Universidade do Algarve

Cátia Alexandra Ferreira de Sousa

Analyzing Oct4 conserved domains in lower vertebrates

Faculdade de Engenharia e Recursos Naturais

Engenharia Biológica com Mestrado Integrado

Universidade do Algarve

FERN

Dissertação Engenharia Biologica com Mestrado integrado

2007/2008

Cátia Alexandra Ferreira de Sousa

Analyzing Oct4 conserved domains in lower vertebrates

Supervisor: Andrew Johnson;
David Tannahill;
Dr. José Belo

Ano Lectivo 2007 to 2008

To my family that was my support and inspiration

Abstract

Oct-4 is a POU-V domain transcription factor which regulates pluripotency in mammals and expressed in ES cells and germ cells, and was believed to be a gene unique to mammals. Recently, it has been demonstrated to be present in the genomes of *Xenopus*, zebrafish, sturgeon and axolotl. It has been shown that Oct-4 has three transactivation domains, (N), POU and (C), with DNA binding mediated through the POU domain. It was unknown how the activity of these domains has been retained through evolution and how they collectively function to control pluripotency.

This study is new and is the first to analyse the functional conservation of these Oct-4 homologues, and their regulation molecular. Three different assays were developed to study Oct4 functionality: 1) A transactivation assay in which the function of Oct4 protein over-expresion on a known Oct4-target sequence was assessed (by luciferase assay using the p6Wtk-luc reporter containing Oct4 binding sites); 2) A heterologous transactivation assay in which the function of specific Oct4 domains by linking them to Gal4-DNA binding domain was specifically assessed (DBD) (by luciferase assay using the pGal4-lux reporter). 3) The subcellular localization of Oct-4 homologues (generating two constructs; either full' length or POU domain, fused to Green fluorescent protein (GFP).

This study showed the coexistence of DBD conflicts with Oct4 resulting in a decrease on its transactivation capacity when compared to their native state.

Oct-4 function generally conserved among species, with *Xenopus* Oct91 being the Oct-4 homologue with a transactivation function more similar to mouse Oct-4. Between (C) and (N) transactivation domains linked to DBD, the (C) domain was the one with more activity. The (C) domain is cell-type specific regulated by phosphorylation events, while the (N) domain suffers sumoylation. These two regulatory mechanisms are shared in all Oct-4 homologues. It was also possible to conclude that Oct-4 protein is nuclei transcribed. This project opens many possible studies in Oct-4 regulation.

Acknowledgments

It was a pleasure and an intellectual challenge to work in the Genetics lab of the Nottingham University, it was an experience without precedent.

I would like to thanks to Dr. José Belo by accepting being my supervisor from Algarve University. I would like to thank to my supervisors, particularly Dr. Andrew Johnson for taking me into his laboratory, and Dr. James Dixon for all his time, concern and also for tutoring me and support me during the project. I would also like to thank Dr. Carla Toro and Dr. David Tannahill for their advices and concern. I am grateful to Dr. Niwa for the gift of p6Wtk-luc that was used on the assays.

Also thanks to e to my family and friends that always support me and were there for me, renewing my energy and strength

Table of contents

Abstract i
Acknowledgments ii
Table of contents iii
Abbreviations v
List of figures vii
List of tables viii
Chapter 1: Introduction and literature review 1
  1.1 Overview 1
  1.2 Pluripotency 1
  1.3 Research in lower vertebrates 4
  1.4 Oct-4 belongs to POU transcription factor family 7
  1.5 Oct-4 domains 8
  1.6 Oct-4 and pluripotency 9
  1.7 Oct4 expression 10
  1.8 Aims and Objectives 11
Chapter 2: Material and Methods 13
  2.1. Construction of expression vectors and reporter plasmids. 13
  2.1.2. pATG 24
  2.1.2. pDBD 25
  2.2.3. pGFP 26
  2.2. Cell culture and transfection. 27
Chapter 3: Results 29
  3.1. Oct4 domain structure. 29
  3.2. Oct4 sequence identity as function prediction 29
  3.3. Amino Acid Composition 32
    3.3.1 N- Terminal (N) 32
    3.3.2 POU domain 33
    3.3.3 C- Terminal (C) 33
  3.4. Alignments 36
    3.4.1. N-terminal (N) 36

3.4.2. POU 37
3.4.3. C-terminal (C) 39
3.5. Cloning 39
3.6. Nuclear localization 40
3.7. Luciferase Assay 42
3.7.1. pATG 42
3.7.2. pDBD 44
Chapter 4. Discussion 50
4.1. Amino Acid Composition 50
4.2. SEB (Serine [S]/ Glutamic acid [E]) box 52
4.3. Nuclear Localization 53
4.4. Luciferase Assay 53
4.5. Functional Conservation 59
Chapter 5: Conclusions 62
5.1. Future work 65
References 66
APPENDICES 72
APPENDIX 1 *Ambystoma mexicanum* 72
APPENDIX 2 *Mus musculus* 73
APPENDIX 3 *Acipenser oxyrinchus oxyrinchus* 74
APPENDIX 4 *Danio rerio* 75
APPENDIX 5 *Xenopus laevis* 76
APPENDIX 6 *Xenopus laevis- Oct60* 77
APPENDIX 7- *Xenopus laevis- Oct25* 78
APPENDIX 8. Primers combination 79
APPENDIX 9. Amino acid composition of Oct-4 N terminal 81
APPENDIX 10. Amino acid composition of Oct-4 POU domain 84
APPENDIX 11. Amino acid composition of Oct-4 C-terminal 87
APPENDIX 12. pGEM®-T Easy vector (Promega, UK) 90
APPENDIX 13. p6Wtk-luc vector 91

Abbreviations

(C)	Carboxyl- terminal domain
(N)	Amino- terminal domain
µg	Microgram (10-6 g)
aa	Amino acid
axC	Axolotl Oct-4 C domain
axN	Axolotl Oct-4 N domain
axOct4	Axolotl Oct-4
Bp	Base pairs
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
CO_2	Carbon dioxide
DAPI	4,′6-diamidino-2-phenylindole
DBD	DNA Binding Domain
DMEM	Dulbecco’s modified eagles medium
DNA	Deoxyribonucleic acid
dpc	Days poscoitum
ES	Embryonic stem
FFT	Fast Fourier Transform
FGF	Fibroblast growth factor
g	Gram, unit of mass
GFP	Green fluorescent protein
HEK 293T	Human Embryonic Kidney 293T
hESc	Human embryonic stem cells
ICM	Inner cell mass
L	Liter, unit of volume
LARII	Firefly luciferase reagent
LB	Lysogeny broth
M	Molar

mC	Mouse Oct-4 C domain
MCS	Multiple cloning site
Ml	Milli liter (10-3 l)
mM	Milli Molar (10-3 M)
mN	Mouse Oct-4 N domain
mOct4	Mouse Oct-4
MW	Molecular Weight
NLS	Nuclear Localization Site
PBS	Phosphate buffered saline
PGCs	Primordial germ cells
SEB	Serine [S]- Glutamic acid [E] box
SOB	Super Optimal Broth
stC	Sturgeon Oct-4 C domain
stN	Sturgeon Oct-4 N domain
stOct4	Sturgeon Oct-4
TAD	Transcriptional activation domain
TFs	Transcription factors
Xl25	*Xenopus* Oct-25
Xl60	*Xenopus* Oct-60
Xl91	*Xenopus* Oct-91
zfC	Zebrafish pou2 C domain
zfN	Zebrafish pou2 N domain
Zfpou2	Zebrafish POU2

List of figures

Figure 1. 1- Different paths of establish pluripotent embryonic stem (ES) and embryonic germ (EG) cell lines.. 2
Figure 1. 2 Model of Oct4, Nanog, Sox2, and FoxD3 interaction during early mouse development (1)... 9
Figure 1. 4- Oct-4 expression in pre-implantation and early post-implantation during mouse development.. 11
Figure 2. 1- Different Oct-4 domain constructs. 13
Figure 2. 2- ClustW amino acids alignments 14
Figure 2. 3- Enzymes used to create the different Axolotl Oct-4 fragments................. 16
Figure 2. 4- Enzymes used to create the different Mouse Oct-4 fragments. 16
Figure 2. 5- Enzymes used to create the different Sturgeon Oct-4 fragments. 17
Figure 2. 6- Enzymes used to create the different Zebrafish Pou2, Xenopus Oct-91, Xenopus Oct-60 and Xenopus Oct-25.. 18
Figure 2. 7- pATG vector used for cloning. 24
Figure 2. 8- pDBD vector used for the cloning. 25
Figure 2. 9-Vector pGFP used in cloning.. 26
Figure 3. 1- Average distance tree using BLOSUM62. 31
Figure 3. 2- CD MAFT alignment.. 35
Figure 3. 3- N-terminal alignment for all species.. 37
Figure 3. 4- Jalview MAFFT alignments for the different POU domains 38
Figure 3. 5- MAFT multiple alignment for the different C-terminal sequences........... 39
Figure 3. 6- 293 T cells transfected with pGFP Oct-4 full length and POU domain. 41
Figure 3. 7- pATG-Oct-4 Full length transactivation in HeLa and HEK 293T cells. 43
Figure 3. 8- pDBD- OCT-4 full lenght activation in HeLa and HEK 293 T cells 44
Figure 3. 9- pDBD-POU-C Oct-4 transactivation in HeLa and HEK 293T cells 45
Figure 3. 10- pDBD- N-POU Oct-4 transactivation in HeLa and HEK 293T cells. 46
Figure 3. 11- pDBD-N transactivation in HeLa and HEK 293T cells. 47
Figure 3. 12- pDBD-POU transactivation in HeLa and in HEK 293 T cells 48
Figure 3. 13- pDBD-C transactivation in HeLa and in HEK 293T cells. 49
Figure 4. 1- Sumoylation sites in Oct-4 homologues 58

List of tables

Table 1.1- Different strategies for reprogramming differentiated cells into a pluripotent state 3
Table 2. 1- Domain lengths of Oct4 in the studied samples. 15
Table 2. 2-- Axolotl Oct4 enzyme combinations used to generate different fragments. 16
Table 2. 3- Mouse Oct4 enzyme combinations used to generate different fragments ... 17
Table 2. 4- Sturgeon Oct4 enzyme combinations used to generate different fragments.17
Table 2. 5- Zebrafish Pou2 enzyme combinations used to generate different fragments. 18
Table 2. 6- Xenopus Oct-91 enzyme combinations used to generate different fragments. 19
Table 2. 7- Xenopus Oct-60 enzyme combinations used to generate different fragments.. 19
Table 2. 8- Xenopus Oct-25 enzyme combinations used to generate different fragments. 19
Table 2. 9- Enzymes and adaptors used to create different Oct4 fragments for each species.. 20
Table 3. 1- Sequence identity of Oct4. 30
Table 3. 2 Prediction of phosphorylation positions in the C domain, for serine; and tyrosine, according to Netphos 2.0 (56). 34
Table 3. 3 Prediction of phosphorylation positions in the C domain, for serine; and tyrosine, according to Netphos 2.0 (56). 35

Chapter 1: Introduction and literature review

1.1 Overview

Biologists have explored embryonic development, from worms to humans, to understand how different complex organisms can be derived from a single cell, the fertilised egg or zygote [1). It has been demonstrated that many conserved genes and pathways are involved in regulating development. The same genes have been linked to the same cell choices through evolution from simple to complex organism development. We also have a good understanding of the way that the embryo repeatedly uses the same strategies for organogenesis, tissue patterning and cellular specialisation [2]. All species use a common strategy for development, the use of a stem cell to generate and maintain a given tissue or organ. During embryogenesis, cells are initially proliferative and pluripotent; with the ability to differentiate into any cell type; they only gradually become restricted to different cell fates [1, 2]. It is of interest to study if gene functionality is also conserved during evolution, and because the manipulation of these genes and pathways may hold the key to curing all known human disorders

1.2 Pluripotency

Pluripotent cells can give rise to all cells of the embryo or the adult organism [3]. In higher vertebrates Pluripotent cells express Oct-4, Nanog and Sox-2, and these factors are used as pluripotency markers. The potential that pluripotent cells have to develop into specialized cells is being researched for cell and tissue replacement therapies in order to treat disorders such as Parkinson’s disease, rheumatoid arthritis, burns and heart disease [1, 4]. Despite the fact that human embryonic stem cells (hESc) might be used to treat a host of diseases, there are ethical issues regarding the use of human embryos, as well as the problem of tissue rejection followed transplantation [5]. To bypass these issues, there is the possibility of reverting somatic patient cells into pluripotent cells by the constant forced over-expression of pluripotency factors [5]. So far the process of reversion to a pluripotent phenotype is not very well understood however this technology holds much promise for regenerative medicine [6].

The demonstration that fully differentiated cells can revert to pluripotency cell, has resulted in a longer number of groups interested in the process and the different pathways involved [6]. Pluripotent cells can be found in two main embryonic sources. Mouse and human ES cells are derived directly from the inner cell mass (ICM) of embryos at the pre-implantation stage [1] (Figure 1. 1-. Primordial gem cells (PGCs) produce mature germ cells and generate functional adult gametes. Mouse PGCs can be isolated from the gonadal ridge of the embryo [1, 7], and when cultured with appropriate factors *in vitro*, can generate embryonic germ (EG) cells [8]. EG cells have many characteristics of ES cells with respect to their differentiation potential [1], and therefore represent an alternative way to study genes that regulate potency and cell differentiation.

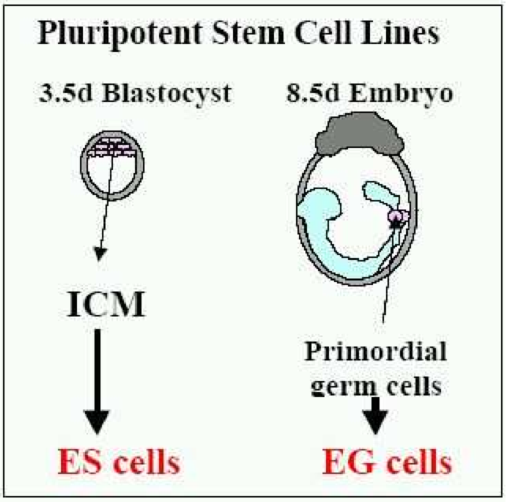


Figure 1. 1- Different paths of establish pluripotent embryonic stem (ES) and embryonic germ (EG) cell lines from the inner cell mass (ICM) of mouse blastocysts and from primordial germ cells, respectively [7].

There are several methods to re-program the nuclei of differentiated cells to an ES cell-like state (Table 1.1) for therapeutic means [9]. But none of these pathways are fully understood, and they still have many restrictions.

Table 1.1- Different strategies for reprogramming differentiated cells into a pluripotent state [9].

Strategies for reprogramming of differentiated cells		
Reprogramming method	Description	Restrictions
Somatic cell nuclear transplantation	Introduction of a somatic cell nucleus into an enucleated unfertilised oocyte. For an increasing number of species, a complete organism can thus be formed by the reconstituted oocyte.	Application may be limited by availability of oocytes and the low cloning efficiency. Furthermore, several developmental abnormalities were observed in cloned animals. Ethical and legal obstacles restrict use of this method for human cells.
Cell-cell fusion	Hybrids of differentiated and pluripotent cells exhibit characteristics of pluripotency.	The reprogrammed cell hybrids contain an additional set of chromosomes. The nucleus of the pluripotent cell may be required for reprogramming.
Treatment with extracts of pluripotent cells	Permeabilised cells are exposed to cell-free extracts of pluripotent cells. Treated cells re-express pluripotency markers and re-differentiate into multiple lineages.	Limited experience with primary cells. Reprogrammed cells will regain only some of the properties of pluripotent cells.
Stable expression of defined factors	Exogenous expression of Sox2, Oct3/4, Kif4, and c-Myc and subsequent selection for pluripotency markers gives rise to cells with similarity to pluripotent cells. Reprogrammed cells can contribute to tissues of all three germ layers in live chimeric mice	Reactivation of transgenes, in particular of the oncogene c-Myc, leads to considerable side effects in offspring oh chimeric mice.

A major breakthrough was made by Takahashi and Yamanaka (2006), who induced pluripotency by retroviral transduction of differentiated cells using four transcription factors: Oct-3/4, Sox-2, c-Myc, and Klf-4 [3, 10]. Later in 2007, Thompson's team screened genes that were highly expressed in hES cells, and found c-Myc and Klf-4 necessary for reprogramming. They were also able to reprogram somatic cells using Oct-4, Sox-2, Nanog, and Lin-28 using lentiviral transduction [10, 11]. In all cases, the re-programmed cells contribute to live chimeric mice and are transmitted via the germline [3]. Although differentiation of cells involves complex genetic and epigenetic changes [3], it is not known how each transcription factor interacts and contributes to pluripotency.

1.3 Research in lower vertebrates

During evolution, most genes have undergone epigenetic modifications. Pluripotent genes are responsible for regulating the differentiation capacity of a cell, and it is likely that these genes might be found conserved during evolution, with small or no epigenetic modifications. In order to understand how pluripotency is regulated, it is easier to study a basic model organism, where the interference of a complex gene network does not occur.

Mammalian systems have a more complex genetic network than amphibians, and mammalian oocytes are also harder to work and have many legal and ethical problems. For this reason *Xenopus* oocytes started to be used to do research on reprogramming by using their germ cell factors [5]. It is also easier to do experiments in *Xenopus* oocytes of than mouse oocytes, as mouse oocytes recover poorly after injection. Both systems have the formation of PGC, and the same genes that are found expressed in inner cell mass (ICM) can also be found in PGC. There are two different ways of generating PCGs.

The germline is created by primordial germ cell (PGC) formation, which may either be initiated by cell-autonomous maternal determinants [12] produced during oogenesis (termed germ plasm), this method is called preformation [13]; or by inductive external signals in the absence of germplasm, termed epigenesis [2].

Frogs, such as *Xenopus laevis,* have germplasm [14], which is seen as dense fibrogranular bodies present in the cytoplasm of oocytes located in the cortex region of the oocyte [13]. During early development there is the inheritance of these maternal determinants [15] leading to the formation of primordial germ cells (PGCs) [13]. By contrast, mammals use Wnt and FGF signalling as key regulators in the inductive methods of PGC formation which takes place later in embryogenesis. The axolotl (*Mexican salamander*), is actively being used to study PGC development by the Johnson laboratory, due to the fact that mouse and axolotl appear to share the same inductive mechanism. Axolotl germ cells are induced in mesodermal tissues during gastrulation, by a very similar process to that of mouse germline formation [14, 15]. This finding brought a new insight in research, not only do they share the same development process but they also have the advantages of amphibian oocytes.

The two different mechanisms of PGC formation is hypothesised to have lead to the appearance of two major amphibian groups; the anurans (frogs and toads), and the urodeles (salamanders) [14]. When the adult morphology of frogs (anuran) or axolotls (urodele) are compared to other vertebrates it is seen that urodeles have a similar skeletal structure to most vertebrates them frogs [14]. More morphological analyses between those two species also show that the frog’s anatomy has become more specialised, with significant variability among species than axolotls. Frogs also go through many development stages that are not necessary for mammals [14]. Despite the fact that axolotl and frogs diverged from a common ancestor [12], axolotls retained more ancestral features and are less variable than frogs, and therefore is likely that one of the mechanisms of PGC specification in urodeles is more conserved than in frogs [14]. A strong link can therefore be drawn between animals that make PGCs by maternal determinants to allow more rapid speciation and evolution (frogs), and animals that epigenetically specify PGCs which are slower to speciate and evolve (axolotl and mouse).

It was assumed that as germplasm determines the formation of PGCs via coding and non-coding maternal RNAs, such as *vasa*, *dazl* genes [12, 15] that these would not be present or expressed in animals showing epigenetic PGC formation, but this was contradicted in 2001, when Johnson *et al* were able to clone the *dazl* sequence from axolotls.

Oct-4 has been shown to be essential for germ cell development in mammals [12] and is responsible for maintaining the ability of a cell to differentiate into all cell type; in the mammalian germ-line, and it was thought that Oct-4 was an evolutionary gene restricted to mammalian systems. However, in 1992, three homologous Oct-4 (family termed POU-V) sequences were isolated in *Xenopus laevis;* Oct-25 (Xl25), Oct-91 (Xl91) and Oct-60 (Xl60) [16]. In 2006, a homologous sequence was found in Zebrafish, Pou2 (zfpou2) [17]. Later, the Johnson Group cloned Oct-4 from axolotls (axOct4) [18], and more recently, in sturgeon (stOct4) (unpublished). The presence of Oct-4 in species that specify PGCs by epigenesis and by preformation suggest that Oct-4 not only is an ancient gene, has but also has an important role in both types of PGC formation.

Other genes, such as the well characterised transcription factors Oct-4, Nanog, c-Myc,Sox2 between others, show a significant role in PGC determination, and are also found in lower vertebrates during early development. It is of interest that when *vasa*, *dazl* and Oct-4 from axolotl, *Xenopus* and mouse are aligned the axolotl sequence shows more similarity to mouse than *Xenopus* sequence [12]. It is likely that in urodeles, like in mammals, RNA binding proteins (*vasa* and *dazl*) and the transcription factor Oct-4, might control the formation of PGC and retain the potency for self renewal [15].

Despite species having different modes of embryonic development, they all conserve the same inductive ancestral genes for PGC formation, whether by epigenesis or by preformation. However, it remains to be tested if their function was also conserved during evolution. The fact that axolotl sequences share a higher homology with mice suggest that they may be a good model organism for studying and understanding the biology of mammalian germ cell development, and the regulatory mechanism to produce stem cells. Gathering all the amphibian advantages, new insights might be gained into Oct-4 function through experiments in axolotls instead of *Xenopus*.

1.4 Oct-4 belongs to POU transcription factor family

The POU family of transcription factors share homology to a domain first found in mammalian transcription factors Pit-1, Oct-1 and a nematode regulatory protein, Unc-86 [23]. POU family transcription factors activate expression of genes by binding to an octameric sequence [AGTC(A/T)AAT] found in the regulatory sequences of cell type-specific as well as ubiquitous genes [1]. The POU domain binds to DNA through recognition of the helix-turn-helix region within the POU domain, with the bases in the DNA major groove at the 3' A/TTTA rich portion of the octamer site [1].

The POU family members share a similar DNA binding domain of approximately 160 residues [19], and contain two structurally independent subdomains, an amino-terminal specific region of 75-amino-acids (POUs) and a carboxy-terminal homeodomain of 60 amino-acids (POU_H) [20] connected by a flexible linker of variable length [1]. The POU domain was highly conserved during evolution – the mouse Oct-4 POU domain shares high similarity with amphibian sequences. It has been shown that the *Xenopus* Oct-4 homologues can replace the mouse POU domain in ES cells to support their self-renewal; zebrafish pou2 is unable to do this [21]. This suggests that the Oct-4 transactivation function has been preserved in some species during evolution.

POU transcription factors are divided into five classes based on the degree of conservation within these domains and linker region [21].The POU-V family are considered to play an important role during embryogenesis, pattern formation and cellular differentiation [22]. Oct-4 is also known as Oct3, Oct-3/4, POU5F1, OTF3, and NF-A3 [23].

Oct-4 is transported to the nucleus, where it binds the octamer motif and starts to transactivate its target genes [24]. The mechanism by which Oct-4 is transported into the nucleus is still unknown, but a conserved sequence necessary for nuclear transportation has been found [24]. The classical nuclear localization signal (NLS) RKRKR was also found within the POU domain [24).

1.5 Oct-4 domains

The POU domain in Oct-4 is surrounded by N- and C- terminal domains that are responsible for the transactivation capacity [1, 15]. Despite both functioning as transactivator domains, the C-terminal (C) shows less activity then the N- terminal (N) when tested in ES cells [16]; however, this has not been tested in other cell types, or by using Oct-4 homologues from other species. The (N), POU and (C) domains by themselves are not sufficient for ES cell self-renewal [16, 21].

The (N) domain is found upstream from the POU domain, and in mouse is 126 aa [21]. Oct-4 like proteins in other species have longer N-domains (e.g. 252 aa for Zebrafish Pou2). Due to the fact that 25% of the domain is proline (13- 60 aa) it is considered a proline-rich transactivation domain [16]. It was shown that the proline-rich region has an important role on Oct-4 transactivation, as when deleted Oct-4 transactivation is reduced [21]. Further studies revealed that deletion of the full N-domain decreases Oct-4 transactivation. Nonetheless, it was still possible to see a significant transcriptional function in ES cells [16]. Mouse (N) domain has a target site for a small ubiquitin-related modifier (SUMO-1) that results in a significant increase of Oct-4 stability and DNA binding [23]. However, if these sumoylation sites are conserved and influence Oct-4 homologues activity, is still not known.

The (C)- domain is found downstream of the POU domain, and in mouse it contains 67 to 95 aa. It has proline, serine and threonine residues [20] and is known as a serine/threonine-rich transactivation domain [16]. The C-domain is cell-type-specific and is mediated by the Oct-4 POU domain [20]. It is believed that (C) might be regulated by phosphorylation events [15, 20].

1.6 Oct-4 and pluripotency

It was shown that in the absence of Oct-4, ES cells lose the ability to self-renew (Niwa, 2002). Oct-4 is not the only gene responsible for maintaining pluripotency - other genes act together at different development stages (Figure 1. 2). In mouse, Oct-4 transactivation is also mediated by directly binding Sox2, a Sry/HMG transcription factor [15, 25]. Oct-4/ Sox-2 co-operate to up-regulate genes during pre-implantation development, such as FGF-4 and osteopontin (OPN), and down-regulate human chorionic gonadotropin (HCG) [15].

Nanog, another critical transcription factor that promotes ES self-renewal, was found by Chambers *et al.* (2003) [26] and Mitsui *et a.l* (2003) [27]. Nanog is a divergent homeobox factor expressed *in vivo* during early development [1]. *In vitro* it is a marker of all pluripotent cell lines (both murine and human) [1, 27]. Nanog and Oct-4 work together in order to maintain a pluripotent phenotype. In mouse, it is known that Nanog and Oct-4 interact directly with each other in a transcriptional complex with chromatin modifiers (James Dixon, personal communication). It is not known whether the ability to bind Nanog is common to all Oct-4 proteins. Nanog regulates Oct-4 and Sox2 levels to preserve self-renewal and to prevent differentiation of the ICM into all three germ layers [28].

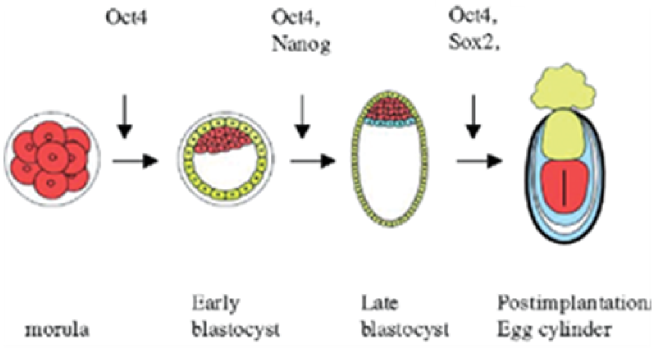


Figure 1. 2 Model of Oct4, Nanog, Sox2, and FoxD3 interaction during early mouse development [1].

Oct4- has been proven to be a transcription factor necessary to maintain pluripotency in embryonic stem cells (ES cells) along with other transcriptional co-regulators, such as *Nanog* and *Sox-2* [1,29], but once cells start to differentiate, it is switched off. Oct-4 is not restricted to ES cells, but is also found in adult tissues, such as bone marrow, intestinal epithelium, brain, liver and hair follicles [29, 30]. To investigate whether Oct4 is necessary for adult tissue renewal, Jaenish in 2007 deleted Oct-4 in differentiated cell lines, and showed that, despite lacking Oct-4 cells where able to renew. This study shows that Oct-4 does not play a self-renewal role in adult cells [29, 30], and highlights some new ideas and concepts. Despite the fact that Oct-4 has a role in maintaining ES pluripotency, it does not have the same role in adult cells and could mean that pluripotency and self-renewal have different regulatory mechanisms [30].

1.7 Oct4 expression

In mouse Oct-4 is a maternally inherited transcription factor that is expressed at low levels in all blastomeres until the 4-cell stage. Afterwards, the gene undergoes zygotic activation, resulting in high Oct-4 protein levels in the nuclei of all blastomeres until their compaction [1, 15] (Figure 1. 3). After cavitation, Oct-4 is only expressed in the inner cells (ICM) of the blastocyst and it is downregulated in the differentiated trophectodrem (TE). It was thought to be one of the first transcription factors to be regulated in early development [15, 20]. But nowadays it is known that Oct-4 requires the interaction with other transactivation factors such as Sox-2, to maintain ES cells in a pluripotent state [25].

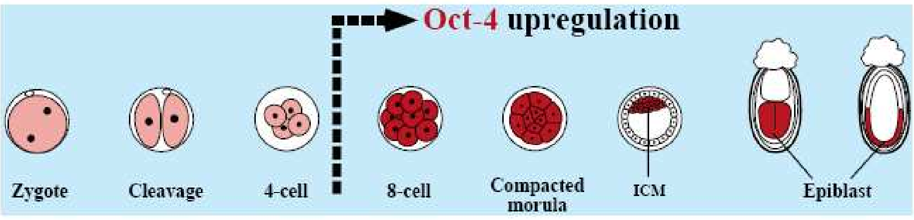


Figure 1. 3- Oct-4 expression in pre-implantation and early post-implantation during mouse development. It begins to be present as maternal transcript in the zygote and remains low until the 8-cell stage when the gene is zygotically activated, eventually becoming restricted to the epiblast [15].

A study with different Oct-4 expression levels revealed that when Oct-4 expression is increased by 50%, it leads ES cells to differentiate into extra-embryonic endoderm and mesoderm. However, a 50% decrease result in trophoectoderm differentiation [31]. Nonetheless, subtle Oct-4 changes do not have a drastic effect. *Xenopus* has a similar expression regulation, when Oct-25 (Xl25), Oct-60 (Xl60) and Oct-91 (Xl91) are over-expressed it suppresses mesoderm formation, while loss of Oct-25 and Oct-60 results in mesoderm differentiation [22]. Oct-4 cell-type specificity and temporal expression suggest unique mechanisms for regulating its expression [32], it is the precise threshold of Oct-4 that determines the three possible cell paths (self-renewal, trophectoderm, or extra-embryonic endoderm and mesoderm) [1].

1.8 Aims and Objectives

Oct-4 passed from a mammalian specific transcription factor to a transcription factor present in at least five embryo vertebrates; mammals; axolotl [12]; sturgeon (not published); in zebrafish and *Xenopus* [21]. Oct-4 in mammals is responsible for maintaining pluripotency, and essential for the formation of PGCs. This gene is found conserved during evolution, and it is of interest to study if it retained the same function to activate Oct-4 target genes from mammals to lower vertebrates.

The aim of the present study was *i*) to see whether the Oct-4 proteins from different species share the same functional domain; *ii)* if their activity was conserved during evolution in differentiated and pluripotent cells; *iii)* to determine how domains influence Oct-4 transactivation in these cells; *iv)* create different molecular tools for Oct-4 genes to allow further functional analysis and *vi)* assess subcellular localization.

This study will not only highlight different aspects of Oct-4 conservation and function, but it will also allow determination which of lower vertebrate share the highest functional conservation to mammals and present an alternative model system to study pluripotency pathways.

These aims were addressed by caring out the following studies:

a) Determination of the different domains for all Oct-4 sequences; analyse their length, amino acid composition and conservation;

b) Cloning of the full length and Oct-4 domains into a general utility vector by PCR;

c) Sub-clone these into over-expression GFP and Gal4DBD vectors;

d) Assess Oct-4 localization.

e) Analysis of the function of Oct-4 to activate target genes by reporter analysis;

f) Analysis of the regulation and conservation of particular Oct-4 domains for their function.

Chapter 2: Material and Methods

2.1. Construction of expression vectors and reporter plasmids.

Oct-4 is divided in three different domains, an amino terminal domain (N-domain) which has a large proportion of proline residues, a carboxyl terminal domain (C-domain), which has a large proportion of proline serine and threonine residues that surrounds the POU domain [16]. Different Oct-4 fragments were constructed for each species, the Oct4 full- length (FL), the N-terminal and POU domain (N-POU), the C terminal and POU domain (POU-C), and also fragments with only the N-domain (N), POUV domain (POU) or C-domain (C), all fragments are described on Figure 2. 1.

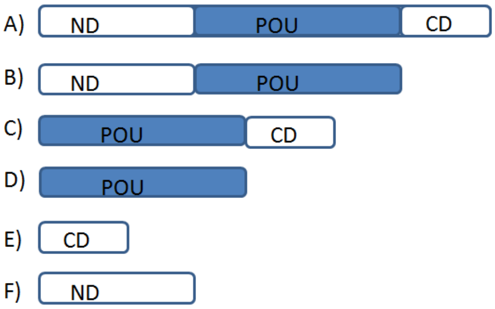


Figure 2. 1- Different Oct-4 domain constructs. A) full length Oct-4 domain; B) N-POU domain, that only has the N-terminal and POU domain; C) POU-C domain, only with the POU domain and C-terminal; D) POU domain, E) C domain, only the C-terminal, F) N domain, only with the N-terminal.

Oct-4 Sturgeon (stOct4) sequence was supplied by James Dixon. Axolotl (axOct4), mouse (mOct4), *Xenopus* (XL91, Xl60, Xl25) and Zebrafish (Pou2) Oct-4 sequences were obtained through NCBI database [33], (sequences and accession numbers are given in appendix 1 to 7).

Brikman in “Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development” [9], describes the domain lengths for mouse Oct-4, *Xenopus* Oct-25, *Xenopus* Oct-91, *Xenopus* Oct-60 among others. This information was not only used for domain constructs, but also, to predict Pou2, stOct-4 and axOct-4 domains by using ClustW multiple alignments on BioEdit v. 7.0.9 [34]. The resulting alignment is shown on Figure 2. 2.

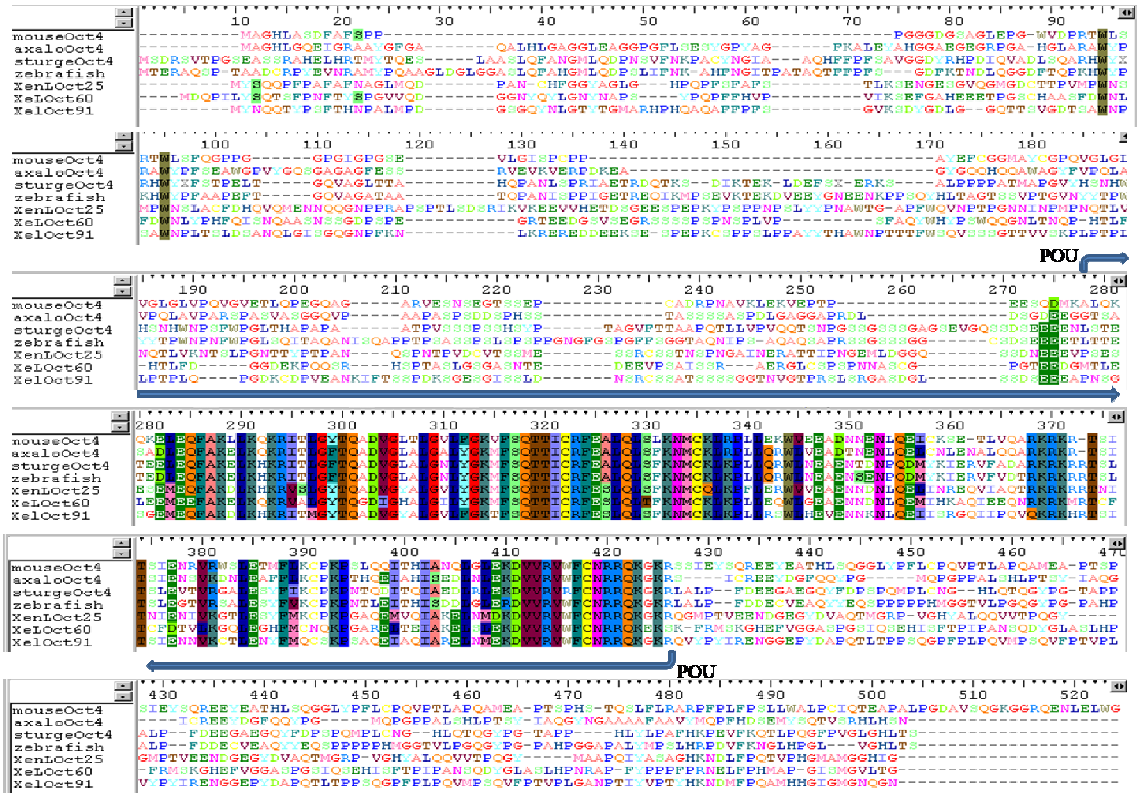


Figure 2. 2- ClustW amino acids alignments, (the lightened area determines the POU domain, the upstream region correspond to N-terminal domain (N) and the downstream region corresponds to the C-terminal domain (C).)

The lengths of each domain to be used to PCR amplification are described in

Table 2. 1.

Table 2. 1- Domain lengths of Oct4 in the studied samples.

Specie	**Domain**	**Length (bp)**
Axolotl Oct4	N domain	1-549
	POU domain	550- 996
	C domain	997- 1197
Mouse Oct4	N domain	1-399
	POU domain	400- 846
	C domain	847- 1131
Sturgeon Oct4	N domain	1- 705
	POU domain	706- 1158
	C domain	1159- 1362
Zebrafish	N domain	1- 756
	POU domain	757- 1209
	C domain	1210- 1419
Xenopus 91	N domain	1- 660
	POU domain	661- 1143
	C domain	1144- 1338
Xenopus 60	N domain	1- 612
	POU domain	613- 1076
	C domain	1077- 1281
Xenopus 25	N domain	1- 690
	POU domain	691- 1140
	C domain	1141- 1347

Each fragment was cloned using different restriction enzyme sites surrounding them. Using NEB cutter Ver.2.0 (35), it was possible to identify which enzymes that do not cut in each sequence, and ones that cut within the fragments. The main restriction enzymes chosen where; Nhe, AgeI, XhoI and HindIII and only when necessary replaced by, other restriction enzymes, Sal 1 and and BglII. Restriction enzyme sequences were added to the primer sequence.

Axolotl Oct4 has a restriction site at the 82 bp for the Xho1 (35), indicated on Figure 2. 3 with the symbol (▲), so only the fragment N-POU needs to be replaced by Sal1 to avoid double digestion. All enzyme combinations for each domain are indicated on Figure 2. 3 and constructs domain lengths for axOct4 are indicated in the Table 2. 2.

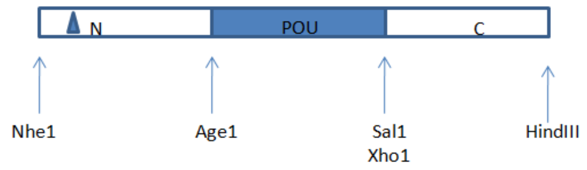


Figure 2. 3- Enzymes used to create the different Axolotl Oct-4 fragments. ▲ indicates where the Xho1 cuts the sequence.

Table 2. 2-- Axolotl Oct4 enzyme combinations used to generate different fragments.

Specie	Restriction Enzymes	Domain	Size (bp)
Axolotl Oct4	Nhe1 + HindIII	Full- Length	1197
	Age1 + HindIII	POU-C	648
	Nhe1 + Sal1	N-POU	996
	Nhe1+ Age1	N domain	549
	Age1 + Xho1	POU domain	447
	Xho1 + HindIII	C domain	201

Mouse Oct4 has a restriction enzyme at the 105 bp for HindIII [35], indicated on Figure 2-4 with the symbol (▲), so the only full- length fragment needs a replacement of the HindIII adaptor for BglII. All enzyme combinations and constructs used to make the different fragments are described on the Figure 2. 4, and the respective domain sizes can be found in Table 2. 3.

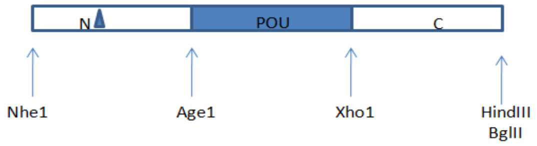


Figure 2. 4- Enzymes used to create the different Mouse Oct-4 fragments. ▲ indicates where the HindIII cuts the sequence.

Table 2. 3- Mouse Oct4 enzyme combinations used to generate different fragments

Specie	**Restriction Enzymes**	**Domain**	**Size (bp)**
Mouse Oct4	Nhe1 + BglII	Full Length	1131
	Age1 + HindIII	POU-C	732
	Nhe1 + Xho1	N-POU	846
	Nhe1+ Age1	N domain	399
	Age1 + Xho1	POU domain	447
	Xho1+ HindIII	C domain	285

Sturgeon Oct-4 has a restriction site at the 1024 bp for Xho1 [35], indicated on Figure 2. 5 with the symbol (▲), therefore the fragments N-POU and POU cannot take the Xho1 adapter, being replaced by Sal1. The enzyme combinations used to make the fragments are described in Table 2. 4.

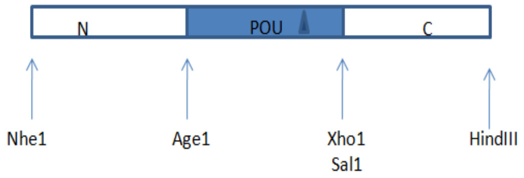


Figure 2. 5- Enzymes used to create the different Sturgeon Oct-4 fragments. ▲ indicates where the Xho1 cuts the sequence.

Table 2. 4- Sturgeon Oct4 enzyme combinations used to generate different fragments.

Specie	Restriction Enzymes	Domain	Size (bp)
Sturgeon Oct4	Nhe1 + HindIII	Full Length	1362
	Age1 + HindIII	POU-C	657
	Nhe1 + Sal1	N-POU	1158
	Nhe1 + Age1	N domain	705
	Age1 + Sal1	POU domain	453
	Xho1+ HindIII	C domain	204

Zebrafish Pou2 and all *Xenopus* (Xl25, Xl60, Xl91) cDNA do not have restriction sites for any of the main enzymes (Nhe1, Age1, Xho1 and HindIII) [35] so all the fragments were inserted with the same adapters but generating different size fragments. The restriction enzymes sites are described on Figure 2-6, and all the different combinations and fragment sizes for zfPou2 are in the Table 2. 5, for Xl91 the fragments are described on Table 2. 6, for Xl60 the fragments are on

Table 2. 7 and for Xl25 are in

Table 2. 8.

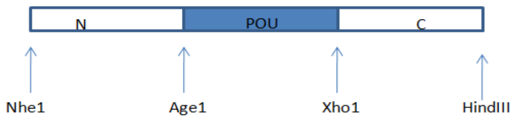


Figure 2. 6- Enzymes used to create the different Zebrafish Pou2, *Xenopus* Oct-91, *Xenopus* Oct-60 and *Xenopus* Oct-25.

Table 2. 5- Zebrafish Pou2 enzyme combinations used to generate different fragments.

Specie	Restriction Enzymes	Domain	Size (bp)
Zebrafish	Nhe1 + HindIII	Full Length	1419
	Age1 + HindIII	POU-C	663
	Nhe1 + Xho1	N-POU	1209

	Nhe1+ Age1	N domain	756
	Age1 + Xho1	POU domain	453
	Xho1 + HindIII	C domain	210

Table 2. 6- *Xenopus* Oct-91 enzyme combinations used to generate different fragments.

Specie	**Restriction Enzymes**	**Domain**	**Size (bp)**
Xenopus Oct-91	Nhe1 + HindIII	Full length	1338
	Age1 + HindIII	POU-C	678
	Nhe1 + Xho1	N_POU	1143
	Nhe1+ Age1	N domain	660
	Age1 + Xho1	POU domain	483
	Xho1 + HindIII	C domain	195

Table 2. 7- *Xenopus* Oct-60 enzyme combinations used to generate different fragments..

Specie	**Restriction Enzymes**	**Domain**	**Size (bp)**
Xenopus Oct-60	Nhe1 + HindIII	Full length	1281
	Age1 + HindIII	POU-C	669
	Nhe1 + Xho1	N-POU	1076
	Nhe1+ Age1	N domain	612
	Age1 + Xho1	POU domain	464
	Xho1 + HindIII	C domain	205

Table 2. 8- *Xenopus* Oct-25 enzyme combinations used to generate different fragments.

Specie	**Restriction Enzymes**	**Domain**	**Size (bp)**

Xenopus Oct-25	Nhe1 + HindIII	Full length	1347
	Age1 + HindIII	POU-C	657
	Nhe1 + Xho1	N-POU	1140
	Nhe1+ Age1	N domain	690
	Age1 + Xho1	POU domain	450
	Xho1 + HindIII	C domain	207

Each fragment was amplified with primers corresponding to specific enzyme with the addition of a random nucleotide sequence: CAGT at the 5’ extremity of each primer. The enzymes and their corresponding sequences that were used are described in Table 2.9. The total oligonucleotide sequence with the adaptor sequences are described in Table 2-10.

Table 2. 9- Enzymes and adaptors used to create different Oct4 fragments for each species.

Fragment	Enzyme	Adaptors	A	M	S	Z	25	60	91
FL	Nhe1 F+ HindIII R	5’CAGT**GCTAGC+ 5’**CAGT**AAGCTT**	+	-	+	+	+	+	+
FL	Nhe1 F+ BglII R	5’CAGT**GCTAGC+ 5’** CAGT**AGATCT**	-	+	-	-	-	-	-
POU-C	Age1 F+ HindIII R	5’CAGT**ACCGGT+ 5’**CAGT**AAGCTT**	+	+	+	+	+	+	+
N-POU	Nhe1 F+ Xho1 R	5’CAGT**GCTAGC +5** CAGT**CTCGAG**	-	+	-	+	+	+	+
N-POU	Nhe1 F+ Sal1 R	5’CAGT**GCTAGC +5’**CAGT**GTCGAC**	+	-	+	-	-	-	-
POU	Age1 F+ Xho1 R	5’CAGT**ACCGGT+ 5’**CAGT**CTCGAG**	+	+	-	+	+	+	+
POU	Age1 F+ Sal1 R	5’CAGT**ACCGGT+5’**CAGT**GTCGAC**	-	-	+	-	-	-	-
CD	Xho1 F+ HindIIIR	5’CAGT**CTCGAG+ 5’**CAGT**AAGCTT**	+	+	+	+	+	+	+
ND	Nhe1F + Age1 R	5’CAGT**GCTAGC+ 5’**CAGT**ACCGGT**	+	+	+	+	+	+	+

Adaptors include the CAGT sequence plus the enzyme sequence. The boldface section corresponds to the enzyme sequence. A- Axolotl, M- Mouse, S- Sturgeon, 25- *Xenopus* Oct25, 60- *Xenopus* Oct60, 91- *Xenopus* Oct91. (+) where the combination was used (-) species without this combination

.

Table 2-10. Oligonucleotides used for creating Oct4 mutants

MOUSE OCT4	
MOCT4-NHE1-F1	5’CAGTGCTAGCATGGCTGGACACCTGGCTTCA

MOCT4-AGE1-R1	5’CAGTACCGGTTTTCATGTCCTGGGACTCCTC
MOCT4-AGE1-F1	5’CAGTACCGGTGCCCTGCAGAAGGAGCTAGAAC
MOCT4-XHO1-R1	5’CAGTCTCGAGACTTGATCTTTTGCCCTTCTG
MOCT4-XHO1-F1	5’CAGTCTCGAGATTGAGTATTCCCAACGAGAA
MOCT4-BGLII-R1	5’CAGTAGATCTACCCCAAAGCTCCAGGTTCTC
AXOLOTL OCT4	
AXOCT4-NHE1-F1	5’CAGTGCTAGCATGGCTGGGCATTTGGGACAG
AXOCT4-AGE1-R1	5’CAGTACCGGTCCCTTCCTCGTCTCCGCTGTC
AXOCT4-AGE1-F1	5’CAGTACCGGTGGGACGTCGGCGGACCTTGAA
AXOCT4-SALI-R1	5’CAGTGTCGACGCTGCGCTTCCCCTTCTGTCG
AXOCT4-XHO1-F1	5’CAGTCTCGAGATTTGCCGGGAGGAGTATGAT
AXOCT4-HINDIII-R1	5’CAGTAAGCTTGTTGGAGTGCAGGTGCCTTCT
STURGEON OCT4	
STOCT4-NHE1-F1	5’CAGTGCTAGCATGTCTGATCGGTCTGTCACC
STOCT4-AGE1-R1	5’CAGTACCGGTATTCTCCTCTTCTTCCGAGTC
STOCT4-AGE1-F1	5’CAGTACCGGTTTGTCCACGGAGGAGCTGGAG
STOCT4-SALI-R1	5’CAGTGTCGACGGCCAGACGCTTCCCCTTCTG
STOCT4-XHO1-F1	5’CAGTCTCGAGCTGCCCTTTGATGAGGAGGGT
STOCT4-HINDIII-R1	5’CAGTAAGCTTGCTGGTCAGGTGTCCCAGCCC
ZEBRAFISH OCT2	
ZFOCT4-NHE1-F1	5’CAGTGCTAGCATGACGGAGAGAGCGCAGAGCCCA
ZFOCT4-AGE1-R1	5’CAGTACCGGTCAGAGTCTCCTCTTCCTCAGA
ZFOCT4-AGE1-F1	5’CAGTACCGGTACTACTGAAGATTTGGAGCAG
ZFOCT4-XHO1-R1	5’CAGTCTCGAGCAAAGCTAGACGCTTTCCCTT
ZFOCT4-XHO1-F1	5’CAGTCTCGAGCCCTTTGATGACGAGTGTGTT
ZFOCT4-HINDIII-R1	5’CAGTAAGCTTGCTGGTGAGATGACCCACCAA
***XENOPUS* OCT91**	
XNOCT91-NHE1-F1	5’CAGTGCTAGC ATGTATAACCAACAGACCTACCCT
XLOCT91-AGE1-R1	5’CAGTACCGGT GGCTTCCTCCTCACTGTCACT
XLOCT91-AGE1-F1	5’CAGTACCGGTCCTAATTCTGGGGAGATGGAG
XLOCT91-XHO1-R1	5’CAGTCTCGAGCTCCCCGCCATTCTCCCTAAT
XLOCT91-XHO1-F1	5’CAGTCTCGAGCCTTATGACGCCCCCCAAACC
XLOCT91-HINDIII-R1	5’CAGTAAGCTTGTTGCCTTGGTTACCCATGCC
***XENOPUS* OCT60**	
XLOCT60-NHE1-F1	5’CAGTGCTAG ATGGACCAGCCCATATTGTACAGC
XLOCT60-AGE1-R1	5’CAGT ACCGGTTCCATCCTCTTCAGTTCCAG
XLOCT60-AGE1-F1	5’CAGT ACCGGTATGACCCTTGAGGAGATGGAA
XLOCT60-XHO1-R1	5’CAGT CTCGAG TTGGACATTCTGAACTTGCTC
XLOCT60-XHO1-F1	5’CAGT CTCGAGGGGGCATGAGTTTGTGGGTGG
XLOCT60-HINDIII-R1	5’CAGT AAGCTTGCCGGTCAGGACCCCCATAGA
***XENOPUS* OCT25**	
XLOCT25-NHE1-F1	5’CAGTGCTAGCATGTACAGCCAACAGCCCTTCCCA
XLOCT25-AGE1-R1	5’CAGTACCGGTGGGAACCTCCTCCTCATTGTC
XLOCT25-AGE1-F1	5’CAGTACCGGTAGCGAATCAGAAATGGAGCAG
XLOCT25-XHO1-R1	5’CAGTCTCGAGTCCCTGGCGCTTGCCTTTCTG

XLOCT25-XHO1-F1	5’CAGTCTCGAGATGCCCACCGTTGAGGAGAAC
XLOCT25-HINDIII-R1	5’CAGTAAGCTTGCCAATGTGGCCCCCCATGGC
XLOCT25-FULL LENGTH F1	5’ATGTACAGCCAACAGCCCTTC
XLOCT25-FULL LENGTH R1	5’TCAGCCAATGTGGCCCCCCAT

The different fragments were generated by PCR from plasmids containing the cDNA templates of the genes to be amplified. This plasmids were supplied by James Dixon and Jodie Edgson. The amplification was done using the primers described above using REDTaq REadyMix (Sigma- Aldrich, UK) according to the manufacturer’s instructions. The amplification was performed on an Eppendorf Mastercycle epGradient S cycler machine. PCR reactions were heated to 94°C for 5 minutes followed by 45 cycles of denaturation at 94°C for 1 minute, annealing at 56°C for 5 1 minute and 30 seconds, and extension at 72°C for 1 minute and 30 seconds. After amplification the PCR samples were loaded on a (0.5X) TAE Agarose gel (1.2%) and separated at 135v. Their sizes where compared to 1kb and 100bp molecular weight markers (Biolabs, New England). After isolating the correct amplified size segments, they were purified from the gel using Qiaspin column and Gel Extraction Kits (QIAGEN, UK) according to the manufacturer’s instructions.

PCR products were ligated into the pGEM®-T Easy vector (Promega, UK) according to the manufacturer’s instructions. The ligation mix was used to transform DH5-α E. coli competent cells. The culture cells were plated in LB agar medium supplemented with (10 mg/ml) Amplicilin and IPTG to 0.5mM and left to grow overnight at 37ºC. Individual colonies were selected and transfered to liquid LB medium with Amplicilin, overnight at 37ºC.

Plasmid DNA was extracted from cultures using QIAprep Spin Miniprep Kit using a microcentrifuge (QIAGEN, UK). To confirm that the fragments were inserted into the pGEM®-T Easy vector, 5 µL of the miniprep DNA was digested with EcoRI (10000U/ml) (BioLabs, New England) for 1 hour and 30 minutes at 37ºC with 2µL of Buffer 2 (BioLabs, New England), 2 µL of 10X BSA, 2µL o EcoRI (BioLabs, New England) and 9µL of purified water. EcoRI cuts on either side of pGEM®-T Easy (appendix 12) vector at the 23 and 70 bp after the transcription site, it also cuts the stOct-4 insert at 398 bp, due to an internal EcoRI site. When the fragments are correctly inserted into pGEM®-T the fragment sizes described in the Table 2-11, should be obtained.

Table 2-11- Size of the different fragments constructed of each species.

Species	Domain	Size (bp)
Axolotl Oct4	Full- Length	1197
	POU-C	648
	N-POU	996
	N domain	549
	POU domain	447
	C domain	201
Mouse Oct4	Full Length	1131
	POU-C	732
	N-POU	846
	N domain	399
	POU domain	447
	C domain	285
Sturgeon Oct4	Full Length	964 and 398
	POU-C	657
	N-POU	1158
	N domain	398 and 310
	POU domain	453
	C domain	204
Zebrafish	Full Length	1419
	POU-C	663
	N-POU	1209
	N domain	756
	POU domain	453
	C domain	210
Xenopus 91	Full length	1338
	POU-C	678
	N_POU	1143

	N domain	660
	POU domain	483
	C domain	195
Xenopus 60	Full length	1281
	POU-C	669
	N-POU	1076
	N domain	612
	POU domain	464
	C domain	205
Xenopus 25	Full length	1347
	POU-C	657
	N-POU	1140
	N domain	690
	POU domain	450
	C domain	207

The primary objective is to generate plasmids for the generation of fragments that can be ligated into specific vectors, pATG and pDBD, for transcriptional assays.

2.1.2. pATG

pATG vector is derived from pEGFP-C1 (Clontech) GenBank accession #: U55763. The vector and the alterations can be seen on Figure 2. 7. The Multiple cloning site (MCS) of pEGFP-C1 vector can be found in appendix 10. The MCS of pATG consists of Nhe1, Age1, Xho1, HindIII, BamHI (James Dixon, personal communication)

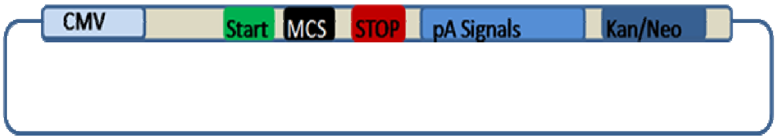


Figure 2. 7- pATG vector used for cloning.

The enzymes in the pATG MCS are compatible with the fragments generated above. In order to clone the fragments into pATG, the vector and the fragments need to be digested with the respective enzymes that are described on Table 2. 2 and on

Table 2. 8. To increase cloning efficiency pATG was also desphosphorylated by alkaline phosphatase, calf intestinal (CIP) (10,000 units/ml) (BioLabs, New England) after a double digestion with the restriction enzymes.

Enzymatic digestion was performed using 300 ng/µL of plasmid, 3 µL of enzyme (1,5 µl each), 3µL of BSA (10X) (BioLabs, New England) and 3µL of the respective buffer to a final volume of 30 µL. The buffers used for plasmid digestion were chosen using BioLab instructions [36]. Digestion was performed at 37ºC for 1h and 30 minutes.

After digestion, the samples were loaded into a (0.5X) TAE 1.2% Agarose Gel and run at 135V. The correct fragments were identified by comparison to 1kb and 100bp molecular weight (MW) makers (BioLabs, NewEngland); extracted and purified using Qiaquick Spin- Gel Extraction Kit (QIAGEN, UK) according to manufacturer's instructions.

2.1.2. pDBD

pDBD- Gal4 vector was prepared by James Dixon (personal communication). It has the same similar MCS as pATG with the exception for Xho1which was replaced by Sal1 (cohesive ends of Xho1 and Sal1) [36]. The vector can be seen in Figure 2. 8.

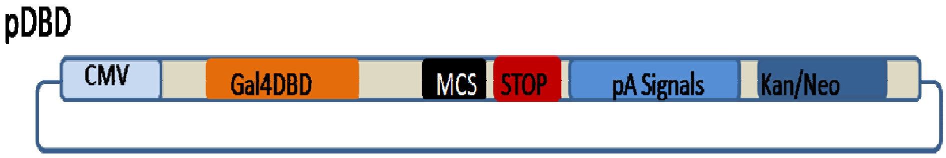


Figure 2. 8- pDBD vector used for the cloning.

The vector was digested with the appropriate enzymes using the buffers suggested by BioLab instruction [36]. To create some of the vectors it necessary to do sequential digestion, due to the incompatible enzyme buffers [36]. The digestions were purified purified using Qiaquick Spin columns (QIAGEN, UK). Each digestion was done at 37ºC for 1h and 30 min. After digestion, the samples and 1Kb and 100bp DNA ladder (BioLabs, New England) were loaded into TAE 1.2% Agarose Gel and fragments separated at 135V. The correct fragments were extracted and purified using Qiaquick Spin- Gel Extraction Kit (QIAGEN, UK) according to manufacturer's instructions.

The ligation of fragments to vector was done to a ratio 5:1 in the presence of T4 DNA ligase (2000,000 units/mL) (BioLabs, New England) in appropriate buffer (BioLabs, New England). The reaction was performed at room temperature for 1h and 30 min. After ligation the mix was used to transform DH5α competent cells. The transformations were performed by adding 100µL of competent cells to the ligation reaction, and leaving on ice for 30 min. Then heatshock was performed at 42ºC for 45 seconds followed by returning to ice for 3 minutes. 500µL of SOB is then added and the cells incubate in a 37ºC shaker for 1 hour and 30 minutes. After transformation, 200µL of the mixture is spread on an Agar (30µg/ml) plate and incubated overnight at 37ºC. Two distinct colonies are then picked and grown overnight in a 37ºC shaker, in liquid LB media supplemented with Kanamycin ((30µg/ml). Plasmid DNA is then extracted using Qiagen Plasmid Mini Kits (QUIAGEN, UK) according to manufacturer's instructions.

2.1.3. pGFP

pGFP vector is identical to pDBD vector but instead of the DNA binding domain is has a green fluorescent protein (GFP) molecule, Figure 2. 9. The ligation and transformation were done in the same way as above described.

Figure 2. 9-Vector pGFP used in cloning.

HeLa cells, were transfected with 0. 25µg/well of pGFP using GeneJuice Trasfection Reagent (Novagen, Germany) and incubated at 35ºC (5% CO_2). Cells were fixed at room temperature for 25 min with 2% formaldehyde in phosphate buffered saline (PBS). Nuclei were stained with 4,′6-diamidino-2-phenylindole (DAPI) and GFP expression was followed by fluorescence microscopy.

2.2. Cell culture and transfection.

HeLa cells and HEK 293T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Sigma) supplemented with 10% fetal calf serum, 1X non-essential amino acids, and 100µg/ml antibiotics (penicillin and streptomycin). For reporter assays, approximately 200,000 HeLa and HEK 293T cells were seeded in 24-well cell culture plates and incubated at 37ºC (5% CO_2) overnight. Cells should be 50-80% confluent before transfection. For the reporter assay two different luciferase reporter were used. The reporter plasmid 6Wtk-luc (appendix 13) that was kindly provided by Dr. Niwa (16), and the Gal4-lux reporter. The plasmids were also co-transfected with pGL4.74 (*hRluc*/TK) (Promega, UK). The transfection was done using GeneJuice Transfection Reagent (Novagen, Germany) according to manufacturer's instructions. Cells were transfected with 0.25µg/well of the luciferase reporter DNA, and with 0.25µg/well of the test expression vector DNA and 0.05µg/well of the RL-TK DNA. Experiments were done in triplicates for each construct and reporter. The cells were left two days at 37ºC (5% CO_2).

The dual-luciferase assays were conducted by James Dixon. Dual luciferase assays were performed using a dual luciferase assay system (Promega, UK) in a Veritas™ Microplate Luminometer (P/N 9100-002) (Promega, UK). Cells were lysed in 200ul passive lysis buffer. For luminescence measurements, 25µl of firefly luciferase reagent (LARII) was added to 25µl of lysate sample, where the firefly luciferase activity is measured. Followed by addition of 25µl of stop and glow (the *Renilla* luciferase reagent and firefly quenching). The results are expressed as the ratio of firefly to *Renilla* luciferase activity, (Fluc/Rluc). Experiments were conducted in triplicate.

Chapter 3: Results

3.1. Oct4 domain structure.

Oct-4 protein can be divided in three distinct functional domains; a POU domain which mediates DNA-binding function; an amino (N)-terminal (N), which has been shown to act as a ubiquitous transcriptional activation domain (TAD), and a carboxyl (C)-terminal (C) as a cell-type specific TAD (16). The importance of these domains has so far been investigated using mouse Oct-4 (mOct4). So far no work has been carried out to determine the domain function of Oct-4 proteins in other species. Before, determining how the domains influence Oct-4 transactivation, and if that function is shared in similar Oct-4 proteins; it was necessary to localise the domains.

As previously shown in Chapter 2, the domains have similar lengths, with the POU domain being the most conserved among all species. By using standard molecular biology techniques and mammalian over-expression analyses Oct-4 domain function through vertebrate evolution was investigated. To achieve this, vectors were created to express Oct-4 proteins by isolating and cloning different combinations of the N, POU and C domains. These fragments were isolated and recombined to form the constructs shown in Figure 2.1. The different vectors were used in different assays.

3.2. Oct4 sequence identity as function prediction

Apart from directly testing Oct-4 function experimentally, it is possible to predict how Oct-4 proteins might function by studying their amino acid sequence; domain conservation and sequences within them. As many Oct-4 studies have been conducted with mOct-4 and *Xenopus* Oct-4 (Xl91, Xl60, and Xl25) proteins, I aimed to predict protein behaviour by comparing the amino acid conservation and similarity of these two species with axolotl (axOct-4), sturgeon (stOct-4), and zebrafish (zfPou2); it was also possible to determine the phylogenetic relationship between Oct-4 family members. This was practically examined by reporter analyses, which will be given later.

The amino acid sequences of the Oct4 proteins were obtained using the cDNA sequences in Appendix 1 to 7, and afterwards were compared to previously published sequences in NCBI Genbank, except for stOct-4. The protein sequences where entered into BioEdit [34] to calculate the identity for the full length sequence, and also each domain individually. According to the values described inTable 3.1, axolot Oct4 has highest identity to mouse Oct4, when compared with all the species.

Table 3. 1- Sequence identity of Oct4.

Protein sequence	A/M	A/S	A/Z	A/25	A/60	A/91	M/S	M/Z	M/25	M/60	M/91
Complete	0.263	0.234	0.224	0.253	0.256	0.249	0.232	0.195	0.222	0.222	0.219
POU	0.469	0.483	0.032	0.273	0.435	0.417	0.423	0.022	0.106	0.376	0.410
(C)	0.298	0.418	0.393	0.350	0.380	0.361	0.277	0.270	0.266	0.242	0.270
(N)l	0.076	0.055	0.063	0.069	0.063	0.068	0.012	0.059	0.056	0.053	0.059
The percentage was measured using BioEdit [34]. A corresponds to Axolotl; M to Mouse; S to Sturgeon; Z to Zebrafish Pou2; 25 to *Xenopus* Oct25; 60 to *Xenopus* Oct60; 25 to *Xenopus* Oct25											

Protein sequence	S/Z	S/25	S/60	S/91	Z/25	Z/60	Z/91	25/60	25/91	60/91
Complete	0.263	0.265	0.246	0.254	0.257	0.236	0.241	0.266	0.297	0.240
POU	0.024	0.052	0.590	0.590	0.029	0.029	0.032	0.058	0.059	0.616
(C)	0.581	0.373	0.411	0.380	0.319	0.360	0.356	0.342	0.443	0.376
(N)	0.051	0.076	0.063	0.089	0.083	0.043	0.059	0.073	0.130	0.100
The percentage was measured using BioEdit (34). A corresponds to Axolotl; M to Mouse; S to Sturgeon; Z to Zebrafish Pou2; 25 to *Xenopus* Oct25; 60 to *Xenopus* Oct60; 25 to *Xenopus* Oct25										

According to Table 3.1, Oct-4 mouse full length has more similarity with axOct4 (0.263) that with any other species and the lowest identity is with Zebrafish Pou2 (0.195). As expected, the POU domain is the most well conserved domain. Despite mouse and axolotl having a high level of similarity, it is not as high as sturgeon and zebrafish, but these two species came from the same ancestor (12). The identity starts to decrease drastically on the surrounding domains, (C) and (N), but even in those domains mouse and axolotl still share the highest similarity value.

Using the full length amino acid sequence it is possible to draw a phylogenetic tree based upon average distance (Figure 3. 1).

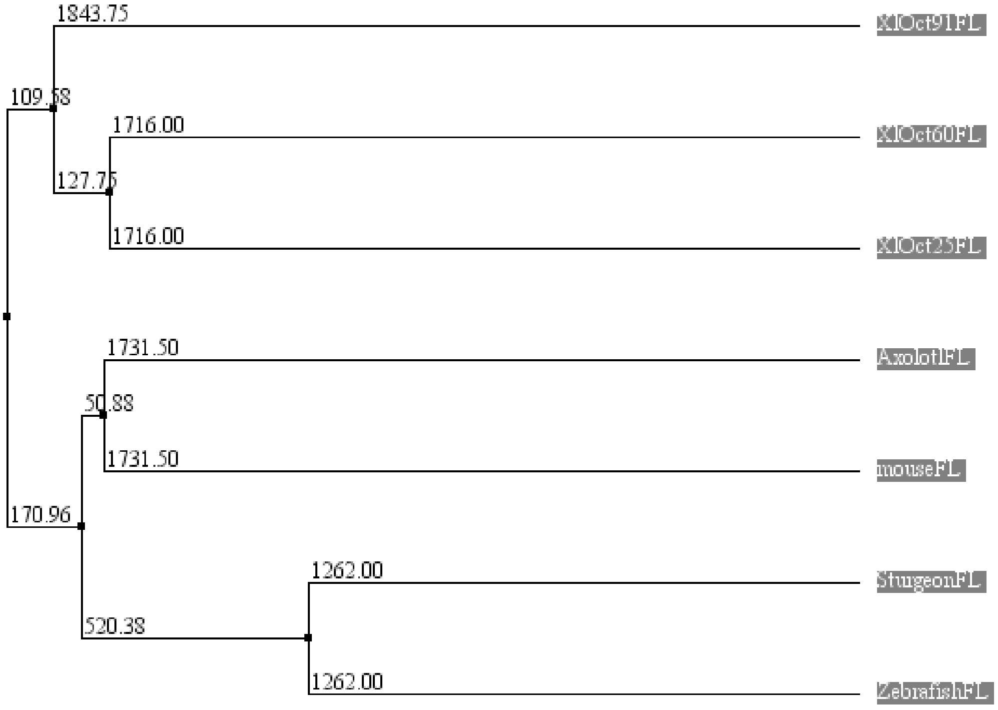


Figure 3. 1- Average distance tree using BLOSUM62 [34].

According to Figure 3.1 there is an evident division of the *Xenopus* Oct expression to all the others. Within the *Xenopus*, the Xl60 is closer to Xl25 than to Xl91. It also shows that axOct4 and mOct4 share more similarities between them, that with any other species. StOct4 and zfpou2 also have a higher similarity between them.

3.3. Amino Acid Composition

It is important to determine the amino acid composition, similarity between the different Oct-4 homologues. It is possible that a difference between and within the domains is based on their different amino acid composition.

The amino acid composition of each species and the respective amino acid graphics can be found in Appendix 9 to 11. The amino acid composition was calculated using BioEdit program [34]. The full length values are not shown because it represents the sum of the domains.

3.3.1 N- Terminal (N)- domain

The N-domain was known by being a proline rich domain, due to the fact that 25% of its constitution is proline [16, 20], and that this amino acid plays an important role in mouse Oct-4 transactivation [16].

According to the amino acid composition predicted by BioEdit [34], mouse (N) does not have 25% of proline, but 14.29%, and also glycine is also the amino acid present in higher quantity (17.29%) followed by proline and then glutamic acid. In general, the proline values oscillate from 8.2% (axolotl) to 14.29% (mouse), the species with similar amount of proline to mouse is *Xenopus* Oct-25 followed by zebrafish.

The amino acid found in higher quantity in axolotl and mouse is glycine; while in sturgeon, Xl91 and Xl60 is serine; in zebrafish Pou2 and in Xl25 is proline. *Xenopus* Oct-4 homologues (Xl25, Xl60, Xl91) have in common proline and serine constitution; and Xl60 is similar on the amount of glutamic acid when compared to mouse.

3.3.2 POU domain

The values and respective graphics for POU amino acid composition are in Appendix 9.

According to the results obtained, the amino acid present in higher amounts in axolotl, mouse, sturgeon, and zebrafish is leucine. In Xl91 and Xl25 is lysine while in Xl60 is glutamic acid.

The second amino acid present in higher amount in axolotl is glutamic acid; in mouse is lysine; in sturgeon and zebrafish are lysine and arginine; in Xl91 is glutamine; in Xl60 is lysine and leucine; and in Xl25 is leucine, glutamine and arginine.

3.3.3 C- Terminal (C)-domain

The amounts of amino acid present on the C-terminal can be found in Appendix 11.

C-domain is known for having high amounts of proline, serine and threonine [16], but according to Table 3.4, proline is in fact present in significant levels, but not serine and threonine, the amino acids that are present in higher amounts are leucine and proline (13.68% and 12.63%). Proline is in fact present in high levels in stOct4 (16.18%), zbpou2 (20.00%), Xl91 (20%) and Xl60 (16.18%). In this same species, the second amino acid present in higher levels is glycine, fact also shared with Xl25.

The tryptophan is the only amino acid present in mOct4 that is not present on the other species. In Nanog, the tryptophan is associated with the development of extra-embryonic tissue (James Dixon, unpublished data).

a) Phosphorylation prediction

If the C-terminal is correlated with phosphorylation events [40], tyrosine, serine and threonine are known to be involved in phosphorylation reactions. According to the table on appendix 11, Xl60 is the one with more serine, but very low amounts of threonine and tyrosine. Axolotl (C) has big quantities of serine and tyrosine. The amino acid present in higher quantities in mOct4 is serine, but it also has threonine and tyrosine. All the other species also have the three amino acids related to phosphorylation.

For phosphorylation events, the amino acid constitution is not enough, it is necessary a favourable composition of the surrounded amino acids. By using NetPhos 2.0 [38] it is possible to predict the probability of phosphorylation events to happen on a given sequence. The probability is scored from 0.5 to 1, only the values above 0.500 are considered. The results are shown in Table 3.2 and Table 3.3.

Table 3. 2 Prediction of phosphorylation positions in the C domain, for serine; and tyrosine, according to Netphos 2.0 [38].

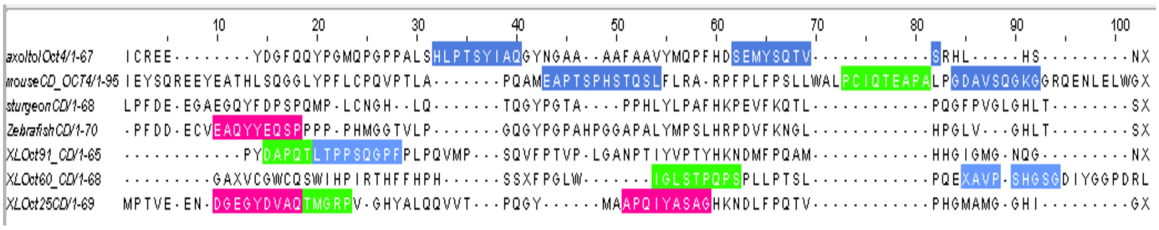
	Serine			Tyrosine		
	Position	Sequence	Score	Position	Sequence	Score
AxOct4	28	HLPTSYIAQ	0.930	6	CREEYDGFQ	0.670
	56	SEMYSQTVS	0.735	55	DSEMYSQTV	0.653
Moct4	5	EAPTSPHST	0.964			
	8	TSPHSTQSL	0.959			
	45	GDAVSQGKG	0.553			
StOct4						
Zbpou2				12	EAQYYEQSP	0.928
XLOct25				12	DGEGYDVAQ	0.860
				43	APQIYASAG	0.746
XlOct60	26	XAVPSHGSG	0.829			
XlOct91	12	LTPPSQGPF	0.890			

Table 3. 3 Prediction of phosphorylation positions in the C domain, for serine; and tyrosine, according to Netphos 2.0 [38].

	Threonine		
	Position	Sequence	Score
AxOct4			
mOct4	34	PCIQTEAPA	0.768
StOct4			
Zbpou2			
XLOct25	17	DVAQTMGRP	0.544
XlOct60	7	IGLSTPQPS	0.561
XlOct91	7	DAPQTLTPP	0.862

According to Netphos 2.0 [38] stOct4 does not have relative phosphorylation sites, despite having threonine. Mouse C-terminal has four phosphorylation sites that might justify the hypothesis that this domain is regulated by phosphorylation events. Nonetheless, axC also has a high number of phosphorylation amino acids, two in serine and two in tyrosine. The *Xenopus* domains have less phosphorylation sites, but with a higher probability of occur.

To know if those regions are conserved among the species in study, the sequence given by Netphosh (Table 3.2; Table 3.3) as phosphorylation place, were localized in a MAFT alignment and the phosphorylation sites where highlighted. The result can be seen in Figure 3.2.



```
axolotlOct4/1-67      ICREE--------YDGFQQYPGMQPGPPALSHLPTSYIAQGYNGAA--AAFAAVYMQPFHDSEMYSQTV------------SRHL-----HS---------NX
mouseCD_OCT4/1-95     IEYSQREEYEATHLSQGGLYPFLCPQVPTLA-------PQAMEAPTSPHSTQSLFLRA-RPFPLFPSLLWALPCIQTEAPALPGDAVSQGKGGRQENLELWGX
sturgeonCD/1-68       LPFDE-EGAEGQYFDPSPQMP-LCNGH--LQ-------TQGYPGTA----PPHLYLPAFHKPEVFKQTL------------PQGFPVGLGHLT--------SX
ZebrafishCD/1-70      -PFDD-ECVEAQYYEQSPPPP-PHMGGTVLP-------GQGYPGPAHPGGAPALYMPSLHRPDVFKNGL------------HPGLV---GHLT--------SX
XLOct91_CD/1-65       ------------PYDAPQTLTPPSQGPFPLPQVMP---SQVFPTVP-LGANPTIYVPTYHKNDMFPQAM------------HHGIGMG-NQG---------NX
XLOct60_CD/1-68       ---------GAXVCGWCQSWIHPIRTHFFHPH------SSXFPGLW-------IGLSTPQPSPLLPTSL------------PQEXAVP-SHGSGDIYGGPDRL
XLOct25CD/1-69        MPTVE-EN-DGEGYDVAQTMGRPV-GHYALQQVVT---PQGY------MAAPQIYASAGHKNDLFPQTV------------PHGMAMG-GHI---------GX
```

Figure 3. 2- C-Domain MAFT alignment with the phosphorylation sites highlightened; in blue is where it might be a phosphorylation in a serine amino acid; in red is where there might be phosphorylation in a threonine amino acid; in green is where phosphorylation might occur in a tyrosine amino acid.

Zebrafish Pou2, Xl91 and Xl25 have similar phosphorylation sites(10- 30 aa), and axolotl and mouse have a similar phosphorylation area (30 and 70 aa).

3.4. Alignments

To study the conservation between the different species, it is necessary to make amino acid alignments. There are several alignment programs such as BioEdit [34]; Jalview [37]; Blast; Emboss:: Neddle; Emboss::Water. All of these have different algorithms for sequence alignment.

MAFFT makes sequence alignments using a fast Fourier transform (FFT) approximation [39]. It is able to align more than 50 sequences with a higher accuracy than ClustalW [39], program that can be found in BioEdit. MAFFT version 5, can be found in Jalview v.2.3, that combined with Jalview tools display information about the quality conservation and consensus within the sequences [37]. The consensus display for the alignment is scored from 1 to 9, where 1 is the lower level, and when there is a 100% match, it is scored with a (*). The consensus graphic combines all the sequences into the amino acid that is present in all sequences for the same position. For being more accurate, and for displaying more information in one data, the program used to make all sequence alignments was MAFFT from Jalview v.2.3. The amino acid constitution from each sequence was also calculated by Jalview v2.3 [37].

3.4.1. N-terminal (N)-domain

The proline region of mOct4 is found between the 13-60 amino acid, and when that region is deleted there is a significant decrease in Oct-4 transactivation [16]. It was possible to see that the proline is not present in similar values for all species. However, it might be possible that proline is concentrated in the same area as mouse. The influence of this region on other species, is not known.

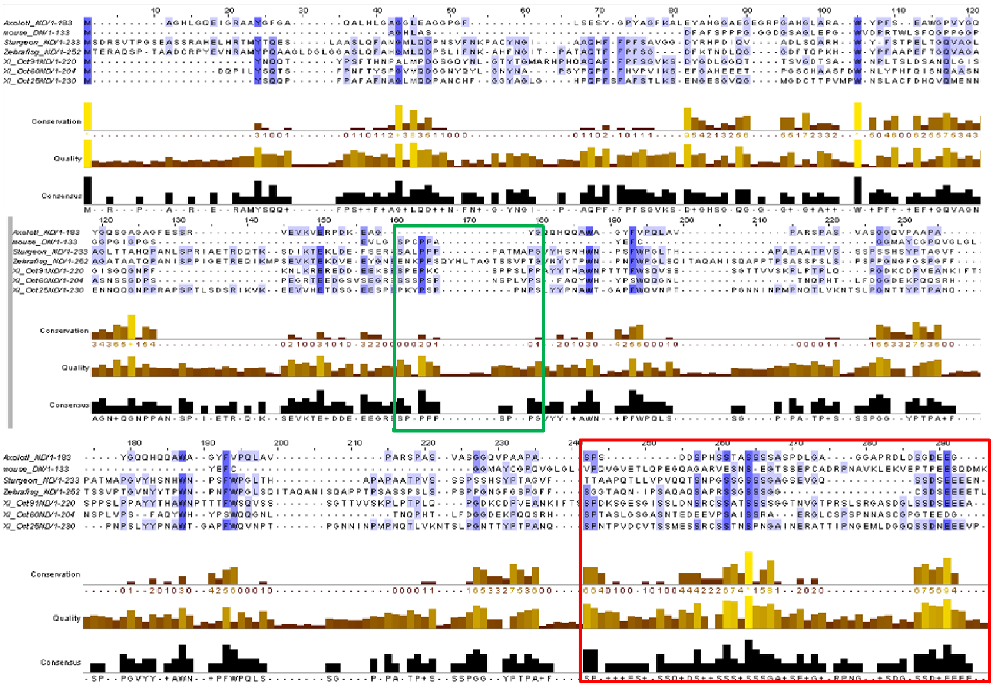


Figure 3. 3- N-terminal alignment for all species. The green delimitation shows where most of the proline is concentrated. The red delimitation shows a conserved area within the N-terminal.

According to Figure 3.3, the proline region is not conserved within species, but there is a region where most of the proline can be found (160 to 180 aa). It is possible to see a different conserved area, where most of the amino acids are serine and glutamic acid, but this similarity is not shared in mouse sequence, now termed SEB Box.

3.4.2. POU-domain

POU alignment can be seen in Figure 3.4. Duanquin Pei *et at.* (24) identified a nuclear localization signal (NLS) in mouse Oct-4. The NLS was found in the Pou domain, the sequence responsible for its localization in the nuclei and required for the transactivation of its target genes is RKRKR [24]. By comparing outa multiple alignment of POU, it was possible to localise the sequence and find out if it is present and conserved for all species.

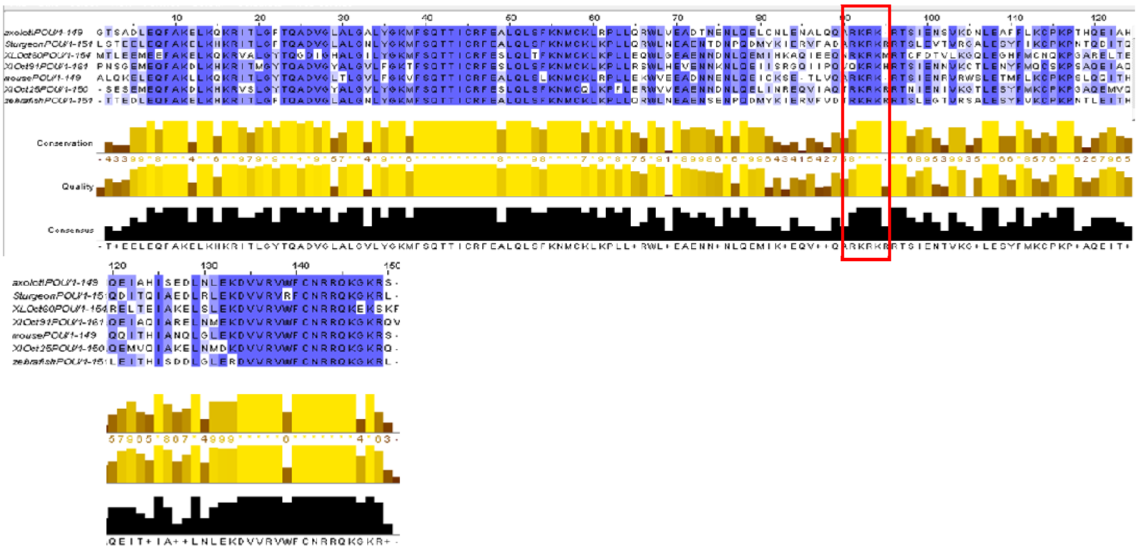


Figure 3. 4- Jalview MAFFT alignments for the different POU domains, the selected area corresponds to NLS.

Figure 3.4 shows how conserved the POU domain is among the different species. In most of the cases it has a 100% conservation and only one region with lower conservation (80 and 90 aa). The NLS, in mOct4 with the sequence: RKRKR. However, in other species it was only found RKRK. It is not known if this sequence is enough to transport the Oct-4 sequence into the nucleus. The pGFP vector with different Oct-4 domain will help to determine whether the Oct-4 is transported into the nuclei or not.

3.4.3. C-terminal (C)- domain

The C-terminal MAFT alignment can be possible to see in Figure 3. 5.

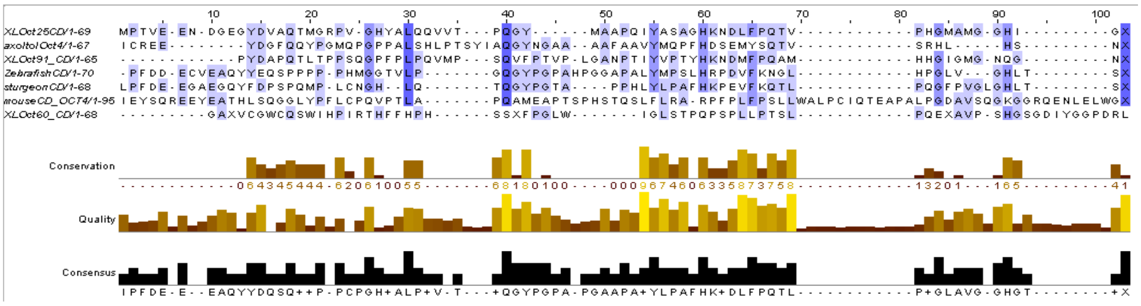


Figure 3. 5- MAFT multiple alignments for the different C-terminal sequences.

The sequence is more conserved in the middle than in the domain extremities. In the middle (between 40- 70 aa) some of the phosphorylation amino acids are also found (Figure 3. 2). If this similarity is enough to regulate the cell type-specificity, in the same way as for mouse, it will be determine by over-expressed vectors,

3.5. Cloning

All full lengths and Oct-4 domains were successfully cloned into pGEM®-T Easy vector (Promega), with the exception of Xl25 that was not possible to isolate the correct band size by PCR. The band size was always 200 bp inferior to that expected; nevertheless this band was inserted into the vector and sent for sequencing. According to the sequencing the primer had a mismatch and this might be the reason why the band was always 200 bp smaller.

Unfortunately, Due to time constraints, the subclonning of the full length and the domains into overexpression vectors was not completed in time for the Luciferase assay. Therefore the assay was not done, nevertheless it was possible to do the assay pATG with the Oct-4 full length.

The subclonning of the full length and the POU domain, was successfully done into pGFP vector, has well the subclonning of the full length into Gal4DBD vector. The (N) from zfpou2; POU from Xl91 and Xl60; (C) from mOct4, zfpou2 and Xl60 were not subcloned on time for the assay.

3.6. Nuclear localization

It is important to determine whether Oct-4 protein is translocated to the nucleous. For that propose, the full length sequences and the POU domains were fused to a GFP signal, and that used for transfect 293T cells. The results can be seen in Figure 3. 6

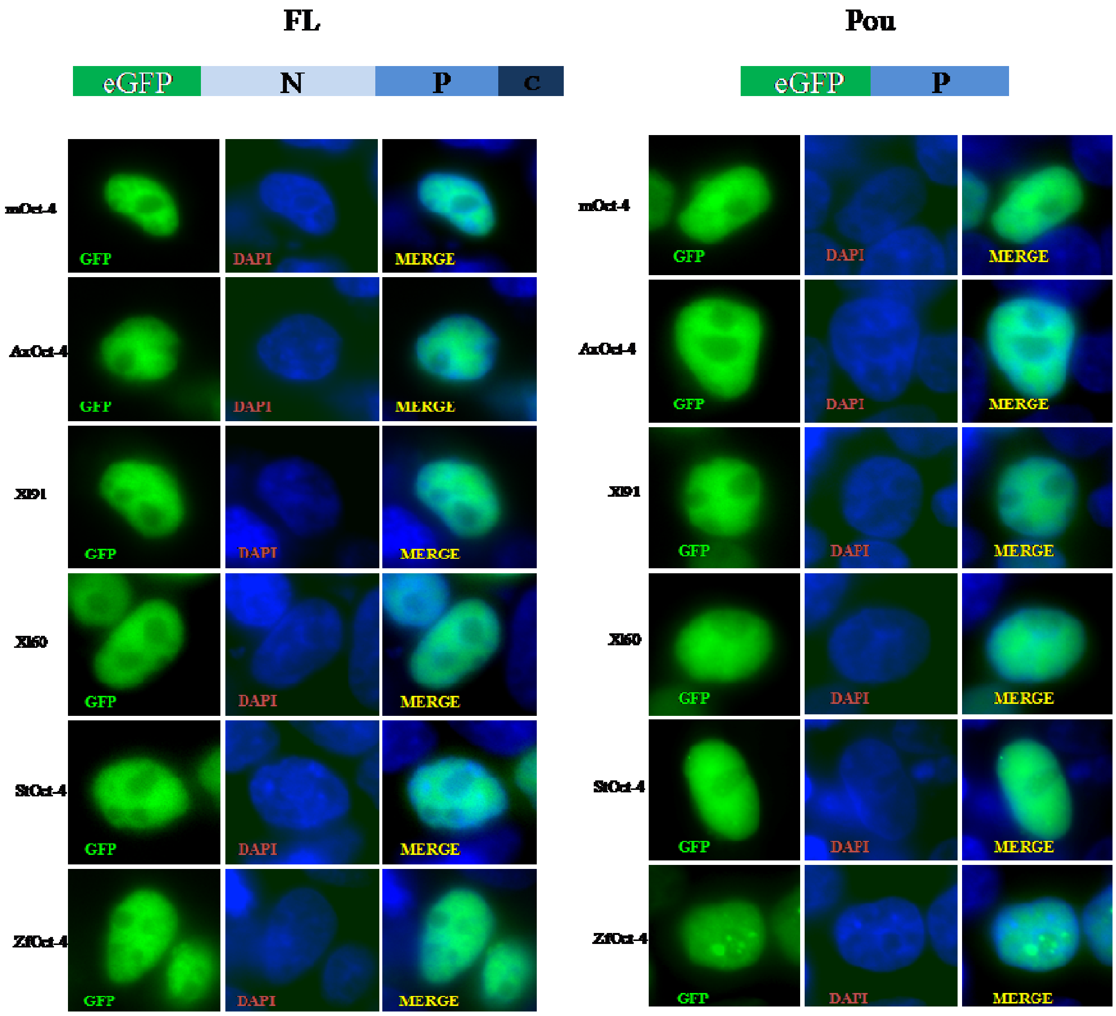


Figure 3. 6- 293 T cells transfected with pGFP Oct-4 full length and POU domain.

For the sequence to be transported into the nuclei, it needs to have the NLS, for mouse this sequence is RKRKR. Only mOct4, has this sequence, while all the other Oct-4 homologues only have RKRK. Despite this, all the Oct-4 was found in the nuclei. Once the POU domain was transported into the nuclei, we can say that the NLS is localized in this domain, and that RKRK, is highly likely to represent their NLS.

3.7. Luciferase Assay

Transcription factors (TFs) domain architecture includes at least a DNA- binding domain (DBD), responsible for binding to a special promoter region, in order for the gene be expressed. To see when the transactivator is being expressed, it needs to be attached to a reporter gene, in this case luciferase.

Reporter genes had been used to analyse the function of genetic elements, such as promoters and enhances. To assess how Oct-4 full length and the different domains function as transactivators, were used used special promoters fused to luciferase gene (6Wtkluc, and Gal4lux reporters). These promoters have multiple copies of the enhancer element where the TF binds. Binding in this region results in the expression of firefly luciferase, an enzyme that can be quantified, giving information about the transactivation activity for each gene.

There can be a lot of variability during the experimental procedures: differences in cell seeding numbers, transfection efficiency. Therefore, it is not very accurate to rely only on the measurement of the luciferase enzyme. The dual-luciferase assay overcomes this problem by using two luciferase reporters, one as the experimental reporter (6Wtkluc and Gal4lux) and another as the normal control (*Renilla* RL-TK*)*. Both the firefly and *Renilla* luciferase proteins have proven to be highly effective as gene reporters.

The transactivation activity was measured in two different systems, in HeLa and in HEK 293T cells, and in ES cells. The transfection in ES cells was not successful and therefore there is not the possibility to study the transactivation capacity in this cell system.

3.7.1. pATG

The transactivation of the Oct-4 full length was measured by a Oct-4 binding site promoter, that has 6 copies of oligonucleotides with octamer-binding motif from mOct4, also known as 6Wtkluc reporter. This plasmid has a transcriptional regulatory element that only requires Oct-4 to be active in ES cells (16)

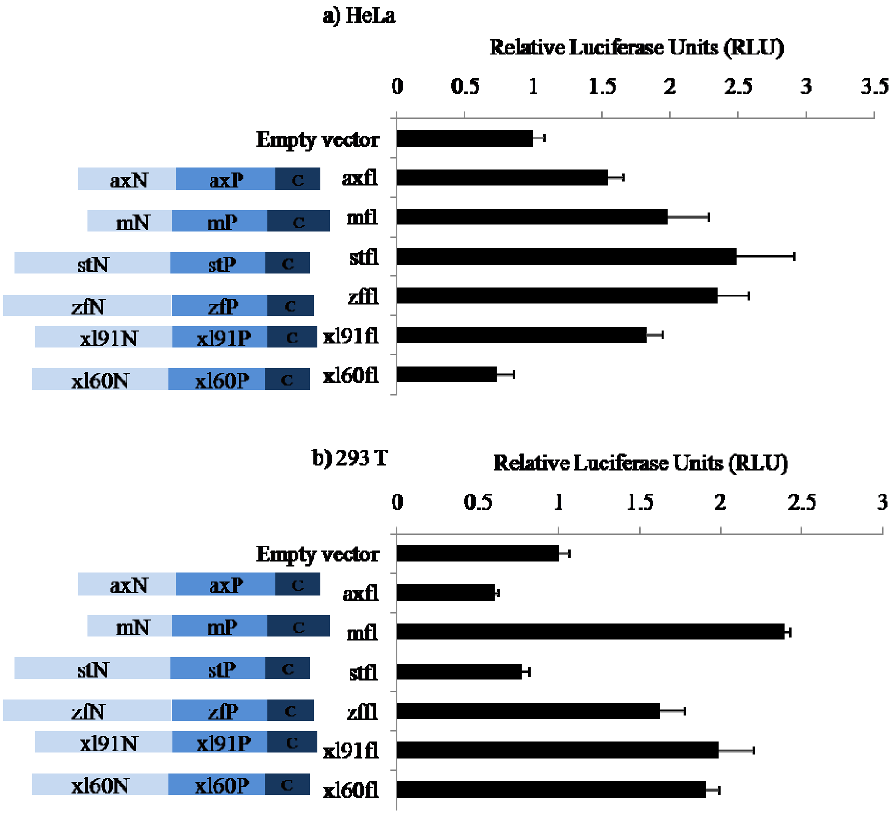


Figure 3. 7- pATG-Oct-4 Full length transactivation in HeLa and HEK 293T cells.

Figure 3.7 shows that the different Oct-4 full lengths have different behaviour depending on the cell-type. In HeLa cells, all full lengths active 6Wtk-luc with exception of Xl60. Axolotl Oct-4 is able to activate the reporter in HeLa cells but inhibits it in 293T cells the same happens to sturgeon, but this last one with higher activation in 293 T cells then axOct4. Mouse Oct-4, and Xl91 has transactivation function in both cell types, but it has higher activity in 293T cells. Zebrafish Pou2 has high activity function in HeLa than in 293T cells.

In HeLa cells the Oct-4 with higher transactivation activity is stOct4, while in 293T cells is mOct4.

3.7.2. pDBD

The DBD of the pDBD and the POU domain are very similar, both with DNA-binding function, and both able to bind to the reporter promoter (Gal4). To test which domain function has a better transactivation function, the Oct-4 full length and domains where fused to DBD.

A) Full length Oct-4 activation

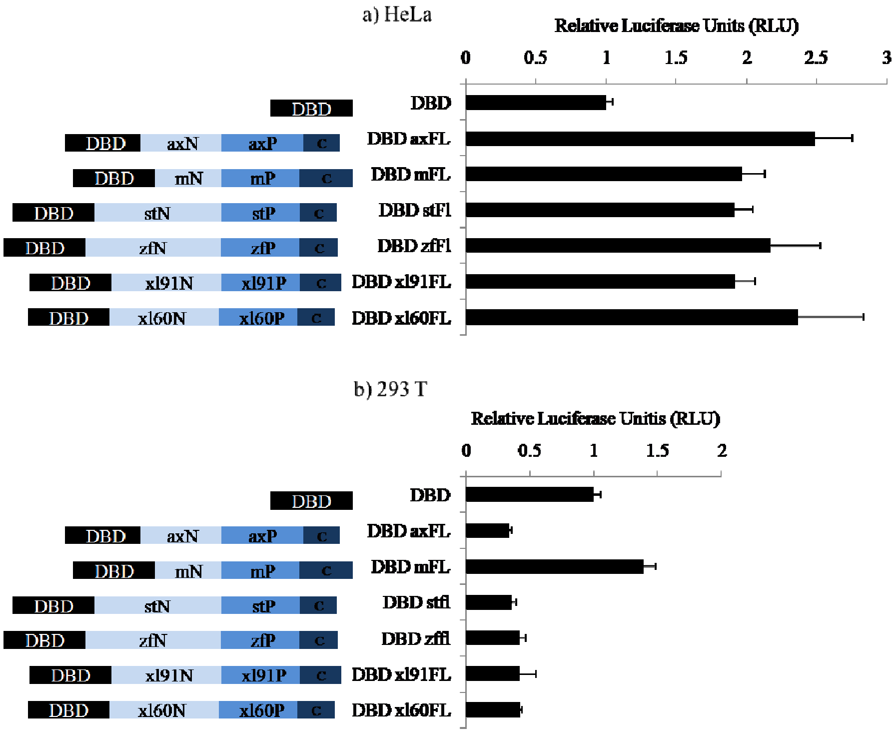


Figure 3. 8- pDBD- OCT-4 full lenght activation in HeLa and HEK 293 T cells

As it is possible to visualize (Figure 3. 8) that all the Oct-4 full lengths were able to activate Gal4-*lux* in HeLa cells but not in 293T cells. The only transactivator able to binds to Gal4-*lux* in 293 T cells is mOct4. Axolotl pDBD-FL is the one with higher transactivation in HeLa cells than any other species. Mouse Oct-4 has activity in both cell types, but has higher expression in HeLa cells than in 293T cells.

B) POU-C constructs

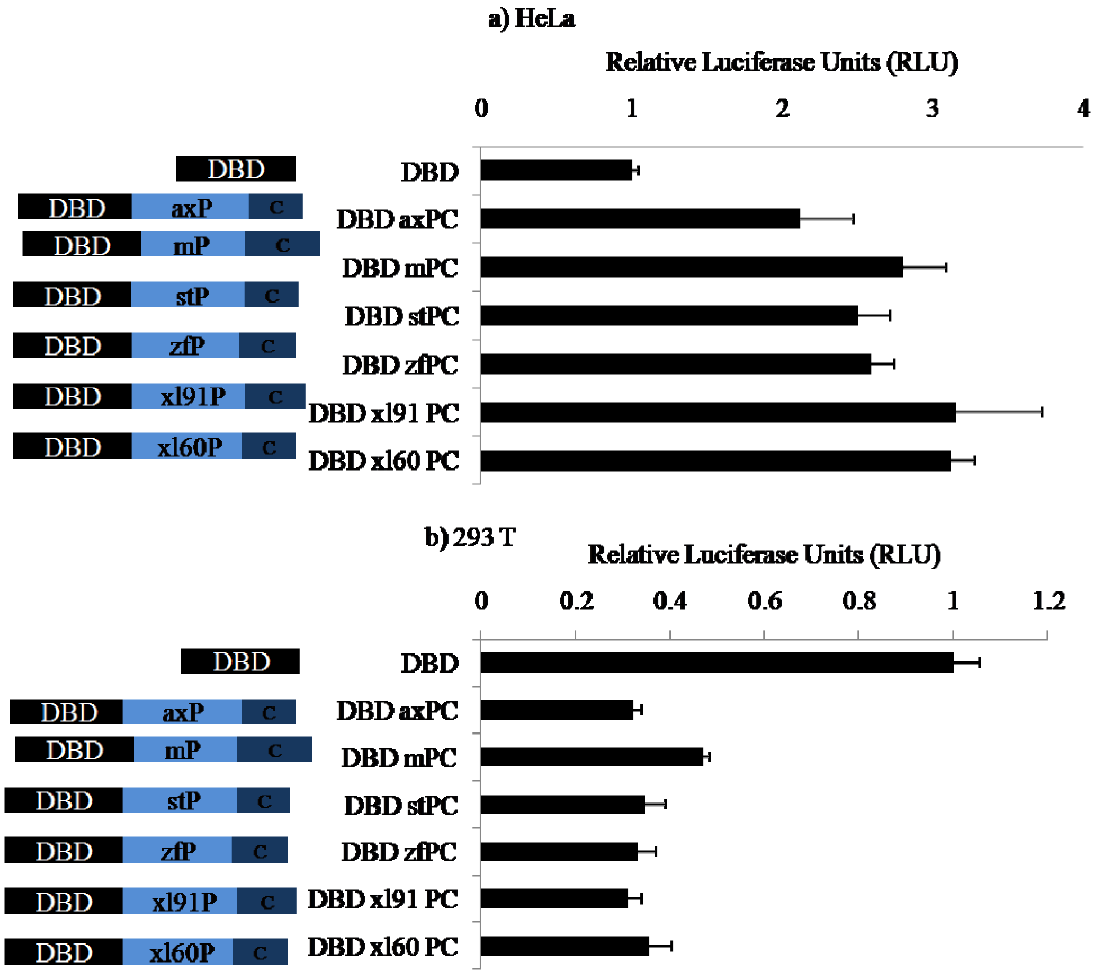
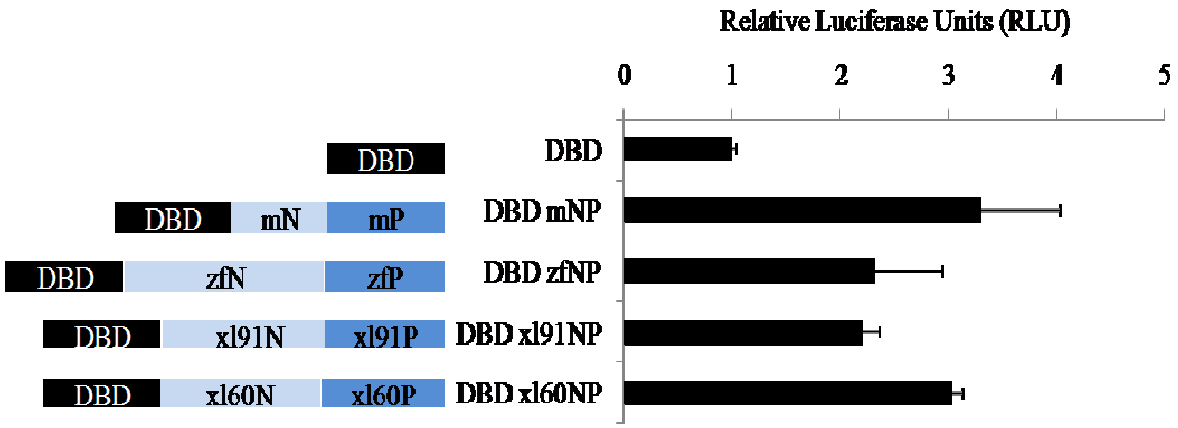


Figure 3. 9- pDBD-POU-C Oct-4 transactivation in HeLa and HEK 293T cells

The different Oct-4 DBD-POU-C all have transactivation function in HeLa cells, but not in 293T cells. Axolotl Oct-4 is the one with lower transcription activity in HeLa while Xl91 was the one with the highest activity, followed by Xl60.

C) N-POU

a) HeLa



b) 293 T

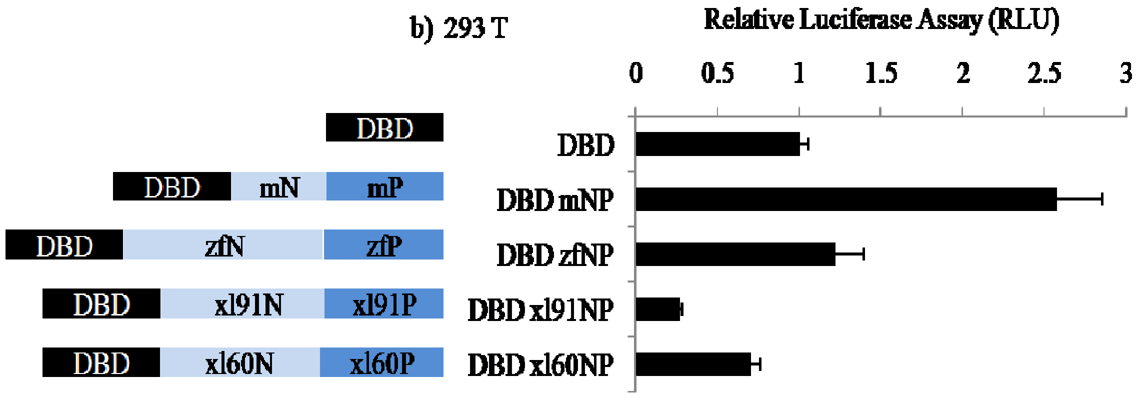


Figure 3. 10- pDBD- N-POU Oct-4 transactivation in HeLa and HEK 293T cells.

According to Figure 3. 10 the species involved in this assay are all active in HeLa cells, where mouse is the one with higher transactivation capacity, followed by Xl60. In 293T cells only mouse and zebrafish are active, and once again with mouse being the one with higher activity. Mouse Oct-4 transactivation activity is better in HeLa cells than in 293 T cells.

D) N-terminal

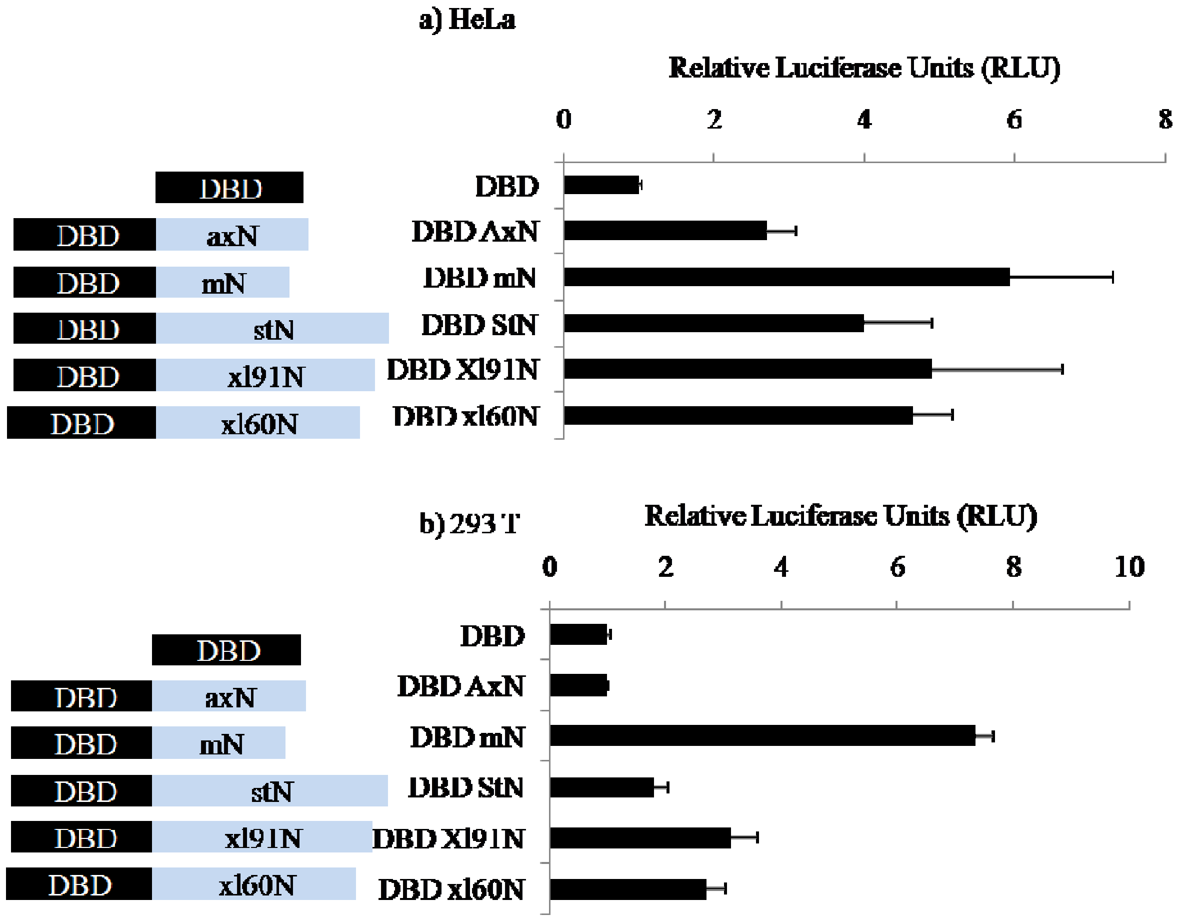


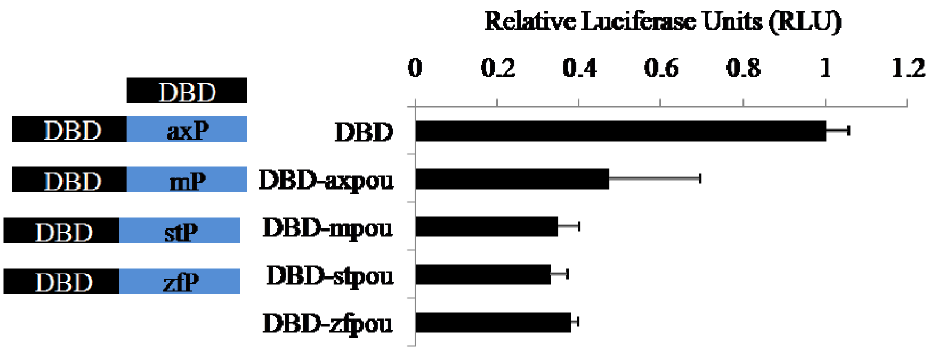
Figure 3. 11- pDBD-N transactivation in HeLa and HEK 293T cells.

According to Figure 3. 11 the transactivator domain (N) was able to activate Gal4-Lux in HeLa cells, and in 293T cells, to exception of axN.

Sturgeon, Xl91, Xl60 are more active in HeLa than in 293 T cells, the opposite to mN that has a higher transactivation function in 293 T than in HeLa cells.

E) POU

a) HeLa



b) 293 T

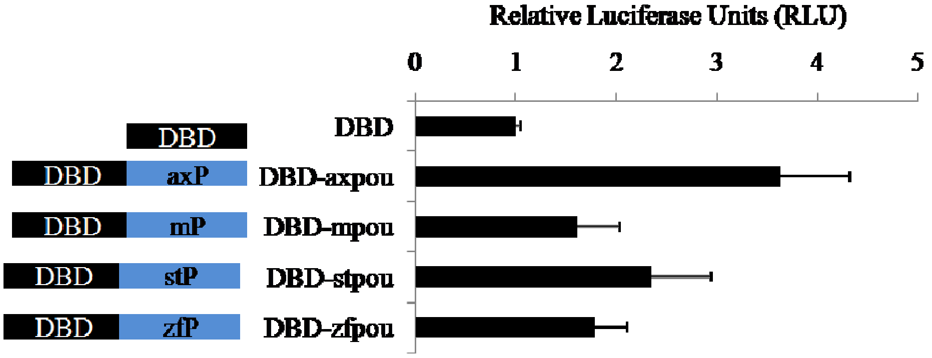


Figure 3. 12- pDBD-POU transactivation in HeLa and in HEK 293 T cells

According to Figure 3. 12, POU inhibits the activity in HeLa cells, but activates Gal4-lux in 293T cells. axPOU is the transactivator domain with higher activity in 293T cells.

F) C-terminal

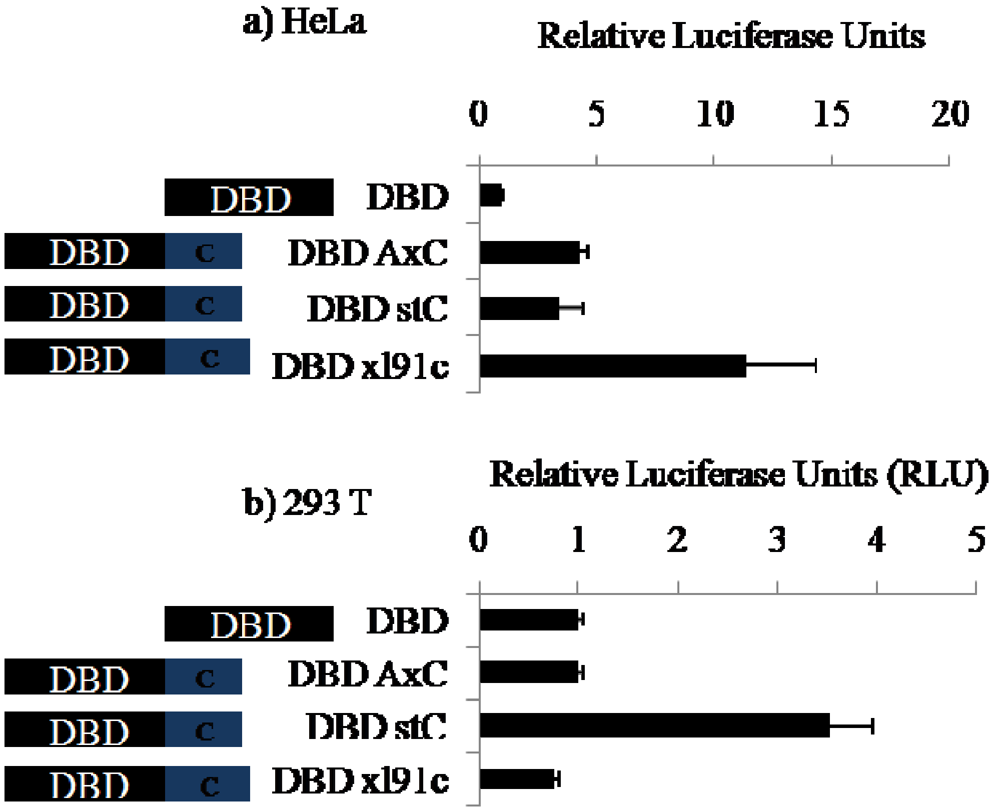


Figure 3. 13- pDBD-C transactivation in HeLa and in HEK 293T cells.

According to Figure 3. 13, axolotl, sturgeon and Xl91 able to activate Gal4-lux in HeLa cells, while in 293 T cells only axolotl and sturgeon have transactivation capacity. Xl 91-(C) in 293 T cells is not active, while in HeLa cells is the one with higher transactivation function.

Chapter 4. Discussion

Stem cells have great potential for future medicine, due to their abilities to regenerate tissues, *i.e.* pluripotency. However, the mechanisms that regulate stem cell potency is largely unknown. Oct-4 protein has attracted a lot of attention as a key regulator for stem cell potency and differentiation [16, 21, 24]. Oct-4 has proven to be one of the key genes necessary for reprogramming somatic cells into pluripotency. Therefore it is important to know which homologue, Oct-4 protein functions better in somatic cells.

Oct-4 belongs to the octamer-binding (Oct) family of transcription factors, which contain the POU domain [40]. Brehm *et al.* in 1997 proved that in somatic cells the individual domains did not have transactivation capacity, only their simultaneous presence resulted in Oct-4 activation [20]. In 2002 Niwa *et al.* demonstrated that the same occurs in ES cells, and that the POU domain of Oct-4 plus the N- or C-terminal were required for maintaining the ES phenotype [16].The present study was aimed to clarify the contribution of the different functional domains for Oct-4 transactivation capacity, and how this function is conserved in other species.

The findings demonstrate that Oct-4 full length and its domains have distinct activities in HeLa and in HEK 293T cells, and that in general Oct-4 mutants show similar activity function.

4.1. Amino Acid Composition

The amino acid composition determines the functionality of a protein; therefore it is important to know if the differences in their activation function are due to differences in its general amino acid composition. It is not yet possible to explain all the functions of a protein by its amino acid sequence, but it is possible to establish a correlation between the structure and the function by its amino acids [41].

According to the amino acid composition of (N) there are two amino acids that are more present; Serines and Prolines. It is possible to notice a division between the species according these two amino acids, Xl60 and Xl91 have similar enrichment for Proline followed by Serine; this similarity is shared by stOct4, when comparing the serine amounts.

In mouse, it has previously been shown that a Proline-rich region in the (N) influences transactivation activity function [16]. If we predict the function of the (N) simply by comparing the enrichment of prolines, the (N) domain that would have the greatest activity would be in the order: mOct4>Xl25>zfpou2>Xl60>stOct4>Xl9>AxOct4. To assess this hypothesis the transcriptional activation mediated only by the (N) domain was analysed. Data displayed in Fig 3.11 for DBD-N, showed that for the N-domain; Xl91>Xl60 and stOct4, suggesting that this simple hypothesis that proline content directly reflect transcriptional activity is not correct. However, prolines are not the only residue enriched in the mOct4 N-domain. I analysed the sequence for the second most enriched amino acid, which was the Serine residues. If Serine richness was a better indicator of transcriptional activity of the (N) domain then Xl91>Xl60>stO4 in terms of transcriptional activation. Data in Fig 3.11 confirms this hypothesis which indicates that not only Proline residues have an important role in activity but also that Serine may promote transactivation mediated by the N-domain.

In the POU domain the amino acid constitution is very similar as this is one of the most conserved domains in proteins generally. Even though it is possible to differentiate them, all *Xenopus* homologues are similar with Xl25 sharing the highest similarity with Xl91 rather than with Xl60. Zfpou2 and stOct4 share higher similarity, followed by the mOct4 and AxOct4.

Even though all proteins containing POU domains share very high amino acid conservation, and all Oct4 homologues have highly conserved (N) and POU domains they have clearly have different molecular function. These differences will be attributable to specific residue differences between homologues which will require further investigation and will alter transactivation activity by affecting interaction with other co-factors.

The (C) domain amino acid sequence is not as conserved as the (N) and POU domains. The amino acid make-up of the (C) domain in some homologues have high amounts of proline, and glycine. As for the hypotheses above for the N-domain proline enrichment in the (C) might influence transactivation activity function. If this is true, then by comparing the enrichment of proline in the (C) domain, then the greatest activity would be: zfpou2≈ Xl91> Xl60> Xl91>axOct4>mOct4. To assess this hypothesis the transcriptional activation mediated only by the (C) domain was analysed. The data displayed in fig. 3.12 for the DBD-C, shows that this hypothesis is not correct Once the (C) domain has been suggested previously to be regulated by phosphorylation [20, 23] on serine, threonine and tyrosine residues [38]. I analysed the sequence for this amino acid enrichment. The amino acids are differently distributed by the (C), but these data set does not give all the information.More tests to establish a possible relation between amino acid and protein function would require a new assay with all the species.

4.2. SEB (Serine [S]/ Glutamic acid [E]) box

Within this study I determined a highly conserved motif in the very N-terminal of the N domain which I termed the SEB (Serine [S]/ Glutamic acid [E] box). This is not the only region on the (N), mouse has a proline region responsible for increasing the transactivation activity [16]. The other (N) domains do not have this proline-rich region, instead they have a SEB box. As was mentioned earlier, according to the proline function and amino acid percentage, serine might have an important role in the transactivation function. Serine and Glutamic acid are chemically similar, a glutamic acid when phosphorylated becomes very similar to serine. Mouse (N) domain has serine where the other species have glutamic acid. This might indicate that the other species can be regulated by phosphorylation. To assess this hypothesis I analysed the transcriptional activation mediated only by the (N) in the two cell types. The data displayed (Fig. 3.11), showed a different regulation in HeLa and in 293T cells, in HeLa all the (N) are active, while in 293T only mouse (N) has activity. These differences in transactivation support the hypothesis that apart from mouse (N), the remaining species are phophosregulated by different co-factors existent in the cell-type. Further tests would be required to assess this hypothesis.

4.3. Nuclear Localization

For the Oct-4 be transcribed it needs to be transported to the nuclei, this transportation is mediated by a nuclear localization signal (NLS). There are two major groups of NLS [42]. The first contain 3-5 amino acids with Lys-Arg/Lys-*X*-Arg/Lys. The other is a bipartite type, containing two clusters of basic regions of 3-4 residues, each separated by approximately 10 amino acids [42]. Pan *et al.* identify the sequence: RKRKR as being the Oct-4 NLS for mouse [24].This NLS corresponds to the first type of NLS, indicating that mouse Oct-4 does not require two different sequences to be transported into the nuclei. This sequence is not found in the other species nevertheless it was found a similar sequence: RKRK in the POU domain, in the same conserved region that mouse. The pGFP assay showed that the POU domain was the one responsible for the transport into the nuclei. Due to the high degree of similarity, it is likely that this sequence (RKRK) corresponds to the NLS for the remaining species. In order to test this hypothesis more assays would be required sing additional constructs.

4.4. Luciferase Assay

Brehm in 1997 did a similar study by analysing transcriptional activation by the Oct-4 (N) and (C) domains in their native protein, using a 6W-37tkCAT reporter and with Oct4 full length and its domains linked to a Gal4 DNA binding domain. According to their results, none of the transactivation domains are able to function alone. In HeLa cells both mutants (POU-C and N-POU) can stimulate transcription, but not as much has the full length Oct-4 protein, indicating that the (N) and (C) domains are required for a full transcription [20]. The present study tested this for the Oct-4 homologues protein, but unfortunately it was not possible to analyse the N-POU and the POU-C native state, instead the same mutants were linked to Gal4-DBD.

pATG

Oct-4 acts together with other transcription factors, such as Sox-2 and Nanog; it requires this interaction in order to regulate genes during early development [8]. The different transcription factors are differently regulated by cofactors such as Fgf-4 (16).

There are at least five different reporters that can be differently activated by Oct-4 and cofactors, such as Fgf-4 which is activated by Oct-4 and Sox-2 and the Rex promoter is activated by Oct-4 and Xox1. The reporter used in this study does not require Sox-2 or any other cofactor to be activated; 6WTK is upregulated by Oct-4 alone [16].

The Oct-4 transactivation function was only possible to study for the full length; no mutants were analysed for the luciferase assay. When analysing the Oct-4 expression, all the species are active in their native state with exception of Xl60 that does not have transactivation activity in HeLa cells (Figure 3.7a) but it has in 293T cells (Fig. 3.7b), and the opposite happens with axOct4 and with stOct4. The activity measured in 293T cells for *Xenopus* is consistent with previous works, where Oct homologues were upregulated until stage 16 cell of *Xenopus*, which is until 18h and 15 min [43]; involving all Oct-4 homologues.

These different activities in the different cell types might be due to specific cofactor interactions differently existed in HeLa and in 293T cells. The different behaviour between Xl91 and Xl60 might be due to their different developmental roles. *Xenopus* Oct-4 homologues are expressed in different stages, and that means different cofactor interactions. Xl60 is maternally transcribed [43]; Xl25 both maternally and zygotically transcribed, and Xl91 is only zigotically transcribed [43]. For instance, Xl25 and Xl60 interact with β-catenin to prevent the formation of mesoderm, and β-catenin exists in HeLa cells. In this example, the maternally transcribed proteins have specific interactions, opening the possibility that Xl91 and Xl60 have different cofactors interactions in HeLa cells, but the nature of these interactions remains unclear.

pDBD

When analysing the Oct-4 expression, all the species are active in their native state with the exception of Xl60 that does not have transactivation activity in HeLa cells (Figure 3.7a). All the Oct-4s are also able to function when linked to the Gal4 DNA-binding domain (DBD) (Figure 3.8a). When comparing the activity levels in Oct-4 in HeLa and 293T cells in its native state (Figure 3.7) with Oct-4 linked to Gal4-DBD (Figure 3.8), it is possible to see in general a decrease of transcriptional activity. The fact that Oct-4 is active in 293T cells without DBD (Fig.3.7b) and inactive in its presence, suggests a significant regulatory interaction between DBD and with any of the transactivation domains of the Oct-4, resulting in the reporter inhibition.

In HeLa cells, the only Oct-4 homologues where Gal4-DBD increased transcription activity were Xl60 and axOct4, suggesting that there is a different interaction between the Oct-4 domains and the Gal4-DBD that result in promoter stimulation. In 293T cells native axolotl (axFL) and sturgeon (stFL) are not active, but their status does not change when linked to the Gal4-DBD. Mouse Oct-4 is the only one with transcriptional activity in 293T cells. To know which domain is responsible for the DBD-mFL transactivation it is necessary to compare mouse POU-C and N-POU. By looking to figure 3.9b and 3.10b we can assume that (mN) is the transcription domain responsible for the transactivation in 293T cells. It is also possible to see that mNP and stNP have a superior transcription activity then their FL and their POU-C mutants, therefore we can assume that their N domains interact both with the POU and Gal4-DBD to stimulates the Gal4-lux reporter in 293T cells, suggesting that there might exist cofactors in 293T that interact with (mN) resulting in its activation

In HeLa cells the both mutants (POU-C and N-POU) linked to the Gal4-DBD were able to stimulate transcription, but the presence of an alternative DNA binding domain presenting opposite results to Brehm study [20]. The separation of transcription domains (N) and (C) resulted in a higher activation of the Gal4-lux reporter, them their full lenghts. mPC is the only mutant which activity is inferior in POU-C them in N-POU, but both higher them the full length. Zebrafish-FL and Xl91FL has a similar activity to N-POU but both activities are inferior to their respective POU-C mutant (figure 3.8a; fig. 3.9a; fig. 3.10a). The presence of an alternative DNA binding domain act as an enhancer when only one of the transcription domains (N) or (C) are present, otherwise it acts as a repressor to the Oct-4 transcriptional activation.

Mechanism of C-domain regulation

POU-C was not able to activate Gal4-lux reporter in 293T cells while in HeLa cells was the construct with higher transactivation activity indicating that this mutant, POU-C, is cell-type specific. It is known that in mouse, the C-domain is cell-type specific (16) and this can explain its different behaviour in HeLa and in 293T cells.

When the (C) is linked to the DBD, it is active in HeLa but not in 293T cells (Fig. 3.13); when (C) is linked to POU and DBD it is also active in HeLa but not in 293T cells; when the (C) is removed and only the POU is linked to the DBD, this mutant is inactive in HeLa but not in 293T cells. Indicating that the (C) controls (POU) by cell-type specific, it is also suggested that differential phosphorylation in HeLa and 293T cells can affect transactivation ability of the (C) domain [23]. This can explain the transactivation activity that is seen in stOct4, it is the only one active when the (C) is linked to DBD in 293T cells (fig. 3.13b). Sturgeon Oct-4 does not have any phosphorylation sites in its (C) domains (fig. 3.2) meaning that it requires its native DNA binding domain (POU) to mediate cell-type specificity.

(C) is not the only factor responsible for regulating POU function. It is known that there are viral and cellular cofactors that modulate the activity of POU transcription factors. These cofactors rather interact with the POU domain than with other transactivation domains and that it might result in negative regulation. Consistent with this, is the activation of the POU domain only in 293T cells, indicating that in HeLa cells might exist different cofactors that does not exist in 293T cells, that interacts with the POU domain resulting in negative regulation.

(N) domain

In HeLa cells, both transactivation domains (N) and (C) are sufficient to stimulate Gal4-lux, whether they are linked to Oct-4 binding domain or to Gal4-DBD, suggesting that the two separated transactivation domains have equivalent function in HeLa cells, but its activity decrease in the simultaneous presence of Oct-4 and Gal4 DNA binding domain, suggesting that (C) and (N) interact in a similar way with DBD and POU.

As was said before, the (C) domain is cell-type specific, but the (N) domain does not exhibit the same regulation. (N) is active in HeLa and in 293T cells independent if it is linked to Oct-4 or Gal4 binding domain. The activities of the two transactivation domains can be differentially affected by protein-protein interaction, consistent with their different amino acid constitutions. The (C) is a serine/threonine-rich transactivation domain [16] while the (N) domain is a proline-rich transactivation domain, and its proline region has proved to be essential for transactivation. This region is not only responsible for transactivation; ubiquination post-translation modifications have also been reported to regulate Oct-4 transcriptional activity [23].

Oct-4 is a target of small ubiquitin-related modifier (SUMO) modification that increases Oct-4 stability; its DNA binding, and transactivation function. There are three possible sumoylation sites, two of them are on the POU domain, and one in the (N) domain. It was already proved, that in mouse the sumoylation occurs in the (N) on Lys118. For sumoylation occur it needs SUMO-1, E1-activating enzyme; E2-conjungating enzyme and E3 ligase [23]. The E3 ligase binds the target protein and provides specificity. Four SUMO homologues have been described in mammals one of them, SUMO-1, shares a sequence identity with ubiquitin [23] and both ubiquitin and various E3 ligases are highly expressed in HeLa cells [44]; another one, SUMO-2 is expressed mainly in the kidneys; 293T are human embryonic kidney cells, and therefore may express SUMO-2.

The sumoylation takes place at a specific target sequence in the substrate protein. The consensus SUMO acceptor site requires a strong hydrophobic amino acid; followed by the lysine (K), then any amino acid, and finally a glutamic acid (E) or a aspartic acid [41]. mOct4 has the SUMO motif ^{117}VKLE^{120}. Analysing the different Oct-4 homologues, it was possible to find similar motifs conserved in stOct4 (^{125}IKTE^{128}); in zbpou2 (^{131}VKTE^{134}) and in Xl91 (^{92}LKRE^{95}). Xl60 and Xl25 have similar motifs but not in the same region as the previous ones. Xl60 has ^{49}IKSE^{52} and Xl25 has ^{44}LKSE^{47}; axOct4 does not have this motif in the (N) were the sumoylation can occurs. These events can explain the different activities exhibited for the multiple (N) domains. This sumoylation sites can be found in fig.4.1.

The other two sumoylation sites found in mOct4, ^{214}CKSE^{217} and ^{243}LKCPK^{247} (23) can be found in the POU domain and surprising, given the relatively high degree of conservation of this domain, they are not as conserved as the previous one. A similar sequence to the second sumoylation site, can only be found is stOct4 (^{314}YKIE^{317}) and in zbpou2 (^{332}YKIE^{335}). The third sumoylation site can be found in axOct-4 (^{294}LKCPK^{298}); stOct4 (^{345I}KCPK^{349}) and zbpou2 (^{364}VKCPK^{367}). This sumoylation sites can be found in Fig. 4.1.

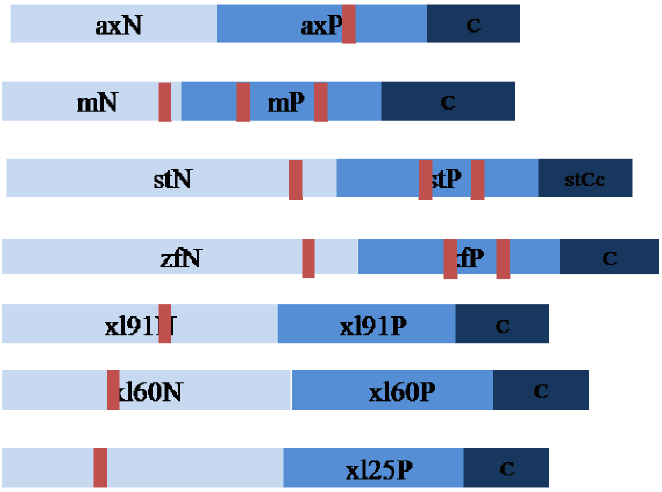


Figure 4. 1- Sumoylation sites in Oct-4 homologues (red).

In HeLa cells it is possible to find two substrates that promote sumoylation, SUMO-1 and E3, while in 293T cells there is the SUMO-2. It is curious to see that all the mutants that have the first sumoylation site (DBD-N) are all active in HeLa and in 293T cells (fig. 3.11), and that the mutants (DBD-N-POU) without sumoylation sites in the POU domain (Xl91, and Xl60) were not active in 293T cells (fig. 3.9b). It might be necessary for an interaction of the SUMO-2 with POU domain sumoylation sites. Consistent with this, is the activation in 293T cells by the POU domain, of the species that have these sumoylation sites (fig. 3.12b). SUMO-2 might bind with the POU domain, resulting in its transactivation activity, where it requires the simultaneous presence of the second and the third sumoylation for stimulating the reporter. This affinity might explain the different activities of the POU domain in HeLa and in 293T cells. It also explains the lack of activity of axN in 293T cells (fig. 3.11 b), which does not have a POU domain thus, the SUMO-2 it is not able to interact with the mutant resulting in its insufficient capacity to bind to Gal4-Lux.

In the presence of two transactivation domains (N and POU), sumoylation might only happen in the (N), and that the SUMO-1 and E3, might have more affinity with the (N) them with the POU. This can explain both activation results that the mutants with (N) are active in HeLa (fig.3.10a and fig.3.11a), and once (N) is removed their lose activity capacity in HeLa cells (fig. 3.12a).

4.5. Functional Conservation

Despite it not being possible to complete the different assays to all the mutants, it was still possible to study the transactivation capacity of Oct-4 domains. It is clear that all species exhibit different transactivation behaviours in HeLa and in 293T cells, therefore their transactivation conservation were individually studied for the different cell types.

In HeLa cell, the different species of Oct-4 have their transactivation function conserved, with exception of Xl60 Oct-4 (fig. 3.7a). In 293T cells, axOct4 and stOct4 were not able to activate the 6Wtk reporter, while all the others did; meaning that despite axOct4 and stOct4 conserved their transactivation function in HeLa cells that conservation is lost in 293T cells, the opposite happed to Xl60. The small difference between Xl60 and Xl91 is enough to give different transactivations activities; this can be somehow related to their different expression stages during *Xenopus* development.

When the same constructs are linked to Gal4 DNA binding domain (DBD), all the species are active in HeLa meaning that they share similar mechanism of interaction with DBD. Despite differences in activation levels, the different mutants all behave the same way in HeLa cells, meaning that their transactivation mechanisms and interactions are conserved in this cell type.

The same does not happen in 293T cells, when the Oct-4 full length is linked to DBD only mOct4 mutant is able to activate the reporter; with this it is possible to see that the different interactions that might exist in 293T cells are only conserved in the other species. The only visible transactivation function that was conserved in all species is their POU domains, this is consistent with the fact that this domain is the most conserved during evolution. According to the amino acid conservation, the (N) domain is more conserved then the (C) domain. When analysing the DND-N it is possible to see a similar transactivation activity between the different species with exception of axN. The DBD-C assay did not include many species, and therefore is not accurate assuming that the functionality, or lack of it, might be conserved.

In DBD-N-POU the regulation mechanisms and the interaction with DBD are only conserved in mOct4 and in stOct4, being the only ones able to stimulate the Gal4-lux reporter. In DBD-POU-C mutants, the transactivation function is conserved in all species where none of the mutants were able to activate the reporter.

The POU domain despite their functional conservation, exhibit different transcriptional levels some of them very superior to mPOU. The fact that one mutant has more transactivation capacity them another is not a good indicator that one species works better than other. In ES cells, it is the precise level of Oct-4 that maintains ES cells in their pluripotent stage [16]. Insufficient or excessive transcriptional levels of Oct-4 promote mesoderm formation [16]. Some of the mutants had high transactivation levels, such as axPOU, that had 3.5 fold activity in 293T cells; and Xl91C that had 11 fold activity whereas axC that had 5 fold both in HeLa cells. For the mutants with (N) or (C) and POU plus DBD, the different transactivation levels also diverge. In HeLa cells, for N-POU mutants, mOct4 is the one with higher activity, followed by Xl60; while for POU-C, the mutant with higher activity level is Xl91>Xl60.

Their transactivation function and its levels are not the only aspects necessary to choose an alternative study system. There are evolutionary divergences that may account for distinct patterns. What might seem a small difference might have a big impact on the embryo development, problems such as severe defect on the axis formation, head structures and neural fold happen with some mutations on Xl25 and Xl60 [9]. Despite Xl91 and Xl60 have transactivation function, it is required the overexpression of all the Oct factors present in *Xenopus* to generate the same phenotype as mOct-4 [9]. Axolotls only have one Oct protein that shares the higher similarity with the mouse than any other species, it is already known that its POU domain can rescue ES cells self-renewal, but it is not know its effects when its phenotype is knockdown, or if they can replace mouse Oct-4 and develop a viable clone.

Now that it is known the different transactivation capacities for each species, it would be necessary to do more assays in order to suggest an alternative model system to study pluripotency regulation. It would be necessary to test with the different mutants in their ability to rescue ES cell self-renewal, and their effects on reprogramming somatic cells.

Chapter 5: Conclusions

This study represents the first investigation of Oct-4 protein domains in lower vertebrates by functional expression analyses. Although previous studies have been carried out by over-expression in *Xenopus* and zebrafish, recently other Oct-4 homologues have been cloned in sturgeon and axolotl, and no functional conservation assays have been done to assess their activity until this study. Different approaches were employed in order to study Oct4 conversation and to understand function of the different domains in regulating activity.

The amino acid analyses are a useful way to begin to relate amino acid conservation to function and transactivation capacity. By comparing sequence analyses and oct4 functional assay data I concluded that the N-terminal domain represents a transcriptional activation domain in Oct4 and that within this domain it is likely that proline and serine residues have important roles on increasing transactivation and promoting the activation of the luciferase reporters and endogenously Oct4 target genes. There are two distinct regions where these two amino acids can be found, in mOct4 there is the proline region whereas in the remaining species there is a serine-rich region, that I termed here as the SEB box.

It is presumed that all Oct4 homologues would be localised to the nucleus subcellularly as this is where transcription factors undertake their role regulating DNA. However this has not been proven for all homologues. mOct4 is localised to the nucleus and requires an NLS sequence in the POU domain for this localisation. It was not determined whether this NLS is present in Oct4 proteins from the lower vertebrates species, or if the POU or NLS in these drove this localisation. Here I showed that all Oct4 proteins analysed fused to GFP are nuclear and the POU domain at least is responsible for transporting Oct-4 into the nuclei. It was possible to highlight a similar NLS sequence to the mOct4 (RKRKR) in the POU of all sequences analysed (24). Due to the high level of similarity between these sequences and mOct4 NLS there is a strong possibility that this sequence is an NLS .

The function for the different Oct4 proteins was tested experimentally. One of the aims was to study function in the native state of the protein in pluripotent cells, but unfortunately due to lack of time, was only possible to study the activity in somatic cells. In HeLa cells, only Xl60 was not able to stimulate 6Wtk reporter, meaning that the transactivation function was conserved to the remaining proteins. In HeLa and 293T cells, the species that had a more similar behaviour to mOct4 was Xl91.

The construction of different mutants lacking N, POU or C domains linked to Gal4-DNA binding domain was used as a novel approach to identify the function of specific domains and such analyses have not been carried out for Oct4 until this study activating a heterologous target. Whenever DBD and two domains are present, the transactivation levels are inferior than when in individual domains where employed. (C) and (N) domains interact in a similar way with POU and DBD, but in the presence of DBD, they present lower activity levels, nevertheless DBD-POU-C has in general more transactivation capacity them DBD-N-POU, suggesting that the C-terminal represent a more potent transcriptional activator. Furthermore the domain with more transactivation capacity when linked to DBD is the (C) domain. The different behaviour exhibited by the mutants in HeLa and in 293T cells will be due to differences in regulation and the interaction of different cofactors in the cells and will need further study to define these differences.

From previous studies it has been shown that the (C)-domain interacts with the DNA binding domain, POU or DBD by differential phosphorylation in HeLa and in 293T cells, resulting in the report activation or inactivation depending on the cell-type. The phosphorylation sites are in general conserved in the different Oct4 proteins, with the notable exception of stOct4 which is the only homologue without predicted phosphorylation sites, and therefore this is an explanation for its non-cell-type specific function, being the only C-domain active in both HeLa and in 293T cells.

With exception of mM, the remaining species might be phosphoregulated by the SEB box, in the different cell-type. mOct4 through modification of the (N) and POU domains has been shown to be regulated by sumoylation and possibly ubiquitination regulating protein stability. There are three sumoylation sites in mOct-4, one in the (N) domain and two in the POU domain. My analyses of the putative sumoylation sites in the homologues show them to be differently conserved, with the only homologues sharing the three sites being mOct4, StO4 and zfPou2. Xl91 and Xl60 only have the putative sumoylation site in the (N); and AxOct4 only has the third putative sumoylation site in the POU domain. These sumoylation sites are regulated by different cofactors, and it has been shown to be cell-type specific.

Even though there are small differences between Oct4 homologues in their activity to transactivate target genes in HeLa and 293T cells, their activity is on conserved. In HeLa cells all the mutants had a similar behaviour, meaning that the transactivation capacity and regulatory mechanisms of each domain are similar. However the Oct4 homologues behave differently in 293T cells where axOct4 and stOct4 have a different transactivation activity to the others.

This study also intended to present an alternative method to study Oct4 function which it achieved, but the assays displayed here for transactivation function are not enough to definitively prove which Oct4 homologue has activity in vivo or to rescue ES self-renewal, or reprogramming somatic cells which are other published assays for Oct4 functionality. What is possible to see from my data is that Oct4 from lower vertebrate shares a similar functional activity to mammalian Oct4 proteins. According to the results, *Xenopus* is the lower vertebrate that has a similar transactivation activity to mouse, more specific: Oct91 protein. This result is consistent with previous studies were Xl91 had the capacity to maintain murine ES cells in the absence of Oct-4 [16], revelling their high homologue function. But in terms of interactions and developmental function, *Xenopus* does not replace and behave the same way was mouse Oct-4, where in order to maintain the same mouse phenotype it is necessary to overexpress all *Xenopus* Oct proteins, Xl91, Xl60 and Xl25 [16]. Not only amphibians are simpler species that share all this functional conservation, but they also are easier to manipulation and have less ethic regulations.

The amphibians in this study were *Xenopus* and axolotl, and axolotl only has one Oct protein whereas *Xenopus* have three; axolotl also have higher similarity mouse than with any other species, due to these reasons I was expecting that axolotl would be the species to share the highest functional conservation to mammals, but unfortunately it was not possible to test the functional activity to all axolotl mutants. By these reason, it would require a more deepl study in axolotl function and I believe that this species promises to be the next species where pluripotent mechanisms are going to be study, and that axolotl Oct-4 might even be possible to reprogram somatic cells with the same efficiency than mouse Oct-4.

5.1. Future work

- It is necessary to redo the luciferase assay with all the different mutants created, and study their regulation without the DBD;
- Study SEB box function, by its ablation;
- Determine the hypothesis for the nuclear localization site for the lower vertebrates; that RKRK represents NLS for Oct-4 homologues;
- Link the clones to a Myc-tag for future protein measurement;
- Test the reprogramming capacity for all the mutants with the other three transcription factors (Sox-2, c-Myc, and Klf-4);
- Produce a viable chimeric by replacing mOct-4 for axOct-4, study effects on the development;
- Combine domains between species and study their activity;
- Study the capacity that the mutants have to maintain ES cells phenotype.

References

1. Lanza, R., Gearhart, J., Hogan, B. (2006). Essentials of Stem Cell Biology. Elsevier ISBN-13: 978-0-12-088442-1

2. Beddinghton, R. S. P., Robertson, E. J. (1989). An assessment of the development potential of embryonic stem cells in the midgestation mouse embryo. *Development* Vol 105, pag: 733-737

3. Lewtzky, M., Yamanaka, S. (2007). Reprogramming somatic cells towards pluripotency by defined factors. *Tissue and cell engineering* Vol 18:467-473

4. United States Department of Health & Human Services. (2001). Human pluripotent stem cell research guidelines. Last revised: October 6, 2001 (acceded 16. 08.2008)

4. United States Department of health & Human Services. 2001. Human pluripotent Stem Cell research guidelines. Webpage: http://www.hhs.gov/news/press/2001pres/01fsstemcell.html last revised: October 6, 2001. Accessed: August 16, 2008

5. Takahashi, K., Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*
Vol 126: 1-14

6. Alberio, R., Campbell, K. H., Johnson, A. D. (2006). Reprogramming somatic cells into stem cells. *Reproduction* Vol 132: 709-720.

7. Docteur en Médecine Thesis, Thérapie cellulaire de línsufficance cardique dans un modéle d’infarctus du myocarde chez le rat. (2004). Université de Genève. Thèse nº Méd. 10358

8. Matsui, Y., Zseba, K., Hogan, B. L. (1992). Derived of pluripotential embryonic stem cells from murine primordial germ cell in culture. *Cell*

Vol 70:841-847

9. Yamanaka, S. (2008). Pluripotency and nuclear reprogramming. *Philosophical transactions of the royal society B* doi:10.1098/rstb.2008.2261

10. Durcova-Hills, G. (2008). Induced reprogramming of human somatic cells into pluripotency: a new way how to generate pluripotent stem cells. *Differentiation* DOI: 10.1111/j.1432-0436.2008.00266.x

11. Junying, Y., Vodayanik, M., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., Thomson, J. A. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* Vol 318: 1917-1920

12 Johnson, A.D., Crother B., White, M. E., Patient, R., Bachvarova, R. F., Drum, M., Masi, T. (2003). Regulative germ cell specification in axolotl embryos: a primitive trait conserved in the mammalian lineage. *Philosophical Transactions of the Royal Society B*. Vol 358: 1371- 1379

13. Eddy, E. M. *Encyclopedia of Life Sciences*, (2006). Germ Plasm and the Molecular Determinants of Germ Cell Fate. John Wiley& Sons, Ltd doi:10.1093/npg.els.0005960

14 Johnson, A.D., Drum, M., Bachvarova, R., Masi , T., White, M. E., Crother B., (2003). Evolution of predetermined germ cells in vertebrate embryos: implications for macroevolution, *Evolution & Development*. Vol 5. Issue 4: 414-431

15. Pesce, M., Gross, M. K., Scholer, H. R. (1998) In line with our ancestors: Oct4 and the mammalian germ. *BioEssays*. Vol 20: 722-732.

16. Niwa, H., S. Masui, I. Chambers, A. G. Smith, J. Miyazaki. (2002) Phenotypoic Complementation Establishes requirements for Specific POU domain and Generic Transactivation Function of Oct-3/4 in Embryonic Stem

Cells. *Molecular and Cellular Biology.* Vol.22 **5:** 1526-1536

17. Fang, W., Scholer, H. R., Atchison, M. L. (2007) Sumoylation of Oct-4 Enhances Its Stability, DNA Binding and Transactivation. *The Journal of Biological Chemistry* Vol 282: 21551-21560

17. National Center for Biotechnology Information, NM_131112. Danio rerio POU [gi:18859248]

< http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nuccore&id=18859248>

Accessed May 4, 2008

18. National Center for Biotechnology Information, AY542376. Ambystroma mexicanum POU domain transcription factor oct-4 mRNA, complete cds. [gi:47028622]

< http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nuccore&id=47028622>

Accessed May 4, 2008

19. Alazard, R., Mourey, L., Ebel C., Konarev P. V., Petoukhov, M. V., Svergun, D. I., Erard, M., (2007) Fine-tuning of intrinsic N-Oct-3 POU domain allostery by regulatory DNA targets. Vol: 35 (13) 4420-32.

20. Brehm, A., Ohbo, K., Scholer H., (1997) The Carboxil-Terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain. *Molecular and Cellular Biology*. Vol 17: 154-162.

21. Morrison, G., Brickman, J. M., (2006) Conserved roles for oct4 homologues in maintaining multipotency during early vertebrate development. *Development.* Vol: 33 2011-2022

22. Cao, Y., Siegel, D., Knochel, W. (2006) *Xenopus* POU factors of subclass V inhibit activin/nodal signaling during gastrulation. *Mechanisms of Development* Vol.123 614-625

24. Pan, G., Qin, B., Liu, N., Scholer, H. R., Pei, D. (2004) Identification of a

Nuclear Localization Signal in Oct4 and Generation of a Dominant Negative Mutant by its Ablation. Vol 279:37013-37020

25. Masui, S., Natakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., Ko, M. S. H., Niwa, H. (2007). Pluripotency governed by *Sox2* via regulation of *Oct3/4* expression in mouse embryonic stem cells. *Nature, cell biology* Vol 9: 625-635

26 Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*. Vol: 113: 643-655.

27. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*. Vol: 113:631-642

28. Levasseur, D. N., Wang, J., Dorschner, M. O., Stamatoyannopoulos, J. A., Orkin, S. H. (2008). Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes & Development*

doi: 10.1101/gad.1606308

29. Lengner, C. J., Camargo, F. D., Hochedlinger, K., Welstead, G. G., Zaidi, S.,Gokhale, S., Sholer, H. R., Tomilin, A., Jaenish, R. (2007). Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell*

Vol 1(4): 403-415

30. Webb, S., (2007). No role for Oct4 in regenerating adult tissues. *Nature Reports Stem Cells*. doi:10.1038/stemcells.2007.103

31. Niwa, H., Miyazaki, J., Smith, A. G. (2000). Quantative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genetic* Vol:24 372-376

32. Yang, H., Do H., Oh, J., Kim, J., Choi, S., Cha, K., Chung, H., Kim,J. (2005). Characterization of Putative Cis-Regulatory Elements that control the transcriptional activity of the human Oct4 promoter. *Journal of Cellular Biochemistry* Vol 96: 821-830

33. National Center for Biotecnology Information, U.S. National Library of Medicine; revised July 23, 2008

    < http://www.ncbi.nlm.nih.gov/>

34. BioEdit Sequence Alignment Editor. Hall, T.A. 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis programs for Windows 95/98/NT. *Nucl. Acids. Symp.* Ser. 41:95-98

35. Vincze, T., Posfai, J. and Roberts, R.J. (2003). NEBcutter: a program to cleave DNA with restriction enzymes Nucleic Acids Res. 31: 3688-3691

    < http://tools.neb.com/NEBcutter2/index.php>

    Accessed May 5, 2008

36. NEB BioLabs Catalog & Technical Reference NEB 2007-08

37. Clamp, M., Cuff, J., Searle, S.M, Barton, G. I. (2004). The Jalview Java Alignment Editor. *Bioinformatics* Vol 20: 426-7

38. Blom, N., Gammeltoft, S., Brunak, S. (1999). Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology* Vol 294 (5) 1352-1363

39. Katoh. K., Kuma, K., Toh, H., Myata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* Vol 33: 511-518

40. Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, Z., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., Yamanaka, S. (2008). *Nature*

Biotechnology Vol 26: 101-106

41. Brody, T. B. The Inter*active* Fly © 1996 Evolutionary homologues < http://chervil.bio.indiana.edu:7092/allied-data/lk/interactive-fly/segment/cubitus2.htm> accessed August 17, 2008

42. Miyamoto, Y., Imamoto, N., Sekimoto, T., Tachibana, T., Seki, T., Tada, S., Enomoto, T., Yoneda, Y. (1997). Differential Modes of Nuclear Localization Signal (NLS) Recognition by Three Distinct Classes of NLS Receptors. Vol 272: 26375-26381

43. Hansis, C., Barreto, G., Maltry, N., Niehrs, C. (2004). Nuclear Reprogramming of Human Somatic Cells by *Xenopus* Egg Extract Requires BRG1. *Current Biology*, Vol 14: 1475-1480

44. Chaudhry, M. A., Chodosh, L.A., McKennal, W. G., Muschel, R. J. (2002). Gene expression profiling of HeLa cells in G1 or G2 phases. *Oncogene* Vol 21: 1934-1942.

APPENDICES

APPENDIX 1 *Ambystoma mexicanum*

Axolotl Oct4 (accession number AY54376)

```
ATGGCTGGGCATTTGGGACAGGAGATTGGGCGGGCTGCCTATGGGTTCGGTGCACAGGC
CTTGCACCTGGGGGCCGGGGGCCTCGAGGCGGGCGGGCCGGGCTTCCTGTCCGAGAGCT
ATGGGCCCTACGCCGGCTTCAAGGCGCTGGAGTATGCCCATGGCGGGGCGGAAGGAGAG
GGCCGACCGGGGGCCCATGGGCTGGCACGGGCCTGGTACCCCTTCTCGGAGGCCTGGGG
CCCTGTGTATGGGCAGAGCGGTGCCGGCGCAGGGTTCGAGAGCAGCCGGGTGGAGGTCA
AGGTGGAGAGGCCCGACAAGGAGGCTGGCTACGGGCAGCAGCACCAGCAGGCCTGGGCT
GGCTACTTCGTGCCCCAGCTGGCAGTGCCCGCCAGGTCGCCTGCGTCCGTGGCCAGCGG
AGGGCAAGTACCGGCCGCACCTGCCAGCCCGTCCGATGACAGCCCGCACAGCAGCACCG
CCAGCAGCAGCAGCGCCAGCCCGGACCTGGGGGCTGGGGGCGCCCCGCGGGACCTGGAC
AGCGGAGACGAGGAAGGGGGACGTCGGCGGACCTTGAACAGTTTGCCAAGGAGCTGAA
GCAGAAGCGCATCACGCTGGGCTTTACGCAGGCGGATGTAGGGCTGGCGCTCGGGGCGC
TGTACGGGAAGATGTTCAGCCAGACGACGATCTGCCGGTTCGAGGCCCTGCAACTGAGC
TTCAAGAACATGTGTAAACTGAGACCCCTGCTCCAGCGCTGGCTGGTCGAGGCCGACAC
CAACGAGAACCTGCAGGAGCTCTGCAACCTGGAGAATGCCCTGCAACAAGCCCGGAAGA
GGAAAAGAACCAGCATCGAGAACAGCGTCAAGGACAACCTGGAGGCCTTCTTCCTGAAG
TGTCCGAAGCCCACCCATCAGGAGATCGCCCACATCTCCGAGGACCTCAATCTGGAGAA
GGACGTGGTCCGCGTCTGGTTCTGCAACAGGCGACAGAAGGGGAAGCGCAGCATTTGCC
GGGAGGAGTATGATGGCTTCCAGCAGTACCCAGGGATGCAGCCGGGCCCACCGGCTCTG
AGCCACCTGCCCACCTCCTACATCGCGCAGGGCTACAACGGCGCCGCCGCCGCCTTCGC
CGCGGTCTACATGCAGCCCTTCCACGACAGCGAGATGTACTCGCAGACCGTGAGCAGGC
ACCTGCACTCCAACTGA
```

White background: N domain

Light gray: POU domain

Dark gray: C domain

APPENDIX 2 *Mus musculus*

Mouse Oct-4 (accession number NM_013633)

```
ATGGCTGGACACCTGGCTTCAGACTTCGCCTTCTCACCCCCACCAGGTGGGGGTGATGG
GTCAGCAGGGCTGGAGCCGGGCTGGGTGGATCCTCGAACCTGGCTAAGCTTCCAAGGGC
CTCCAGGTGGGCCTGGAATCGGACCAGGCTCAGAGGTATTGGGGATCTCCCCATGTCCG
CCCGCATACGAGTTCTGCGGAGGGATGGCATACTGTGGACCTCAGGTTGGACTGGGCCT
AGTCCCCCAAGTTGGCGTGGAGACTTTGCAGCCTGAGGGCCAGGCAGGAGCACGAGTGG
AAAGCAACTCAGAGGGAACCTCCTCTGAGCCCTGTGCCGACCGCCCCAATGCCGTGAAG
TTGGAGAAGGTGGAACCAACTCCCGAGGAGTCCCAGGACATGAAAGCCCTGCAGAAGGA
GCTAGAACAGTTTGCCAAGCTGCTGAAGCAGAAGAGGATCACCTTGGGGTACACCCAGG
CCGACGTGGGGCTCACCCTGGGCGTTCTCTTTGGAAAGGTGTTCAGCCAGACCACCATC
TGTCGCTTCGAGGCCTTGCAGCTCAGCCTTAAGAACATGTGTAAGCTGCGGCCCCTGCT
GGAGAAGTGGGTGGAGGAAGCCGACAACAATGAGAACCTTCAGGAGATATGCAAATCGG
AGACCCTGGTGCAGGCCCGGAAGAGAAAGCGAACTAGCATTGAGAACCGTGTGAGGTGG
AGTCTGGAGACCATGTTTCTGAAGTGCCCGAAGCCCTCCCTACAGCAGATCACTCACAT
CGCCAATCAGCTTGGGCTAGAGAAGGATGTGGTTCGAGTATGGTTCTGTAACCGGCGCC
AGAAGGGCAAAAGATCAAGTATTGAGTATTCCCAACGAGAAGAGTATGAGGCTACACAC
CTTTCCCAGGGGGGGCTGTATCCTTTCCTCTGCCCCCAGGTCCCCACTTTGGCACCCCA
GGCTATGGAAGCCCCCACTTCACCACACTCTACTCAGTCCCTTTTCCTGAGGGCGAGGC
CTTTCCCTCTGTTCCCGTCACTGCTCTGGGCTCTCCCATGCATTCAAACTGAGGCACCA
GCCCTCCCTGGGGATGCTGTGAGCCAAGGCAAGGGAGGTAGACAAGAGAACCTGGAGCT
TTGGGGTTAA
```

White background: N domain

Light gray: POU domain

Dark gray: C domain

APPENDIX 3 *Acipenser oxyrinchus oxyrinchus*

Sturgeon Oct-4 (not published)

```
ATGTCTGATCGGTCTGTCACCCCGGGTTCGGAGGCTTCCAGTCGGGCACACGAGCTCCA
CCGGACGATGTACACGCAGGAGAGCCTAGCCGCTTCTCTTCAGTTCGCCAACGGGATGC
TACAAGACCCGAACTCCGTGTTCAATAAACCCGCCTGCTACAACGGTATCGCCGCCCAA
CATTTCTTCCCGTTTTCTGCAGTCGGCGGTGACTATCGACATCCCGATATTCAGGTAGC
GGACCTGAGCCAAGCTAGACACTGGTACCCNTTCTCTACCCCGGAGCTCACGGGCCAAG
TAGCGGGACTGACCACAGCCCACCAACCGGCGAATCTGAGCCCACGCATCGCAGAAACC
CGGGATCAGACCAAAAGCGACATTAAAACCGAAAAACTGGACGAATTCTCGCNTGAAAG
AAAATCAGCGCTGCCACCGCCGCCCGCTACCATGGCCCCCGGGGTTTATCACTCAAACC
ACTGGAACCCTTCGTTCTGGCCCGGCTTGACCCACGCCCCAGCCCCCGCCGCCACTCCG
GTCTCCTCTTCGCCCAGTAGCCACAGTTACCCGACGGCTGGTGTCTTCACGACGGCAGC
GCCCCAAACGCTACTGGTCCCAGTGCAACAGACCTCGAACCCCGGGAGCAGCGGCTCCT
CCAGTGGCGCGGGCAGTGAAGTGGGGCAGTCTAGCGACTCGGAAGAAGAGGAGAATTTG
TCCACGGAGGAGCTGGAGCAGTTTGCCAAGGAACTGAAACACAAGCGGATCACCCTGGG
ATTCACGCAGGCTGATGTGGGGCTGGCGCTGGGCAACCTCTACGGGAAGATGTTCAGTC
AGACCACGATCTGCCGGTTCGAAGCTCTGCAGCTGAGCTTCAAGAACATGTGCAAGCTG
AAGCCACTGCTGCAACGCTGGCTGAACGAGGCTGAGAACACAGACAACCCGCAGGATAT
GTATAAGATTGAGCGTGTGTTTGCTGACGCCCGGAAGAGGAAGCGCAGGACCAGCCTGG
AGGTGACCGTGCGCGGCGCTCTCGAGTCCTACTTCATCAAGTGCCCGAAGCCCAACACC
CAGGACATCACGCAGATCGCAGAGGACCTCCGCCTGGAGAAAGATGTGGTGCGGGTGCG
GTTCTGTAATCGACGTCAGAAGGGGAAGCGTCTGGCCCTGCCCTTTGATGAGGAGGGTG
CAGAGGGGCAGTATTTCGATCCCAGCCCACAAATGCCCCTGTGCAACGGGCATCTCCAA
ACTCAGGGGTACCCTGGCACTGCACCCCCACACCTCTACCTACCCGCCTTCCACAAGCC
AGAGGTCTTCAAGCAGACCCTGCCCCAGGGATTCCCTGTGGGGCTGGGACACCTGACCA
GCTAG
```

White background: N domain

Light gray: POU domain

Dark gray: C domain

APPENDIX 4 *Danio rerio*

Zebrafish Oct4 (accession number NM_131112)

ATGACGGAGAGAGCGCAGAGCCCAACAGCAGCAGACTGCAGACCCTATGAGGTCAACAG
GGCCATGTATCCTCAAGCCGCGGGCCTGGATGGACTTGGCGGAGCGTCCTTGCAGTTTG
CGCACGGTATGCTTCAGGATCCAAGTCTGATTTTTAACAAGGCCCATTTCAACGGAATC
ACCCCCGCGACAGCCCAGACCTTCTTTCCATTTTCAGGCGATTTTAAAACGAACGATTT
GCAAGGTGGCGACTTTACGCAGCCCAAACACTGGTACCCGTTTGCGGCCCCCGAGTTCA
CTGGGCAGGTTGCAGGAGCGACGGCCGCCACTCAGCCGGCGAACATCAGCCCTCCTATC
GGCGAGACTAGAGAGCAAATTAAGATGCCATCTGAGGTCAAAACCGAGAAAGATGTTGA
AGAATACGGGAATGAAGAAAACAAGCCGCCGTCACAATATCACCTCACCGCTGGAACAT
CTTCCGTCCCCACCGGGGTGAACTACTACACGCCATGGAACCCTAATTTCTGGCCTGGA
CTGTCCCAAATTACGGCCCAAGCTAATATTTCCCAAGCTCCCCCAACTCCCTCCGCTTC
ATCCCCATCTCTGTCTCCGTCTCCCCCTGGAAATGGGTTCGGAAGCCCAGGATTTTTTA
GCGGAGGCACCGCGCAAAACATTCCCTCAGCTCAAGCGCAAAGTGCACCCCGGAGCAGT
GGGTCCTCCAGTGGAGGATGCAGTGATTCTGAGGAAGAGGAGACTCTGACTACTGAAGA
TTTGGAGCAGTTTGCGAAAGAGCTTAAACACAAGCGCATCACTCTGGGCTTCACGCAGG
CAGATGTGGGACTCGCGCTTGGAAACTTGTATGGCAAAATGTTCAGTCAGACGACAATC
TGCCGCTTTGAGGCTCTCCAACTTAGTTTCAAGAACATGTGCAAACTGAAGCCGTTGTT
GCAGAGGTGGTTGAACGAGGCCGAAAACTCCGAGAACCCTCAGGATATGTACAAAATTG
AACGGGTGTTTGTCGACACGCGAAAAAGAAAACGAAGGACCAGCTTGGAAGGCACAGTC
CGTTCTGCTCTAGAGTCGTACTTCGTGAAGTGCCCCAAACCCAACACTCTGGAGATAAC
GCACATATCCGATGATCTAGGCCTGGAGAGAGATGTAGTGCGTGTATGGTTCTGCAACC
GTAGACAGAAGGGAAAGCGTCTAGCTTTGCCCTTTGATGACGAGTGTGTTGAAGCACAG
TATTACGAGCAGAGTCCACCACCTCCACCCCACATGGGTGGCACTGTGCTCCCAGGTCA
AGGCTATCCTGGACCAGCCCATCCTGGAGGAGCCCCTGCCTTATACATGCCATCCCTCC
ACCGACCAGATGTCTTCAAAAACGGCTTGCACCCTGGTTTGGTGGGTCATCTCACCAGCT
AA

White background: N domain

Light gray: POU domain

Dark gray: C domain

APPENDIX 5 *Xenopus laevis*

Xenopus laevis Oct-91 (Accesion number M60077)

ATGTATAACCAACAGACCTACCCTTCCTTTACCCACAACCCAGCCCTGATGCCAGATGG
CAGCGGACAGTATAACTTGGGCACCTACACCGGCATGGCCAGGCACCCCCACCAAGCTC
AAGCATTCTTTCCTTTCTCTGGGGTGAAATCGGACTATGGGGACCTTGGGGGGCAAACT
ACTAGCGTGGGTGATACTTCTGCCTGGAACCCCCTAACTTCTCTGGATTCTGCCAACCA
GTTGGGCATCTCCGGCCAAGGGAACCCGTTTAAGAACTTAAAAAGGGAGAGAGAAGATG
ATGAGGAGAAATCAGAATCTCCTGAGCCCAAATGCAGCCCACCGTCTCTTCCACCCGCC
TATTACACCCATGCGTGGAATCCCACCACCACCTTCTGGTCTCAGGTCTCCTCAAGTGG
GACCACGGTCGTGTCCAAGCCTTTACCCACCCCACTCCAGCCCGGTGATAAATGTGACC
CCGTGGAGGCAAATAAAATTTTTACCAGTAGCCCCGATAAGTCAGGAGAGAGTGGGATC
TCCAGCTTGGACAACAGTCGCTGTAGCAGCGCCACCAGCAGCTCATCTGGTGGGACCAA
CGTGGGGACCCCTAGAAGCCTTTCCAGGGGAGCTAGTGATGGGCTGAGCAGTGACAGTG
AGGAGGAAGCCCCTAATTCTGGGGAGATGGAGCAGTTTGCCAAGGATCTGAAGCACAAA
CGGATCACAATGGGCTACACTCAGGCGGATGTTGGCTACGCTCTAGGAGTGCTTTTTGG
TAAAACATTCAGCCAGACCACCATCTGCCGATTCGAGTCCCTGCAGCTCTCCTTCAAGA
ACATGTGCAAGCTGAAACCTCTGCTGCGGAGTTGGCTCCATGAGGTTGAAAACAATAAA
AACCTCCAAGAAATCATCAGCCGAGGGCAGATAATTCCTCAGGTTCAGAAGAGAAAGCA
CAGGACCAGCATCGAGAACAACGTGAAATGCACCTTGGAGAACTATTTCATGCAGTGCT
CAAAGCCAAGTGCCCAGGAGATCGCCCAGATTGCGAGAGAGCTGAACATGGAGAAAGAC
GTGGTGAGAGTCTGGTTCTGCAACCGGCGGCAGAAGGGCAAGCGCCAGGTGTACCCCTA
TATTAGGGAGAATGGCGGGGAGCCTTATGACGCCCCCCAAACCTTGACACCCCCTTCTC
AAGGTCCTTTTCCATTACCCCAAGTGATGCCCTCACAAGTTTTTCCCACGGTCCCACTG
GGTGCCAACCCTACGATCTACGTTCCGACGTATCACAAAAACGACATGTTCCCCCAGGC
AATGCATCATGGGATTGGCATGGGTAACCAAGGCAACTAG

White background: N domain

Light gray: POU domain

Dark gray: C domain

APPENDIX 6 *Xenopus laevis- Oct60*

Xenopus laevis Oct-60 (accesion number: M60075)

ATGGACCAGCCCATATTGTACAGCCAAACCTCCTTCCCCAACTTCACCTACAGCCCAGG
AGTGGTGCAAGACGGGGGCAATTACCAGTATTTGGGCAACTACAATGCCCCGTCCTACC
CGCAGCCGTTCTTCCATGTCCCTGTTATAAAATCAGAATTTGGTGCCCATGAGGAAGAA
ACGCCGGGGAGTTGCCATGCTGCTTCCTTTGACTGGAACCTGTACCCTCACTTTCAGAT
CTCTAACCAGGCGGCTTCCAACAGTTCTGGAGATCCAAGTCCAGAGGGAAGAACTGAGG
AGGATGGTTCTGTCAGTGAAGGGAGGTCCTCCAGTTCCCCTTCCCCCAATTCTCCCCTG
GTGCCTTCCTTTGCCCAATATTGGCATTATCCCTCCTGGCAGCAGGGGAACCTAACCAA
CCAACCTCACACTCTTTTTGATGGGGGTGATGAGAAGCCCCAACAGTCTCGTCACAGTC
CAACGGCCTCGCTAGGGAGTGGGGCGTCCAACACCGAGGATGAGGAGGTCCCCAGTGCC
ATCTCCAGTAGAGCAGAAAGAGGTCTCTGTAGTCCCTCTCCTAATAATGCCTCATGTGG
CCCTGGAACTGAAGAGGATGGAATGACCCTTGAGGAGATGGAAGAGTTTGCCAAGGAGC
TGAAACAGAAGCGGGTGGCACTGGGTTATACCCAAGGAGACATTGGCCACGCCCTGGGA
ATATTATACGGGAAGATGTTCAGCCAGACGACTATCTGCCGCTTTGAGTCCCTGCAACT
GACCTTCAAAAATATGTGTAAACTCAAACCCCTATTGGAGCAGTGGCTGGGAGAGGCGG
AGAATAACGACAACCTACAGGAGATGATCCACAAGGCCCAGATTGAGGAGCAGAACCGC
AAGCGGAAGATGAGGACCTGCTTTGATACTGTTCTAAAGGGCCAACTAGAGGGCCACTT
CATGTGCAATCAGAAACCTGGTGCCAGGGAGCTGACGGAAATTGCCAAAGAACTGAGTC
TGGAGAAAGATGTGGTGAGGGTCTGGTTCTGCAACCGGCGGCAGAAGGAGAAGAGCAAG
TTCAGAATGTCCAAGGGGCATGAGTTTGTGGGTGGTGCCAGTCCTGGATCCATCCAATC
AGAACACATTTCTTTCACCCCCATTCCAGCTAATTCCCAGGACTATGGATTGGCCTCTC
TACACCCCAACCGAGCCCCCTTCTACCCACCTCCCTTCCCCAGGAATGAGCTGTTCCCT
CACATGGCTCCGGGGATATCTATGGGGGTCCTGACCGGCTGA

White background: N domain

Light gray: POU domain

Dark gray: C domain

APPENDIX 7- *Xenopus laevis- Oct25*

Xenopus laevis Oct-25 (Accesion number M60074)

```
ATGTACAGCCAACAGCCCTTCCCAGCCTTCGCTTTCAACGCCGGACTCATGCAGGATCC
CGCCAACTGTCATTTTGGGGGTTACGCGGGTTTAGGACACCCCCAGCCCTTCTCCTTCG
CCTTCTCTACGCTGAAATCGGAAAACGGAGAGTCTGGAGTTCAGGGTATGGGGGACTGT
ACGACTCCTGTGATGCCCTGGAACTCACTGGCGTGTTTCGATCACCAGGTCCAGATGGA
GAACAACCAGCAAGGGAATCCGCCCAGAGCCCCAAGTCCGACTCTCAGCGACTCCAGGA
TTAAGGTCAAAGAGGAGGTTGTCCATGAAACTGACAGCGGAGAAGAGTCCCCAGAACCC
AAATACCCCAGCCCCCCTAATCCCTCTCTCTACTACCCCAACGCATGGACTGGCGCCCC
TTTCTGGCAAGTGAACCCCACGCCGGGCAATAACATCAACCCAATGCCCAACCAGACTC
TTGTGAAAAACACCAGCCTACCGGGGAACACCACCTACCCCACCCCAGCAAACCAAAGC
CCCAATACCCCAGTAGACTGTGTGACCTCCAGTATGGAAAGCAGCAGATGCAGCAGCAC
CAACTCCCCCAATGGGGCAATTAATGAACGGGCCACCACTATCCCTAATGGAGAGATGC
TTGATGGGGGGCAATCCAGCGACAATGAGGAGGAGGTTCCCAGCGAATCAGAAATGGAG
CAGTTTGCCAAAGATCTGAAGCACAAGCGAGTGTCCCTGGGCTACACACAGGCGGATGT
TGGCTACGCACTCGGGGTCCTGTATGGCAAGATGTTCAGTCAGACGACAATCTGTCGCT
TCGAGTCGCTGCAGCTCAGCTTCAAGAACATGTGTCAACTGAAACCTTTCCTGGAGCGC
TGGGTGGTGGAGGCAGAGAACAACGACAACCTGCAGGAGTTGATCAACCGGGAGCAGGT
CATTGCCCAAACACGGAAGAGAAAAAGGAGGACGAACATAGAGAATATAGTGAAGGGGA
CCCTGGAGAGTTACTTCATGAAATGTCCCAAGCCGGGCGCCCAGGAGATGGTGCAGATC
GCCAAGGAACTGAACATGGACAAAGATGTGGTCCGGGTCTGGTTTTGCAATCGGCGGCA
GAAAGGCAAGCGCCAGGGAATGCCCACCGTTGAGGAGAACGACGGCGAAGGCTACGATG
TTGCACAGACCATGGGTCGCCCTGTTGGACACTATGCGCTGCAGCAGGTGGTGACCCCA
CAAGGTTACATGGCAGCTCCGCAGATTTACGCCTCGGCGGGCCACAAAAACGACCTGTT
CCCCCAGACGGTCCCACACGGAATGGCCATGGGGGGCCACATTGGCTGA
```

White background: N domain

Light gray: POU domain

Dark gray: C domain

APPENDIX 8. Primers combination

List of the combination of the primers for al fragments and their respective fragment sizes

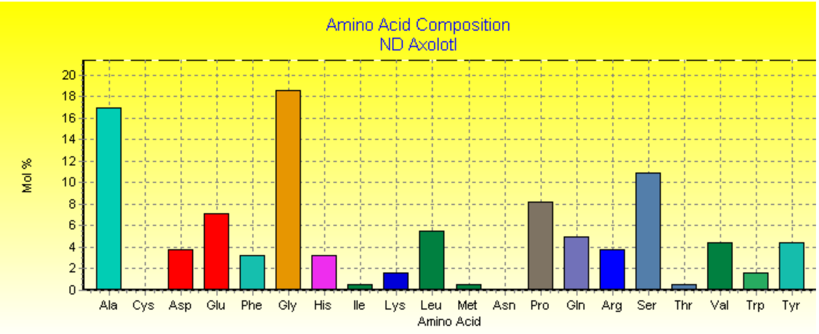
Specie	Primer	Domain	Size (bp)
Axolotl Oct4	Nhe1 F+ HindIII R	Full- Lenght	1197
	Age1 F+ HindIII R	POU-C	648
	Nhe1 F+ Sal1 R	N-POU	996
	Nhe1+ Age1 R	N domain	549
	Age1 F+ Xho1 R	POU domain	447
	Xho1 F+ HindIIIR	C domain	201
Mouse Oct4	Nhe1 F+ BglII R	Full Length	1131
	Age1 F+ HindIII R	POU-C	732
	Nhe1 F+ Xho1 R	N-POU	846
	Nhe1+ Age1 R	N domain	399
	Age1 F+ Xho1 R	POU domain	447
	Xho1 F+ HindIIIR	C domain	285
Sturgeon Oct4	Nhe1 F+ HindIII R	Full Length	1362
	Age1 F+ HindIII R	POU-C	657
	Nhe1 F+ Sal1 R	N-POU	1158
	Nhe1F + Age1 R	N domain	705
	Age1 F+ Sal1 R	POU domain	453
	Xho1 F+ HindIIIR	C domain	204
Zebrafish	Nhe1 F+ HindIII R	Full Lenght	1419
	Age1 F+ HindIII R	POU-C	663
	Nhe1 F+ Xho1 R	N-POU	1209
	Nhe1+ Age1 R	N domain	756
	Age1 F+ Xho1 R	POU domain	453
	Xho1 F+ HindIIIR	C domain	210
Xenopus 91	Nhe1 F+ HindIII R	Full length	1338
	Age1 F+ HindIII R	POU-C	678
	Nhe1 F+ Xho1 R	N_POU	1143
	Nhe1+ Age1 R	N domain	660
	Age1 F+ Xho1 R	POU domain	483
	Xho1 F+ HindIIIR	C domain	195
Xenopus 60	Nhe1 F+ HindIII R	Full length	1281
	Age1 F+ HindIII R	POU-C	669
	Nhe1 F+ Xho1 R	N-POU	1076
	Nhe1+ Age1 R	N domain	612
	Age1 F+ Xho1 R	POU domain	464

	Xho1 F+ HindIIIR	C domain	205
Xenopus 25	Nhe1 F+ HindIII R	Full length	1347
	Age1 F+ HindIII R	POU-C	657
	Nhe1 F+ Xho1 R	N-POU	1140
	Nhe1+ Age1 R	N domain	690
	Age1 F+ Xho1 R	POU domain	450
	Xho1 F+ HindIIIR	C domain	207

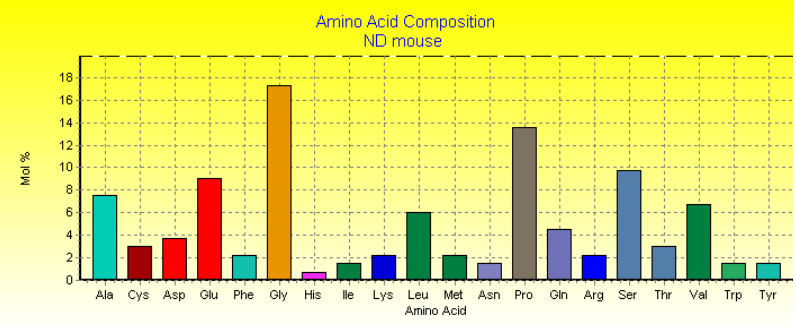
APPENDIX 9. Amino acid composition of Oct-4 N terminal

	Axolotl		Mouse		Sturgeon		Zebrafish		Xl 91		Xl 60		Xl 25	
Amino Acid	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %
Alanine	31	16.94	10	7.52	26	11.06	27	10.71	12	5.45	12	5.88	13	5.65
Cysteine	0	0	4	3.01	1	0.43	2	0.79	3	1.36	3	1.47	5	2.17
Aspartic Acid	7	3.83	5	3.76	9	3.83	8	3.17	13	5.91	9	4.41	8	3.48
Glutamic Acid	13	7.1	12	9.02	13	5.53	15	5.95	11	5	15	7.35	15	6.52
Phenylaline	6	3.28	4	3.01	9	3.83	14	5.56	7	3.18	9	4.41	9	3.91
Glycine	34	18.58	23	17.29	15	6.38	25	9.92	21	9.55	19	9.31	18	7.83
Histidine	6	3.28	1	0.75	10	4.26	4	1.59	4	1.82	7	3.43	4	1.74
Isoleucine	1	0.55	2	1.5	4	1.7	8	3.17	3	1.36	4	1.96	4	1.74
Lysine	3	1.64	3	2.26	5	2.13	7	2.78	9	4.09	2	0.98	5	2.17
Leucine	10	5.46	8	6.02	13	5.53	10	3.97	13	5.91	8	3.92	9	3.91
Methionine	1	0.55	3	2.26	4	1.7	4	1.59	3	1.36	1	0.49	8	3.48
Asparagine	0	0	2	1.5	9	3.83	13	5.16	11	5	13	6.37	23	10
Proline	15	8.2	19	14.29	23	9.79	29	11.51	21	9.55	23	11.27	31	13.48
Glutamine	9	4.92	6	4.51	13	5.53	17	6.75	10	4.55	12	6.37	13	5.65
Arginine	7	3.83	3	2.26	8	3.4	5	1.98	6	2.73	15	2.45	4	1.74
Serine	20	10.93	11	8.27	33	14.04	28	11.11	37	16.82	33	16.18	24	10.43
Threonine	1	0.55	4	3.01	18	7.66	20	7.94	19	8.64	9	4.41	17	7.39
Valine	8	4.37	9	6.77	11	4.68	6	2.38	7	3.18	7	3.43	11	4.78
Tryptophan	38	1.64	2	1.5	3	1.28	3	1.19	3	1.36	3	1.47	3	1.3
Tyrosine	8	4.37	2	1.5	6	2.55	7	2.78	7	3.18	9	4.41	6	2.61

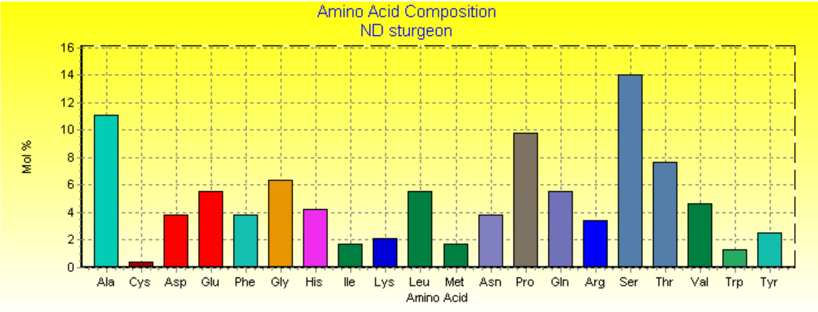
Axolotl



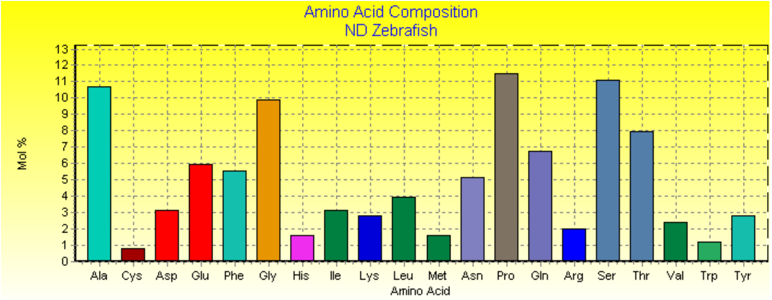
Mouse



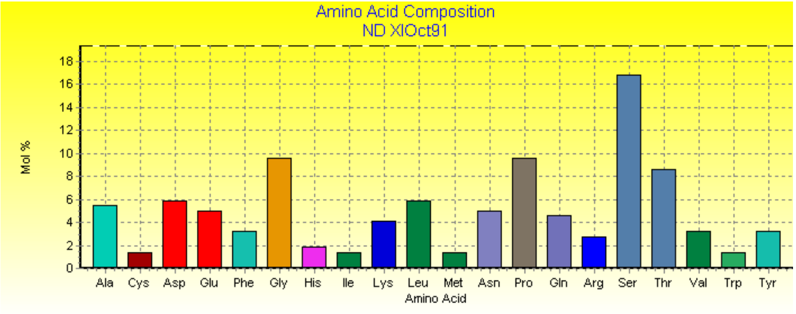
Sturgeon



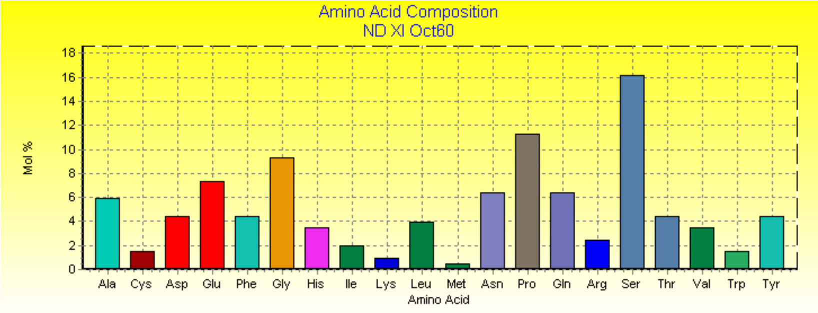
Zebrafish



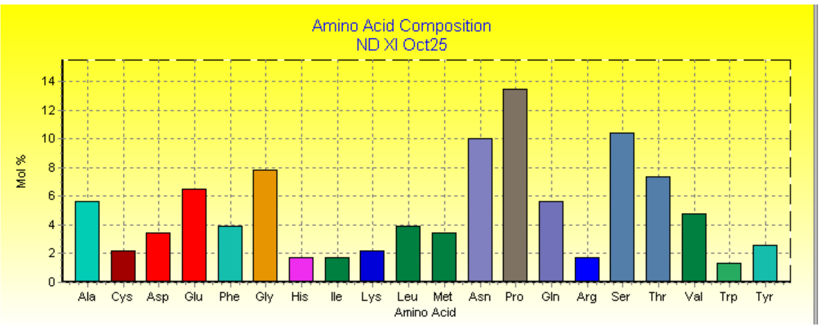
Xl Oct91



Xl Oct60


Amino Acid Composition
ND Xl Oct60
Mol %
0
2
4
6
8
10
12
14
16
18
Ala
Cys
Asp
Glu
Phe
Gly
His
Ile
Lys
Leu
Met
Asn
Pro
Gln
Arg
Ser
Thr
Val
Trp
Tyr
Amino Acid

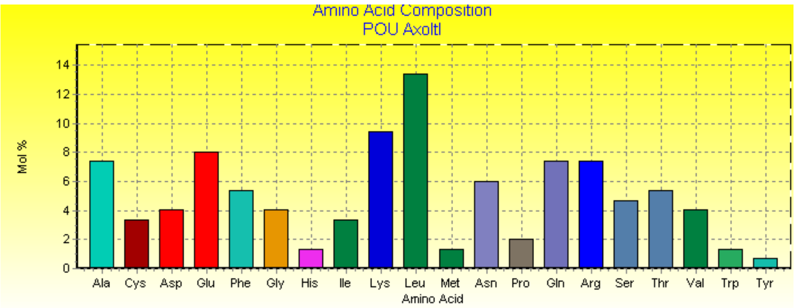
Xl Oct25


Amino Acid Composition
ND Xl Oct25
Mol %
0
2
4
6
8
10
12
14
Ala
Cys
Asp
Glu
Phe
Gly
His
Ile
Lys
Leu
Met
Asn
Pro
Gln
Arg
Ser
Thr
Val
Trp
Tyr
Amino Acid

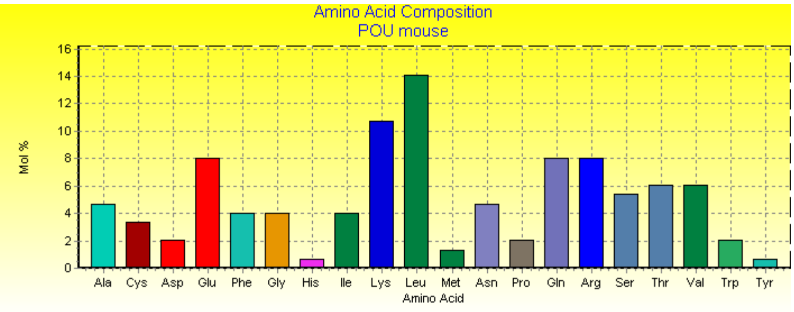
APPENDIX 10. Amino acid composition of Oct-4 POU domain

	Axolotl		Mouse		Sturgeon		Zebrafish		Xl 91		Xl 60		Xl 25	
Amino Acid	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %
Alanine	11	7.38	7	4.7	10	6.62	7	4.64	6	3.97	7	4.55	7	4.67
Cysteine	5	3.36	5	3.36	4	2.65	4	2.65	5	3.31	5	3.25	4	2.67
Aspartic Acid	6	4.03	3	2.01	7	4.64	7	4.64	3	1.99	4	2.6	5	3.33
Glutamic Acid	12	8.05	12	8.05	12	7.95	12	7.95	11	7.28	18	11.69	13	8.67
Phenylaline	8	5.37	6	4.03	8	5.3	8	5.3	7	4.64	8	5.19	7	4.67
Glycine	6	4.03	6	4.03	6	3.97	7	4.64	7	4.64	9	5.84	8	5.33
Histidine	2	1.34	1	0.67	1	0.66	2	1.32	3	1.99	3	1.95	1	0.67
Isoleucine	5	3.36	6	4.03	6	3.97	5	3.31	9	5.96	6	3.9	6	4
Lysine	14	9.4	16	10.74	15	9.93	14	9.27	15	9.93	17	11.04	15	10
Leucine	20	13.42	21	14.09	18	11.92	19	12.58	12	7.95	17	11.04	12	8
Methionine	2	1.34	2	1.34	3	1.99	3	1.99	5	3.31	8	5.19	6	4
Asparagine	9	6.04	7	4.7	7	4.64	7	4.64	12	6.62	7	4.55	9	6
Proline	3	2.01	3	2.01	4	2.65	4	2.65	4	2.65	2	1.3	3	2
Glutamine	11	7.38	12	8.05	9	5.96	7	4.64	13	8.61	11	7.14	12	8
Arginine	11	7.38	12	8.05	15	9.93	14	9.27	11	7.28	10	6.49	12	8
Serine	7	4.7	8	5.37	5	3.31	7	4.64	9	5.96	5	3.25	7	4.67
Threonine	8	5.37	9	6.04	10	6.62	11	7.28	7	4.64	8	5.19	6	4
Valine	6	4.03	9	6.04	7	4.64	8	5.3	9	5.96	5	3.25	11	7.33
Tryptophan	2	1.34	3	2.01	1	0.66	2	1.32	2	1.32	2	1.3	2	1.33
Tyrosine	1	0.67	1	0.67	3	1.99	3	1.99	3	1.99	2	1.3	4	2.67

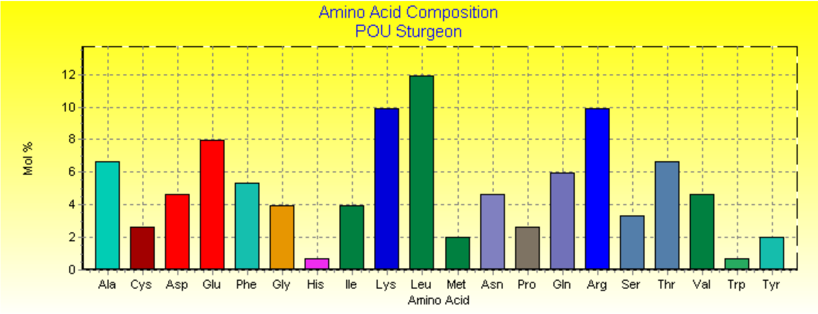
Axolotl



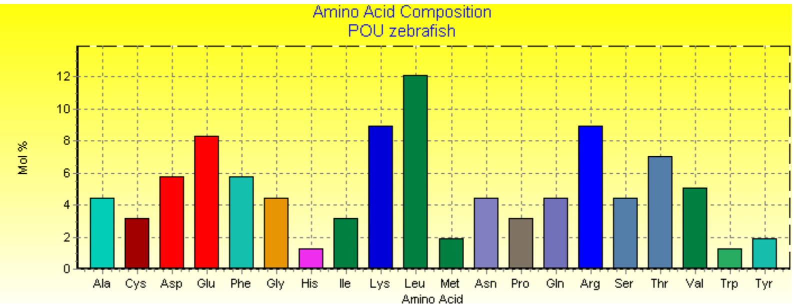
Mouse



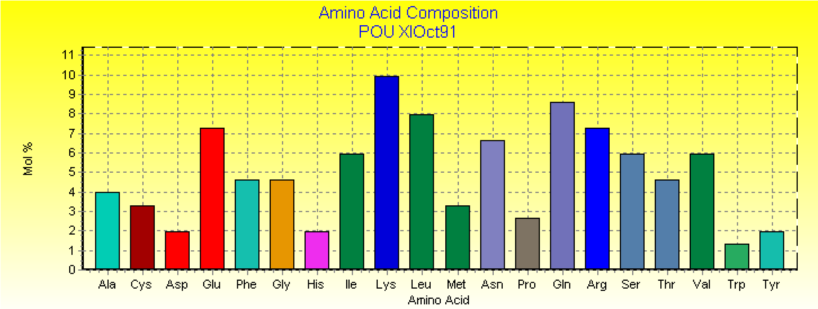
Sturgeon



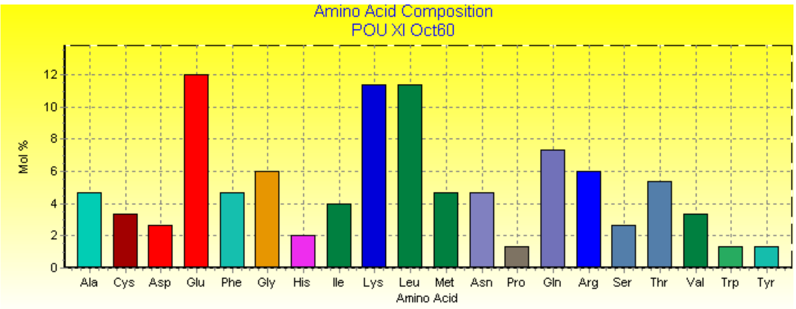
Zebrafish



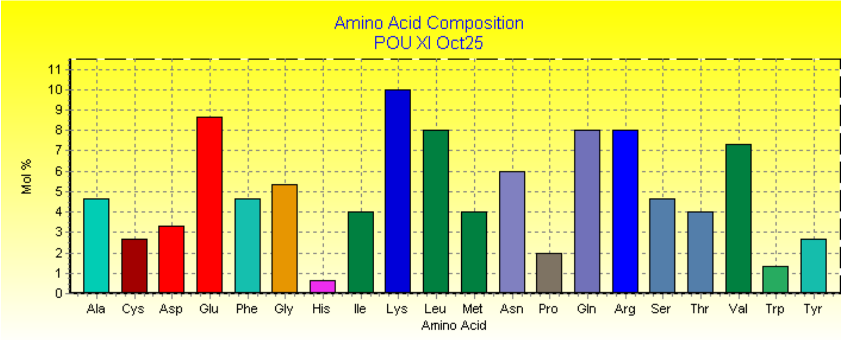
Xl Oct91



Xl Oct60



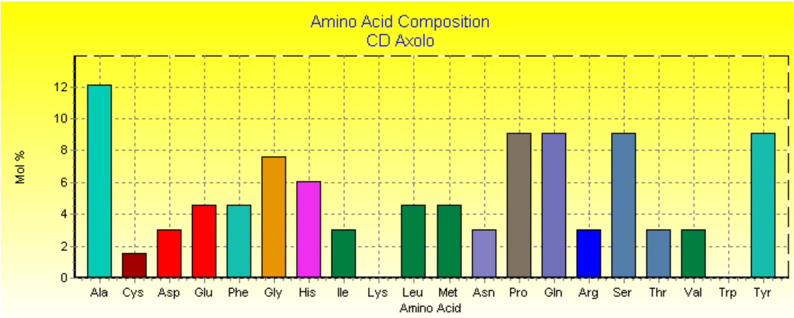
Xl Oct25



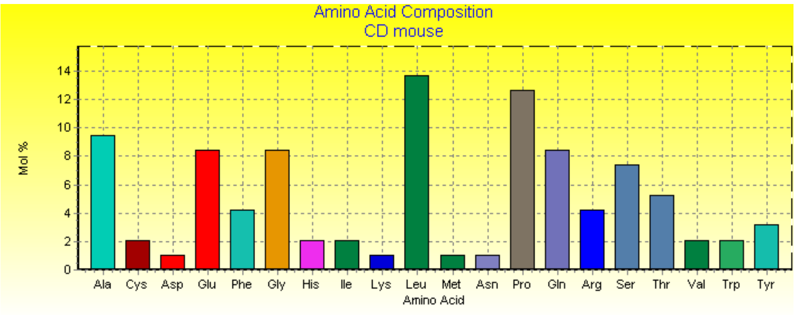
APPENDIX 11. Amino acid composition of Oct-4 C-terminal

	Axolotl		Mouse		Sturgeon		Zebrafish		Xl 91		Xl 60		Xl25	
Amino Acid	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %	Nº	Mol %
Alanine	8	11.94	9	9.47	3	4.41	4	5.71	3	4.62	5	7.35	7	10.14
Cysteine	1	1.49	2	2.11	1	1.47	1	1.43	0	0	0	0	0	0
Aspartic Acid	2	2.99	1	1.05	2	2.94	3	4.29	2	3.08	1	1.47	3	4.35
Glutamic Acid	3	4.48	8	8.42	4	5.88	3	4.29	0	0	3	4.41	3	4.35
Phenylaline	3	4.48	4	4.21	5	7.35	2	2.86	3	4.62	5	7.35	1	1.45
Glycine	5	7.46	7	7.37	8	11.76	10	14.29	6	9.23	8	11.76	10	14.49
Histidine	4	5.97	2	2.11	4	5.88	5	7.14	3	4.62	4	5.88	4	5.8
Isoleucine	2	2.99	2	2.11	0	0	0	0	2	3.06	4	5.88	2	2.9
Lysine	0	0.00	1	1.05	2	2.94	1	1.43	1	1.54	0	0	1	1.45
Leucine	3	4.48	13	13.68	8	11.76	6	8.57	3	4.62	4	5.88	2	2.9
Methionine	3	4.48	1	1.05	1	1.47	2	2.86	4	6.15	2	2.94	5	7.25
Asparagine	2	2.99	1	1.05	1	1.47	1	1.43	4	6.15	3	4.41	2	2.9
Proline	6	8.96	12	12.63	11	16.18	14	20	13	20	11	16.18	6	8.7
Glutamine	6	8.96	8	8.42	6	8.82	3	4.29	6	9.23	2	2.94	6	8.7
Arginine	2	2.99	4	4.24	0	0	1	1.43	0	0	2	2.94	1	1.45
Serine	6	8.96	7	7.37	2	2.94	3	4.29	2	3.08	7	10.29	1	1.45
Threonine	2	2.99	5	5.26	4	5.88	2	2.86	5	7.69	2	2.94	4	5.8
Valine	2	2.99	2	2.11	2	2.94	4	5.71	4	6.15	2	2.94	6	8.7
Tryptophan	0	0.00	2	2.11	0	0	0	0	0	0	0	0	0	0
Tyrosine	6	8.96	3	3.16	3	4.41	4	5.71	3	4.62	2	2.94	4	5.8

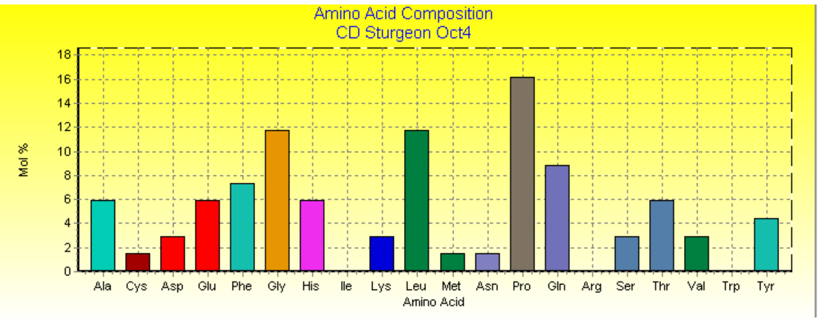
Axolotl



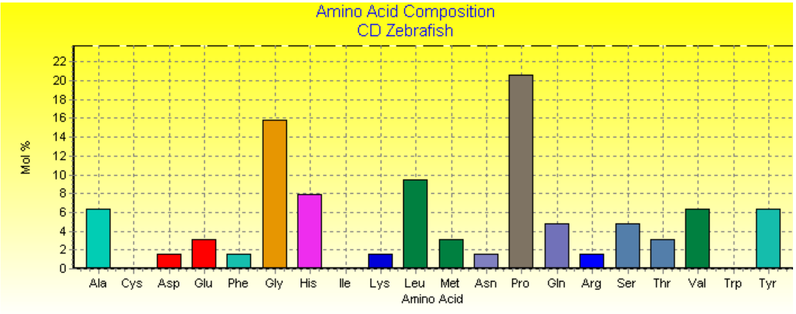
Mouse



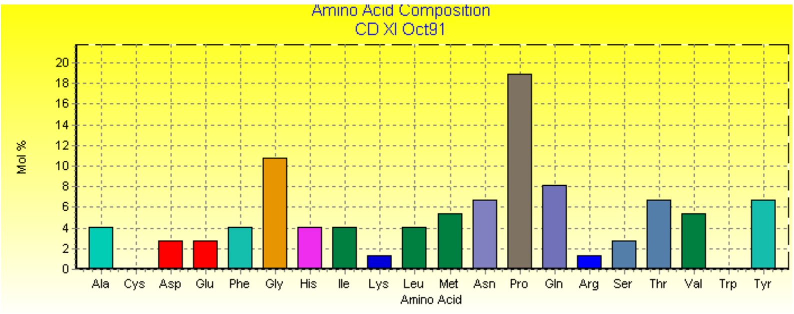
Sturgeon



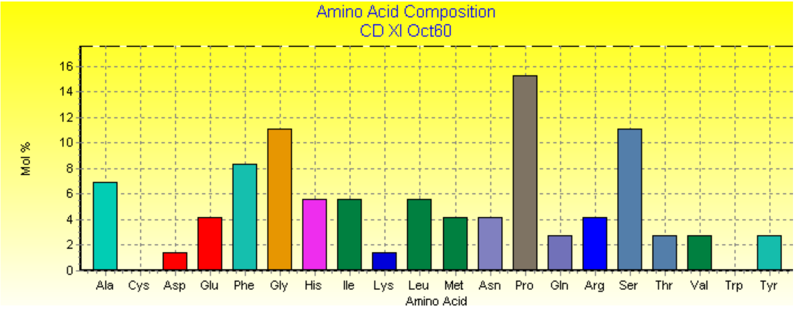
Zebrafish



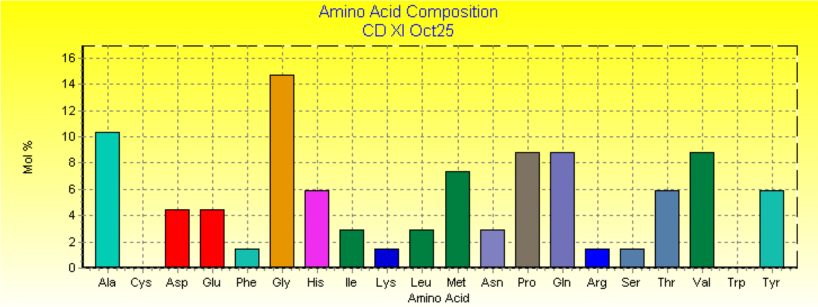
Xl Oct91



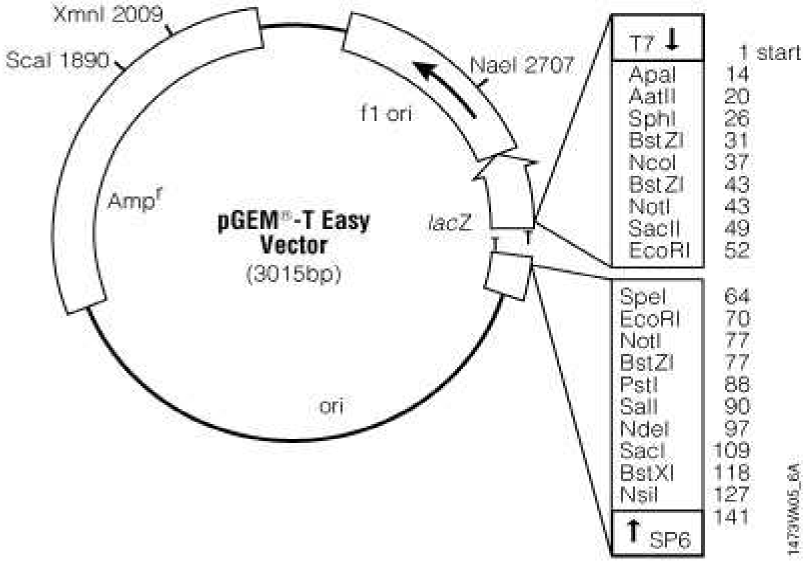
Xl Oct60



Xl Oct25

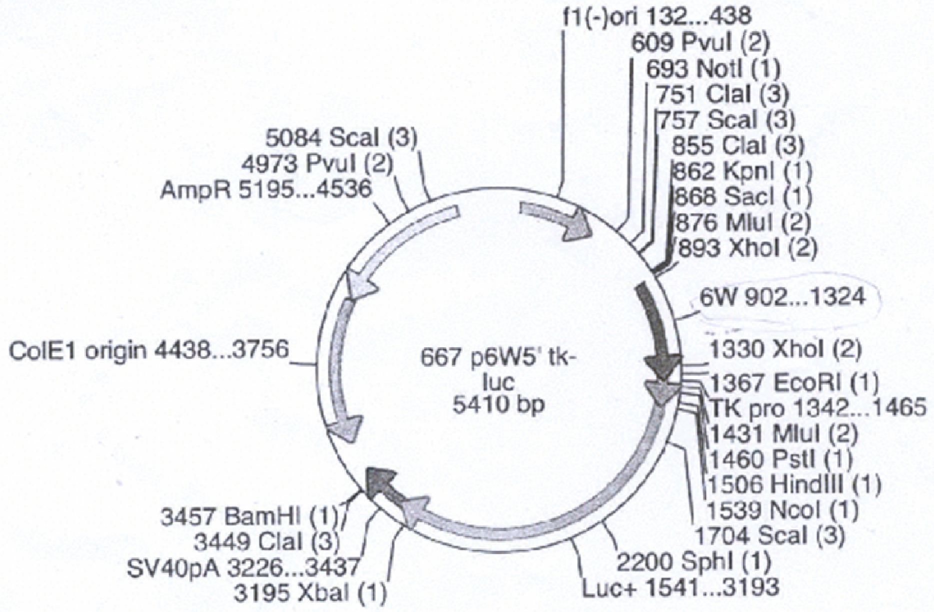


APPENDIX 12. pGEM®-T Easy vector (Promega, UK)



APPENDIX 13. p6Wtk-luc vector

No.	667	Name	p6W5'tk-luc
Size:	5410	bp	



	Enzyme		Size
	NotI	3B	5410
	KpnI	1B	5410
	SacI	1B	5410
	EcoRI	Ecof	5410
	PstI	3B	5410
	HindIII	2	5410
	NcoI	3	5410
	SphI	2	5410
	XbaI	2B	5410
	BamHI	3B	5410

	Enzyme		Size1	Size2
	PvuI	3B	4364	1046