Expression, purification, and characterization of recombinant human pancreatic duodenal homebox-1 protein in *Pichia pastoris*

Shi-Wu Li, Yuping Sun, William Donelan, Hongfang Yu, Joanna Scian, Dongqi Tang, Li-Jun Yang*

Department of Pathology, Immunology, and Laboratory Medicine, University of Florida College of Medicine, Gainesville, FL 32610, USA

**Article info**

Article history:
Received 11 October 2009
and in revised form 2 April 2010
Available online 8 April 2010

**Keywords:**
Pancreatic duodenal hemedomain protein-1 (PDX1)
Diabetes
Protein expression in yeast
*Pichia pastoris*

**Abstract**

Pancreatic duodenal homebox-1 (PDX1) is essential for the development of the embryonic pancreas and plays a key role in pancreatic β-cell differentiation, maturation, regeneration, and maintenance of normal pancreatic β-cell insulin-producing function. Purified recombinant *PDX1* (rPDX1) may be a useful tool for many research and clinical applications, however, using the *Escherichia coli* expression system has several drawbacks for producing quality *PDX1* protein. To explore the yeast expression system for generating rPDX1 protein, the cDNA coding for the full-length human PDX1 gene was cloned into the secreting expression organism *Pichia pastoris*. SDS–PAGE and western blotting analysis of culture medium from methanol-induced expression yeast clones demonstrated that the rPDX1 was secreted into the culture medium, had a molecular weight by SDS–PAGE of 50 kDa, and was glycosylated. The predicted size of the mature unmodified *PDX1* polypeptide is 31 kDa, suggesting that eukaryotic post-translational modifications are the result of the increased molecular weight. The recombinant protein was purified to greater than 95% purity using a combined ammonium sulfate precipitation with heparin–agarose chromatography. Finally, 120 μg of the protein was obtained in high purity from 1 L of the culture supernatant. Bioactivity of the rPDX1 was confirmed by the ability to penetrate cell membranes and activation of an insulin-luciferase reporter gene. Our results suggest that the *P. pastoris* expression system can be used to produce a fully functional human rPDX1 for both research and clinical application.

© 2010 Elsevier Inc. All rights reserved.

**Introduction**

The incidence of both Type 1 and especially Type 2 diabetes is increasing dramatically and diabetes now affects roughly 8% of the U.S. population [1]. Therefore, more than ever, there is an urgent need for new treatment modalities for diabetes. Although the two forms of diabetes have different pathophysiologic mechanisms, a shared feature is insulin insufficiency due to β-cell deficiency [2]. To cure diabetes, researchers are pursuing strategies for restoring β-cell mass by searching for factors that either promote endogenous pancreatic β-cell regeneration or reprogram non-pancreatic cells into insulin-producing cells (IPCs) [3]. Until recently, over-expression of key pancreatic transcription factor (PTF) genes by means of viral vectors was the most effective way to stimulate pancreatic-cell differentiation into β-cells and to reprogram liver cells/liver stem cells into IPCs. Several key PTF proteins including PDX1, Ngn3, NeuroD, and Pax4 are critical for pancreatic β-cell differentiation and maturation and they all contain a special amino acid sequence called a protein transduction domain (PTD) [4]. The most effective and well characterized PTDs are positively charged cationic sequences that allow proteins to rapidly enter living cells or whole organs [5]. It has been suggested that molecules containing these types of PTDs transduce cells by lipid raft-mediated macropinocytosis and activate the transcription of their target genes [6,7]. This strategy opens a new avenue for reactivating β-cell development or for directing stem-cell differentiation with protein therapy. *PDX1* contains a highly basic sequence of 16 amino acids, which constitutes an antennapedia-like PTD that facilitates penetration of plasma membranes [8]. This PTD allows *PDX1* to cross the cell and nuclear membranes, eliciting biological effects independent of endocytosis [8].

The PDX1 gene consists of two exons coding for a protein of 283 amino acids with a predicted molecular mass of 31 kDa. *PDX1* is now widely regarded as a master transcriptional regulator in the pancreas and is critical for the development [9–11], regeneration [12,13], and maintenance of β-cell function [13,14]. During embryogenesis, the PDX1 gene is expressed in
all progenitor cells differentiating toward the exocrine, pancreatic ducts, and endocrine pancreas. In adults, PDX1 expression is restricted mainly to β-cells and plays a key role in insulin gene expression [15,16].

Our recently published data offer a proof-of-principle demonstration that treatment of streptozotocin-induced diabetic mice with recombinant PDX1 protein (rPDX1) promotes β-cell regeneration and transient liver-cell reprogramming, leading to restoration of normoglycemia [17]. Although rPDX1 treatment of diabetic mice is a promising avenue, the in vivo biologic activity of bacterially-expressed rPDX1 protein is much lower than transgene-expressed rPDX1. This is at least partially due to the limitations of the prokaryotic bacterial expression system since the rPDX1 lacks post-translational modifications. In addition, an additional tag is needed for protein affinity purification, which may affect a protein’s biological function, creating immunogenic epitopes that can stimulate antibody production.

The use of Escherichia coli as an expression system has drawbacks when used to manufacture rPDX1 [18]. PDX1 is a large protein that is glycosylated post-translationally. There is evidence that this modification plays a specific role in the binding of PDX1 to DNA and it directly correlates with glucose-stimulated insulin secretion in β-cells [19]. Since the E. coli expression system lacks the ability to glycosylate proteins post-translationally [20], we explored the yeast expression system. Pichia pastoris is widely used as an expression system with the capacity to generate post-translationally modified proteins [21]. Moreover, P. pastoris can express proteins extracellularly, eliminating the need to use a His-tag for protein purification. Also, P. pastoris does not secrete many intrinsinc proteins, simplifying the process of purification of the recombinant protein from the culture medium. Finally, P. pastoris is a methylotrophic organism that can be cultured to high cell densities at relatively low cost [21]. In the present study, our aim was to construct an efficient system for the expression and purification of rPDX1 in P. pastoris.

Materials and methods

Expression plasmid construction

Human PDX1 cDNA (283 amino acids) was amplified by PCR on a plasmid encoding full-length human PDX1 purchased from OriGene technologies (USA). The sequence of the amplified gene was analyzed and confirmed by an ABI 3130xl Genetic Analyzer. In order to express the native N-terminus of PDX1, an XhoI site was introduced to allow in-frame cloning behind the α-mating factor pre-secretion signal of pPICZαA and a nucleotide sequence encoding the KeX2 gene cleavage site was placed upstream of the PDX1 gene. The PCR products were cut with XhoI and XbaI and ligated into this site downstream of the alcohol oxidase 1 promoter (AOX1) in pPICZαA (Invitrogen, USA).

Transformation of P. pastoris and selection of transformants

Pichia pastoris X-33 was transformed with a linearized expression vector by digestion with SacI. The transformation was performed using the lithium chloride method following the kit manual (Pichia Easycomp Transformation Kit, Invitrogen, USA). The transformant cells were plated on YPDS (1% yeast extract, 2% peptone, 2% dextrose, and 1 M o-sorbitol) plates containing 100 μg/ml of Zeocin. Approximately thirty Zeocin-resistant colonies were replated on a YPD (1% yeast extract, 2% peptone, and 2% dextrose) plate containing 500 μg/ml of Zeocin. After incubation at 30 °C for 2–3 days, several colonies appeared on the YPD plate and eight of the larger colonies were selected for protein expression.

Expression and purification of human PDX1

Eight transformants were screened for expression of secreted rPDX1 by western blot analysis of culture medium, and the S1 clone was chosen for large scale expression and purification of rPDX1. S1 cells were grown in 500 ml of BMGY (100 mM potassium phosphate, pH 6.0, 1.34% YNB, 4 × 10^{-5} biotin, and 1% glycerol) for 48 h at 28 °C. The cells were harvested by centrifugation (1000 rpm, 25 min at 4 °C) and resuspended in 1.0 L of BMMY (100 mM potassium phosphate, pH 6.0, 1.34% YNB, 4 × 10^{-5} biotin, and 0.5% methanol) in a 2 L flask for induction of AOX1. The yeast cells were grown in BMMY media for 96 h at 28 °C with shaking. Every 24 h methanol (100%) was added to a final concentration of 0.5% to maintain induction. Cell culture supernatant was harvested by centrifugation and was precipitated with 60% saturated ammonium sulfate [22]. The precipitated proteins were collected and dissolved in buffer A (25 mM Heps, pH 7.9, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, and 0.5 mM DTT), and dialyzed against the same buffer. The resulting sample was loaded onto a heparin–agarose column (Bio–Raid). The column was first washed with 0.2 M KCl in buffer A, and proteins were eluted with 1.5 M KCl in buffer A. The elution fractions were analyzed using SDS–PAGE followed by Coomassie brilliant blue staining and western blotting. The purified PDX1 was dialyzed against 10% glycerol in phosphate-buffered solution (PBS) and stored at −80 °C.

N-glycanase digestion and western blotting analysis

The rPDX1 expression sample was digested for 5 h at 37 °C with recombinant N-glycanase (PNGase F, New England Biolabs) according the manufacturer’s instructions. The digested and undigested controls were subjected to SDS–PAGE on a 12% gel and transferred to a polyvinylidine fluoride (PVDF) membrane. The blot was probed with rabbit anti-PDX1 polyclonal antibodies and developed using the Amersham Pharmacia Biotech ECL detection system.

Insulin-reporter luciferase assay

The rat insulin II promoter-luciferase reporter construct (pGL2-RIP2-Luc) was a gift from Dr. Carlotti [23]. Human hepatocellular carcinoma Huh-7 cells were seeded onto a 12-well plate and grown for 24 h in Dulbecco’s modified Eagle’s medium (DMEM) containing 10% PBS at 5% CO2 and 37 °C. One microgram of pGL2-RIP2-Luc reporter plasmid and 0.01 μg of internal control vector HSV-TK-Rluc (Renilla luciferase, pRL-TK, Promega) were used to co-transfect these cells using Lipofectamine 2000 reagent (Invitrogen, USA). Following 12 h transfection, the medium was replaced with medium containing rPDX1 protein or PBS with 10% glycerol (protein storage solution) for negative controls. After 24 h protein treatment, cell lysates were collected and assayed for luciferase activities using Dual-Glo- Assay–System (Promega).

Results

Optimization of rPDX1 expression conditions and strains

The resulting plasmid from the construction, pPICZα-rPDX1, was transformed into P. pastoris. Of the 15 single colonies that grew on YPD containing 500 μg/ml Zeocin, eight clones were selected for testing the capacity of rPDX1 protein expression by western blot analysis (Fig. 1). All selected clones expressed rPDX1 with
a molecular weight of 50 kDa, which is slightly higher in yeast compared with native PDX1 expressed in the rat insulinoma β-cell line INS-1 (46 kDa) (Fig. 1, lane C), suggesting that the protein expressed in yeast might be glycosylated. There was no protein that reacted with anti-PDX1 antibody in untransformed P. pastoris X-33 (data not shown). Fig. 1 shows that the S1 clone has the highest expression of the rPDX1 and therefore was used for the subsequent time-course study.

To determine the optimal time for rPDX1 expression, the S1 clone was seeded in the culture medium and samples were collected at 12, 24, 48, 72, and 96 h. Fig. 2A shows that the intensity of rPDX1 protein increased over time, with the strongest signal intensity at 96 h. Therefore, a 96-h incubation time was identified as optimal (Fig. 2A). To determine the secretory ability of the rPDX1 in the yeast expression system, we examined the protein levels both in the culture medium and in the yeast cells by western blotting. Fig. 2B reveals that the majority of the rPDX1 protein was cell-associated whereas only 10% of rPDX1 was secreted into the culture medium.

**Large-scale fermentation and purification of rPDX1**

To obtain the optimal fermentation conditions to scale-up protein production, we used the S1 clone after 96 h incubation in the presence of 1% methanol. The rPDX1 protein in the supernatant was purified by a two-step method [22]. First, rPDX1 was precipitated with 60% saturated ammonium sulfate. The pellet containing rPDX1 was dissolved in column buffer and subjected to heparin-gel affinity chromatography. The rPDX1 was eluted from the column with elution buffer containing 1.5 M KCl. The purity of the rPDX1 was assessed by SDS–PAGE (Fig. 4A). An average of about 120 μg of purified rPDX1 was recovered from 1 L of culture medium. The purity of the purified rPDX1 reached nearly 95% as confirmed by SDS-gel staining and western blotting with anti-PDX1 antibodies (Fig. 4B, lanes 2 and 3).

**Glycosylation analysis of purified rPDX1**

To examine whether the rPDX1 harvested from the yeast expression system underwent post-translational glycosylation, the purified rPDX1 from supernatant was treated with PNGase F to remove N-linked glycosylation and subjected to SDS–PAGE and western blotting analysis. As shown in Fig. 3, the pretreated rPDX1 has a predicted molecular weight of approximately 50 kDa, whereas the treated rPDX1 exhibited two distinct bands, one large band at 46 kDa and one faint band at 31 kDa (Fig. 3). The shift in the band post-PNGase F treatment indicates that the rPDX1 was indeed glycosylated in the yeast expression system. PNGase F is known to strictly cut N-linked glycans. As previously mentioned, protein expressed in yeast is glycosylated by both O- or N-glycosidic linkages. We interpret that the emergence of two

![Image](image-url)
bands is due to the ability of PNGase F to only remove the N-form of glycosylation from the rPDX1 protein. It is also possible that the emergence of two bands indicates that the PNGase F may have digested rPDX1 only partially. Regardless, it is clear that rPDX1 was secreted from the P. pastoris and was glycosylated post-translationally.

Biological activity of rPDX1

There are several direct target genes for PDX1, including the insulin gene [24,25]. To determine the biologic function of rPDX1, we transfected a rat insulin II promoter-luciferase reporter gene construct were treated with or without the rPDX1 protein for 24 h. Cell lysates were collected for determination of luciferase activity. Remilia luciferase was used as an internal control for transfection efficiency. Bars represent means plus SD and this result is one of the three independent experiments. The difference between control and treated groups was significant (p < 0.001, t-test).

Discussion

Our data show that P. pastoris yeast can successfully be used as an expression system for generating biologically functional rPDX1 protein, serving as an alternative to prokaryotic expression systems. We demonstrate here that the P. pastoris yeast expression system has the ability to secrete glycosylated and biologically active rPDX1 into the culture medium. Furthermore, we developed a purification strategy and protocol for rPDX1 purification from the culture medium of P. pastoris yeast cells. Although P. pastoris cells expressed a large quantity of rPDX1, the quantity of secreted rPDX1 protein was less than that we had hoped, posing a challenge for rPDX1 generation using this system.

It may be possible to improve the expression level by varying the nature of the signal peptide [30], deleting the nuclear localization signal (NLS) [31], chaperone co-expression [27–29,32–35] or varying expression conditions [36], which have proven effective in other cases.

In addition to its roles in embryonic development of the pancreas and the maintenance of β-cells in the adult pancreas, PDX1 over-expression induces the adoption of an insulin-producing cell phenotype in hepatocytes [8]. Its ability to direct β-cell differentiation and transdifferentiation to the β-cell phenotype gives it the potential for use in β-cell replacement therapy for Type 1 diabetes [37]. Inducing the over-expression of PDX1 and other pancreatic transcription factors to study their effects typically requires the use of viral (e.g. lentiviral) vectors for delivery into cells and incorporation into the genome. This approach inevitably entails some risk of insertional mutagenesis [38,39] and raises issues for potential clinical application. PDX1’s PTD, therefore, makes it especially attractive for use as a differentiation reagent for generating patient-specific pluripotent stem cells to differentiate into pancreatic β-cells because it can translocate directly into cells without requiring the use of viral vectors.

Protein therapy using PDX1 protein for diabetes provides hope for enhanced safety. Accumulating evidence shows that reactivating the developmental program for reprogramming non-pancreatic β-cells or for differentiating pancreatic stem/progenitor cells into insulin-secreting β-like cells has great clinical potential in diabetes. The ability to obtain biologically functional, near-native rPDX1 protein in good yields is important for clinical application and for understanding the reprogramming process in pancreatic β-cells. Our current work provides a starting point for future efforts to optimize the expression, secretion, and purification of biologically active rPDX1 for the therapy of diabetes.

Acknowledgments

This work was supported in part by grants from the National Institutes of Health, NIDDK DK071831 (L.J. Yang), from the University of Florida, OTL Research Fund, and Sponsored Research from Transgenon Therapeutics, Inc.

References


![Fig. 5. Biological function of rPDX1. Huh-7 cells transfected with the insulin promoter-luciferase reporter gene construct were treated with or without the rPDX1 protein for 24 h. Cell lysates were collected for determination of luciferase activity. Renilla luciferase was used as an internal control for transfection efficiency. Bars represent means plus SD and this result is one of the three independent experiments. The difference between control and treated groups was significant (p < 0.001, t-test).](image)


