Design and Construction of Two Yeast Shuttle Vectors Containing Human Procollagen Genes Expression Cassette for Expression in Yeast

Baharak Abdemami 1,2, Mohammad Ali Shokrgozar 1*, Hossein Khanahmad Shahreza 3*, and Mehdi Ghavami 1,2

1. National Cell Bank of Iran, Pasteur Institute of Iran, Tehran, Iran
2. Science and Research Branch, Islamic Azad University, Tehran, Iran
3. Pasteur Research and Production Complex, Karaj, Iran

Abstract
Collagens are the most abundant proteins in the human body. Their main function is to provide structural and mechanical support for the tissues, but they are also involved in a number of other biological functions including cell attachment, migration and differentiation. Collagens and gelatins are widely used in pharmaceutical and medical applications. Every year, more than 50,000 tons of collagen and gelatin are used in medical applications. These materials may have some viral and prion impurity and/or stimulate allergic response in human body. Therefore, scientists have produced human collagen in recombinant systems. In this study we have constructed two yeast shuttle vectors containing human procollagen genes expression cassette for expression in yeast. Total RNA was extracted from human skin fibroblast cell line, and cDNA synthesis was done by oligo dt. Then gene fragments were amplified from the cDNA with the necessary changes by Polymerase Chain Reaction (PCR). Finally they were cloned in yeast vector pPICZαA containing regulatory sequences for expressing and secreting the polypeptide product. Two yeast shuttle vectors containing human COL1A1 and COL1A2 expression cassettes were created. Final constructs were confirmed by enzymatic digestion, PCR of desired fragment and sequencing. The yeast shuttle vectors containing human COL1A1 and COL1A2 can be transferred into the yeast in the later stages to determine the scale of expression.

Keywords: Collagen, Fibroblasts, Yeasts

Introduction
Collagens are the most abundant protein in body, which support most of the tissues and form the cells (1). Collagen has high elasticity and is capable of forming connective tissues. Collagen is the major component in extracellular matrix of cartilage, tendons, bone and skin (2,3). Individual characteristic of connective tissues are because of collagen structure, function and its correlation with other molecules (4). Collagen is essential for healthy and firm skin, so decrease in quantity of collagen in ageing skin leads to loss of elasticity. Collagen tightens the blood vessels and plays a role in tissue development (2).

Collagen molecules consist of three polypeptide chains, called α chains that are coiled around each other into a triple helix. These chains differ for each collagen type, collagen type I contains two α1 and one α2 chains (5). Distinct characteristic of collagen polypeptide chains is repeating triplet sequence Gly-X-Y, in which the X position amino acid is often
proline and Y position amino acid is often 4-hydroxyproline (6). The collagen superfamily in vertebrates consists of at least 28 proteins encoded by 43 specific genes (7-9).

Collagen is widely used in medical and pharmaceutical applications such as cell scaffolds in tissue engineering, biomaterials in drug delivery, implants, hemostats, wound dressings, vascular graft coatings, capsules and tablets. Moreover, collagen has been used in cosmetic surgery, reconstructive bone surgery, dentistry, orthopedics and composition of artificial skin. In addition, denaturated and partially degraded collagen, gelatin, plays an important role in setting the weight and appetite reducing as a protein, and gelatin is also used in photography and as a gelling in food industry (10-14).

Sources of collagens and gelatins for these applications were extracted from animals, mainly from bovine and porcine skin. The use of animal derived collagen and gelatin comprise several concerns. These extracted proteins have the potential to cause immune reactions and may also carry disease causing contaminants such as viruses or prions (8,9). Production of recombinant human collagen and gelatin could overcome this problem. Recombinant production systems are provided to generate engineered collagen products for different purposes (6).

Several expression systems have been tested in producing human recombinant collagens such as transfected mammalian cells, insect cells, yeast, Escherichia coli (E. coli), transgenic tobacco, mice, silkworms and barley (15-17). These experiences showed that yeast expression system is the best system for producing of collagen. Yeasts are single-celled microorganisms that are easy to manipulate and culture and are suitable for genetic research as a eukaryotic organism (15,16).

Among the yeast systems, Pichia pastoris (P. pastoris) is highly successful for the production of a wide variety of recombinant proteins. Several types of collagen have been produced in P. pastoris (8). Since 1984, more than 300 proteins such as Hepatitis B surface antigen, Bovine herpes virus and FMD viral proteins have been successfully expressed in P. pastoris. Among these proteins, several types of collagen proteins are expressed with similar characteristics to human collagen (18,19).

More than 50,000 metric tons of collagen and gelatin are used in research and medical applications annually (2). Considering the strategic importance of collagen in health and pharmaceutical industries, we embarked on recombinant production of collagen for the first time in Iran. In this study, the cDNA encoding COL1A1 and COL1A2 genes were cloned and the recombinant plasmids were analysed by restriction digestion and sequencing.

Materials and Methods
Vector design

According to the objective of this study, pPICZαA vector was chosen as the final expression vector obtained from Recombinant Gene Bank of Iran (Pasteur Institute of Iran). This vector has the alcohol oxidase strong promoter. The AOX1 promoter has been utilized for production of heterologous protein in P. pastoris. Several copies of pPICZαA vector (Figure 1) are introduced into yeast genome, so it will have high expression of heterologous protein.

According to this vector, the restriction enzymes and primers were designed. To re-
place the collagen gene in frame, multiple cloning sites of vector was exchanged. In addition to Zeocin resistant gene, the Ampicillin resistant gene was also introduced into vector as a cheap bacterial selection marker. In order to generate a second construct, *Pichia* expression vector was modified by replacing Zeocin resistant gene with Hygromycin resistant gene. Gene Runner software was used to evaluate these gene fragments.

**Fibroblast cultures**

Human Skin Fibroblast Cell line (HSF) was purchased from National Cell Bank of Iran (NCBI). The cell line was cultured in Dulbecco’s modified Eagle’s medium (DMEM) (Gibco, U.S.A.) supplemented with 10% fetal bovine serum (FBS) (Gibco, U.S.A.), 1% p/s (100 U/ml penicillin, 100 μg/ml streptomycin) and was incubated at 37°C in 5% CO2 incubator. Cultured cells were passaged using trypsin containing EDTA (Gibco, U.S.A.).

**Total RNA extracted from fibroblast cells**

Total RNA was isolated from fibroblasts using TRIZOL reagent (Gibco BRL). Fifty μl total RNA was extracted from 10^5 cells by trizol according to the manufacturer’s instruction.

**cDNA synthesis**

Total cDNA was synthesized using the first strand cDNA synthesis kit (Fermentas, Germany) on total RNA. According to the protocol, total RNA was incubated for 5 min at 65°C at first because the desired genes were rich in %G+C and then 60 min incubation with MuLV reverse transcriptase at 42°C was done. The product of the first strand cDNA synthesis reaction was used directly in PCR.

**Primer design and proliferation**

The sequences were obtained from NCBI and primers were designed using Gene Runner software with required cut sites. These primers were constructed by Gene Fanavar Company (Table 1). COL1A1 and COL1A2 gene fragments were amplified by PCR from total cDNA. Ampicillin and Hygromycin resistant genes were also proliferated by PCR from pTZ57R and pCDNA3.1Hygro vectors, respectively.

Due to the long size of pPICZaA backbone, this gene fragment was amplified using Long PCR Enzyme Mix kit (Fermentas Co.). The size of this fragment was 4120 bp. At the end of the amplification, the size of the PCR product was determined to compare with 1 kb DNA marker, using 1% agarose gel electrophoresis followed by staining with ethidium bromide. Then, these PCR products were purified using DNA gel extraction kit (Fermentas) and then used for cloning.

**Cloning steps**

pPICZαA vector was digested with XhoI and XbaI restriction enzymes and was replac-

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
<th>Enzyme site</th>
<th>Fragment length</th>
<th>PCR conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F-COL1A2</strong></td>
<td>5’GGTACCCAGTATGGAAAGAGGAGTTGG3’</td>
<td>KpnI</td>
<td>3120</td>
<td>Annealing: 60°C</td>
</tr>
<tr>
<td><strong>R-COL1A2</strong></td>
<td>5’AGATCTAGCCCTGTAAGATCTCCATC3’</td>
<td>BglII</td>
<td>Extension: 6 min 10 s</td>
<td></td>
</tr>
<tr>
<td><strong>F-COL1A1</strong></td>
<td>5’GAATTCACGCTTATGGATGACAGGACTG3’</td>
<td>EcoRI</td>
<td>3183</td>
<td>Annealing: 68°C</td>
</tr>
<tr>
<td><strong>R-COL1A1</strong></td>
<td>5’AGATCTAGCCCGGTAGTGCG3’</td>
<td>BglII</td>
<td>Extension: 6 min 20 s</td>
<td></td>
</tr>
<tr>
<td><strong>F-Amp</strong></td>
<td>5’AGATCTATATATATGAACTTGTGCTGACAG3’</td>
<td>BglII</td>
<td>923</td>
<td>Annealing: 45°C</td>
</tr>
<tr>
<td><strong>R-Amp</strong></td>
<td>5’AGATCTAAAATATTGAAAAGGAGGATAGTG3’</td>
<td>BglII</td>
<td>Extension: 120 s</td>
<td></td>
</tr>
<tr>
<td><strong>F-Oligo</strong></td>
<td>5’TCCAGAAAAGAGAGGCTGAGGCTGGAAGCC3’</td>
<td>Xhol</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R-Oligo</strong></td>
<td>5’CTAGAAGGTTACAGCTTTGACGCTCTTTTCC3’</td>
<td>XbaI</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F-Backbone</strong></td>
<td>5’AATCTTAACGCGAGCTCGAGGCCC3’</td>
<td>AflII</td>
<td>4120</td>
<td>Annealing: 68°C</td>
</tr>
<tr>
<td><strong>R-Backbone</strong></td>
<td>5’ATTAGCTAGCCGGTTATGCTCCACCTTTTGCTATTATCTATG3’</td>
<td>NheI</td>
<td>Extension: 4 min 20 s</td>
<td></td>
</tr>
<tr>
<td><strong>F-Hygro</strong></td>
<td>5’ATTAGCTAGCATGAAAAAGCCCTGA ACTCTGC3’</td>
<td>NheI</td>
<td>1126</td>
<td>Annealing: 56°C</td>
</tr>
<tr>
<td><strong>R-Hygro</strong></td>
<td>5’ATCTTAAAGCTTATCCCTTTGGCCTCG3’</td>
<td>AflII</td>
<td>Extension: 65 s</td>
<td></td>
</tr>
</tbody>
</table>
ed by the previously designed oligonucleotide containing these cut sites. Then, pPICZ_{α}A vector was linearized with BgIII and AmpR gene was ligated into the linearized vector. COL1A1 and COL1A2 fragments were flanked by artificial 5'_{EcoRI} and 3'_{KpnI} sites, and 5'_{KpnI} and 3'_{BglII} sites, respectively. Then, COL1A1 PCR products were digested with EcoRI and KpnI and co-ligated into EcoRI - KpnI digested pPICZ_{α}A, which led to the production of pPICZ_{α}A construct containing the COL1A1 gene.

In order to produce the second construct containing COL1A2, Hygromycin resistant gene and also pPICZ_{α}A(AmpR) which didn’t include zeocin, were both flanked by artificial cut sites, and then digested with NheI and AflII. These two fragments were generated by PCR, such that the second one was extended from the last codon of the Zeocin resistant gene up to its initiation codon. Then, these PCR products were co-ligated; and ZeoR gene was replaced by HygroR. Finally, COL1A2 was digested with BglII - KpnI and cloned into BglII - KpnI digested AmpR hygroR pPICZ_{α}A vector in order to generate the second construct.

All plasmids were transformed into the E.coli Top10F’ strain to proliferate. Bacteria were cultured in LB-Broth medium (Merck-Germany) containing 100 μg/ml ampicillin.

Schematic diagram of cloning steps is shown in figure 2.

**Confirmation**

All clones and subclones were confirmed by clony PCR, digestion and sequencing. The constructs were directly sequenced by the dideoxy chain termination method (Gene Fanavaran Company) using designed primers.

**Results**

**Design result**

Final shuttle vector was designed so COL1A1 and COL1A2 genes were placed to correct the position of the share between the desired and correct translation frame in the two structures separately. In addition, the Ampicillin resistance gene was placed into BglII restriction site of pPICZ_{α}A vector and in the second structure of pPICZ_{α}A vector, Zeocin resistance gene was replaced by Hygromycin resistance gene. The size of AmpR gene was 923 bp and HygroR gene previously proliferated by PCR was 1126 bp.

**Total RNA extraction result**

If total RNA is extracted properly, two bands in 28S and 18S related to ribosomal RNAs would be observed by gel electrophoresis. Electrophoresis of the product was determined with the Total RNA fully obtained (Figure 3).
cDNA synthesis and PCR results

cDNA synthesis protocol properly was done, then complete cDNA was confirmed by PCR. COL1A1 and COL1A2 gene segments were amplified by PCR. The size of collagen fragments are 3183 bp (COL1A1) and 3120 bp (COL1A2) (Figure 4).

Final shuttle vector confirmation

Final vectors were analyzed by digestion, PCR and sequencing. The relevant parts of all results indicated that all of these steps were performed correctly.

In regards to the three enzymes BglII cut sites existing in the final structure (two locations on both sides of AmpR gene and a location at the 3’ end of COL1A1 and COL1A2 genes), this enzyme was used to confirm the cloning. In cutting with this enzyme, the structure contains COL1A1; if the parts are correctly cloned, three bands of approximately 900 bp, 2500 bp and 4500 bp will be observed (Figure 5A) and in structure containing COL1A2, if the gene entering is correct, three bands of approximately 900 bp, 3000 bp and 4400 bp will be made (Figure 5B).

Results of enzymatic digestion with BglII, indicate correct entry into COL1A1 and COL1A2, so two vectors containing COL1A1 and COL1A2 were produced (Figures 6A and B). The results of sequencing were analyzed by chromas and were aligned by Mega4 and there were not any mutation, deletion or frame shift.

Discussion

Due to the complexity of collagen biosynthesis, it is a challenge to produce it in recombinant systems. Besides, several post translational modifications are required and its expression in bacterial systems is not possible. Several systems are used to produce recombinant collagen and gelatin. Several systems have been developed such as mammalian
cells, insect cells, yeast and E.coli and transgenic animals such as mice, silkworms, tobacco and barley plants (15,16).

In 1998, Fichard et al and Tyler et al began collagen production in cell lines, and despite their efforts, the collagen which was produced, did not have enough quality and due to the high cost of cell culture, it was not suitable for commercial production (20). The genomic fragment of collagen type I was cloned in mice by Toman in 1999 and the collagen was produced in mouse milk. The resulted collagen had much lower hydroxy proline in its structure and was not desirable (21).

One year later Ruggiero cloned collagen type I gene into tobacco plants and collagen was expressed in tobacco (22), and Human collagen type I homotrimer was produced in tobacco by Merle, but in both of them, hydroxy proline amount was low (23). Human collagen type III was expressed in silkworm, but the produced collagen fragments were not hydroxylated, and collagen triple helix was not formed (24). Ritala tried to express collagen in barley cell cultures in 2008, but their efforts did not produce satisfactory results and triple helix did not formed (25).

Production in microorganism systems is the most desirable method of protein production. Thus, the efforts to produce collagen were carried out in microbial systems. In non-biological methods, production rate is much higher and the cost is lower (2,8).

The efforts to produce human recombinant collagen in E.coli eventually succeeded to produce gelatin fragments (26). In 2000, recombinant human type I procollagen trimers was expressed in Saccharomyces cerevisiae (27). The best results were obtained when Vuorela et al expressed collagen type III in P.pastoris, in 1997 (9). Four years later collagen type I was produced as a complete triplet and with high yields in P.pastoris (12).

P.pastoris has the ability to express several foreign genes simultaneously, and this is one of the most important characteristics which is needed to produce collagen (28). Production of stable recombinant collagen triple helix structure requires prolyl 4-hydroxylase enzyme (29) and P.pastoris expressing prolyl 4-hydroxylase enzyme is engineered to produce high (g/15/1-1) hydroxylated triple helixes of recombinant human collagen type I, II and III (8,30). The recombinant human collagen produced in P.pastoris has a similar structure and content of hydroxy proline to compare with human wild collagen molecules (2). Therefore, this is why we have followed this system to produce collagen; thus, we have designed two yeast shuttle vectors containing the gene expression cassettes of human collagen type I.

Using optimized synthetic genes for expression in P.pastoris increased 25% in the amount of products (2). Researchers’ results showed that N-propeptide is not necessary for the collagen triple helix formation. When N-propeptide was separated from the gene, the triple helix formed and the amount of human recombinant collagen type III was increased about 25% in comparison to the presence of N-propeptide (31). These results indicate that the N-propeptides do not play a role in the accumulation of procollagen molecules within...
the endoplasmic reticulum lumen. Therefore, to make gene cassettes containing the coding sequences of \( \alpha_1 \) and \( \alpha_2 \) collagen chains, according to previous studies for optimizing gene, regions of the genes were amplified without N-propeptide ends using PCR.

Resistant to Zeocin antibiotic was the selective marker gene for COL1A1 construction. In order to determine entry of both final constructions into the yeast, COL1A2 construction’s selection marker was changed by Hygromycin.

**Conclusion**

As already mentioned, by cell culture, we gained a number of human skin fibroblast cells. We extracted total RNA with sigma protocol from cells. After synthesis of cDNA, we amplified COL1A1 and COL1A2 fragments by designed primers. Two desired constructs were made by several clones and subclones. As mentioned above prolyl 4-hydroxylase is needed to produce stable collagen protein, so in this project we had to construct two additional vectors containing expression cassette of \( \alpha \) and \( \beta \) genes of prolyl 4-hydroxylase. In future steps of this study, *P. pastoris* will be transformed by these four vectors. At the end, collagen expression will be characterized by SDS-PAGE and western.

**Acknowledgement**

Finally, the authors express their gratitude to Drs. Noori Inanlu, Ahangary, Movassagh and colleagues in B.C.G and Cell Bank of Pasture Institute of Iran for their cooperation during the procedure performed in the project.

**References**


