THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 44, pp. 33554–33565, November 3, 2006


# The Human OCT-4 Isoforms Differ in Their Ability to Confer Self-renewal*

Received for publication, April 25, 2006, and in revised form, August 25, 2006 Published, JBC Papers in Press, September 1, 2006, DOI 10.1074/jbc.M603937200

**Jungwoon Lee**[‡1], **Hye Kyoung Kim**[‡], **Jeung-Yon Rho**[§], **Yong-Mahn Han**[§], **and Jungho Kim**[‡2]

*From the* [‡]*Laboratory of Molecular and Cellular Biology, Department of Life Science, Sogang University, Seoul 121-742 and the* [§]*Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea*

**OCT-4 transcription factors play an important role in maintaining the pluripotent state of embryonic stem cells and may prevent expression of genes activated during differentiation. Human OCT-4 isoform mRNAs encode proteins that have identical POU DNA binding domains and C-terminal domains but differ in their N-terminal domains. We report here the cloning and characterization of the human OCT-4B isoform. Human OCT-4B cDNA encodes a 265-amino acid protein with a predicted molecular mass of 30 kDa. Embryonic stem (ES) cell-based complementation assays using ZHBTc4 ES cells showed that unlike human OCT-4A, OCT-4B cannot sustain ES cell self-renewal. In addition, OCT-4B does not bind to a probe carrying the OCT-4 consensus binding sequence, and we demonstrate that two separate regions of its N-terminal domain are responsible for inhibiting DNA binding. We also demonstrate that OCT-4B is mainly localized to the cytoplasm. Overexpression of OCT-4B did not activate transcription from OCT-4-dependent promoters, although OCT-4A did as reported previously. Furthermore, transcriptional activation by human OCT-4A was not inhibited by co-expression of OCT-4B. Taken together, these data suggest that the DNA binding, transactivation, and abilities to confer self-renewal of the human OCT-4 isoforms differ.**

The *oct-4* gene, also referred to as *oct-3*, encodes a nuclear protein that belongs to a family of transcription factors containing the POU DNA binding domain (1–6). It is normally found in the pluripotent stem cells of pregastrulation embryos, including oocytes, early cleavage-stage embryos, and the inner cell mass of the blastocyst (1, 3, 7, 8). Its expression is down-regulated during differentiation, and knock-out of *oct-4* causes early lethality in mice because of the absence of an inner cell mass (9). These results suggest that OCT-4 plays a pivotal role in mammalian development (10) and in the self-renewal of embryonic stem cells (11). During human development, OCT-4 is expressed at least until the blastocyst stage where it regulates gene expression (12).

OCT-4 is a transcriptional regulator of genes involved in maintaining the undifferentiated pluripotent state and may also prevent expression of genes activated during differentiation (13). It activates transcription via octamer motifs located proximal or distal to transcriptional start sites (14). The POU domain of OCT-4 is a conserved DNA binding domain that binds as a monomer to the octamer sequence motif 5′-ATGCAAAT-3′ (15). This *cis*-acting element is important in controlling the activity of many promoters and enhancers of housekeeping and cell type-specific genes (16). OCT-4-binding sites have been found in various genes, including *fgf-4* (fibroblast growth factor-4), *pdgfαr* (platelet-derived growth factor-α receptor), osteopontin, and *Nanog* (17–21). In addition, genes expressed in the trophectoderm but not in the embryo prior to blastocyst formation, such as *IFN-τ* (τ-interferon) and the α and β subunits of chorionic gonadotropin, may be targets for silencing by OCT-4 (22–24). This suggests that OCT-4 functions as a master switch during differentiation by regulating cells that have pluripotent potential or can develop such potential (25, 26).

Transcriptional regulation by OCT-4 is complex. In embryonic stem cells, the octamer sequence motif is active irrespective of its distance from the site of transcriptional initiation (2, 28). However, in differentiated cells, OCT-4 can transactivate only when the octamer motif is in a proximal position (1, 13, 29); to be active from distal sites, it requires stem cell-specific bridging factors that link it to the transcription initiation site (29). A number of factors such as Sox2, high mobility group, E7, and E1A are known to influence the ability of OCT-4 to act as activator or repressor (15, 29–32). Recently, physical association of OCT-4 with Ewing's sarcoma protein was documented, suggesting that Ewing's sarcoma protein may also play a role in regulating OCT-4 (35).

Although only a single form of OCT-4 mRNA has been identified in embryonic mouse tissues, two forms, *i.e.* OCT-4A and OCT-4B, generated by alternative splicing, were identified in the RT[3]-PCR products from adult human pancreatic islets (36). Compared with human OCT-4A, little is known about the properties of OCT-4B. To identify the biochemical functions of

* This work was supported in part by Grant SC2090 from the Stem Cell Research Center of the 21st Century Frontier Research Program funded by the Ministry of Science and Technology, Republic of Korea, by the Advanced Basic Research Laboratory Program Grant R14-2002-012-01002-0 of the Korea Research Foundation, and by Sogang University Special Research Grant 2005-1511. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Recipient of a Seoul Science fellowship.

[2] To whom correspondence should be addressed. Fax: 82-2-716-2092; E-mail: jkim@sogang.ac.kr.

[3] The abbreviations used are: RT, reverse transcription; ES, embryonic stem cells; EGFP, enhanced green fluorescent protein; EMSA, electrophoretic mobility shift assay; PBS, phosphate-buffered saline; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; CTD, C-terminal domains; NTD, N-terminal domains; NLS, nuclear localization signal; GST, glutathione *S*-transferase.

the human OCT-4B isoform, we performed RT-PCR and sequenced OCT-4 cDNAs from human ES cells. This revealed a novel alternative spliced variant of OCT-4 mRNA in which exon 1a is replaced by exon 1b. The DNA binding and C-terminal domains of OCT-4B are identical to the corresponding domains of OCT-4A, but it lacks the sequences necessary for transactivation. Moreover, it does not bind DNA and mainly localizes to the cytoplasm. We also found that, unlike OCT-4A (35), it cannot stimulate transcription from OCT-4-dependent promoters, nor does it antagonize the induction of gene expression by OCT-4A. In addition, ectopic expression of human OCT-4B in ZHBTc4 ES cells, unlike that of human OCT-4A, was not sufficient to maintain stem cell self-renewal and permit them to display differentiated ES cell phenotypes. These data imply that the DNA binding and transactivation properties of the human OCT-4 isoforms and their abilities to confer self-renewal differ. Thus, the different polypeptides encoded by the human *OCT-4* gene may have different targets as well as different roles as regulators of human ES cells.

## EXPERIMENTAL PROCEDURES

*Materials and General Methods*—Restriction endonucleases, calf intestinal alkaline phosphatase, the Klenow fragment of DNA polymerase I, and T4 DNA ligase were purchased from New England Biolabs. *Pfu* Turbo polymerase was purchased from Stratagene, and [$\gamma$-$^{32}$P]ATP (3000 Ci/mmol) was obtained from PerkinElmer Life Sciences. Preparation of plasmid DNA, restriction enzyme digestion, agarose gel electrophoresis of DNA, DNA ligation, bacterial transformations, and SDS-PAGE of proteins were carried out by standard methods (37). Subclones generated from PCR products were sequenced by the chain termination method with double-stranded DNA templates to ensure the absence of mutations.

*Plasmid Construction*—To construct the human OCT-4 isoforms, total RNA was prepared from human ES cells, and cDNA was synthesized using the Superscript™ first-strand synthesis system (Invitrogen), as described previously (35). The construction of pcDNA3-hOCT-4A has been described previously (35). To construct pcDNA3-OCT-4B, the human ES cell-derived cDNAs were amplified with primer 5′-hOCT4B-1 (5′-GATCGGATCCATGCACTTCTACAGACTATTCCTTGGGGCC-3′ (a BamHI site is underlined)) and 3′-hOCT-4B-40 (5′-pCTGGATTTTAAAAGGCAG-3′), and primers 5′-hOCT-4B-41 (5′-pTCCCAGGACATCAAAGCT-3′) and 3′-OCT-4B-265 (5′-GATCGAATTCTCAGTTTGAATGCATGGG-3′ (an EcoRI site is underlined)). The two PCR products were ligated, digested with BamHI and EcoRI, and cloned into the same sites of pcDNA3.

pCAG-IP/EGFP, pCAG-IP/mOCT-4-EGFP, pCAG-IP/hOCT-4A-EGFP, and pCAG-IP/hOCT-4B-EGFP were generated as follows. (*a*) For pCAG-IP/EGFP, vector pEGFP N1 (Clontech) was digested with XhoI and NotI and cloned into the same sites of vector pCAG-IP. (*b*) For pCAG-IP/mOCT-4-EGFP, mouse OCT-4 was amplified from pcDNA3/mOCT-4 by PCR using primers 5′-mOCT-4-1 (5′-GATCGGATCCATGGCTGGACACCTGGCT-3′, a BamHI site is underlined) and 3′-OCT-4-351 (5′-GATCACCGGTGCTCCGTTTGAATGCATGGG-3′, an AgeI site is underlined), digested with BamHI and AgeI, and cloned into the same sites of pEGFP N1 to generate pEGFP-mOCT-4. This was then digested with XhoI and NotI and cloned into the same sites of vector pCAG-IP. (*c*) For pCAG-IP/hOCT-4A-EGFP, human OCT-4A was amplified from pcDNA3/hOCT-4A by PCR using primers 5′-hOCT-4A-1 (5′-GATCGGATCCATGGCGGGACACCTGGCT-3′, a BamHI site is underlined) and 3′-OCT-4-351, digested with BamHI and AgeI, and cloned into the same sites of pEGFP N1 to generate pEGFP-hOCT-4A. This in turn was digested with SalI and NotI and cloned into the XhoI and NotI sites of pCAG-IP. (*d*) For pCAG-IP/hOCT-4B-EGFP, human OCT-4B was amplified from pcDNA3/hOCT-4B by PCR using primers 5′-hOCT-4B-1 (5′-GATCGGATCCATGCACTTCTACAGACTATTCCTTGGGGCC-3′, a BamHI site is underlined) and 3′-OCT-4-351, digested with BamHI and AgeI, and cloned into the same sites of pEGFP N1 to generate pEGFP-hOCT-4B, which was digested with SalI and NotI and cloned into XhoI and NotI sites of vector pCAG-IP.

GST fusion OCT-4A and GST fusion OCT-4B, and their derivative plasmids GST-OCT-4A(POU$^{A}$), GST-OCT-4B(POU$^{B}$), GST-OCT-4B(N$^{B}$P$^{B}$), GST-OCT-4B(P$^{B}$C), GST-OCT-4A(N$^{B}$P$^{A}$C), GST-OCT-4B(N$^{A}$P$^{B}$C), GST-OCT-4B (N$^{B}$P$^{B}$)$\Delta$(21–40), and GST-OCT-4B (N$^{B}$P$^{B}$)$\Delta$(1–20), were generated by the following steps. (*a*) For GST-OCT-4A, plasmid pcDNA3/OCT-4A was digested with BamHI and EcoRI and cloned into the same sites of pGEX (4T-1) vector (Amersham Biosciences). (*b*) For GST-OCT-4B, plasmid pcDNA3/OCT-4B was digested with BamHI and EcoRI and cloned into the same sites of pGEX (4T-1). (*c*) For GST-OCT-4A (POU$^{A}$), the POU domain of OCT-4A was amplified from pcDNA3/OCT-4A by PCR using primers 5′-hOCT4A-134 (5′-GATCGGATCCGAGGAGTCCCAGGACATC-3′, a BamHI site is underlined) and 3′-hOCT4A-289 (5′-GATCGAATTCGCTTGATCGCTTGCCCTT-3′, an EcoRI site is underlined), digested with BamHI and EcoRI, and cloned into the same sites of pGEX (4T-1). (*d*) For GST-OCT-4B (POU$^{B}$), the POU domain of OCT-4B was amplified from pcDNA3/OCT-4B by PCR using primers 5′-hOCT4B-41 (5′-GATCGGATCCATCCAGTCCCAGGACATC-3′, a BamHI site is underlined) and 3′-hOCT4A-289 (5′-GATCGAATTCGCTTGATCGCTTGCCCTT-3′, an EcoRI site is underlined), digested with BamHI and EcoRI, and cloned into the same sites of pGEX (4T-1). (*e*) For GST-OCT-4B (N$^{B}$P$^{B}$), the N-terminal and POU domain of OCT-4B was amplified from pcDNA3/OCT-4B by PCR using primers 5′-hOCT4B-1 (5′-GATCGGATCCATGCACTTCTACAGACTATTCCTTGGGGCC-3′, a BamHI site is underlined) and 3′-hOCT4A-289 (5′-GATCGAATTCGCTTGATCGCTTGCCCTT-3′, an EcoRI site is underlined), digested with BamHI and EcoRI, and cloned into the same sites of pGEX (4T-1). (*f*) For GST-OCT-4B(P$^{B}$C), the POU and C-terminal domain of OCT-4B was amplified from pcDNA3/OCT-4B by PCR using primers 5′-hOCT4B-41 (5′-GATCGGATCCATCCAGTCCCAGGACATC-3′, a BamHI site is underlined) and 3′-hOCT4A-360 (5′-GATCGAATTCTCAGTTTGAATGCATGGG-3′, an EcoRI site is underlined), digested with BamHI and EcoRI, and cloned into the same sites of pGEX (4T-1). (*g*) For GST-OCT-4A(N$^{B}$P$^{A}$C), the N-terminal domain of OCT-4B or the POU and C-terminal domains of OCT-4A were amplified from pcDNA3/OCT-4B or pcDNA3/OCT-4A

by PCR using primers 5′-hOCT4B-1 and 3′-hOCT4B-40 (5′-CTCCTCCTGGATTTTAAAAGGCAGAAGA-3′) or primers 5′hOCT4A-134B (5′-ATCCAGGAGGAGTCCCAGGACATCAAAG-3′) and 3′-hOCT4A-360, respectively. The secondary PCR was performed using gel-purified primary PCR products as templates and primers 5′-hOCT4B-1 and 3′-hOCT4A-360. The secondary PCR products were digested with BamHI and EcoRI and cloned into the same sites of pGEX (4T-1). (*e*) For GST-OCT-4B($N^AP^BC$), the N-terminal domain of OCT-4A or the POU and C-terminal domains of OCT-4B were amplified from pcDNA3/OCT-4A or pcDNA3/OCT-4B using primers 5′-hOCT4A-1 and 3′-hOCT4A-137 (5′-CTGGGACGGGTTTTGCTCCAGCTTCTCC-3′) or primers 5′-hOCT4B-41B (5′-AACCCGTCCCAGGACATCAAAGCTCTGC-3′) and 3′-hOCT4A-360, respectively. The secondary PCR was performed using gel-purified primary PCR products as templates and primers 5′-hOCT4A-1 and 3′-hOCT4A-360. The secondary PCR products were digested with BamHI and EcoRI and cloned into the same sites of pGEX (4T-1). (*f*) For GST-OCT-4B ($N^BP^B$)Δ(21–40), a fragment of the human OCT-4B sequence extending from the N-terminal domain through the POU DNA binding domain was cloned into pKSII vector (Stratagene). A deletion of amino acids 21–40 was introduced by site-directed mutagenesis using a mutagenic primer set (5′-CTTGAATCCCGAATGGAAATCCCAGGACATCAAAGCTC-3′ and 5′-GAGCTTTGATGTCCTGGGATTTCCATTCGGGATTCAAG-3′) and the QuikChange kit™ (Stratagene). Then the OCT-4B ($N^BP^B$)Δ(21–40) fragment of the pKSII-OCT-4B ($N^BP^B$)Δ(21–40) construct was digested with BamHI and EcoRI and cloned into the same sites of pGEX (4T-1). (*g*) For GST-OCT-4B ($N^BP^B$)Δ(1–20), the ($N^BP^B$)Δ(1–20) fragment of OCT-4B was amplified from pcDNA3/OCT-4B by PCR using primers 5′-hOCT4B-21 (5′-GATCGGATCCGGGGAGATTGATAACTGG-3′, a BamHI site is underlined) and 3′-hOCT4–289, digested with BamHI and EcoRI, and cloned into the same sites of pGEX (4T-1).

The reporter construct pOCT-4(10x)TATAluc has been described (35). The FOR promoter (−508 to +1) was cloned by PCR amplification from pGL3-FOR(3.5 kb).[4] The primers used for this purpose were 5′-FOR(−508) (5′-GATCGGTACCCCATGCCCCCTT-3′, a KpnI site is underlined) and 3′-FOR(+1) (5′-GATCTTCGAATAACCCCTATTCG-3′, a HindIII site is underlined). The PCR product was digested with KpnI and HindIII and cloned into the same sites of the promoterless pGL3 basic vector (Promega).

*Cell Culture*—Human ES cells (Miz-hES1, SNU-hESC3, and Cha-hESC3) were grown as described previously (38, 39). Briefly, they were cultured in Dulbecco's modified Eagle's medium/F12 medium with 20% knock-out serum replacement, 1 mM L-glutamine, 1% nonessential amino acids, 0.1 mM β-mercaptoethanol, and 4 ng/ml basic fibroblast growth factor (Invitrogen) on mitomycin C-treated mouse embryonic fibroblast feeders at 37 °C and 5% $CO_2$. After 5 days of culture, colonies were detached mechanically from the feeder cells with a micropipette, and individual colonies were mechanically divided into four or five pieces. These ES cell clumps were then separately plated on fresh feeder cell layers. HEK293T or NIH3T3 cells were maintained in Dulbecco's modified Eagle's medium with 10% heat-inactivated fetal calf serum (Invitrogen), penicillin, and streptomycin.

*RT-PCR and Alkaline Phosphatase Staining*—Total RNA was prepared from human ES cells using an RNeasy mini kit (Qiagen) with on-column DNase treatment, and messenger RNA was purified with an Oligodex-dT mRNA mini kit (Qiagen) followed by cDNA synthesis using a Superscript first-strand synthesis system for RT-PCR (Invitrogen), as instructed by the manufacturers. To detect human OCT-4 isoforms in human ES cells, human OCT-4A and OCT-4B cDNAs were amplified with human OCT-4A primers [5′e1a (5′-GATCGGATCCATGGCGGGACACCTGGCT-3′) and 3′e1 (5′-CCTTCCCAAATAGAACCC-3′)], and human OCT-4B primers [5′e1b (5′-GATCGGATCCATGCACTTCTACAGACTATTCCTTGGGGCC-3′) and 3′e1], respectively.

RT-PCRs for *Oct-4* downstream target genes and differentiation marker genes were performed with gene-specific primer sets as described previously (9, 11, 30, 41). Alkaline phosphatase was stained with an AP staining kit (Sigma).

*Quantitative Real Time PCR*—Quantitative real time PCR was performed with an Applied Biosystems 7500 Fast real time PCR system (Applied Biosystems) and SYBR Green Master Mix (Applied Biosystems), as described previously (43). As a control, the level of GAPDH mRNA was determined in the real time PCR assay of each RNA sample and was used to correct for experimental variation. The following primer sequences were used: hOCT-4A forward primer was 5′-CTCCTGGAGGGCCAGGAATC-3′, and hOCT-4A reverse primer was 5′-CCACATCGGCCTGTGTATAT-3′. The hOCT-4B forward primer was 5′-ATGCATGAGTCAGTGAACAG-3′, and the hOCT-4B reverse primer was 5′-CCACATCGGCCTGTGTATAT-3′. The GAPDH forward primer was 5′-GAAGGTGAAGGTCGGAGTC-3′, and the GAPDH reverse primer was 5′-GAAGATGGTGATGGGATTTC-3′. Quantitation of the relative expression levels of the human OCT-4 isoforms was achieved by normalizing for the endogenous GAPDH using the $\Delta C_T$ method of quantitation.

*ES Cell-based Complementation Assay*—ES cell-based complementation assays were performed with ZHBTc4 ES cells as described previously (44). Fluorescence was detected with a fluorescence microscope (Olympus, 1X51) equipped with a CoolSNAP digital camera (Olympus).

*Expression and Purification of GST Fusion Proteins*—GST-OCT-4 proteins were generated in *Escherichia coli* as described previously (45). After binding to glutathione-Sepharose, the proteins were washed and eluted with reduced glutathione (Sigma). Protein concentrations were determined by the method of Bradford (Bio-Rad). The purity and size of the eluted proteins were evaluated by Coomassie staining of SDS-polyacrylamide gels.

*Western Blot Analysis*—Western blot analysis was performed using an anti-OCT-4 antibody (C-20; Santa Cruz Biotechnology), and reactive bands were detected by chemiluminescence using Western Lightening (PerkinElmer Life Sciences).

[4] J. Lee and J. Kim, manuscript in preparation.

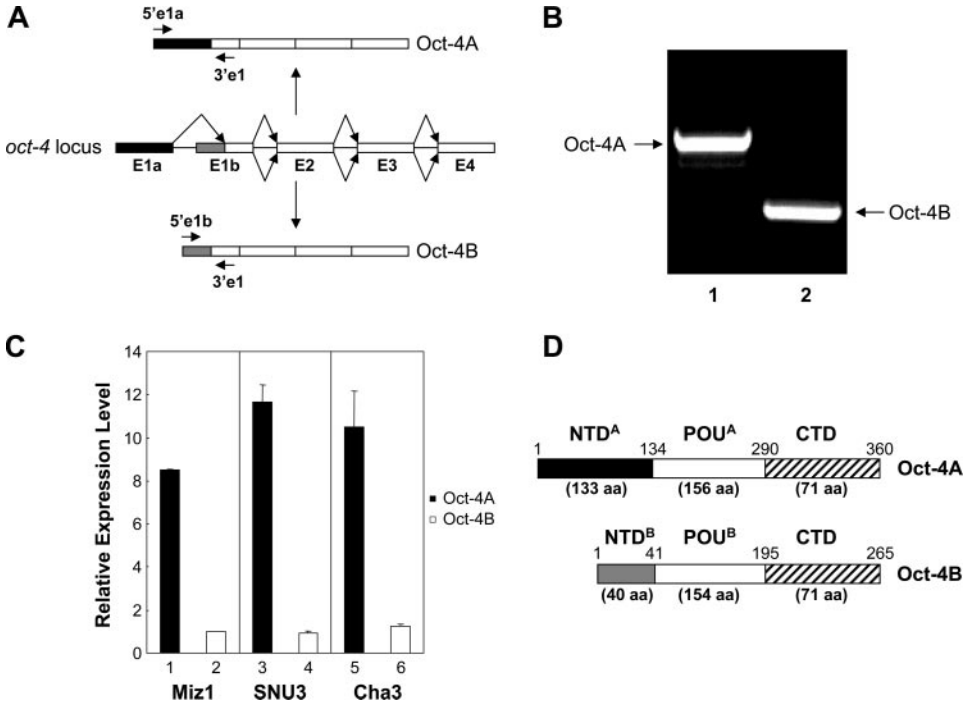


FIGURE 1. **Expression of OCT-4 isoforms in human ES cells.** *A,* schematic representation of human OCT-4 transcripts. OCT-4A and OCT-4B are derived from the same gene by alternative splicing. The N-terminal domain of human OCT-4A is indicated in *black,* and the 40 amino acids at the unique N terminus of OCT-4B in are shown in *gray*. *B,* expression of OCT-4A and OCT-4B mRNAs in human ES cells. The novel human *OCT-4B* transcript was identified from human ES cells by RT-PCR using the 5′e1b and 3′e1 primer pair. The amplified product was resolved on a 2% agarose gel stained with ethidium bromide, and the negative image of the photograph is shown. *Lane 1,* RT-PCR for human *OCT-4A* transcript using the 5′e1a/3′e1 primer pair; *lane 2,* RT-PCR for human *OCT-4B* transcript using the 5′e1b/3′e1 primer pair. *C,* quantitative real time PCR analyses of the expression of human OCT-4 isoform mRNAs. Three different human ES cell lines (Miz-hES1, SNU-hESC3, and Cha-hESC3) were analyzed for levels of expression of OCT-4A (*black bars*) and OCT-4B (*white bars*) mRNAs as described under "Experimental Procedures." GAPDH was used as a control to normalize the quantitative real time PCR results. Values are expressed as means ± S.E. relative to the expression of human OCT-4B in Miz-hES1, which was set at 1. *Miz1*, Miz-hES1; *SNU3*, SNU-hESC3; *Cha3*, Cha-hESC3. *D,* schematic representation of the protein domains of the human OCT-4 isoforms. The N-terminal domains (*NTD*), POU DNA binding domains (*POU*), and C-terminal domains (*CTD*) of the isoforms are depicted together with the amino acid (*aa*) positions. The unique 133 amino acids at the N terminus of OCT-4A are shown in *black* and the unique 40 amino acids at the N terminus of OCT-4B in dark *gray*.

*Electrophoretic Mobility Shift Assays (EMSAs)*—Probes for EMSAs were prepared from synthetic oligonucleotides whose sequences have been described previously (46). The probes were prepared by end labeling annealed complementary oligonucleotides with [$\gamma$-$^{32}$P]ATP using T4 polynucleotide kinase. EMSAs were performed with recombinant GST-OCT-4 proteins (OCT-4A, OCT-4B, and their derivatives) for 30 min at 4 °C in binding buffer containing 20 mM HEPES (pH 7.4), 40 mM KCl, 1 mM $MgCl_2$, 0.1 mM EDTA, 0.1% Nonidet P-40, and 0.2 $\mu$g of poly(dI·dC)·(dI·dC). Following binding, the reaction mixtures were run on 4% polyacrylamide gels (acrylamide/bisacrylamide ratio of 37:1) in 0.5× TBE (44.5 mM Tris-HCl, 44.5 mM boric acid, 1 mM EDTA) buffer at 150 V for 2–3 h at 4 °C. The gels were dried and exposed to Kodak X-Omat film at −70 °C.

*Transfection and Reporter Assays*—Cells were transiently transfected by electroporation with the Gene Pulser II RF module system (Bio-Rad), as instructed by the manufacturer. Luciferase assays were performed with the dual-luciferase assay system (Promega). *Renilla* luciferase activity was used to normalize transfection efficiencies.

*Subcellular Localization Experiments*—Full-length human OCT-4A or OCT-4B cDNAs were subcloned into the BamHI/EcoRI sites of pBabePuro. NIH3T3 cells were plated on glass coverslips and infected with high titer retrovirus stocks produced by transient transfection of Phoenix cells (47). Immunocytochemical analyses were performed as described previously (45). Briefly, the infected cells were washed in phosphate-buffered saline (PBS) and fixed for 10 min at −20 °C with acetone/methanol (1:1, v/v). To detect human OCT-4A or OCT-4B, we used anti-OCT-4 antibody (C-20; Santa Cruz Biotechnology) and horseradish peroxidase-conjugated secondary antibody (Santa Cruz Biotechnology), and fluorescence was detected with a confocal laser scanning microscope (LSM5 Pascal; Carl Zeiss Co., Ltd.).

To localize the human OCT-4 isoforms in ES cells, pCAG-IP/EGFP, pCAG-IP/mOCT-4-EGFP, pCAG-IP/hOCT-4A-EGFP, or pCAG-IP/hOCT-4B-EGFP-transfected ZHBTc4 ES cells were washed in PBS and fixed for 10 min at −20 °C in a mixture of acetone and methanol (1:1, v/v). The coverslips were mounted with 50% glycerol/PBS, and green fluorescence was detected with a fluorescence microscope (IX51; Olympus, Tokyo, Japan) equipped with a CoolSNAP digital camera (Olympus).

## RESULTS

*Identification of OCT-4 Isoforms in Human Embryonic Stem Cells*—Transcripts of the human OCT-4 isoforms (OCT-4A and OCT-4B) were first observed as RT-PCR products during a search for transcription factors containing POU-domains in adult human pancreatic islets (36). We examined OCT-4 isoforms in human embryonic stem cells by RT-PCR using total human ES cell (Miz-hES1) RNA with specific oligonucleotide primers flanking the coding region corresponding to the N-terminal domain of OCT-4 proteins. As shown in Fig. 1*A*, OCT-4A mRNA was generated from exon 1a (labeled *E1a*), the 3′-half of exon 1b (*E1b*), and exons 2–4 (*E2, E3,* and *E4*) using an internal splicing acceptor site in exon 1b. OCT-4B mRNA contains exons 1b to 4 (Fig. 1*A*). PCR products of the expected size of 532 and 247 bp derived from human OCT-4A (Fig. 1*B, lane 1*) and OCT-4B (*lane 2*) mRNA, respectively, were observed using human ES cell RNA. These RT-PCR products were sequenced and confirmed to represent nucleotides 1–532 and 1–247 of the reported human OCT-4A or OCT-4B cDNA sequences, respectively (data not shown). These results point to the existence of two isoforms of human OCT-4 differing in

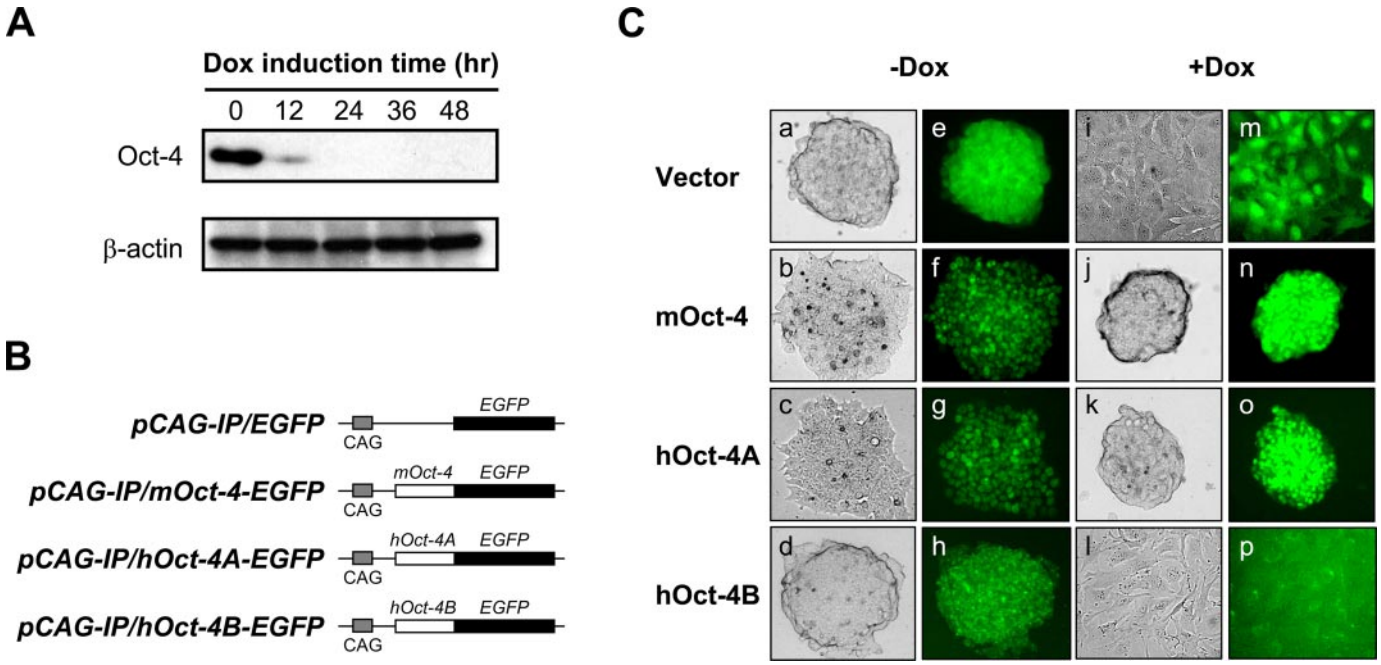


FIGURE 2. **Differential abilities of the human OCT-4 isoforms to rescue self-renewal of embryonic stem cell.** *A,* Western blot analysis of OCT-4 expression in ZHBTc4 ES cells harboring the mouse *Oct-4* gene under the control of the inducible tetracycline promoter. Total cell extracts were prepared from ZHBTc4 ES cells 0, 12, 24, 36, and 48 h following doxycycline treatment. The extracts were resolved by 15% SDS-PAGE, transferred to a polyvinylidene difluoride membrane, and immunoblotted with anti-OCT-4 antibody (C-10; Santa Cruz Biotechnology). *B,* schematic representation of the OCT-4 isoform EGFP fusions used in this study. Expression vectors pCAG-IP/mOCT-4-EGFP, pCAG-IP/hOCT-4A-EGFP, and pCAG-IP/hOCT-4B-EGFP correspond to mouse OCT-4, human OCT-4A, and human OCT-4B fused to EGFP, respectively. The pCAG-IP/EGFP expression vector was used as a control. CAG expression units (CAGs) (44) are indicated by *shaded boxes*, and mouse OCT-4, human OCT-4A, and human OCT-4B are represented by *white boxes*, and EGFP is indicated by *solid boxes*. *C,* human OCT-4B is defective in ability to confer self-renewal. ZHBTc4 ES cells were transfected with pCAG-IP/EGFP (vector), pCAG-IP/mOCT-4-EGFP (mOCT-4), pCAG-IP/hOCT-4A-EGFP (hOCT-4A), or pCAG-IP/hOCT-4B-EGFP (hOCT-4B), and cultured in the absence (−Dox, *left two panels*) or in the presence (+Dox, *right two panels*) of doxycycline. Phase-contrast (*panels a–d* and *i–l*) and fluorescence views (*panels e–h* and *m–p*) are shown.

their N-terminal domain, as reported previously (36). Using quantitative real time PCR, we were also able to quantify levels of mRNA for the human OCT-4 isoforms. In human ES cells (Miz-hES1) human OCT-4A mRNA was 8-fold more abundant than human OCT-4B (Fig. 1*C, bars 1* and *2*). Similar results were obtained in other human ES cell lines (SNU-hESC3 (Fig. 1*C, bars 3* and *4*) and Cha-hESC3 (Fig. 1*C, bars 5* and *6*)).

*Structural Features of Human OCT-4B*—Human OCT-4A and OCT-4B isoforms are composed of 360 and 265 amino acids, respectively, of which the 225 amino acids at the C termini are identical. As schematically shown in Fig. 1*D*, human OCT-4B shares two conserved regions with OCT-4A. These regions correspond to the DNA binding domain (POU) and the C-terminal domain (CTD). The CTDs (the 71 amino acids) of human OCT-4A and OCT-4B are identical, but OCT-4B POU domain ($POU^B$, total 154 amino acids) lacks two amino acids of the N-terminal sequence of the OCT-4A POU domain ($POU^A$, total 156 amino acids). Outside these regions OCT-4B has little similarity to OCT-4A. In particular it lacks a recognizable transactivation domain, present at the N terminus of human OCT-4A. Thus, the transactivation and self-renewal properties of OCT-4B may not be similar to those of OCT-4A protein.

*Different Abilities of the Human OCT-4 Isoforms to Confer Self-renewal*—To investigate the abilities of the human OCT-4 isoforms to confer self-renewal on ES cells, we performed ES cell-based complementation assays (44) using ZHBTc4 ES cells. This ES cell line has both endogenous alleles of OCT-4 inactivated by gene targeting and harbors the tetracycline-repressible mouse *Oct-4* transgene (11). It can be propagated as a mouse ES cell line in the absence of tetracycline, in which condition the *Oct-4* transgene is active, but not in the presence of tetracycline, which represses the transgene (44). In this system, the tetracycline analogue, doxycycline, can be used to prevent expression of OCT-4. Addition of doxycycline to the growth medium of ZHBTc4 ES cells resulted in rapid repression of OCT-4 expression, as determined by Western blotting of total cell extracts (Fig. 2*A*).

To test whether the stem cell phenotype can be rescued by transfections of human OCT-4 isoforms, OCT-4A or OCT-4B cDNAs under the control of the constitutive CAG expression unit were transfected into ZHBTc4 ES cells, respectively. To identify the transfected ES cells, we constructed plasmids expressing green fluorescent protein fusions of the human OCT-4 isoforms. The structures of expression vectors used are shown schematically in Fig. 2*B*. Consistent with the previous reports (11, 44), the mouse OCT-4-transfected ZHBTc4 ES cells were able to differentiate because of superthreshold production of OCT-4 (Fig. 2*C, panels b* and *f*). In addition, growth of the transfectants in the presence of doxycycline rescued their self-renewal ability and stem cell phenotype (Fig. 2*C, panels j* and *n*). As a control, ZHBTc4 ES cells transfected with EGFP and cultured in the presence of doxycycline underwent differentiation (Fig. 2*C, panels i* and *m*).

To investigate the behavior of the OCT-4 gene products of different species, we introduced an expression construct for human OCT-4A into the ZHBTc4 ES cells and performed the

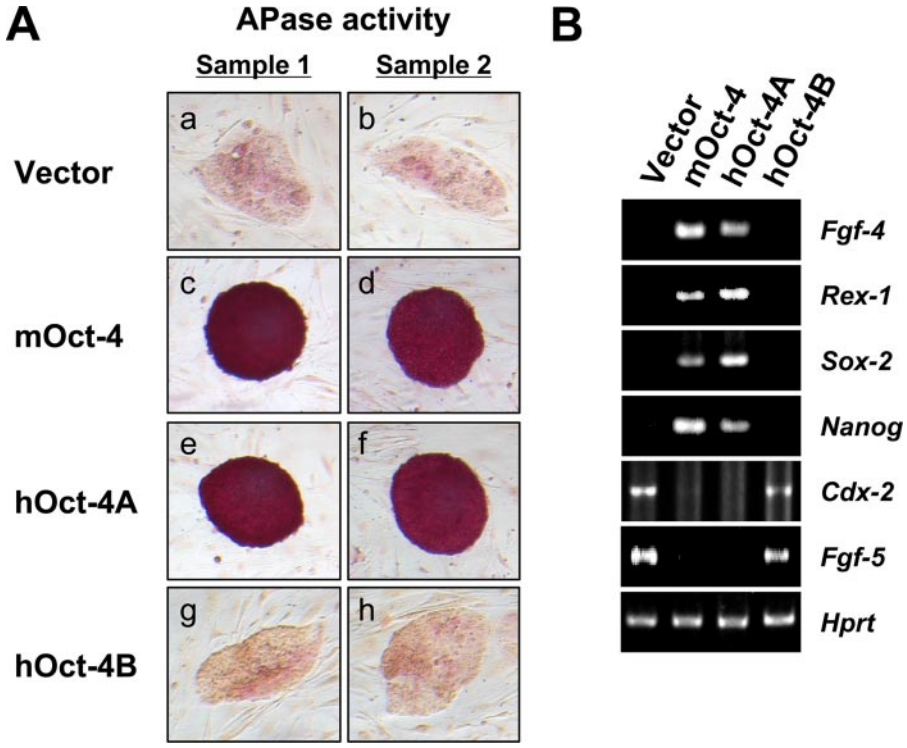


FIGURE 3. **Characterization of ZHBTc4 ES cells expressing human OCT-4 isoforms.** *A*, expression of alkaline phosphatase. Alkaline phosphatase activity was assessed in ZHBTc4 ES cells expressing human OCT-4 isoforms. The expression vectors used in this experiment were the same as those in Fig. 2*B*. *B*, expression of OCT-4 downstream target genes and lineage-specific markers. RT-PCR analyses of *Fgf-4*, *Rex-1*, *Sox-2*, *Nanog*, *Cdx-2*, and *Fgf-5* mRNAs were performed in ZHBTc4 ES cells expressing vector, mouse OCT-4, or human OCT-4 isoforms. *Hprt* was used as a control to quantify the RT-PCR results. Following amplification, an aliquot of each product was analyzed by staining the gel with ethidium bromide. The ES cell lines from which the input RNAs used in the RTs are derived are shown *above* the panel.

ES cell-based complementation assay. The ZHBTc4 ES cells cultured in the absence of doxycycline again differentiated because of superthreshold production of OCT-4 (Fig. 2*C*, *panels c* and *g*), and their self-renewal ability was rescued in the presence of doxycycline, indicating that human OCT-4A protein is active in mouse ZHBTc4 ES cells (Fig. 2*C*, *panels k* and *o*). However, contrary to expectation, when human OCT-4B was introduced into ZHBTc4 ES cells, it failed to rescue stem cell renewal in the presence of doxycycline (Fig. 2*C*, *panels l* and *p*). These data indicate that the abilities of human OCT-4A and OCT-4B isoforms to confer self-renewal on ES cells differ.

*Different Abilities of Human OCT-4 Isoforms to Maintain the Undifferentiated State*—The flat morphology of doxycycline-treated ZHBTc4 ES cells expressing human OCT-4B (Fig. 2*C*) suggested failure to maintain the undifferentiated state of the ES cells. To verify this hypothesis, we examined known molecular markers of undifferentiated ES cells. To stain for alkaline phosphatase activity (a marker of pluripotent cells of embryonic origin) (48), the pCAG-IP/EGFP, pCAG-IP/mOCT-4-EGFP, pCAG-IP/hOCT-4A-EGFP, and pCAG-IP/hOCT-4B-EGFP constructs (Fig. 2*B*) were linearized with PvuI, and 50 $\mu$g of each linearized plasmid DNA was transfected into ZHBTc4 ES cells ($1 \times 10^7$) using a Gene Pulser II RF module system. 48 h post-electroporation, puromycin was added to the medium at a final concentration of 1 $\mu$g/ml to select transfected ZHBTc4 ES clones. After selection of pCAG-IP/EGFP-, pCAG-IP/mOCT-4-EGFP-, pCAG-IP/hOCT-4A-EGFP-, or pCAG-IP/hOCT-4B-EGFP-transfected ZHBTc4 cells, $4 \times 10^3$ transfectants were seeded onto 35-mm dishes containing mitomycin C-treated mouse embryonic fibroblast feeder layers. After 2 days of culture, the cells were transferred to medium containing 1 $\mu$g/ml doxycycline to repress the tetracycline-repressible mouse *Oct-4* transgene. After another 4 days, colonies resistant to puromycin were tested for OCT-4-EGFP protein expression by examining green fluorescence and staining the cells for alkaline phosphatase. ZHBTc4 ES cells expressing human OCT-4A contained alkaline phosphatase activity (Fig. 3*A*, *panels e* and *f*) comparable with that in ZHBTc4 ES cells expressing mouse OCT-4 (Fig. 3*A*, *panels c* and *d*). However, the clones expressing human OCT-4B (Fig. 3*A*, *panels g* and *h*) or EGFP alone (Fig. 3*A*, *panels a* and *b*) had lost this characteristic, pointing to failure to maintain the undifferentiated state.

To investigate further the features of the ZHBTc4 ES cells transfected with the human OCT-4 isoforms, we investigated levels of expression of known OCT-4 downstream target genes such as *Fgf-4* (20), *Rex-1* (50), *Sox-2* (51), and *Nanog* (17, 18) by RT-PCR. As shown in Fig. 3*B*, expression of all four *Oct-4* downstream target genes was detected in the tetracycline-treated ZHBTc4 ES cells transfected with mouse OCT-4 (*2nd lane*) or human OCT-4A (*3rd lane*) cDNAs. However, these genes were down-regulated in tetracycline-treated ZHBTc4 ES cells expressing human OCT-4B (Fig. 3*B*, *4th lane*). Tetracycline-treated ZHBTc4 ES cells expressing EGFP served as a negative control (Fig. 3*B*, *1st lane*).

We also evaluated the expression of markers of differentiated cells. As shown in Fig. 3*B*, expression of *Cdx-2* mRNA, which is implicated in trophoblast differentiation (52), was detected in ZHBTc4 ES cells expressing human OCT-4B or vector but was not present in ZHBTc4 ES cells expressing mouse OCT-4 or human OCT-4A. *Fgf-5* mRNA, a marker of primitive ectoderm (53), also appeared in ZHBTc4 ES cells expressing human OCT-4B, indicating formation of the ectoderm lineage. These properties all suggest that human OCT-4A-expressing ES cells expanded normally and remained in an undifferentiated status, whereas human OCT-4B-expressing ES cells did not.

*DNA Binding Properties of Human OCT-4 Isoforms*—Because the results presented above indicate that human OCT-4B is not able to induce stem cell self-renewal, we examined its biochemical characteristics. In order see whether it can bind to an authentic OCT-4 DNA recognition sequence, we studied the DNA binding properties of the human OCT-4 isoforms using

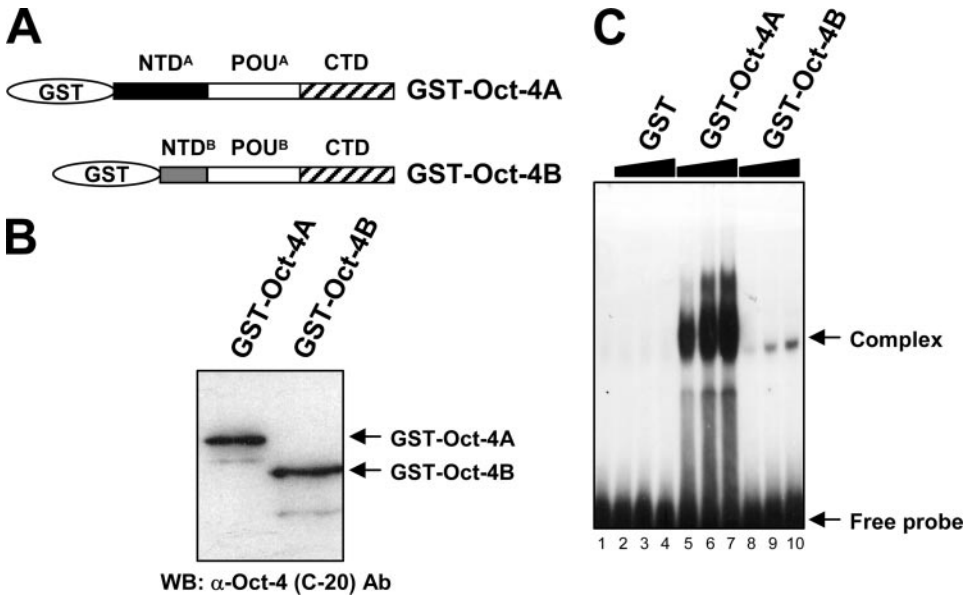


FIGURE 4. **Binding of human OCT-4 isoforms to oligonucleotides.** *A,* schematic representation of the GST fusion OCT-4 isoforms used in this study. Constructs GST-OCT-4A and GST-OCT-4B represent the GST domain fused to the OCT-4A and OCT-4B, respectively. The functional domains of the OCT-4 isoforms are also shown. *B,* immunoblot analysis of OCT-4 isoforms to quantify the amounts of GST fusion proteins. The GST fusion OCT-4 isoforms utilized in the EMSAs were fractionated on 12% SDS-PAGE and visualized by Western blotting (*WB*) with anti-OCT-4 antibody (C-20; Santa Cruz Biotechnology). *C,* DNA binding by the OCT-4 isoforms. Following preparation of radiolabeled probes, EMSAs were performed with either no recombinant protein (*lane 1*), recombinant GST (*lane 2*, 50 ng; *lane 3*, 100 ng; *lane 4*; 150 ng), GST-OCT-4A (*lane 5*, 50 ng; *lane 6*, 100 ng; *lane 7*; 150 ng), or GST-OCT-4B (*lane 8*, 50 ng; *lane 9*, 100 ng; *lane 10*, 150 ng) as described under "Experimental Procedures." The recombinant proteins used in each EMSA is indicated *above* the gel. Protein-DNA complexes were resolved on nondenaturing 4% polyacrylamide (acrylamide:bisacrylamide ratio, 37:1) gels run at 4 °C in 0.5× TBE (44.5 mM Tris-HCl, 44.5 mM boric acid, 1 mM EDTA). The positions of the free probe and protein-DNA complexes are indicated.

EMSAs. An oligonucleotide containing the consensus OCT-4 DNA-binding sequence was synthesized and used as target in the binding reactions. The structures of the OCT-4 isoform GST fusions used in this study are shown schematically in Fig. 4*A*. They were expressed in *E. coli*, purified, and coupled to glutathione-Sepharose beads. The affinity-purified OCT-4 isoforms were fractionated by SDS-PAGE, transferred to a polyvinylidene difluoride membrane, and immunoblotted with an anti-OCT-4 antibody (C-20; Santa Cruz Biotechnology) to quantify the amount of protein in each sample. As shown in Fig. 4*B*, the purified fractions contained equal amounts of protein.

EMSAs were performed with the concentration of the OCT-4 probe kept constant, and the amount of input protein varied. GST-OCT-4A bound to the DNA (Fig. 4*C, lanes 5–7*), and binding was specific as it was displaced by an excess of cold oligonucleotide containing the OCT-4-binding site but not by an excess of cold mutant oligonucleotide containing a mutant OCT-4-binding sequence not recognized by OCT-4 (data not shown). However, the OCT-4B isoform hardly bound at all (Fig. 4*C, lanes 8–10*). These results indicate either that the N-terminal domain of OCT-4A is required for efficient DNA binding or that the 40 amino acids present in the N-terminal domain of OCT-4B inhibit binding.

*The N-terminal Domain of the OCT-4B Isoform Inhibits DNA Binding*—To distinguish between these possibilities, we produced a number of mutants with deletions of the N or C termini of OCT-4B, and we evaluated their DNA binding properties. The structure of the OCT-4B deletion and domain-swapping mutants is shown schematically in Fig. 5*A*. The superscripts "A" and "B" indicate whether a particular domain was derived from OCT-4A or OCT-4B. As a result of alternative splicing, the OCT-4B POU domain (total 154 amino acids) lacks two amino acids of the N-terminal sequence of the OCT-4A POU domain (total 156 amino acids) (Fig. 1*D*) (36). EMSAs with the POU domains of OCT-4A and OCT-4B on their own revealed that both bound the probe (Fig. 5*B*, *lanes 5–7* and *8–10*). Moreover, the addition of the C-terminal domain of OCT-4B to the POU DNA binding domain of OCT-4B (OCT-4B ($P^BC$)) produced a recombinant protein still capable of forming protein-DNA complex (Fig. 5*C, lanes 8–10*), whereas the addition of the N-terminal domain of OCT-4B (OCT-4b ($N^BP^B$)) essentially abolished DNA binding (Fig. 5*C, lanes 5–7*). These results show that the N-terminal domain of OCT-4B inhibits POU DNA binding.

To further test the contribution of the N-terminal domain of OCT-4B to DNA binding, we generated a series of chimeric proteins in which the order of the domains remained constant, *i.e.* NTD-POU-CTD, but domains were swapped between OCT-4A and OCT-4B. As shown in Fig. 5*D*, OCT-4A ($N^BP^AC$) was unable to bind the OCT-4 probe, whereas OCT-4B ($N^AP^BC$) was able to bind it. This confirms that the N-terminal domain of OCT-4B contains some sequence or sequences that inhibit DNA binding and that this is active on $POU^A$ as well as $POU^B$.

*Two Separate Regions in OCT-4B NTD Are Responsible for Inhibiting OCT-4B DNA Binding*—Fusion of the N-terminal domain of OCT-4B to the POU and C-terminal domains of OCT-4A revealed inhibition of DNA binding by the 40-amino acid sequence of OCT-4B (Fig. 5*D*). To define the minimum inhibitory domain, we tested partial deletions of the N terminus in which either amino acids 21–40 or 1–20 were deleted. The structure of the OCT-4B deletion mutants used is shown in Fig. 6*A*. These truncation mutants were expressed as GST fusion proteins in *E. coli* and purified to near homogeneity in roughly equal yield (Fig. 6*B*). As shown in Fig. 6*C*, neither partial deletion mutant was able to bind DNA. These results demonstrate that the N-terminal domain of OCT-4B possesses at least two independently acting sequences that inhibit POU DNA binding.

*OCT-4A Is Nuclear whereas OCT-4B Is Cytoplasmic*—We also determined the intracellular locations of OCT-4A and OCT-4B. For this purpose, we infected NIH3T3 cells with a retroviral expression construct, pBabePuro, containing the

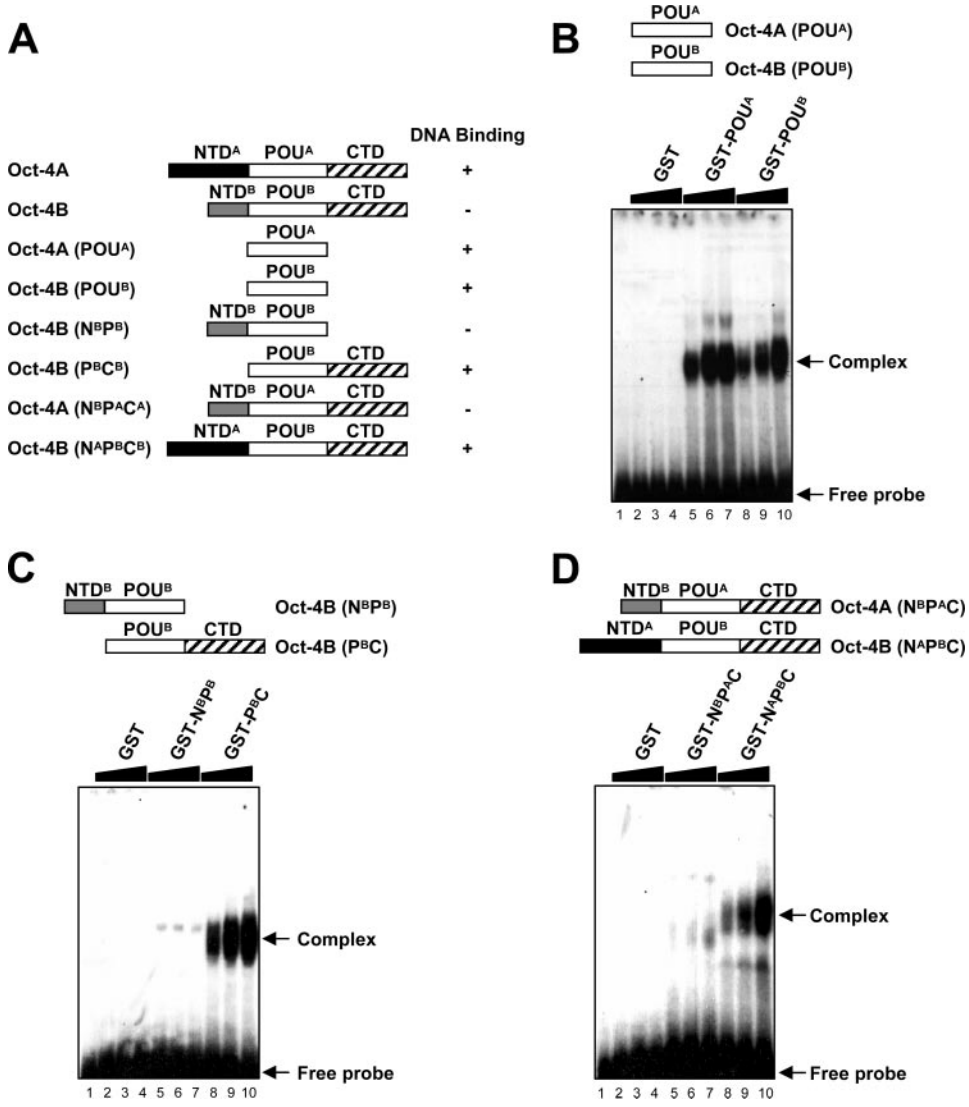


FIGURE 5. **Deletion of the N-terminal domain permits OCT-4B to bind DNA.** *A,* schematic representation of the GST fusion OCT-4A, GST fusion OCT-4B, and derivatives in which domains were deleted or swapped and their ability to bind to DNA. The functional domains located within the OCT-4 isoforms are indicated. (*NTD*$^A$ indicates the N-terminal domain of OCT-4A, etc). The + or − symbols to the *right* refer to the binding ability of the GST fusions to DNA (+, binding; − no binding). *B,* DNA binding activity of OCT-4A (POU$^A$) and OCT-4B (POU$^B$). EMSAs were performed with either no recombinant protein (*lane 1*), recombinant GST (*lane 2*, 50 ng; *lane 3*, 100 ng; *lane 4*; 150 ng), GST-OCT-4A (POU$^A$) (*lane 5*, 50 ng; *lane 6*, 100 ng; *lane 7*; 150 ng), or GST-OCT-4B (POU$^B$) (*lane 8*, 50 ng; *lane 9*, 100 ng; *lane 10*, 150 ng) as described under "Experimental Procedures." The recombinant proteins used in the EMSAs are indicated *above* the gel. Protein-DNA complexes were resolved as in Fig. 4*C*. *C,* DNA binding activity of OCT-4B deletion mutants. EMSAs were performed with either no recombinant protein (*lane 1*), recombinant GST (*lane 2*, 50 ng; *lane 3*, 100 ng; *lane 4*; 150 ng), GST-OCT-4B (N$^B$P$^B$) (*lane 5*, 50 ng; *lane 6*, 100 ng; *lane 7*; 150 ng), or GST-OCT-4B (P$^B$C) (*lane 8*, 50 ng; *lane 9*, 100 ng; *lane 10*, 150 ng) as described under "Experimental Procedures." *D,* DNA binding of OCT-4A (N$^B$P$^A$C) and OCT-4B (N$^A$P$^B$C) chimeric molecules. EMSAs were performed with either no recombinant protein (*lane 1*), recombinant GST (*lane 2*, 50 ng; *lane 3*, 100 ng; *lane 4*; 150 ng), GST-OCT-4A (N$^B$P$^A$C) (*lane 5*, 50 ng; *lane 6*, 100 ng; *lane 7*; 150 ng), or GST-OCT-4B (N$^A$P$^B$C) (*lane 8*, 50 ng; *lane 9*, 100 ng; *lane 10*, 150 ng) as described under "Experimental Procedures."

entire OCT-4 coding region. A pBabePuro retroviral vector lacking OCT-4 cDNA sequences was used as negative control. The cells were infected with either empty pBabePuro expression vector (data not shown), pBabePuro-OCT-4A (Fig. 7*A*, *panels a* and *b*), or pBabePuro-OCT-4B (Fig. 7*A*, *panels c* and *d*) and processed for immunofluorescence. We previously showed that mouse OCT-4 protein is localized to the nucleus (35). In accordance with this previous result, human OCT-4A was clearly localized to the nucleus (Fig. 7*A*, *panel a*). However, contrary to our expectation, we found that most of the human OCT-4B protein was cytoplasmic (Fig. 7*A*, *panel c*). We also transfected COS-7 cells with pcDNA3-OCT-4B or pEGFP-OCT-4B with similar results (data not shown).

To confirm these results in cells possessing stem cell properties, the subcellular distribution of the human OCT-4 isoforms was also determined in ZHBTc4 ES cells (Fig. 7*B*). pCAG-IP/EGFP, pCAG-IP/hOCT-4A-EGFP, and pCAG-IP/hOCT-4B-EGFP constructs that express EGFP, OCT-4A-EGFP (GFP fusion human OCT-4A), and OCT-4B-EGFP (GFP fusion human OCT-4B), respectively (Fig. 2*B*), were transfected into ZHBTc4 ES cells, and the subcellular locations of these proteins were detected by fluorescence microscopy. In the ZHBTc4 ES cells, EGFP-tagged OCT-4A was clearly localized to the nucleus (Fig. 7*B*, *panels c* and *d*), whereas most of the EGFP-tagged OCT-4B was cytoplasmic (Fig. 7*B*, *panels e* and *f*). EGFP alone was found in both the nucleus and cytoplasm of ES cells (Fig. 7*B*, *panels a* and *b*). Taken together, these data suggest that OCT-4A is a nuclear protein whereas OCT-4B is a cytoplasmic protein.

*OCT-4B Is Unable to Activate Transcription of OCT-4-responsive Genes*—OCT-4B lacks the N-terminal sequence that mediates the transcriptional activity of OCT-4 (13). In addition, its N-terminal domain inhibits DNA binding (Figs. 5 and 6), and it is cytoplasmic in location (Fig. 7). All these properties suggest that it may not be a transcriptional activator. To address this question, we performed transient transfection assays. 293T cells were transfected with a luciferase reporter plasmid with 10 consensus OCT-4-binding sites in its promoter. Co-transfected OCT-4A increased the transcription of this reporter 50-fold (Fig. 8*A*, *black bars*), whereas OCT-4B had no effect (Fig. 8*A*, *white bars*). Similarly, co-expression of OCT-4A stimulated gene expression from the FOR (farnesoid X receptor-like orphan receptor) promoter containing two OCT-4-binding sites,[4] whereas co-expression of OCT-4B did not (Fig. 8*B*).

To determine whether OCT-4B can inhibit transactivation by OCT-4A, we co-transfected the expression vectors for

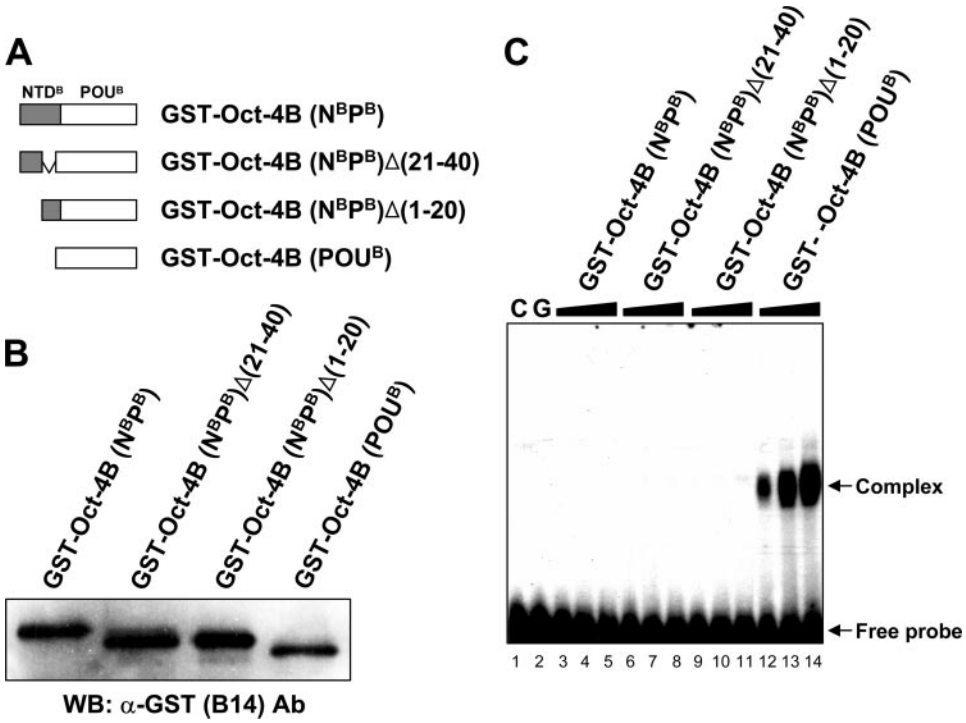


FIGURE 6. **Two independent N-terminal regions of OCT-4B inhibit its DNA binding.** *A*, schematic diagram of the OCT-4B cDNA fragments fused in-frame to the GST gene in pGEX vector. The N-terminal domain (*NTD*$^B$) and POU DNA binding domain (*POU*$^B$) of OCT-4B are shown in *dark gray* and *white boxes*, respectively. *B*, immunoblot analysis of the OCT-4B mutants to quantify the amounts of proteins. The GST fusion OCT-4B proteins used in the EMSAs were fractionated on 12% SDS-PAGE and visualized by Western blotting with anti-GST antibody (B-14; Santa Cruz Biotechnology). *C*, inhibition of the DNA binding of OCT-4B (POU$^B$) by two independent regions in OCT-4B NTD. EMSAs were performed with either no recombinant protein (*C, lane 1*), recombinant GST (*G, lane 2*, 150 ng), GST-OCT-4B (N$^B$P$^B$) (*lane 3*, 50 ng; *lane 4*, 100 ng; *lane 5*; 150 ng), GST-OCT-4B (N$^B$P$^B$)Δ(21–40) (*lane 6*, 50 ng; *lane 7*, 100 ng; *lane 8*, 150 ng), GST-OCT-4B (N$^B$P$^B$)Δ(1–20) (*lane 9*, 50 ng; *lane 10*, 100 ng; *lane 11*, 150 ng), or GST-OCT-4B (POU$^B$) (*lane 12*, 50 ng; *lane 13*, 100 ng; *lane 14*, 150 ng) as described under "Experimental Procedures."

OCT-4B and OCT-4A with an OCT-4-dependent reporter gene. In the absence of OCT-4B, OCT-4A increased reporter gene activity severalfold, as expected (Fig. 8*C, bar 2*), and increasing input levels of OCT-4B had no inhibitory effect (Fig. 8*C, bars 3* and *4*). Hence, we may conclude that overexpression of OCT-4B does not interfere with transactivation by OCT-4A.

## DISCUSSION

Numbers of stem cells, and their decisions to differentiate, must be tightly controlled during embryonic development and in the adult animal to avoid premature senescence or tumor formation. Embryonic and adult stem cells share the properties of self-renewal and multiple developmental potential, suggesting the presence of common cellular machinery. Thus, greater understanding of the molecular determinants responsible for these properties is desirable. Accordingly, there is growing interest in the functional characterization of OCT-4 in embryonic stem cells.

The results reported here begin to characterize the human OCT-4 isoforms. Two isoforms of human OCT-4 are generated by alternative splicing (Fig. 1*A*). *Oct-4* encodes a POU transcription factor that is expressed by all pluripotent cells during embryogenesis and is also abundantly expressed by ES, embryonic germ, and embryonic carcinoma cell lines (2, 3, 7, 57, 58). Differentiation of pluripotent cells to somatic lineages occurs at the blastocyst stage and during gastrulation, coincident with down-regulation of OCT-4. The unique *oct-4* expression pattern in the mouse embryo also leads to the hypothesis on the pluripotent cycle (27). Consistent with this speculation, *oct-4*-null embryos die at implantation because of a failure to form the inner cell mass (9). These results demonstrate that *oct-4* is required to prevent somatic differentiation of the inner cell mass and is important for maintaining the undifferentiated state during embryonic development.

Despite the fact that OCT-4A function is critical for maintaining the pluripotency of embryonic stem cells and promoting tumorigenesis in human tissues, little is known of the biological properties of the OCT-4B isoform. First, we investigated the existence of the two OCT-4 isoforms in human ES cells by RT-PCR, because although two human OCT-4 isoforms have been reported to be expressed in adult human pancreatic islets (36), it was not known whether they were also expressed in human ES cells. Two OCT-4 isoforms were indeed detected in the human ES cells. In addition, human OCT-4A was more abundant in human ES cells than OCT-4B (Fig. 1).

OCT-4 is structurally and functionally divided into three domains (see Fig. 1*D*). The N-terminal 133 amino acid residues of OCT-4A encompass a transcriptional activation region that is active in various cultured cell types. Amino acid residues 134–289 form the central POU domain of the protein that binds to DNA in a sequence-specific fashion. The third domain between amino acid residues 290 and 360 also controls the transactivation function of OCT-4, but its activity is cell type-specific (13). As shown in Fig. 1*D*, human OCT-4A and OCT-4B mRNA encode proteins that share POU DNA binding and C-terminal domains but differ in sequence at their N termini. Because the N-terminal domain of OCT-4 functions as a transactivation domain, we measured the transactivation potential of OCT-4B. OCT-4A functioned as a transcriptional activator as reported previously (33), whereas OCT-4B did not appear to do so (Fig. 8).

We also demonstrated that the N-terminal domain of OCT-4B inhibits the sequence-specific DNA binding of OCT-4 protein, via the central POU domain (Figs. 4 and 5). Moreover, regions containing amino acids 1–20 and 21–40 both inhibited POU sequence-specific DNA binding, indicating that both regions independently maintain OCT-4B in the latent state with little or no affinity for its target sequences (Fig. 6).

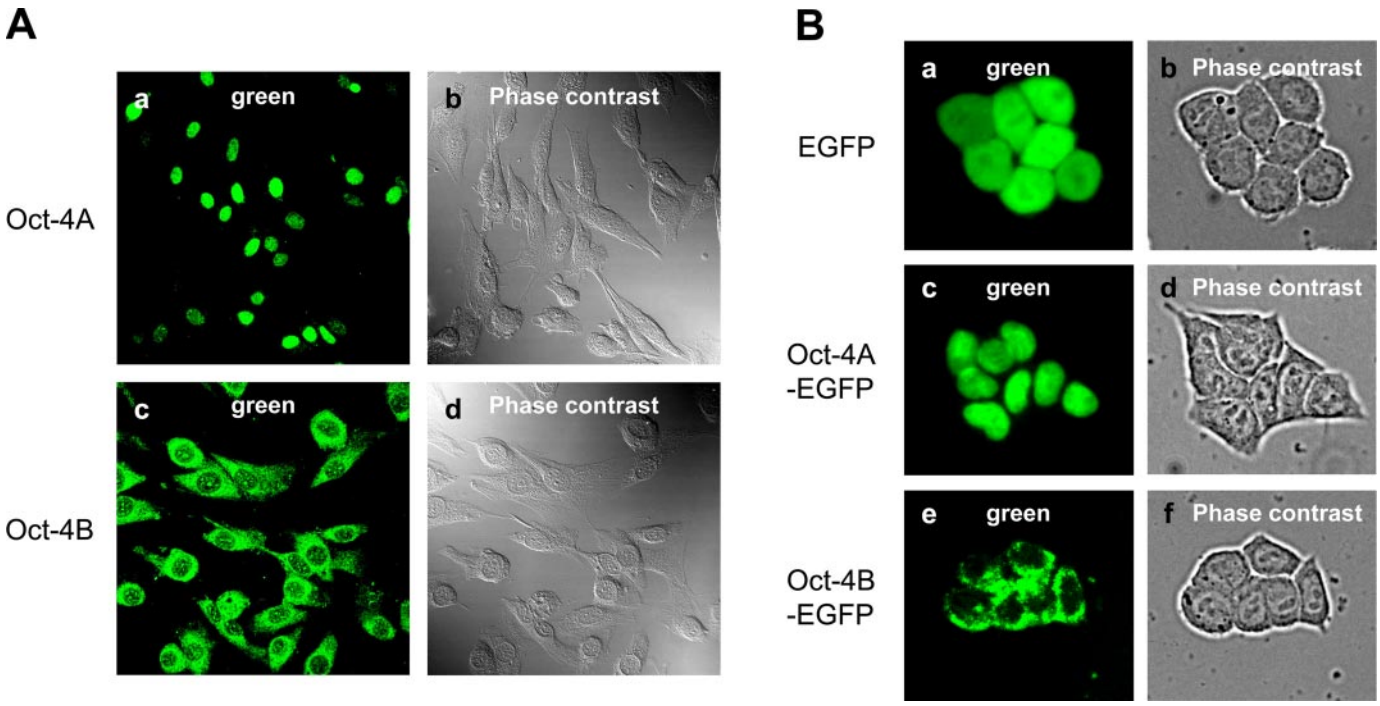


FIGURE 7. **Subcellular localization of OCT-4 isoforms in cells.** *A,* subcellular distribution of human OCT-4 isoforms in NIH3T3 cells. NIH3T3 cells grown on coverslips were infected with a retroviral expression construct, pBabePuro, containing the OCT-4A (*panels a* and *b*) or OCT-4B (*panels c* and *d*) coding regions. The subcellular distributions of OCT-4A (*panel a*) or OCT-4B (*panel c*) in infected NIH3T3 cells were analyzed with anti-OCT-4 antibody (C-20; Santa Cruz Biotechnology). *B,* subcellular distribution of human OCT-4 isoforms in ES cells. ZHBTc4 ES cell lines expressing EGFP (panels *a* and *b*), human OCT-4A-EGFP (panels *c* and *d*), or human OCT-4B-EGFP (panels *e* and *f*) were cultured on coverslips under low density conditions. The cells were fixed with an acetone/methanol mixture and analyzed for EGFP by fluorescence microscopy.

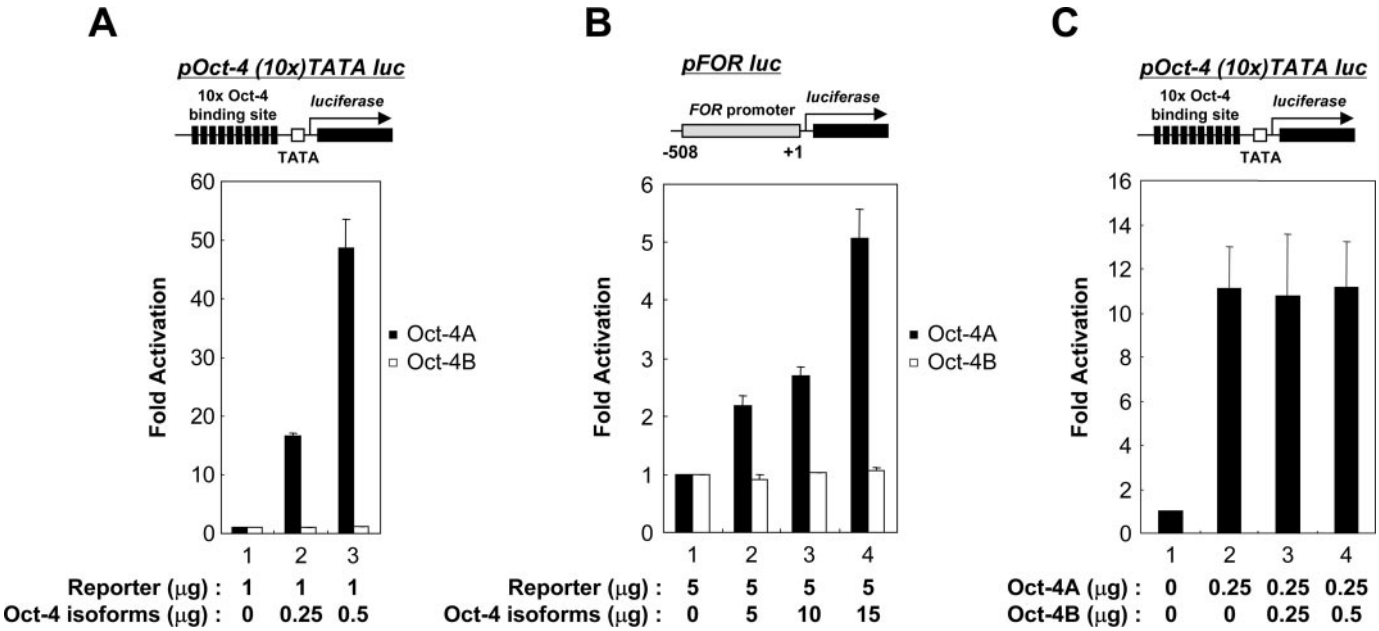


FIGURE 8. **Transactivation properties of the OCT-4 isoforms.** *A,* transcriptional activation by the OCT-4 isoforms using a pOCT-4(10x)TATAluc reporter vector. 293T cells were co-transfected with expression vectors encoding the OCT-4A (*black bars*) or OCT-4B (*white bars*) isoforms, the pOCT-4(10x)TATA-luc reporter vector, and the *Renilla* luciferase normalizing vector. Firefly luciferase activity was normalized with *Renilla* luciferase activity to correct for transfection efficiencies. Each transfection was performed at least three times independently, and the mean value is plotted with the standard deviation (*vertical bars*). Fold induction is relative to the empty expression vector. *B,* transcriptional activation by the human OCT-4 isoforms using pFORluc reporter vector. 293T cells were co-transfected with expression vectors encoding the OCT-4A (*black bars*) or OCT-4B (*white bars*) isoforms, the pFORluc reporter vector, and the *Renilla* luciferase normalizing vector. Firefly luciferase activity was normalized with *Renilla* luciferase activity to correct for transfection efficiencies. Each transfection was performed at least three times independently, and the mean value is plotted with the standard deviation (*vertical bars*). Fold induction is relative to the empty expression vector. *C,* human OCT-4B isoform does not interfere with transcriptional activation by OCT-4A. The pOCT-4(10x)TATAluc reporter was co-transfected into 293T cells with 0.25 μg of human OCT-4A and the indicated amounts of OCT-4B isoform. The average fold induction of transcription (normalized firefly luciferase activity) and standard error of three independent experiments are presented.

Classical NLS sequences contain regions rich in basic amino acids and generally conform to one of three motifs (34). The first type of NLS consists of a continuous stretch of four basic amino acids (lysine or arginine) or three basic amino acids together with a histidine or proline. The second type of NLS starts with a proline and is followed within three residues by an amino acid sequence containing three out of four basic residues. The third type of NLS, known as a bipartite motif, consists of two basic amino acids, a 10 amino acid spacer and a 5 amino acid sequence containing at least three basic residues. It has been reported recently that OCT-4A harbors a conserved nuclear localization signal RKRKR in its POU DNA binding domain (40). Therefore, one consequence of fusing the NTD$^{B}$ to the POU DNA binding domain should be nuclear targeting. Surprisingly, however, OCT-4B mostly localized to the cytoplasm, whereas OCT-4A localized to the nucleus, as expected (Fig. 7). These data suggests two possible models. One is that the NTD of OCT-4B protein contains a nuclear export signal that is recognized by transport receptors that carry it from the nuclear envelope to the cytoplasm. The other is that the NLS within the POU domain is blocked or buried by the OCT-4B NTD. Additional experiments will be required to distinguish between these possibilities.

There are several *oct-4*-like genes in the mouse and human genomes. The human *OCT-4* gene has been mapped to the region of the major histocompatibility complex on chromosome 6 and spans about 7 kb (36). The cognate mouse *Oct-4* gene is on chromosome 17 in a region that is syntenic with the major histocompatibility complex region of human chromosome 6 (1, 49). The pseudogene-like sequence for human *OCT-4* has been mapped to human chromosome 8 and is classified as a retroposon (36). Pseudogenes arise from the action of reverse transcriptases; cellular mRNAs can be copied into DNA by this enzyme, and the resulting DNA fragments can be reintegrated into the genome at a low rate. These pseudogenes have in common that they lack introns and have an additional poly(A) sequence (54, 55). Interestingly, the sequence of the *oct-4* pseudogene homologous to *oct-4* mRNA has 97.5% nucleotide sequence identity with the sequence of *oct-4* cDNA. However, although human *OCT-4* consists of five exons, the *oct-4* pseudogene lacks introns. In addition, it has a poly(A) track at its 3′-end and flanking direct repeats (5′-GAAAAGTAACATAATT-3′) at both ends of the gene, indicating that it is a retroposon (36). A comparison of the sequences of human *OCT-4* and of the OCT-4A and OCT-4B cDNA clones clearly indicates that OCT-4A and OCT-4B mRNAs are derived from the *oct-4* gene by alternative splicing, not from the *oct-4* pseudogene (36).

It is not easy to evaluate the role of human OCT-4B in embryonic stem cells. The N-terminal domain of human OCT-4A, unlike that of human OCT-4B, is rich in glycine and proline residues and appears to have a transcriptional activation domain. A search of OCT-4B (NTD) on line failed to reveal any functional motifs or homology with other proteins. However, analysis of its protein sequence revealed several potential serine, threonine, and tyrosine phosphorylation sites. It is possible that extranuclear OCT-4B is modified by several signaling molecules. Consistent with this speculation, it has been reported that OCT-4 exists as a phosphoprotein in embryonic carcinoma cells (13). Thus, it will be important to assess whether human cytoplasmic OCT-4B is involved in signal transduction.

In a broader context, our findings suggest that the OCT-4B isoform may play a different role than the unique function of the OCT-4A protein in self-renewal. Consistent with this speculation, expression of human OCT-4B alone is not sufficient to maintain stem cell self-renewal (Fig. 2*C*) and undifferentiated state (Fig. 3). These failures can be explained by the loss of transactivation ability (Fig. 8), which can in turn be accounted for by its localization in the cytoplasm (Fig. 7) as well as its inability to bind to the OCT-4 consensus motif (Fig. 4). Because OCT-4 is expressed in human testicular germ cell tumors and its expression transforms and endows tumorigenicity in nude mice (42, 56), it would also be interesting to determine whether OCT-4B can collaborate with OCT-4A in transforming nontumorigenic cells.

*Acknowledgments—We thank Dr. Jerry Pelletier for helpful discussions, and Dr. Hitoshi Niwa for providing ZHBTc4 ES cells and the pCAG-IP vector.*

## REFERENCES


1. Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H., and Gruss, P. (1990) *EMBO J.* **9,** 2185–2195
2. Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., and Hamada, H. (1990) *Cell* **60,** 461–472
3. Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirier, F., Rigby, P. W., and Staudt, L. M. (1990) *Nature* **345,** 686–692
4. Goto, T., Adjaye, J., Rodeck, C. H., and Monk, M. (1999) *Mol. Hum. Reprod.* **5,** 851–860
5. Hansis, C., Grifo, J. A., and Krey, L. C. (2000) *Mol. Hum. Reprod.* **6,** 999–1004
6. Burdon, T., Smith, A., and Savatier, P. (2002) *Trends Cell Biol.* **12,** 432–438
7. Scholer, H. R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990) *Nature* **344,** 435–439
8. Yeom, Y. I., Ha, H. S., Balling, R., Scholer, H. R., and Artzt, K. (1991) *Mech. Dev.* **35,** 171–179
9. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998) *Cell* **95,** 379–391
10. Pesce, M., Wang, X., Wolgemuth, D. J., and Scholer, H. (1998) *Mech. Dev.* **71,** 89–98
11. Niwa, H., Miyazaki, J., and Smith, A. G. (2000) *Nat. Genet.* **24,** 372–376
12. Abdel-Rahman, B., Fiddler, M., Rappolee, D., and Pergament, E. (1995) *Hum. Reprod.* **10,** 2787–2792
13. Brehm, A., Ohbo, K., and Scholer, H. (1997) *Mol. Cell. Biol.* **17,** 154–162
14. Lamb, K. A., and Rizzino, A. (1998) *Mol. Reprod. Dev.* **51,** 218–224
15. Scholer, H. R. (1991) *Trends Genet.* **7,** 323–329
16. Herr, W., and Cleary, M. A. (1995) *Genes Dev.* **9,** 1679–1693
17. Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H., and Robson, P. (2005) *J. Biol. Chem.* **280,** 24731–24737
18. Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S. Y., Suemori, H., Nakatsuji, N., and Tada, T. (2005) *Mol. Cell. Biol.* **25,** 2475–2485
19. Kraft, H. J., Mosselman, S., Smits, H. A., Hohenstein, P., Piek, E., Chen, Q., Artzt, K., and van Zoelen, E. J. (1996) *J. Biol. Chem.* **271,** 12873–12878
20. Ambrosetti, D. C., Basilico, C., and Dailey, L. (1997) *Mol. Cell. Biol.* **17,** 6321–6329
21. Vigano, M. A., and Staudt, L. M. (1996) *Nucleic Acids Res.* **24,** 2112–2118
22. Ezashi, T., Ghosh, D., and Roberts, R. M. (2001) *Mol. Cell. Biol.* **21,** 7883–7891
23. Liu, L., Leaman, D., Villalta, M., and Roberts, R. M. (1997) *Mol. Endocrinol.* **11,** 1651–1658

24. Liu, L., and Roberts, R. M. (1996) *J. Biol. Chem.* **271,** 16683–16689
25. Pesce, M., and Scholer, H. R. (2001) *Stem Cells* (*Dayton*) **19,** 271–278
26. Ovitt, C. E., and Scholer, H. R. (1998) *Mol. Hum. Reprod.* **4,** 1021–1031
27. Pesce, M., Gross, M. K., and Scholer, H. R. (1998) *BioEssays* **20,** 722–732
28. Scholer, H. R., Balling, R., Hatzopoulos, A. K., Suzuki, N., and Gruss, P. (1989) *EMBO J.* **8,** 2551–2557
29. Scholer, H. R., Ciesiolka, T., and Gruss, P. (1991) *Cell* **66,** 291–304
30. Chew, J. L., Loh, Y. H., Zhang, W., Chen, X., Tam, W. L., Yeap, L. S., Li, P., Ang, Y. S., Lim, B., Robson, P., and Ng, H. H. (2005) *Mol. Cell. Biol.* **25,** 6031–6046
31. Butteroni, C., De Felici, M., Scholer, H. R., and Pesce, M. (2000) *J. Mol. Biol.* **304,** 529–540
32. Brehm, A., Ohbo, K., Zwerschke, W., Botquin, V., Jansen-Durr, P., and Scholer, H. R. (1999) *Mol. Cell. Biol.* **19,** 2635–2643
33. Niwa, H. (2001) *Cell Struct. Funct.* **26,** 137–148
34. Nakai, K., and Kanehisa, M. (1992) *Genomics* **14,** 897–911
35. Lee, J., Rhee, B. K., Bae, G. Y., Han, Y. M., and Kim, J. (2005) *Stem Cells* (*Dayton*) **23,** 738–751
36. Takeda, J., Seino, S., and Bell, G. I. (1992) *Nucleic Acids Res.* **20,** 4613–4620
37. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, 3 Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
38. Aamlid, K. H., Hough, L., and Richardson, A. C. (1990) *Carbohydr. Res.* **202,** 117–129
39. Park, J. H., Kim, S. J., Oh, E. J., Moon, S. Y., Roh, S. I., Kim, C. G., and Yoon, H. S. (2003) *Biol. Reprod.* **69,** 2007–2014
40. Pan, G., Qin, B., Liu, N., Scholer, H. R., and Pei, D. (2004) *J. Biol. Chem.* **279,** 37013–37020
41. Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005) *Cell* **123,** 917–929
42. Abate-Shen, C. (2003) *Cancer Cell* **4,** 329–330
43. Rho, J. Y., Yu, K., Han, J. S., Chae, J. I., Koo, D. B., Yoon, H. S., Moon, S. Y., Lee, K. K., and Han, Y. M. (2006) *Hum. Reprod.* **21,** 405–412
44. Niwa, H., Masui, S., Chambers, I., Smith, A. G., and Miyazaki, J. (2002) *Mol. Cell. Biol.* **22,** 1526–1536
45. Lee, H. J., Kim, S., Pelletier, J., and Kim, J. (2004) *FEBS Lett.* **564,** 188–198
46. Botquin, V., Hess, H., Fuhrmann, G., Anastassiadis, C., Gross, M. K., Vriend, G., and Scholer, H. R. (1998) *Genes Dev.* **12,** 2073–2090
47. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D., and Lowe, S. W. (1997) *Cell* **88,** 593–602
48. Evans, M. J., and Kaufman, M. H. (1981) *Nature* **292,** 154–156
49. Siracusa, L. D., Rosner, M. H., Vigano, M. A., Gilbert, D. J., Staudt, L. M., Copeland, N. G., and Jenkins, N. A. (1991) *Genomics* **10,** 313–326
50. Ben-Shushan, E., Thompson, J. R., Gudas, L. J., and Bergman, Y. (1998) *Mol. Cell. Biol.* **18,** 1866–1878
51. Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M., and Okuda, A. (2002) *Nucleic Acids Res.* **30,** 3202–3213
52. Beck, F., Erler, T., Russell, A., and James, R. (1995) *Dev. Dyn.* **204,** 219–227
53. Hebert, J. M., Boyle, M., and Martin, G. R. (1991) *Development* (*Camb.*) **112,** 407–415
54. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y. J. (2001) *Nature* **409,** 860–921
55. Li, W. H., Gu, Z., Wang, H., and Nekrutenko, A. (2001) *Nature* **409,** 847–849
56. Gidekel, S., Pizov, G., Bergman, Y., and Pikarsky, E. (2003) *Cancer Cell* **4,** 361–370
57. Palmieri, S. L., Peter, W., Hess, H., and Scholer, H. R. (1994) *Dev. Biol.* **166,** 259–267
58. Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H. R. (1996) *Development* (*Camb.*) **122,** 881–894