DNA-dependent Oct4–Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy

Chen Sok LAM*, Tapan Kumar MISTRI†‡, Yong Hwee FOOt, Thankiah SUDHAHARAN*, Hui Theng GAN*, David RODD Az, Leng Hiong LIM‡§, Chai CHOUI, Paul ROBSON‡§, Thorsten WOHLAND† and Sohail AHMED*†

*Neural Stem Cell Laboratory, Institute of Medical Biology, Singapore 138678, †Biophysical Fluorescence Laboratory, Department of Chemistry, National University of Singapore, Singapore 117543, ‡Developmental Cellomics Laboratory, Genome Institute of Singapore, Singapore 138672, †Department of Biological Sciences, National University of Singapore, Singapore 117543, and ‡Program in Neuroscience, DUKE-NUS Graduate Medical School, Singapore 169857

Oct4 and Sox2 are two essential transcription factors that co-regulate target genes for the maintenance of pluripotency. However, it is unclear whether they interact prior to DNA binding or how the target sites are accessed in the nucleus. By generating fluorescent protein fusions of Oct4 and Sox2 that are functionally capable of producing iPSCs (induced pluripotent stem cells), we show that their interaction is dependent on the presence of cognate DNA-binding elements, based on diffusion time, complex formation and lifetime measurements. Through fluorescence correlation spectroscopy, the levels of Oct4 and Sox2 in the iPSCs were quantified in live cells and two diffusion coefficients, corresponding to free and loosely bound forms of the protein, were distinguished. Notably, the fraction of slow-diffusing molecules in the iPSCs was found to be elevated, similar to the profile in embryonic stem cells, probably due to a change in the nuclear milieu during reprogramming. Taken together, these findings have defined quantitatively the amount of proteins pertinent to the pluripotent state and revealed increased accessibility to the underlying DNA as a mechanism for Oct4 and Sox2 to find their target binding sites and interact, without prior formation of heterodimer complexes.

Key words: diffusion coefficient, fluorescence correlation spectroscopy (FCS), fluorescence lifetime microscopy (FLIM), Förster resonance energy transfer (FRET), transcription factor, stem cell.

INTRODUCTION

Oct4 and Sox2 are two TFs (transcription factors) essential for maintaining the pluripotent cell state both in vitro in ESCs (embryonic stem cells) and in vivo within the developing mouse embryo [1,2]. By modulating the level of both of the TFs in ESCs using RNAi (RNA interference) or gene overexpression, it has been found that distinct cell fates pertinent to the germ layers could be acquired [3,4]. For instance, by increasing the expression level of Oct4 artificially above a certain threshold, ESCs were driven to endoderm and mesoderm lineages, whereas lowering the levels resulted in the loss of pluripotency and differentiation to trophectoderm [3]. Similarly, small increases in the level of Sox2 triggered the differentiation of ESCs into a wide range of differentiated cell types [4], whereas reducing the level of Sox2 promoted the differentiation of ESCs into trophectoderm-like cells [5]. Thus a narrow range for both Oct4 and Sox2 is required to maintain pluripotency, which otherwise will lead to the acquisition of divergent cell fates. The need for sustained ectopic expression of Oct4 and Sox2, together with other TFs, in somatic fibroblasts to obtain iPSCs (induced pluripotent stem cells) [6,7], further attest to a requirement for an appropriate expression level of the proteins to establish pluripotency. Intriguingly, little is known of the protein levels of Oct4 and Sox2 that constitutes the pluripotent cell state.

The identification of several downstream molecular targets which are co-regulated by Oct4 and Sox2 underscores the importance of the interaction of these two proteins in establishing tightly regulated transcriptional networks associated with pluripotency [8]. Fgf4 (fibroblast growth factor 4) [9–11], Nanog [12–14] and Zfp206 [15] are salient examples of target genes that are directly regulated by Oct4 and Sox2. These genes possess consensus motifs that lie in close proximity for the binding of Oct4 and Sox2, leading to the idea that these two factors could interact with each other and synergize to control gene expression. In support of this idea, several thousand gene regulatory sites in the ESC genome were uncovered to be co-targeted by both of the TFs through ChIP (chromatin immunoprecipitation) studies [8,16]. Essentially, these target genes constitute one tightly controlled circuit, as the Oct4–Sox2 complex was found to activate Oct4 expression and Oct4 regulated Sox2 activity [5,17]. In turn, the level of Sox2 functioned as a molecular rheostat to control Oct4–Sox2 target genes [18]. Although the key function of the conserved POU and HMG (high-mobility group) domains in Oct4 and Sox2 respectively is to mediate DNA binding, these two domains are also known to selectively interact with one another so that a high-affinity ligand-binding complex is formed [19,20]. Moreover, specific residues on both Oct4 and Sox2 have been identified that affect its assembly and transcriptional activity [21,22]. Thus it is conceivable that Oct4–Sox2 heterodimerization could occur...
prior to DNA binding. However, crystal structures of POU–HMG–DNA complexes have revealed differential assembly of Oct4 and Sox2 on the UTF1 (undifferentiated embryonic cell TF 1) and FGF4 enhancer [23], suggesting that the factors occupied the binding sites sequentially in order to determine the extent of co-operativity. As the description of these interactions is, in general, qualitative, the mechanism behind heterodimerization is still unclear, warranting further studies to define the proteins and protein complexes in a spatially and temporally resolved quantitative manner.

In the present study, we generated the fluorescent fusions GFP [green FP (fluorescent protein)]–Oct4 and mRFP (monomeric red FP)–Sox2 to investigate the protein–protein interaction and to examine their intracellular dynamics using FRET ( Förster resonance energy transfer), FCS (fluorescence correlation spectroscopy), FCCS (fluorescence cross-correlation spectroscopy) and FLIM (fluorescence lifetime microscopy imaging), collectively called the F-techniques [24]. The fusion proteins displayed functional properties resembling the native proteins. Furthermore, they were able to initiate somatic reprogramming when introduced into MEFs (mouse embryonic fibroblasts). On the basis of diffusion time, complex formation and lifetime measurements, the interaction between Oct4 and Sox2 was found to be DNA-dependent. Using FCS to investigate dynamics in the living cells, we defined the level of exogenous fusion proteins in the reprogrammed iPSCs to be in the micromolar range. In addition, both GFP–Oct4 and mRFP–Sox2 showed distinct fast and slow diffusion coefficients, probably corresponding to free and transiently bound forms of the protein respectively. Notably, the fraction of slow diffusing molecules in somatic fibroblasts was elevated in iPSCs, comparable with the levels observed in pluripotent ESCs. Thus the findings of the present study support the DNA-dependent interaction of Oct4 and Sox2 and reveals a novel feature of the reprogramming process, associated with a change in dynamics of protein function due to increased accessibility of the underlying DNA.

EXPERIMENTAL

Cloning and constructs

The cDNAs of mouse Oct4 and Sox2 were amplified using sequence-specific primers and cloned directionally into XhoI and BglII sites of the expression vector pXJ40:GFP and pXJ40:mRFP respectively. The cloned product with the FP tagged at the N-terminus was sequence-verified in both orientations. Translation in silico revealed a complete reading frame with an intervening linker sequence of six amino acid residues. The tandem fusion mRFP–GFP was generated by inserting GFP between BamHI and NotI in pXJ40:mRFP. The reprogramming factors pMXs-Oct3/4, pMXs-Sox2, pMXs-Klf4 and pMXs-cMyc in retroviral vectors were produced according to the protocol (Millipore). E14 (E is embryonic day) mouse ESCs were grown on dishes coated with mitotically inactivated primary mouse feeder fibroblasts in ES media [DMEM (Dulbecco’s modified Eagle’s medium)-HG (high glucose) (Gibco), 15% heat-inactivated FBS, 2 mM l-glutamine, 1× non-essential amino acids, 1× sodium pyruvate, 1× penicillin/streptomycin, 0.1 mM 2-mercaptoethanol and 1000 units/ml LIF (leukaemia-inhibitory factor)]. For the differentiation of ESCs/iPSCs, cells were trypsinized, pre-absorbed on a dish coated with gelatin to remove feeder cells and incubated in six-well low-adherent culture plates (Costar) containing ES media in which LIF was removed. Embryoid bodies were then collected at different time points for further assays. Growth conditions were at 37°C, 5% CO2 and 95% humidity. CHO cells were transfected via electroporation using GenePulser Xcell (Bio-Rad Laboratories). Plasmids were added to suspended cells in F-12K medium in a 0.2 cm cuvette and electroporation was performed at a pulse length of 15 ms at 160 V.

Preparation of nuclear extracts

Nuclear extracts from CHO cells and from E14 mouse ESCs grown on fibroblasts feeders were prepared according to the method described by Dignam et al. [26]. Briefly, cells were harvested in PBS by scraping, resuspended in five pellet volumes of buffer A [10 mM Heps (pH 7.9), 1.5 mM MgCl2, 10 mM KCl, 0.5 mM dithiothreitol and 1% protease inhibitor mixture] and incubated on ice for 10 min. Cells were vortex mixed for 10 s and centrifuged (580 g for 5 min at 4°C). The supernatant was discarded and the pelleted nuclei were resuspended in 0.6 pellet volumes of buffer C [20 mM Heps (pH 7.9), 25% glycerol, 420 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 0.5 mM dithiothreitol and 1% protease inhibitor mixture] and incubated at 4°C with rotation for 30 min. After centrifugation (15682 g for 5 min at 4°C), the supernatant was dialysed against dialysis buffer [20 mM Heps (pH 7.9), 20% glycerol, 100 mM KCl, 0.83 mM EDTA, 1.66 mM dithiothreitol and 1% protease inhibitor mixture] at 4°C for 2 h. Extracts were either stored at −80°C or processed for further assays.

Virus production and iPSC induction

Retroviral particles that confer expression of the four reprogramming factors were produced according to the protocol of Takahashi et al. [27] with slight modification. BL6 fibroblasts were seeded at a density of 2.67×105 per 6 cm dish of inactivated CF-1 feeder layer. Lentiviral particles expressing the fusion proteins GFP–Oct4 and mRFP–Sox2 were produced using the 293FT cell line following the manufacturer’s instructions (Invitrogen). Medium was changed daily and iPSC colonies were monitored from day 5 post-infection. Putative colonies were picked manually and passed once to ensure that they could self-renew and propagate. Subsequently, colonies were transferred by passaging to a 3 cm sterile FluoroDish with a glass bottom (World Precision Instruments) coated with inactivated MEFs for FCS imaging. For verification, iPSC clones were stained for the expression of alkaline phosphatase using the AP detection kit (Millipore).

EMSA (electrophoretic mobility-shift assay)

PAGE-purified Nanog double-stranded oligonucleotides labelled with Cy5 (iodocarbocyanine) at the 5′ termini of both strands (Proligo) were used for the assay. The sense and anti-sense
sequences were Cy5-NOS-F, 5′-CTTACAGCTTTTGGATTACATGCTCAGTTGGA-3′ and Cy5-NOS-R, 5′-TCCACTTCAGCATTGTTGGTTGAGCTTGTAAG-3′. The DNA-binding reaction at a final volume of 10 μl consisted of 2 μl of nuclear extract (approximately 16 μg), 4 μl of 1.66× dialysis buffer, 50 nM Cy5-labelled oligonucleotide and 5 μg of poly(dG/dC) (Amersham Biosciences). Binding reactions were incubated for 20 min at room temperature (25°C). After incubation, 2 μl of anti-Oct4 (Santa Cruz Biotechnology, catalogue number sc-9081) or anti-Sox2 (Santa Cruz Biotechnology, catalogue number sc-17320) antibody were added to the reaction mix and incubated for a further 20 min. Binding reactions were resolved on pre-run 6% native PAGE gels (18.5 cm × 20 cm) in 0.5× TBE [Tris/borate/EDTA (1×TBE = 45 mM Tris/borate and 1 mM EDTA)] for 3 h at 300 V. Gels were then imaged directly using a Typhoon 9100 Phosphor Imager (Amersham Biosciences). The emission for Cy5 (670 nm), RFP (610 nm) and GFP (507 nm) were obtained sequentially, pseudo-coloured and merged to produce multi-coloured composite images.

**Luciferase reporter assay**

Quantification of luciferase activity was performed using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer’s instructions. In total, 5 μg each of the expression construct and reporter plasmid were transfected into CHO cells via electroporation. For an internal control and normalization, the Renilla luciferase plasmid (pRL-TK, Promega) was co-transfected at 1 μg. After 24 h in culture, cells were harvested using the passive lysis buffer supplied in the kit. Activity of the lysate was measured in triplicate for each condition using a GloMax 96-well plate-reading luminometer (Promega) and the experiment was repeated independently three times.

**Co-immunoprecipitation**

Cells were lysed in 1.0 ml of Nonidet P40 lysis buffer [20 mM Tris/HCl (pH 7.5), 150 mM NaCl, 1% Nonidet P40 and 5 mM EDTA] for 20 min and centrifuged at 16168 g at 4°C for 15 min. Protein lysates (500 μl) were diluted to 1 ml with immunoprecipitation buffer [10 mM Tris/HCl (pH 8.0), 150 mM NaCl, 0.1% Nonidet P40 and 2 mM EDTA] and incubated overnight at 4°C with 5 μg each of either anti-Oct3/4 (Santa Cruz Biotechnology, catalogue number sc-8628) or anti-Sox2 (Santa Cruz Biotechnology, catalogue number sc-17320) antibodies, and 25 μl of Dynabeads® Protein G beads (Invitrogen). The beads were washed three times with immunoprecipitation buffer. Captured proteins were eluted by incubating with 1× Laemmli buffer [6× Laemmli buffer is 125 mM Tris/HCl (pH 6.8), 2% SDS, 20% glycerol, 0.2% Bromophenol Blue and 4% 2-mercaptoethanol] at 95°C for 5 min. Eluted proteins were analysed by Western blot. The secondary antibody used was HRP (horseradish peroxidase)-conjugated donkey anti-goat IgG (Santa Cruz Biotechnology, catalogue number sc-2033) and the signal was developed using SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific).

**FCS, SW-FCCS (single-wavelength FCCS), FRET, FLIM instrument setup and analysis**

FCS was performed using a commercial laser-scanning confocal microscope FV300 (Olympus), which was coupled to a custom-built FCS attachment. GFP was excited with an argon ion 488 nm laser line (Melles Griot) at 4 μW, whereas mRFP was excited with a helium neon 543 nm laser line at 10 μW. Band-pass filters 510AF23 and 615DF45 (Omega Optical) were used for GFP and mRFP respectively. An optical fibre-coupled avalanche-photodiode detector (SPCM-AQR-14-FC, Pacer) was used to collect the photons. Autocorrelations were computed on-line by a hardware correlator (Flex02-01D, http://Correlator.com). Curve fitting was performed by a self-written program in Igor Pro 6.0 (WaveMetrics) using eqn (S2) in the Supplementary Material (at http://www.BiochemJ.org/bj/448/bj4480021add.htm). The effective volume V_eff of the 488 nm and 543 nm laser lines were determined to be 0.53 ± 0.03 fl and 0.69 ± 0.08 fl using fluorescein and Atto565 respectively.

SW-FCCS was performed using a single 514 nm laser line at 30 μW with the set-up described above but with two detection channels separated by a 560DCLP dichroic mirror (Omega Optical). Emission filters used were 545AF35 and 615DF45 (Omega Optical) in the green and red detection channels respectively. The theory and experimental details [28] of SW-FCCS were performed as described in the Supplementary Materials and methods section (at http://www.BiochemJ.org/bj/448/bj4480021add.htm).

AP-FRET (acceptor-photobleaching FRET) was performed on a Zeiss LSM 510 confocal microscope [28]. Briefly, images were acquired using a C-Apochromat 63× 1.2 W objective and GFP/mRFP was excited using 488 and 561 nm laser lines. The emission was detected via a 490/565 dichroic mirror, filtered spectrally by passing through either a GFP (BP 505-550) or RFP (LP 575) emission filter. Selected regions of interest co-expressing the fusion proteins were photobleached using 50 iterations of the 561 nm laser at 75% power. Upon mRFP bleaching, FRET will be eliminated and the donor GFP signal will increase, with the change in the donor signal being a measure of FRET efficiency [29].

The FRET efficiency of GFP was calculated as:

\[
\frac{I_{\text{post}} - I_{\text{pre}}}{I_{\text{pre}}} \times 100
\]

where \(I_{\text{pre}}\) is the pre-bleach intensity and \(I_{\text{post}}\) is the post-bleach intensity.

Time-domain FLIM experiments were performed using the time-correlated single photon counting system (PicoQuant) attached to an Olympus FV-1000 confocal microscope with a 60× 1.2 W objective. The excitation light source was a 485 nm pulsed-diode laser-controlled by a Sepia II driver (PicoQuant) having a dichroic mirror of 488/559 and 520/30 emission filter. Individual photon arrivals were detected using a SPAD (single-photon avalanche diode) detector and events were recorded by a PicoHarp 300 TCSPC module. Lifetime analysis was carried out using Sympotyme (PicoQuant) software, mono- and bi-exponential fittings were obtained for GFP alone and in the presence of mRFP respectively.

**Statistical analysis**

Raw data was exported to Excel 2003 (Microsoft) and charts were plotted using Prism 5.01 software (GraphPad). Values were presented as means ± S.D. and the significance of differences was calculated using Student’s unpaired t test.

**RESULTS**

In order to use the F-techniques to examine interaction and intracellular dynamics of Oct4 and Sox2, two spectrally distinct TF–FP fusion constructs were designed by tagging GFP and mRFP to the N-terminus of mouse Oct4 and Sox2 respectively (Supplementary Figure S1 at http://www.BiochemJ.org/bj/448/bj4480021add.htm). To assess whether the constructs produced...
functional GFP–Oct4 and mRFP–Sox2 proteins, we analysed the fusion proteins in detail using several biochemical approaches.

**Molecular properties of GFP–Oct4 and mRFP–Sox2**

To examine DNA-binding activity of the TF–FPs, an EMSA was performed using enhancer sequences containing the oct:sox recognition motif from the Nanog promoter [12,14]. Nuclear lysates from mouse ESCs and transfected CHO cells were combined with a Cy5-labelled Nanog DNA oligonucleotide probe. The identity of protein–DNA complexes detected in recognition motif from the TFs from ESCs, indicated the presence of monomeric and heterodimeric complexes. Importantly, addition of anti-Oct4 antibody specifically retarded Oct4–Sox2 heterodimer and Oct4 monomer, whereas Sox2 monomers remained intact (Figure 1A, lane 5). Conversely, addition of anti-Sox2 antibody resulted in super-shifted Oct4–Sox2 heterodimer and Sox2 monomer, whereas Oct4 monomer remained unchanged (Figure 1A, lane 6). In support of this, gel shifts performed using lysates from individually transfected cells (Supplementary Figure S2 at http://www.BiochemJ.org/bj/448/bj4480021add.htm) showed a specific band for the protein–DNA interaction, indicating that binding of the TF–FPs to the DNA can occur independently of each other. Thus the ability of Oct4 and Sox2 to form a stable ternary complex with Nanog DNA remains unperturbed in the TF–FPs.

To determine whether the TF–FPs heterodimerize prior to DNA binding, co-immunoprecipitation was performed using whole-cell lysates of GFP–Oct4 and mRFP–Sox2-tagged fusion proteins to Cy5-labelled DNA, which also migrated slower than the native TFs from ESCs, indicated the presence of monomeric and heterodimeric complexes. Furthermore, the iPSC colonies were developmentally potent (Figure 1E) transduced fibroblasts, consistent with the iPSC-derived colonies were morphologically indistinguishable from iPSC and mESC (mouse ESC) controls. In addition, the TF–FP-derived iPSCs were positive for alkaline phosphatase reactivity (inset of Figure 1E), an early marker of pluripotency [31]. Furthermore, the iPSC colonies were developmentally potent as, upon in vitro embryoid body formation and differentiation, cells could be induced to express markers characteristic of the three embryonic germ lineages: ectoderm [Map2 (microtubule-associated protein 2) and Pax6 (paired box gene 6)], mesoderm (Brachyury) and endoderm (Sox17) (Supplementary Figure S5 at http://www.BiochemJ.org/bj/448/bj4480021add.htm). Thus transcriptional activity of the TF–FPs, despite differing slightly in magnitude in comparison with the native TFs, did not perturb production of functional iPSCs. Taken together, these molecular features strongly demonstrate that the TF–FPs retained the functional properties of their untagged counterparts.

**Interaction of Oct4 and Sox2 fusion proteins require the presence of DNA in vitro**

Although co-immunoprecipitation enabled detection of interacting complexes, it remains unclear whether the protein–protein interaction is direct. To analyse steady-state molecular interactions, we used FCS to probe the diffusion of the TF–FPs. The normalized autocorrelation curves of GFP and GFP–Oct4 showed differential mobilities directly related to the mass of the molecule. GFP (27 kDa), the smallest of the proteins analysed, exhibited the fastest diffusion time, whereas GFP–Oct4 (~65 kDa) diffused at a lower rate due to the increase in mass (Supplementary Figure S6 at http://www.BiochemJ.org/bj/448/bj4480021add.htm). Accordingly, a higher diffusion coefficient (D) was measured in the lysates of GFP in comparison with GFP–Oct4 transfected cells (Table 1). In buffered solutions of pure GFP molecules, D was determined to be 95 μm²/s [32]. Importantly, GFP–Oct4 + mRFP–Sox2 nuclear lysates did not show a further increase in diffusion time and D was comparable with the mobility of GFP–Oct4 alone, suggesting that the fusions are not diffusing in concert as a complex. However, upon the addition of 20 nM of an oct:sox cis-element-containing oligonucleotide to GFP–Oct4 + mRFP–Sox2 lysates, D was
Figure 1  Biochemical characterization and functional reprogramming of GFP–Oct4 and mRFP–Sox2

(A) EMSA of nuclear lysates from ESCs and transfected CHO cells with Cy5-labelled Nanog DNA oligonucleotide. Unlabelled nuclear lysates from ESCs showed several retarded protein–DNA complexes (lane 1). In the presence of anti-Oct4 (lane 2) or anti-Sox2 (lane 3) antibodies, specific shifts in the band of the monomeric (labelled by green and red asterisks for Oct4 and Sox2 respectively) and heterodimeric complex (labelled by a white asterisk) were detected. Similarly, addition of anti-Oct4 (lane 5) and anti-Sox2 (lane 6) antibodies to GFP–Oct4 + mRFP–Sox2 CHO nuclear lysates resulted in a shift in the monomeric and heterodimeric forms of the complex. The gel was scanned sequentially for detection of the emissions using 488, 532 and 633 nm excitation lasers. (B) Western blot analysis (WB) of Oct4 and Sox2 in lysates of ESCs and CHO cells overexpressing (O/E) Oct4 and Sox2, and CHO cells alone. (C) Co-immunoprecipitation (IP) of transfected CHO cells with either IgG, anti-Oct4 or anti-Sox2 antibodies. In cells transfected with native Oct4 and Sox2, immunoprecipitates of Oct4 showed very low levels of Sox2-interacting partner. Lysates of GFP–Oct4- and mRFP–Sox2-transfected cells yielded similar results. Predicted mass of GFP–Oct4 = 64.32 kDa and mRFP–Sox = 60.12 kDa. (D) Dual-luciferase reporter assay of the fusions in CHO cells. Similar to native Oct4 and Sox2, GFP–Oct4 and mRFP–Sox2 were able to activate Nanog:luciferase reporter (pNanog:luc). A mutant construct (pNanogO/S:luc) lacking the oct:sox motif showed significantly reduced reporter activity (** P < 0.05). Values are means ± S.E.M. of normalized luciferase units (NLU) from triplicate samples for each condition and the experiment was repeated three times. (E) Morphology of control iPSC clones transduced with Oct4/Sox2/Klf4/cMyc and ESCs appeared identical. Similarly, iPSC clones derived via replacement of Oct4 with GFP–Oct4 and Sox2 with mRFP–Sox2 exhibited indistinguishable morphology. The fluorescent iPSC colonies (at passage two) were alkaline-phosphatase-positive (inset), whereas the feeder layer in the background remained unstained. Scale bar = 100 μm.
Table 1  FCS reveals a DNA-dependent decrease in diffusion coefficient of GFP–Oct4 and mRFP–Sox2 complexes

<table>
<thead>
<tr>
<th>Lysates</th>
<th>Mw (kDa)</th>
<th>D (μm²/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>27</td>
<td>64.79 ± 5.23</td>
</tr>
<tr>
<td>GFP–Oct4</td>
<td>64.3</td>
<td>38.93 ± 3.73</td>
</tr>
<tr>
<td>mRFP–Sox2</td>
<td>60.1</td>
<td>N.D.</td>
</tr>
<tr>
<td>GFP–Oct4 + mRFP–Sox2</td>
<td>124.4</td>
<td>40.59 ± 4.12</td>
</tr>
<tr>
<td>GFP–Oct4 + mRFP–Sox2 + DNA (20 nM)</td>
<td>147.3</td>
<td>27.00 ± 3.83</td>
</tr>
<tr>
<td>GFP–Oct4 + mRFP–Sox2 + DNA (40 nM)</td>
<td>147.3</td>
<td>24.42 ± 2.53</td>
</tr>
</tbody>
</table>

Mw, predicted molecular mass of the protein; D, diffusion coefficient of lysates containing fluorophores emitting in the GFP channel. D for mRFP–Sox2 was not determined (N.D.). Values represent means ± S.D. from at least four measurements for each sample from two experiments.

reduced by a factor of 1.5 in comparison with lysates containing only GFP–Oct4 + mRFP–Sox2. Moreover, upon doubling the concentration of DNA to 40 nM, a further, but slight, decrease in D was observed, probably due to the amount of bound complexes approaching saturation. Thus, from the mean values of D measured from the green fluorescent components (Table 1), formation of an Oct4–Sox2 heterodimer complex appears to be dependent on the availability of DNA.

To confirm the requirement of DNA a priori for Oct4–Sox2 interaction, we used SW-FCCS [33] to measure the extent of complex formation between the two TF–FPs. From nuclear lysates of CHO cells transfected with the relevant constructs, signals detected in the two channels were cross-correlated to derive the CCF (cross-correlation function), which was then used to compute the amount of complex formation (see the Supplementary Materials and methods section). Lysates from GFP–Oct4 and mRFP–Sox2 co-transfected cells showed very low levels of complex formation (2.1 ± 1.9%, Figure 2A), comparable with the negative control measured using lysates from mRFP and GFP co-transfected cells (3.7 ± 2.4%, Figure 2B). However, upon the addition of an exogenous oct:sox cis-element-containing DNA (8 nM) to the GFP–Oct4 and mRFP–Sox2 lysates, the amount of complex detected from the CCF was augmented to 26% (Figure 2C), reasonably close to the complex percentage of 39% derived from the tandem mRFP–GFP fusion positive control (Figure 2D). This result parallels the finding obtained using FCS, supporting a requirement of cognate DNA-binding elements for protein interaction and complex formation.

Interaction of GFP–Oct4 and mRFP–Sox2 in situ

To examine interaction of the proteins in a cellular context, CHO cells which do not express endogenous Oct4 and Sox2 (Figure 1B) were transfected with both GFP–Oct4 and mRFP–Sox2. Expression of the TF–FPs was detected specifically in the nucleus, typical of the distribution of TFs. In a number of doubly transfected cells, co-localization of the TF–FPs was also detected (Figure 3A). To demonstrate whether the TF–FPs interact, cells in which both proteins are expressed were selected for AP-FRET. The FRET efficiency after photobleaching of the acceptor was determined by measuring the fluorescence intensity of the donor GFP–Oct4 emission in comparison with pre-bleach levels. As a positive control, the mRFP–GFP tandem protein was used. After bleaching of mRFP–Sox2 (Figure 3B), an increase in the intensity of GFP–Oct4 was detected due to the interaction of the fusions in close proximity. The average FRET efficiency of GFP–Oct4 and mRFP–Sox2 was 15% (n = 28), comparable with the value (16%, n = 22) obtained with the mRFP–GFP positive control (Figure 3C).

To further investigate the interaction of the two fusions in live cells, an alternative approach using time-domain FLIM to map the nanosecond lifetime of GFP–Oct4 was performed. In order to extract lifetime information, time-correlated single-photon counting images were acquired, emitted photons in response to pulsed excitation were sorted into a histogram and tail fits of the data to a bi-exponential decay function were analysed for every pixel inside the nucleus. The lifetime of GFP–Oct4 was lower in cells showing co-localization of the two fusions than cells transfected with GFP–Oct4 alone (Figure 4A). The mean lifetime of the pixels in GFP–Oct4-transfected cells was 2.44 ± 0.087 ns, but in the presence of mRFP–Sox2 acceptor, the lifetime was reduced significantly to 2.27 ± 0.031 ns. In contrast, cells co-expressing GFP–Oct4 and mRFP, or mRFP–Sox2 and GFP, did not exhibit attenuation of the lifetime (Figure 4B), indicating that the interaction between GFP–Oct4 and mRFP–Sox2 was specific. Similar results were obtained using frequency-domain FLIM analysis (Supplementary Figure S7 at http://www.BiochemJ.org/bj/448/bj4480021add.htm).

Although FRET was detected between the two fusions in the nucleus, it is still unclear whether the interaction was dependent on DNA binding as revealed by the in vitro analyses. Thus we next addressed whether the fusions are bound to DNA in the nuclear compartment. Previously, a high-affinity nucleic acid dye Sytox Orange was demonstrated to function as a FRET acceptor, enabling the determination of the association of GFP-tagged histone H2B with DNA [34] and interaction of YFP (yellow FP)-tagged polypyrimidine-tract-binding protein with RNA [35]. In fixed CHO cells labelled with Sytox Orange, the average lifetime of GFP alone measured with time-domain FLIM was 2.25 ± 0.015 ns. After DNaseI treatment, the lifetime remained relatively constant at 2.24 ± 0.014 ns, indicating the absence of a non-specific interaction between GFP and the DNA dye. In contrast, the lifetime of GFP–Oct4 in combination with Sytox Orange showed a significantly reduced lifetime of 2.17 ± 0.006 ns compared with control GFP (Figure 5A). Upon DNaseI digestion, the lifetime was restored to the natural state, suggesting that the occurrence of FRET between GFP–Oct4 and DNA-bound Sytox Orange was effectively abolished. In the presence of a non-interacting acceptor (mRFP), a similar trend was observed; FRET between GFP–Oct4 and DNA-labelled Sytox Orange remained unchanged, signifying the specificity of the bound fusion to intact DNA. Most saliently, in cells co-transfected with GFP–Oct4 and mRFP–Sox2 stained with Sytox Orange, a further reduction in lifetime to 2.01 ± 0.015 ns was observed (a consequence of two or more acceptors in close proximity). When genomic DNA was digested following DNaseI treatment, the pronounced FRET could no longer be detected as the lifetime of GFP–Oct4 recovered almost completely to 2.23 ± 0.012 ns.

As shown in Figure 5(B), cells co-expressing the fusions (the majority of GFP–Oct4-expressing cells were also mRFP–Sox2 positive; Supplementary Figure S8 at http://www.BiochemJ.org/bj/448/bj4480021add.htm) and stained with Sytox Orange displayed bright orange nuclei as well as some low-level cytoplasmic staining when excited with a 559 nm laser line. After DNaseI treatment, cytoplasmic staining of RNA can still be visualized, but nuclear expression was reduced substantially. As a result, the lifetime of donor GFP–Oct4 in doubly transfected cells recovered to non-FRET levels (compare colour-coded lifetime images ± DNaseI treatment; Figure 5B), comparable with the lifetimes observed in DNaseI-treated GFP or GFP–Oct4-transfected samples (Figure 5A). Thus these results support the interaction of GFP–Oct4 and mRFP–Sox2 in a DNA-dependent manner and not heterodimerization a priori in the nucleus.
Oct4–Sox2 interaction and diffusion properties

Figure 2 DNA-dependent interaction of GFP–Oct4 and mRFP–Sox2 measured via SW-FCCS

Graphs show the auto-correlation function (red and green traces for signals detected in the mRFP and GFP channel) and CCF (blue traces) of interacting complexes derived from SW-FCCS. The plots below each graph depict the fits and the corresponding steady-state intensity of the molecules. (A) In lysates of GFP–Oct4 and mRFP–Sox2 co-transfected cells, the CCF showed a low amplitude and the amount of interacting complex was approximately 2.1 %, comparable with the lysates from (B) non-interacting mRFP + GFP negative control. (C) In the presence of DNA, the CCF measured from GFP–Oct4 and mRFP–Sox2 lysates showed elevated amplitude and the amount of complex formed was approximately 26 %. (D) Signals from the lysates of mRFP–GFP-transfected cells served as a positive control and the theoretical maxima for complex formation was approximately 39 %. Values are means ± S.D. from eight measurements from three independent experiments.

Distinct diffusion properties of Oct4 and Sox2 in somatic and pluripotent cells

Although the interaction of Oct4 and Sox2 requires the presence of available binding sites, how the TFs seek and bind specific sites to elicit global transcriptional changes remain poorly understood. For TFs to access the underlying binding sites in the genome, the chromatin state must necessarily be permissive, achieved for instance by hyperdynamic chromatin proteins that were demonstrated to bind loosely to chromatin in ESCs but not upon differentiation or in other lineage-committed cells [36]. Therefore we next asked whether there is a difference in the mobility of the fusions owing to the accessibility of the chromatin in pluripotent or somatic cells by extracting the diffusion properties using...
FCS. Analysis of the autocorrelation function measured from pluripotent (ESCs) and somatic cells (CHO and MEFs) revealed that freely diffusing GFP and mRFP in the nucleoplasm fitted a single component model with a fast diffusion coefficient ($D_1$). However, the curves for both GFP–Oct4 and mRFP–Sox2 fitted best to a two-component model, enabling a fast ($D_1$) and slow ($D_2$) component to be distinguished, which probably corresponds to free and bound (but still mobile) protein respectively [37,38]. The curves depicting mobilities of GFP–Oct4 and mRFP–Sox2 in somatic and pluripotent cell types are shown in Figure 6, and $D$ values for the two diffusing components are listed in Table 2.

In ESCs, the $D_1$ component for GFP–Oct4 and mRFP–Sox2 was $14.66 \pm 3.33 \mu m^2/s$ and $10.06 \pm 8.74 \mu m^2/s$ respectively, which is slightly slower in mobility in comparison with free GFP and mRFP molecules. In contrast, mobility of the $D_2$ component for GFP–Oct4 and mRFP–Sox2, which is slower by approximately 50-fold relative to $D_1$, was $0.26 \pm 0.06 \mu m^2/s$ and $0.22 \pm 0.18 \mu m^2/s$ respectively. Consistent with our data, the behaviour of Brca2 (breast cancer 2)–EGFP (enhanced GFP) in the nucleoplasm also showed a fast (14.8 $\mu m^2/s$) and slow (0.45 $\mu m^2/s$) diffusion coefficient [38]. In addition, the slow kinetics of other functionally unrelated GFP-labelled nuclear proteins such as HMG-17 (0.45 $\mu m^2/s$), the SF2 (splicing factor 2)/ASF (alternative splicing factor) splicing factor (0.24 $\mu m^2/s$) and the fibrillarin ribonucleoprotein (0.53 $\mu m^2/s$) measured using fluorescence recovery after photobleaching [39] was also comparable with the $D_2$ components of Oct4 and Sox2 fusions. Remarkably, the $D_2$ fraction (percentage of $D_2$) for both GFP–Oct4 and mRFP–Sox2 was 48% and 47% respectively of the total molecules detected, suggesting that a high proportion of the fusions are in the bound state, also taking into consideration competition from endogenous Oct4 and Sox2.

In the TF–FP-derived iPSCs, the dynamics of both $D_1$ and $D_2$ components bear a strong resemblance to the profile observed in ESCs. Most saliently, the behaviour of GFP–Oct4 $D_1$ components, with a diffusion coefficient of 0.24 $\mu m^2/s$ representing 49% of the total detected molecules is almost identical with the measurements found in ESCs. Similarly, the mRFP–Sox2 $D_2$ fraction in both the iPSC and ESC populations represented approximately half of the total detected molecules (46% and 47% respectively). Thus the diffusion property and the fraction of transiently bound molecules are comparable in both ESCs and iPSCs, corroborating the high fraction of mobile chromatin proteins that keeps the chromatin ‘breathing’ for the maintenance of plasticity in the pluripotent state [36]. In addition, the concentration of GFP–Oct4 and mRFP–Sox2 molecules in several independently derived iPSCs, measured from the amplitudes of the autocorrelation function, was estimated to be 1.28 $\pm 0.47 \mu M$ and 1.01 $\pm 0.36 \mu M$ respectively. Coincidentally, these values reflect the most efficient stoichiometry of Oct4 and Sox2, in at least the vector dosage required to obtain iPSCs [30].

In CHO cells, the $D_1$ component of freely diffusing GFP (37.23 $\pm 3.72 \mu m^2/s$) exhibited greater mobility than GFP–Oct4 (6.88 $\pm 1.49 \mu m^2/s$) molecules, recapitulating the larger $D$ measured in vitro for GFP compared with GFP–Oct4 (Table 1). A comparable $D_1$ was also detected for mRFP (33.72 $\pm 5.56 \mu m^2/s$) and mRFP–Sox2 (6.99 $\pm 4.53 \mu m^2/s$). Although the $D_2$ components for GFP–Oct4 and mRFP–Sox2 showed some variability, the diffusion coefficients closely resembled the values obtained with ESCs. Importantly, the average fraction of the $D_2$ component for GFP–Oct4 was only 15%, whereas mRFP–Sox2 represented 37% of the total bound molecules detected. Similar results were also obtained using MEFs, in which the average fraction of the
DISCUSSION

Oct4 and Sox2 are important TFs deemed as master regulators of pluripotency. By generating functional fusions and using fluorescence spectroscopy techniques with single-molecule sensitivity, we report in the present paper the first quantitative analysis of the Oct4–Sox2 interaction and their dynamics. Specifically, we showed the following: (i) heterodimerization is dependent on DNA-harbouring cognate cis-binding elements; (ii) diffusion coefficients in the nucleus of live cells associated with the pluripotent state; (iii) the fraction of bound and free TFs distinguishes the somatic and reprogrammed genomic state; and (iv) the concentration of proteins in iPSCs that were reprogrammed from somatic fibroblasts.

DNA-dependent Oct4–Sox2 interaction

For our in vitro analyses, we have used the nuclear extracts of functionally validated GFP–Oct4 and mRFP–Sox2 to examine their interaction. The nuclear extract format is advantageous as it is a homogeneous mixture in which only the ensemble effects of the fluorescent molecules are considered, circumventing problems associated with heterochromatic or euchromatic regions. Moreover, it avoids the variability in expression inherent of transiently transfected samples and can be used directly without biochemical purification. On the basis of the diffusion time and complex formation derived using FCS and SW-FCCS respectively, our data revealed the lack of pre-formed GFP–Oct4 and mRFP–Sox2 complex in the absence of cognate DNA sequences. Accordingly, binding of the two TFs to DNA must have taken place independently and is not based on prior heterodimerization. This mechanism would aid in the search of specific binding sites by the individual TF and serve to facilitate co-operativity between the interacting TFs to control gene expression, especially of target genes such as fgf4 and utf1, whose promoters possess binding...
Figure 5  Intracellular interaction of GFP–Oct4 and mRFP–Sox2 is DNA-dependent

(A) In the presence of the fluorescence acceptor Sytox Orange–DNA complex, the lifetime of GFP–Oct4 in GFP–Oct4 and mRFP–Sox2 co-transfected cells showed a significantly reduced lifetime as compared with GFP–Oct4-transfected cells alone. Upon DNaseI digestion, FRET is abolished and the lifetime of GFP–Oct4 returned to levels similar to GFP. The lifetime of GFP molecules that do not interact with DNA is independent of the effect of DNaseI. GFP (n = 25), GFP–Oct4 (n = 37), GFP–Oct4 + mRFP (n = 19) and GFP–Oct4 + mRFP–Sox2 (n = 19). (B) Confocal images of GFP–Oct4 + mRFP–Sox2-transfected CHO cells treated with or without DNaseI and stained with Sytox Orange. GFP–Oct4 and mRFP–Sox2 or Sytox Orange–DNA complex was excited at 488 and 559 nm respectively. The average lifetimes of donor GFP–Oct4 was increased upon DNaseI digestion, strongly suggesting that the interaction between GFP–Oct4 and mRFP–Sox2 is dependent on intact DNA in the nucleus. Inset shows colour-coded lifetime values ranging from 1.5 to 2.5 ns. **P < 0.0001 (two-tailed Student's t test).

Figure 6  Diffusion properties of the fusions in somatic and pluripotent cells

Graphs show typical auto-correlation function (ACF) of (A) GFP–Oct4 and (B) mRFP–Sox2 measured in live CHO cells, MEFs, iPSCs and ESCs using FCS. The ACF in the different cell types were normalized to 1/N. The axes G(τ) and τ represent the correlation function and lag time respectively. There is a clear shift of the ACF at the low τ region in iPSCs and ESCs, resulting in an increase in F2 in the pluripotent cell type in comparison with the somatic cells.
sites that are spaced differently [23]. Although evidence from the co-immunoprecipitation experiments suggested interaction of the proteins albeit at a low fraction, this is probably a consequence of the use of less stringent whole-cell lysates that may contain proteins or protein complexes tethered to DNA. Indeed, when the underlying DNA is removed upon treatment with benzonase nuclelease, a reduction in the level of interacting proteins of immunoprecipitated Oct4 was reported [40].

For our intracellular analyses, we have combined the use of AP-FRET and FLIM-FRET (frequency and time domain) to examine the interaction of Oct4 and Sox2. This allowed a robust conclusion to be reached on direct protein–protein interaction of the two TFs in the nuclear compartment. The FRET efficiency between GFP–Oct4 and mRFP–Sox2 using AP-FRET and frequency domain FLIM in fixed cells was ~15 %, whereas it was ~7 % using time domain FLIM in live cells. A plausible explanation for the higher FRET efficiency in fixed samples may be due to the effects of cross-linking that inadvertently drew donor and acceptor molecules closer together, resulting in increased energy transfer. Nevertheless, a FRET efficiency of 15 % and 7 % converts into a molecular distance of 6.2 nm and 7.2 nm respectively; a marginal difference considering the inverse 6th power relationship relative to a Förster distance of ~4.7 nm for a mRFP–GFP FRET pair [41]. As the Oct or Sox DNA element of target genes such as Utf1, Sox2, Fbx15, Pou5f1 and Nanog are either not separated (15 bp/5.1 nm) or have three nucleotide spacing between them, such as Fgf4 (18 bp/6.1 nm), the calculated distance based on the overall FRET efficiency in the nucleus thus fits reasonably well with the description of the binding of Oct4 and Sox2 to at least those motifs [12]. Importantly, we could show that, upon digestion of genomic DNA, the occurrence of FRET between GFP–Oct4 and the acceptor(s) mRFP–Sox2 or Sytox Orange nuclear dye was effectively abolished. Together with the in vitro data, these findings clearly demonstrate that the interaction of GFP–Oct4 and mRFP–Sox2 was DNA-dependent.

### Table 2: Concentration and diffusion coefficients of GFP–Oct4 and mRFP–Sox2 in iPSCs, ESCs, CHO cells and MEFs

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Protein</th>
<th>Concentration (µM)</th>
<th>(D_1(\mu m^2/s))</th>
<th>(D_2(\mu m^2/s))</th>
<th>(F_2(%))</th>
<th>(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPSC</td>
<td>GFP–Oct4</td>
<td>1.28 ± 0.47</td>
<td>16.83 ± 6.18</td>
<td>0.24 ± 0.07</td>
<td>49 ± 0.09</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>mRFP–Sox2</td>
<td>1.01 ± 0.36</td>
<td>10.22 ± 6.29</td>
<td>0.14 ± 0.11</td>
<td>46 ± 0.09</td>
<td>25</td>
</tr>
<tr>
<td>ESC</td>
<td>GFP–Oct4</td>
<td>1.00 ± 0.21</td>
<td>14.66 ± 3.33</td>
<td>0.26 ± 0.06</td>
<td>48 ± 0.14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>mRFP–Sox2</td>
<td>0.32 ± 0.20</td>
<td>10.06 ± 8.74</td>
<td>0.22 ± 0.18</td>
<td>47 ± 0.16</td>
<td>30</td>
</tr>
<tr>
<td>CHO</td>
<td>GFP–Oct4</td>
<td>0.05 ± 0.03</td>
<td>23.45 ± 6.06</td>
<td>N.A.</td>
<td>N.A.</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>mRFP</td>
<td>2.23 ± 0.13</td>
<td>13.78 ± 7.15</td>
<td>N.A.</td>
<td>N.A.</td>
<td>36</td>
</tr>
<tr>
<td>MEF</td>
<td>GFP–Oct4</td>
<td>0.09 ± 0.04</td>
<td>6.88 ± 1.49</td>
<td>0.17 ± 0.15</td>
<td>15 ± 0.05</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>mRFP–Sox2</td>
<td>0.17 ± 0.06</td>
<td>6.99 ± 4.53</td>
<td>0.30 ± 0.27</td>
<td>37 ± 0.09</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>GFP</td>
<td>0.06 ± 0.04</td>
<td>37.23 ± 3.72</td>
<td>N.A.</td>
<td>N.A.</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>mRFP</td>
<td>0.09 ± 0.05</td>
<td>33.72 ± 5.56</td>
<td>N.A.</td>
<td>N.A.</td>
<td>18</td>
</tr>
</tbody>
</table>

Differential mobilities of Oct4 and Sox2 in somatic and pluripotent cells

The detection of a DNA-dependent protein–protein interaction in the nuclear compartment of CHO cells suggests that abundant DNA-binding sites are available. Since CHO cells, unlike ESCs, do not express endogenous Oct4 or Sox2, pluripotency-associated genes should accordingly be inactivated and rendered inaccessible to the trans-activating factors. However, binding of these factors to DNA, typically with high dissociation constants, can still occur in a non-specific manner in regions of loose chromatin conformation [42]. In support of the fusions binding transiently to DNA, results from FCS showed fractions (\(F_2\)) of GFP–Oct4 and mRFP–Sox2 that have slow diffusion coefficients in both CHO and pre-reprogrammed MEFs, consistent with the values obtained from other studies involving DNA–protein interaction in live cells [37,38]. As mobility of the \(D_2\) components was almost two orders of magnitude slower than freely moving \(D_1\) components in the nucleoplasm, these may represent species that are interacting with sites on the DNA. In fact, the smaller fraction of \(D_2\) components in MEFs and CHO cells in contrast with ESCs and iPSCs, may be attributed to a direct consequence of chromatin proteins that are loosely bound to chromatin as a small mobile fraction could still be detected in differentiated ESCs [36], allowing the TF–FPs to access the underlying DNA. However, most of the interaction of TFs with DNA in the nucleus, such as for glucocorticoid receptors [43] or Hox proteins [42], were non-specific in nature, characterized by rapid association/dissociation events. Assuming that the search for a specific binding site occurs in a stochastic manner [39] governed by constant on/off rates, ensemble movement of the fusions would enable FRET to occur between GFP–Oct4 and mRFP–Sox2.

To resolve kinetics of bound/unbound proteins in the nucleus, fluorescence decay after photoactivation has been employed. Using this assay, two distinct immobile fractions for Oct4 that preceded any morphologically distinguishable signs of lineage segregation in the early mouse embryo have been reported [44]. Although cells in the pre-compaction stage destined for the extra-embryonic lineage have a small immobile fraction (~15 %), those that will contribute to the pluripotent lineage have a large immobile fraction (~40 %), due largely to Oct4 bound to DNA. Thus accessibility to the underlying DNA plays a major role in cell-fate determination and also suggests that somatic or non-pluripotent cells in general are not entirely refractory to the binding of pluripotent TFs. Although the bound/immobile fraction will be rendered ‘invisible’ by FCS [42], the findings of the present study showed a corresponding low and high \(F_2\) in differentiated and pluripotent cells respectively for both of the fusions. These two fractions may be directly related to the bound/immobile fraction as observed in the kinetics of Oct4 segregating the extra-embryonic or pluripotent lineage, because the slowly diffusing mobile components precede the stably bound

\[ \text{Oct4–Sox2 interaction and diffusion properties} \]

\[ \text{Table 2: Concentration and diffusion coefficients of GFP–Oct4 and mRFP–Sox2 in iPSCs, ESCs, CHO cells and MEFs} \]

\[ D_1 \text{ and } D_2 \text{ are the fast and slow diffusion coefficients respectively, and } F_2 \text{ is the fraction of molecules derived from } D_2. \text{ N.A., not applicable as the autocorrelation curves from the FCS fits a single diffusion time. Values represent means ± S.D.} \]
state and are subjected to the same changes in nuclear landscape and chromatin structures that determines accessibility and thus the search for specific target sites [36,45]. Collectively, our analysis supports the notion that open and accessible DNA, more abundant in pluripotent than somatic cells, may be the mechanism behind the recruitment of Oct4 and Sox2 to specific regulatory elements and their subsequent regulation of gene expression.

Quantification of Oct4 and Sox2 protein levels in the pluripotent cell state

The demonstration that Oct4 and Sox2 along with other factors can induce differentiated somatic cells to the pluripotent state shows how important the levels of key reprogramming proteins are in developmental decisions [31,46]. Thus determining the intracellular levels of proteins is going to be crucial in defining the fundamental mechanisms in developmental biology. In the present study, taking advantage of the reprogramming activity of the fusions, we directly quantified the amount of protein in the iPSCs and found that cells leading up to the pluripotent state have concentrations of approximately 1.28 \(\mu\)M and 1.01 \(\mu\)M for Oct4 and Sox2 respectively. As the diffusion properties of the fusion derived from measurements in the same cells are comparable with ESCs, these values may reflect the concentration of Oct4 and Sox2 necessary to elicit changes in the nuclear milieu for the acquisition of the pluripotent cell state. Alternatively, the continued expression of exogenous proteins in the iPSCs may be a result of remnant transgene activity and thus regarded as partially reprogrammed, since silencing of transgene expression, at least with the use of retroviral vectors, is one of the features to mark the pluripotent state [7,47]. However, with lentiviral vectors, despite being commonly used to induce mouse and human iPSCs [46,48], silencing is less established, confounded further by previous studies in ESCs and transgenic mice that they may escape silencing altogether [49,50]. Moreover, the silencing of residual lentiviral factor expression was found to be established over an extended period (12–15 passages) and the lack of complete silencing did not hinder the induction, maintenance or directed differentiation of human iPSCs, leading to the suggestion that vector silencing may in fact be an epiphenomenon rather than a factor contributing to the pluripotent state [30]. Nevertheless, further studies will entail generation of GFP/mRFP knock-in animals at the Oct4 and Sox2 locus to determine endogenous protein levels that will be useful for correlating the exogenous protein expression during the process of reprogramming.

In conclusion, in the present study we have demonstrated, through several approaches, quantitative information about two key regulators of pluripotency. In vitro analyses showed that Oct4 and Sox2 require DNA for complex formation. Indeed, Oct4 and Sox2 heterodimerization could be detected in a DNA-dependent manner in the nucleus. The fraction of bound proteins in live cells showed a slow diffusion coefficient as revealed by FCS. When fibroblasts were reprogrammed to iPSCs, the concentration of exogenous Oct4 and Sox2 was found to be in the low micromolar range, and the fraction of slow-diffusing molecules was augmented, mirroring the levels observed in pluripotent ESCs. Thus the transition from a somatic to pluripotent state features a precise amount of Oct4 and Sox2 at the level of expression and distinct kinetics. Taken together, the present study has highlighted the amenability of the various fluorescence techniques for the quantification of protein interactions, concentrations and dynamics of Oct4 and Sox2, the findings of which can be used as predictive tools for determining pluripotency.

Author contribution

Chen Sok Lam, Tapan Kumar Mistri and Yong Hwee Foo conducted the FCS experiments and analyzed the data. Chen Sok Lam, David Rodda, Leng Hiong Lim and Hui Theng Gan performed biochemical characterization of the fusions and iPSC studies. Chen Sok Lam, Thanhkian Sudhahan, Chai Chou, Paul Robson, Thorsten Wohland and Sohail Ahmed designed the study and assisted in the preparation of the paper.

Acknowledgements

We thank Sem Kai Ping for assistance on cloning and Western blotting, and Davor Solter and Barbara Knowles for critical reading of the paper before submission.

Funding

This work was supported by the Singapore Bioimaging Consortium (SBIC) and Singapore Stem Cell Consortium (SSCC) under the Agency of Science, Technology and Research (grant number RP C-001/2007).

References

18 Boer, B., Kopp, J., Malliana, S., Desler, M., Chakravarty, H., Wilder, P.J., Bernadt, C.

Received 30 April 2012/10 August 2012; accepted 22 August 2012
Published as BJ Immediate Publication 22 August 2012, doi:10.1042/BJ20120725

© The Authors Journal compilation © 2012 Biochemical Society
SUPPLEMENTARY ONLINE DATA

DNA-dependent Oct4–Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy

Chen Sok LAM*, Tapan Kumar MISTRIT‡, Yong Hwee FOO†, Thankiah SUDHAHARAN*, Hui Theng GAN*, David RODDA‡, Leng Hiong LIM†‡, Chai CHOU†, Paul ROBSON‡‡, Thorsten WOHLAND† and Sohail AHMED*†‡

*Neural Stem Cell Laboratory, Institute of Medical Biology, Singapore 138678, †Biophysical Fluorescence Laboratory, Department of Chemistry, National University of Singapore, Singapore 117543, ‡Developmental Cellomics Laboratory, Genome Institute of Singapore, Singapore 138672, †Department of Biological Sciences, National University of Singapore, Singapore 117543, and ‡Program in Neuroscience, DUKE-NUS Graduate Medical School, Singapore 169857

MATERIALS AND METHODS

Quantitative real-time PCR

RNA was isolated from embryoid bodies derived from Oct4/Sox2/Klf4/cMyc-, GFP–Oct4/Sox2/Klf4/cMyc- and Oct4/mRFP–Sox2/Klf4/cMyc-induced iPSCs at 0, 4 and 6 days post-differentiation using TRIzol® reagent (Invitrogen). The RNA was purified using an RNeasy kit (Qiagen) and quantified using Nanodrop. The High-Capacity cDNA Reverse Transcription kit was used for first strand cDNA synthesis. The concentration can then be determined by eqn (S6):

\[ C = \frac{N}{N_A V_{eff}} \]

where \( N \) is the average number of particles in the observation volume; \( \omega_0 \) and \( z_0 \) are the radial and axial distances where the excitation intensity reaches \( 1/e^2 \) of its value from the centre of the observation volume respectively; \( F_{sep} \) is the fraction of the particles that have entered the triplet state; \( \tau_{sep} \) is the triplet state relaxation time; \( \tau_{in} \) and \( F_i \) is the diffusion time and fraction of the component \( I_i \); \( G(\infty) \) is the convergence value of the ACF for long times. The diffusion time \( \tau_0 \) reflects the size of the fluorescent-tagged molecule and, in the case of binding, the size of the complex. It is related to the diffusion coefficient \( D \) by (eqn S3):

\[ \tau_0 = \frac{\omega_0^2}{4D} \] (S3)

The amplitude of the ACF is indirectly proportional to the average number of particles \( N \) by (eqn S4):

\[ G(0) = \frac{1}{N} + G(\infty) \] (S4)

The effective observation volume \( V_{eff} \) can be determined by using a fluorescent dye of known \( D \) with the use of eqns (S2), (S3) and (S5):

\[ V_{eff} = \pi^{3/2} \omega_0^2 z_0 \] (S5)

The concentration can then be determined by (eqn S6):

\[ C = \frac{N}{N_A V_{eff}} \]

The concentration can then be determined by eqn (S6):

\[ G(0) = \frac{1}{N} + G(\infty) \] (S4)

In SW-FCCS [1], the fluorescence fluctuations from fluorophores diffusing in and out of an observation volume, defined by a focused single 514 nm laser line at 30 \( \mu \)W and a pinhole, are autocorrelated to extract information such as diffusion time and the average number of fluorophores passing through the observation volume. The normalized ACF (autocorrelation function) is given by eqn (S1):

\[ G(\tau) = \frac{\langle F(t) F(t+\tau) \rangle}{\langle F(t) \rangle^2} \] (S1)

where \( F(t) \) is the fluorescence intensity at time \( t \) and \( \tau \) is the time delay. The experimental ACF is fitted using a multi-component diffusion model with triple state contribution (eqn S2):

\[ G(\tau) = \frac{1}{N} \left[ 1 + \frac{F_{sep}}{1 - F_{sep}} \exp \left( -\frac{\tau}{\tau_{sep}} \right) \right] \sum_i F_i \left( 1 + \frac{\tau}{\tau_{in}} \right)^{-1} \]

\[ \times \left[ 1 + \left( \frac{\omega_0}{z_0} \right)^2 \frac{\tau}{\tau_{in}} \right]^{-1/2} + G(\infty) \] (S2)

1 To whom correspondence should be addressed (email sohail.ahmed@imb.a-star.edu.sg).
\[ G_{G}(0) = \frac{(\eta_{G,G})^2 C_G + (\eta_{R,G})^2 C_R + (q_G \eta_{G,R} + q_R \eta_{G,G})^2 C_{GR}}{N_A V_{eff} \left[ \eta_{G,G} C_G + \eta_{G,R} C_R + (q_G \eta_{G,R} + q_R \eta_{G,G}) C_{GR} + B_G / (N_A V_{eff}) \right]} \] (S7)

\[ G_{R}(0) = \frac{(\eta_{G,R})^2 C_G + (\eta_{R,R})^2 C_R + (q_G \eta_{G,R} + q_R \eta_{R,R})^2 C_{GR}}{N_A V_{eff} \left[ \eta_{G,R} C_G + \eta_{R,R} C_R + (q_G \eta_{G,R} + q_R \eta_{R,R}) C_{GR} + B_R / (N_A V_{eff}) \right]} \] (S8)

\[ G_{GR}(0) = \frac{\eta_{G,G} \eta_{R,R} C_G + \eta_{G,R} \eta_{R,G} C_R + (q_G \eta_{G,R} + q_R \eta_{G,G}) (q_G \eta_{G,R} + q_R \eta_{R,R}) C_{GR}}{N_A V_{eff} \left[ \eta_{G,G} C_G + \eta_{G,R} C_R + (q_G \eta_{G,R} + q_R \eta_{G,G}) C_{GR} + B_G / (N_A V_{eff}) \right]} \times \left[ \eta_{G,R} C_G + \eta_{R,R} C_R + (q_G \eta_{G,R} + q_R \eta_{R,R}) C_{GR} + B_R / (N_A V_{eff}) \right]^{-1} \] (S9)

where \( G_{G}(0) \) and \( G_{R}(0) \) are the amplitudes of the ACFs in the green (EGFP) and red (mRFP) channel respectively, and \( G_{GR}(0) \) is the amplitude of the CCF; \( B_G \) and \( B_R \) are the uncorrelated background count rate in the green and red channels respectively; \( \eta_{G,G} \) and \( \eta_{R,R} \) are the cps (counts per particle per second) of green labelled molecules in the green and red channels respectively; \( \eta_{G,R} \) and \( \eta_{R,G} \) are the cps of the red-labelled molecules in the green and red channels respectively; \( q_G \) and \( q_R \) are correction factors that account for changes in cps during binding via processes such as quenching or FRET for the green and red particles respectively. All of the parameters mentioned here can be obtained by doing control experiments. The details have been discussed in previous reports [2,3]. The output from the equations gives the values of \( C_G, C_R \) and \( C_{GR} \) which are the concentrations of the green free, red free and complex respectively. The values are used to generate the complex percentage (Complex %), which indicates the amount of either the green or red molecules in the complex (eqns S10 and S11):

\[ \frac{C_{GR}}{C_G + C_{GR}} \times 100 \] (S10)

\[ \frac{C_{GR}}{C_G + C_{GR}} \times 100 \] (S11)

**Frequency domain FLIM**

Frequency domain FLIM was performed with the LIFA system (Lambert Instruments) on an inverted wide-field fluorescence microscope (Olympus IX71) with a 60 × 1.35 oil-immersion objective. A sinusodially modulated 4 mW 470 nm light-emitting diode at 40 MHz was used to excite GFP and the emission signals were detected by an intensified CCD camera. For calibration, FITC with a standard lifetime of 4 ns was used as the reference. For each measurement, 12 phase and modulation-shifted images were acquired and fitted with a sine function for the calculation of lifetime using the software provided by Lambert Instruments. Regions of interest from different cells were selected to obtain the average lifetime and S.D.
Figure S1  Construction of full-length GFP–Oct4 and mRFP–Sox2 fusion proteins

Cloned sequences were translated and queried using NCBI Conserved Domain Search. High E-values for the conserved domains (A) Pou (2e-34) and Homeodomain (4e-11) in GFP–Oct4 (1743 bp) and (B) SOX-TCF_HMG-box (2e-26) in mRFP–Sox2 (1569 bp) were returned. The linker sequence (white) in the schematic diagram for both of the fusions comprise six amino acid residues GSKLLE. Fusion proteins are not drawn to scale.
Figure S2 Binding of EGFP–Oct4 and mCherry–Sox2 to Cy5-labelled Nanog DNA

In this experiment, GFP and mRFP were replaced with EGFP and mCherry respectively for Oct4 and Sox2 as they are more photo-stable and resistant to photobleaching. Lysates of EGFP–Oct4- or mCherry–Sox2-transfected cells were reacted with Cy5-labelled DNA and processed for gel-shift analysis. Protein–DNA complexes arising from the interaction of the fusions with cognate Nanog DNA elements were detected by scanning the gel at the respective wavelengths. Open arrows indicate EGFP–Oct4 + Cy5–DNA or mCherry–Sox2 + Cy5–DNA complexes, asterisks indicate unbound proteins and arrows show free Cy5–DNA.

Figure S3 Lack of interaction between GFP–Oct4 and mRFP–Sox2 in nuclear lysates of CHO cells

CHO cells transfected with the respective constructs were harvested for the dual luciferase reporter assay. The combined expression of Oct4 and Sox2 elicited higher levels of luciferase activity than the factors expressed alone, indicating synergistic action on the Nanog promoter. A comparable level of luciferase activity was induced through co-expression of GFP–Oct4 + mRFP–Sox2. However, synergy between the two fusions is not observed due to a less effective activity of mRFP–Sox2. The y-axis indicates firefly luciferase activity (pNanog:luc) normalized against Renilla luciferase (pRL-TK). pGL3-Basic is a promoter-less vector serving as a background control. The expression constructs were transfected at 1 μg each, together with 0.5 μg of pNanog:luc and 0.1 μg of pRL-TK. Results represent means ± S.D. from three independent replicates.

Figure S4 Transcriptional activity of native and FP-tagged Oct4 and Sox2

CHO cells transfected with the respective constructs were harvested for the dual luciferase reporter assay. The combined expression of Oct4 and Sox2 elicited higher levels of luciferase activity than the factors expressed alone, indicating synergistic action on the Nanog promoter. A comparable level of luciferase activity was induced through co-expression of GFP–Oct4 + mRFP–Sox2. However, synergy between the two fusions is not observed due to a less effective activity of mRFP–Sox2. The y-axis indicates firefly luciferase activity (pNanog:luc) normalized against Renilla luciferase (pRL-TK). pGL3-Basic is a promoter-less vector serving as a background control. The expression constructs were transfected at 1 μg each, together with 0.5 μg of pNanog:luc and 0.1 μg of pRL-TK. Results represent means ± S.D. from three independent replicates.

Figure S5 DNA binding of EGFP–Oct4 and mCherry–Sox2

Cells co-transfected with equimolar amounts of GFP–Oct4 and mRFP–Sox2 constructs were processed for co-immunoprecipitation (IP) and Western blotting (WB). Both the immunoprecipitates of anti-Oct4 and anti-Sox2 do not show Sox2 or Oct4 protein respectively. The expected molecular mass for GFP–Oct4 and mRFP–Sox2 is 64.3 kDa and 60.1 kDa respectively. O/E, overexpression; I, Input.
Oct4–Sox2 interaction and diffusion properties

Figure S5 Differentiation assay of iPS colonies via quantitative real-time PCR at 2, 4 and 6 days post-differentiation

The spontaneous acquisition of the three germ layers: ectoderm (Map2, Pax6), mesoderm (Brachyury) and endoderm (Sox17) lineages were analysed for (A) positive control, (B) mRFP–Sox2- and (C) GFP–Oct4-derived iPS colonies after embryoid body differentiation. By day 4, all iPSC clones showed up-regulation of the markers for ectoderm, mesoderm and endoderm lineages. Data represent relative fold changes calculated using the comparative Ct method. Values were obtained by normalization against actin and transcript levels were expressed relative to day 0 post-differentiation. Measurements were performed in triplicate from two independent RNA isolations. D, day.

Figure S6 FCS revealed shifts in diffusion time of GFP, GFP–Oct4 and GFP–Oct4 + mRFP–Sox2 in the absence or presence of Nanog DNA

Graphs show normalized autocorrelation curves of the various FPs derived from CHO nuclear lysates using FCS. The axes G(τ) and τ(S) represent the correlation function and lag time respectively. The increase in mass of the interacting complex GFP–Oct4 + mRFP–Sox2 + DNA resulted in longer diffusion times, shifting the curve to the right in comparison with GFP or GFP–Oct4.
Figure S7 Interaction of GFP–Oct4 and mRFP–Sox2 in the nucleus measured via frequency-domain FLIM

FRET in the nucleus of GFP–Oct4- and mRFP–Sox2-transfected cells showed a significant decrease in lifetime in comparison with cells transfected with GFP–Oct4 alone or GFP–Oct4 + mRFP. (A) Images show typical localization of proteins and a corresponding heat-map of the lifetimes of the region of interests in the nucleus (marked by boxes). (B) Whereas cells expressing GFP–Oct4 alone, GFP–Oct4 + mRFP or with a pair of non-interacting proteins (GFP + mRFP) revealed comparable lifetimes, GFP–Oct4 + mRFP–Sox2-expressing cells showed significantly reduced lifetimes (**P < 0.05). Moreover, the average lifetime of the interacting fusions is 2.1 ns, which is almost similar to the mRFP–GFP positive control (2.0 ns). Data represent mean ± S.E.M. from duplicate experiments. GFP + mRFP (n = 34), GFP–Oct4 + mRFP–Sox2 (n = 45).
Figure S8  Comparison of the expression of GFP–Oct4 and mRFP–Sox2 expression in transfected CHO cells

Images are confocal sequential acquisitions of GFP–Oct4- and mRFP–Sox2-transfected cells excited with 488 and 559 nm lasers respectively. The merged channel shows that almost all GFP–Oct4 cells are also mRFP–Sox2 positive.

REFERENCES


Received 30 April 2012/10 August 2012; accepted 22 August 2012
Published as BJ Immediate Publication 22 August 2012, doi:10.1042/BJ20120725