCAGCGCCTC
BC loop start reverse	GAGGCGCTGCTCTAAGTCCCGGGTGATGTT
BC loop end forward	GACTTGGAGAAGCTGAACGCCACCGCCGAC
BC loop end reverse	GTCGGCGGTGGCGTTCAGCTTCTCCAAGTC

Table 6: Primers used in the generation of the OSM BC loop chimera.

Discussion

The generation of chimeric proteins constitutes a versatile technique, which is able to go beyond the limits of truncated proteins to address questions such as the modularity of cytokine-receptor binding domains[13]. The design of chimeras is a key step in this kind of studies, and requires careful consideration. Preliminary studies to establish functional domains will generally require substitution of broad regions in a first phase, while smaller replacements of variable lengths are more suited to detailed studies of a single region. Special attention should be given to the presence of small conserved motifs within a protein family in this step, since these are often indicative of functional sites[31,32]. Personal experience indicates that more than one round of chimeric protein design can be necessary to narrow down a key functional region, with each round requiring significant time (weeks to months) from initial design to functional assay testing.

As long as there exists a structurally similar protein to the protein of interest, but possessing diverging biological functions, the method is applicable to any sequence of interest, although it has to be optimized for each particular gene due to its reliance on PCR amplification. Particularly, genes possessing GC-rich regions might prove particularly challenging targets, since these types of sequences are known to reduce the efficiency of the amplification[33]. These issues can usually be solved by different means, such as the addition of different additives

(*e.g.* betaine) to the reaction, the use of specialized DNA polymerase buffers, or the modification of the annealing parameters[34]. Hence, it will generally require some trial and error before adequate conditions for the gene of interest are found.

The protocol provided is based on classic restriction enzyme-based cloning methods, which are generally accessible to every type of laboratory, but it can be further adapted to take advantage of more advanced cloning techniques. For example using gateway cloning, which facilitates cloning the same insert in several different vectors (*e.g.* if different expression systems are to be tested in parallel), would merely require particular *attB* recombination sites in place of the restriction sites detailed in this protocol[35]. Other newer cloning methodologies can bypass the need for a second PCR reaction (*e.g.* USER[36] or Gibson assembly[37]) and ligation (*e.g.* sequence and ligation-independent cloning (SLIC)[38] or In-fusion assembly[39]). While requiring different reagents and primer design strategies, readers with access to these methods are encouraged to apply them to significantly speed up the generation of chimeric constructs after following the basic design principles detailed in step 1 of this protocol.

Overall, the application of this method can supply valuable insight regarding the mechanisms by which other protein biological functions take place, in particular involving protein-protein or protein-nucleic acid interactions, and constitutes a useful tool to identify and specify unique structure-function relationships within a protein family[6].

Disclosures

The authors have nothing to disclose.

Acknowledgements

This work was supported by the Max Planck Society and the Schüchtermann-Clinic (Bad Rothenfelde, Germany). Part of this research was originally published in the Journal of Biological Chemistry. Adrian-Segarra, J. M., Schindler, N., Gajawada, P., Lörchner, H., Braun, T. & Pöling, J. The AB loop and D-helix in binding site III of human Oncostatin M (OSM) are required for OSM receptor activation. J. Biol Chem 2018; 18:7017-7029. © the Authors.

References

1. Huising, M. O., Kruiswijk, C. P., & Flik, G. Phylogeny and evolution of class-I helical cytokines. *The Journal of Endocrinology.* **189** (1), 1-25 (2006).
2. Brocker, C., Thompson, D., Matsumoto, A., Nebert, D. W., & Vasiliou, V. Evolutionary divergence and functions of the human interleukin (IL) gene family. *Human Genomics.* **5** (1), 30-55 (2010).
3. Bravo, J., & Heath, J. K. Receptor recognition by gp130 cytokines. *The EMBO Journal.* **19** (11), 2399-411 (2000).
4. Schneider, G., & Fechner, U. Computer-based *de novo.* design of drug-like molecules. *Nature Reviews Drug Discovery.* **4** (8), 649-63 (2005).
5. Heydenreich, F. M., Miljuš, T., Jaussi, R., Benoit, R., Milić, D., & Veprintsev, D. B. High-throughput mutagenesis using a two-fragment PCR approach. *Scientific Reports.* **7** (1), 6787 (2017).
6. Adrian-Segarra, J. M., Schindler, N., Gajawada, P., Lörchner, H., Braun, T., & Pöling, J. The AB loop and D-helix in binding site III of human Oncostatin M (OSM) are required for OSM receptor activation. *The Journal of Biological Chemistry.* **293** (18), 7017-7029 (2018).
7. Wang, Y., & Pallen, C. J. Expression and characterization of wild type, truncated, and mutant forms of the intracellular region of the receptor-like protein tyrosine phosphatase HPTP beta. *The Journal of Biological Chemistry.* **267** (23), 16696-702 (1992).
8. Lim, J., Yao, S., Graf, M., Winkler, C., & Yang, D. Structure-function analysis of full-length midkine reveals novel residues important for heparin binding and zebrafish embryogenesis. *The Biochemical Journal.* **451** (3), 407-15 (2013).
9. Kim, K.-W., Vallon-Eberhard, A., *et al.In vivo.* structure/function and expression analysis of the CX3C chemokine fractalkine. *Blood.* **118** (22), e156-67 (2011).
10. Chollangi, S., Mather, T., Rodgers, K. K., & Ash, J. D. A unique loop structure in oncostatin M determines binding affinity toward oncostatin M receptor and leukemia inhibitory factor receptor. *The Journal of Biological Chemistry.* **287** (39), 32848-59 (2012).
11. Aasland, D., Schuster, B., Grötzinger, J., Rose-John, S., & Kallen, K.-J. Analysis of the leukemia inhibitory factor receptor functional domains by chimeric receptors and cytokines. *Biochemistry.* **42** (18), 5244-52 (2003).
12. Hermanns, H. M., Radtke, S., *et al.* Contributions of leukemia inhibitory factor receptor and oncostatin M receptor to signal transduction in heterodimeric complexes with glycoprotein 130. *Journal of Immunology.* **163** (12), 6651-8 (1999).
13. Kallen, K. J., Grötzinger, J., *et al.* Receptor recognition sites of cytokines are organized as exchangeable modules. Transfer of the leukemia inhibitory factor receptor-binding site from ciliary neurotrophic factor to interleukin-6. *The Journal of Biological Chemistry.* **274** (17), 11859-67 (1999).
14. Gibrat, J. F., Madej, T., & Bryant, S. H. Surprising similarities in structure comparison. *Current Opinion in Structural Biology.* **6** (3), 377-85 (1996).
15. Madej, T., Lanczycki, C. J., *et al.* MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Research.* **42** (Database issue), D297-303 (2014).
16. Berman, H., Henrick, K., & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural Biology.* **10** (12), 980 (2003).
17. Biasini, M., Bienert, S., *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research.* **42** (Web Server issue), W252-8 (2014).
18. Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols.* **4** (1), 1-13 (2009).
19. Maglott, D. R., Katz, K. S., Sicotte, H., & Pruitt, K. D. NCBI's LocusLink and RefSeq. *Nucleic Acids Research.* **28** (1), 126-8 (2000).
20. Davis, M. W. *ApE, A Plasmid Editor.* http://jorgensen.biology.utah.edu/wayned/ape/. (2018).
21. Sievers, F., Wilm, A., *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology.* **7**, 539 (2011).

22. Niwa, H., Yamamura, K., & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene.* **108** (2), 193-9 (1991).
23. Barbas, C. F., Burton, D. R., Scott, J. K., & Silverman, G. J. Quantitation of DNA and RNA. *Cold Spring Harbor Protocols.* **2007**, pdb.ip47 (2007).
24. Sambrook, J., & Russell, D. W. Preparation and Transformation of Competent E. coli Using Calcium Chloride. *Cold Spring Harbor Protocols.* **2006** (1) (2006).
25. Sambrook, J., & Russell, D. W. Preparation of Plasmid DNA by Alkaline Lysis with SDS: Minipreparation. *Cold Spring Harbor Protocols.* **2006** (1) (2006).
26. Green, M. R., & Sambrook, J. Preparation of Plasmid DNA by Alkaline Lysis with Sodium Dodecyl Sulfate: Maxipreps. *Cold Spring Harbor Protocols.* **2018** (1), pdb.prot093351 (2018).
27. Hopkins, R. F., Wall, V. E., & Esposito, D. Optimizing transient recombinant protein expression in mammalian cells. *Methods in Molecular Biology.* **801**, 251-68 (2012).
28. Wang, X., Lupardus, P., Laporte, S. L., & Garcia, K. C. Structural biology of shared cytokine receptors. *Annual Review of Immunology.* **27**, 29-60 (2009).
29. Deller, M. C., Hudson, K. R., Ikemizu, S., Bravo, J., Jones, E. Y., & Heath, J. K. Crystal structure and functional dissection of the cytostatic cytokine oncostatin M. *Structure.* **8**, 863-74 (2000).
30. Huyton, T., Zhang, J.-G., *et al.* An unusual cytokine:Ig-domain interaction revealed in the crystal structure of leukemia inhibitory factor (LIF) in complex with the LIF receptor. *Proceedings of the National Academy of Sciences of the United States of America.* **104**, (31), 12737-42 (2007).
31. Oezguen, N., Kumar, S., Hindupur, A., Braun, W., Muralidhara, B. K., & Halpert, J. R. Identification and analysis of conserved sequence motifs in cytochrome P450 family 2. Functional and structural role of a motif 187RFDYKD192 in CYP2B enzymes. *The Journal of Biological Chemistry.* **283** (31), 21808-16 (2008).
32. Wong, A., Gehring, C., & Irving, H. R. Conserved Functional Motifs and Homology Modeling to Predict Hidden Moonlighting Functional Sites. *Frontiers in Bioengineering and Biotechnology.* **3**, 82 (2015).
33. McDowell, D. G., Burns, N. A., & Parkes, H. C. Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. *Nucleic Acids Research.* **26** (14), 3340-7 (1998).
34. Mamedov, T. G., Pienaar, E., *et al.* A fundamental study of the PCR amplification of GC-rich DNA templates. *Computational Biology and Chemistry.* **32** (6), 452-7 (2008).
35. Park, J., Throop, A. L., & LaBaer, J. Site-specific recombinational cloning using gateway and in-fusion cloning schemes. *Current Protocols in Molecular Biology.* **110**, 3.20.1-23 (2015).
36. Bitinaite, J., Rubino, M., Varma, K. H., Schildkraut, I., Vaisvila, R., & Vaiskunaite, R. USER friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Research.* **35** (6), 1992-2002 (2007).
37. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods.* **6** (5), 343-5 (2009).
38. Li, M. Z., & Elledge, S. J. Harnessing homologous recombination *in vitro.* to generate recombinant DNA via SLIC. *Nature Methods.* **4** (3), 251-6 (2007).
39. Zhu, B., Cai, G., Hall, E. O., & Freeman, G. J. In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *BioTechniques.* **43** (3), 354-9 (2007).

Copyright © 2018 Journal of Visualized Experiments