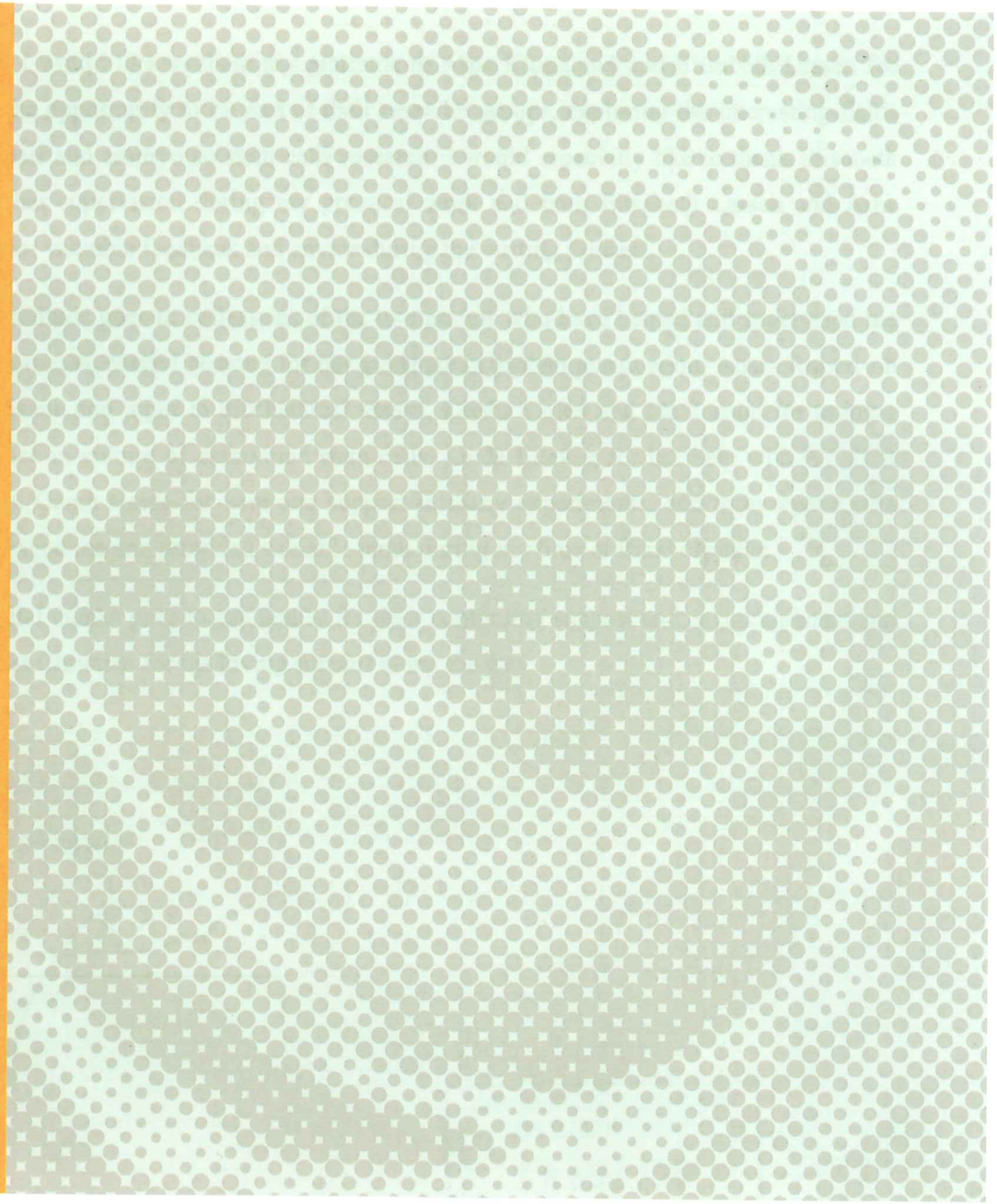
Mastergradsoppg. 2010

Characterization of *Escherichia coli* AlkB Homologue 1

Gry Asker

Norwegian University of Life Sciences
Department of Chemistry, Biotechnology and Food Sciences
Master Thesis 60 credits 2010



Acknowledgement

This master thesis completes my two year Master's Degree in Biotechnology at the Norwegian University of Life Sciences. It is a result of scientific experiments carried out at the Laboratory for Genome Repair and Regulation headed by Prof. Arne Klungland and at the Laboratory for Embryonic Stem Cell Research headed by Postdoc. Elisabeth Larsen in the period from August 2009 to May 2010. Both laboratories are under the Institute of Medical Microbiology, Section for Molecular Biology located at Rikshospitalet in Oslo. My main supervisor has been Elisabeth Larsen and my internal supervisor at UMB has been Prof. Dzung B. Diep.

First of all, I would like to share my gratitude to Arne Klungland and Elisabeth Larsen for giving me the opportunity to be part of their research groups, exploring the interesting fields of DNA repair and embryonic stem cells research. I would especially like to thank Elisabeth Larsen for her supervision during my practical work and both Elisabeth Larsen and Arne Klungland for valuable feedback and advice during the writing process.

A special gratitude goes to colleagues of the two groups who have helped and supervised me during this process. Marivi Nabong for her guidance in protein purification, Adam Robertson for his guidance and input in biochemical assays, Leisha Colen and Ida Jonson for training and assistance in culture of embryonic stem cells, Linda Ellevog for her help with RT-PCR experiments, Gaute Nesse for help and support in many experiments, and to everyone who have encouraged me and provided valuable input during my time at Rikshospitalet.

A very special thanks goes to my family for their continuing support and encouragement and to my dearest Christian for his loving support along the way.

Oslo, May 18th 2010

Gry Asker

Summary

The DNA molecule is exposed to alkylating agents produced endogenously and in the environment. Cells have therefore developed several responses to alkylating damage including direct repair by oxidative demethylation. The *Escherichia coli* AlkB protein is a Fe(II)- and 2-oxogulatarate dependent oxygenase that demethylates N1-methyladenine and N3-methylcytosine lesion in DNA and RNA.

Eight mammalian AlkB homologues (ALKBH1-8 in humans and Alkbh1 in mice) have been identified. The first human AlkB homologue (ALKBH1) has the highest sequence similarity with *E. coli* AlkB and can partially rescue AlkB deficient *E. coli* against the methyl methanesulfonate (MMS) alkylating agent. Despite these findings it is not yet confirmed that ALKBH1 can repair DNA *in vivo*.

Embryonic stem (ES) cells have potential of self-renewal and the ability to develop into any differentiated cell type. A network of transcription factors, including OCT4, NANOG, and SOX2, maintain pluripotency in both human and mouse ES cells. OCT4 and NANOG have been demonstrated to bind the promoter sequence of *ALKBH1*, thus ALKBH1 might have a role in ES cell self-renewal and pluripotency.

An aim of this study was to identify the biochemical function of ALKBH1 and to characterize a potential role in ES cells. The recombinant ALKBH1 protein was successfully purified, however, its biochemical function could not be determined. Two *in vitro* protein-protein interaction assays demonstrated that ALKBH1 interacts with core regulators of ES cell pluripotency and with core histones.

In vitro differentiation of mouse ES cells using two different approaches were investigated and Real-time PCR analysis showed that expression of NANOG, OCT4 and SOX2 was down-regulated, whereas the expression of mesoderm and endoderm markers was up-regulated. The expression of Alkbh1 remained nearly constant during differentiation.

Results from this study indicate that ALKBH1 might play a role in ES cell pluripotency, possibly acting as a histone demethylase.

Sammendrag

DNA-molekylet utsettes kontinuerlig for skade indusert av endogene og eksogene agens. Disse inkluderer blant annet alkylerende stoffer, som fører til methyleringsskader. For å beskytte mot slike skader har cellene utviklet reparasjonsmekanismer, deriblant direkte demetylering. I *Escherichia coli* (*E. coli*) reparerer AlkB proteinet N1-metyladenin og N3-metylcytosin skader i DNA og RNA ved oksidativ demetylering ved bruk av 2-oksogluterat og Fe^{2+}.

I pattedyr er det identifisert åtte AlkB homologer (ALKBH1-8 i mennesker og Alkbh1-8 i mus). Den første humane AlkB homologen (ALKBH1) har høyest sekvenslikhet med *E. coli* AlkB og kan komplementere metyleringssensitiviteten til AlkB mutert *E. coli*. Til tross for dette har den enzymatiske aktiviteten til ALKBH1 ennå ikke blitt identifisert.

Embryonale stamceller (ES-celler) er uspesialiserte (pluripotente) celler som har evne til å utvikle seg til alle kroppens over 200 celletyper. Et nettverk av transkripsjonsfaktorer, blant annet NANOG, OCT4 og SOX2, hindrer at ES-cellene begynner å spesialisere seg. Det er tidligere vist at NANOG og OCT4 binder promotorsekvensen til *ALKBH1* genet, noe som indikerer at ALKBH1 også kan være med i å opprettholde ES-cellers pluripotens.

Et formål med denne oppgaven var å identifisere den biokjemiske funksjonen til ALKBH1 og karakterisere hvilken rolle enzymet kan ha i ES-celler. Rensing av rekombinant ALKBH1 protein var vellykket, men, i likhet med mange andre, lykkes vi ikke med å påvise biokjemisk aktivitet. To ulike protein:protein interaksjonsassay viste at ALKBH1 interagerer med flere, pluripotensfaktorer og med histoner.

To ulike protokoller for *in vitro* differensiering av ES-celler ble testet ut, og Real-time PCR viste at ekspresjonen til NANOG, OCT4 og SOX2 ble nedregulert, mens ekspresjonen til mesoderme og endoderme markører ble oppregulert. Uttrykket av ALKBH1 endret seg lite i løpet av differensieringen.

Resultatene fra denne oppgaven indikerer at ALKBH1 kan ha en rolle in ES-cellers pluripotens, muligens ved å fungere som en histon demetylase.

Table of Contents

1 Introduction

1.1 Genomes and DNA

All living organisms have genomes that contain the biological information needed to maintain their structure and functions as well as harboring the organism's hereditary information. Most genomes are made up of DNA (deoxyribonucleic acid) except for some viruses such as the HIV retrovirus, the poliovirus and the influenza virus, which have RNA (ribonucleic acid) as their genetic material. The DNA molecule is a linear polymer made up of four monomeric subunits named nucleotides composed of a 2'-deoxyribose (five-carbon sugar), a nitrogenous base: cytosine (C); thymine (T); adenine (A); or guanine (G), and a phosphate group (figure 1.1). Two DNA chains running in opposite directions are intertwined to form a right handed double helical structure. These two strands interact by the formation of hydrogen bonds (H-bonds) between complementary bases – two H-bonds between an A and a T and three H-bonds between a G and a C. These are known as Watson-Crick base pairs (bp) named after James D. Watson and Francis Crick who proposed the DNA double helix structure in 1953 (Brown, 2007).

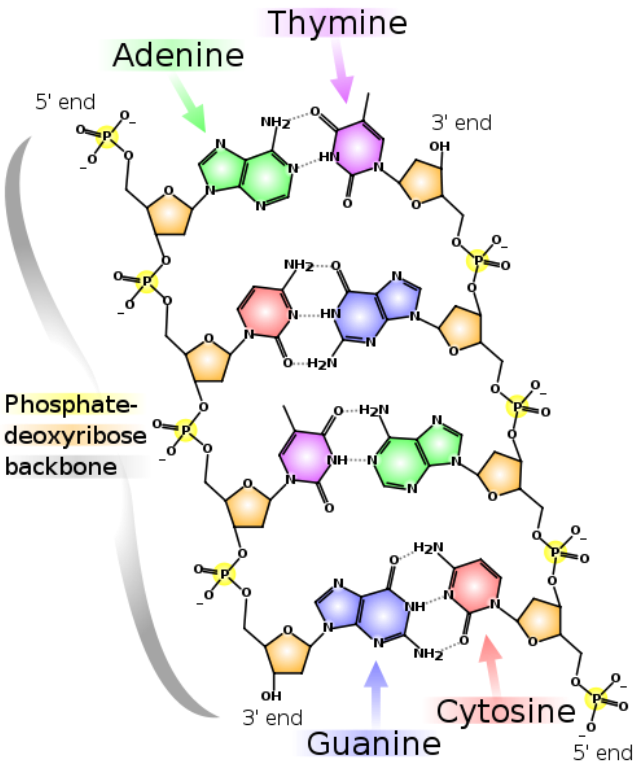


Figure 1.1: Phosphates and 2-dioxyribose sugars make up the backbone in the DNA helix. They are joined together by forming a phophodiester bond between the third and fifth carbon atom of adjacent sugar rings. The direction of the nucleotides in one strand is opposite to the direction in the opposite strand, i.e. they strands are antiparallel. This means that the ends of the two strands are asymmetric with a 5' (five prime) end and a 3' (three prime) end having a terminal phosphate and a terminal hydroxyl group (OH), respectively. H-bonds between complementary bases hold the two strands together (Brown, 2007). Figure is taken from Wikipedia (2010a).

The limitations in base pairing, i.e. A with T and G with C, is of biological importance in DNA replication where two perfect copies can be made using the nucleotide sequence of the

original DNA molecule as a template. During transcription, the process in which DNA encoded genes are transcribed into messenger RNA (mRNA), the same base pairing properties are used. However, this is not as stringent as in DNA replication (Brown, 2007).

1.2 Mutations and Damage to Macromolecules

Genomes are prone to mutations. Mutations can result from errors in DNA replication and may change the nucleotide sequence. Chemical or physical agents can also react with DNA and cause a change in the structure of individual nucleotides. Mutations and chemical alternations can affect the base-pairing capabilities of the altered nucleotide, leading to miscopying of the template DNA strand during DNA replication or block DNA and RNA polymerases that can lead to DNA strand breaks (Brown, 2007).

It is essential to repair DNA lesions as degradation of a damaged DNA molecule is unfavorable as in most situations there is no backup copy (Falnes et al., 2007). Cells have therefore developed many different DNA repair mechanisms that work to minimize the number of lesions (Lindahl and Wood, 1999). If the DNA damage is too extensive to be repaired prior to cell division, apoptosis – programmed cell death – is the outcome. In the case of damages to RNA and protein molecules the outcome is different as these can be newly synthesized based on the information in DNA. Faulty and damaged RNA and protein molecules are therefore degraded continuously in cells. There are, however, indications of existing repair mechanisms for RNA and proteins as well. If a certain modified/damaged RNA base is frequent it may be preferable to have a repair system for that specific damage. Some lesions, however, might be impossible to reverse, thus degradation of the molecule is the outcome (Falnes et al., 2007).

1.2.1 Alkylating Agents and Alkyl Lesions

Cells are exposed to alkylating compounds produced endogenously and in the environment. Alkylating agents are electrophiles that react with the nucleophillic centers of DNA and RNA molecules (figure 1.2). The nucleophillic centers of nucleotide bases are oxygen- (O) and nitrogen- (N) atoms (Begley and Samson, 2003; Drabløs et al., 2007). In a reaction between a partially positively charged electrophile and a partially negatively charged nucleophile, the electrophile accepts an electron pair from the nucleophile. A bond is then made between the two (Nakamura and Miyoshi, 2010). Alkyl lesions in DNA and RNA molecules are therefore

caused by alkylating agents that add alkyl groups to O- and N-atoms in nucleotide bases (Begley and Samson, 2003; Drabløs et al., 2007). An alkyl group consists only of carbons and hydrogens and is usually part of a larger molecule. Methyl, CH_3, is the smallest alkyl group consisting of only one carbon atom. Following methyl is ethyl (2 carbons), propyl (3 carbons), butyl (4 carbons) and so on.

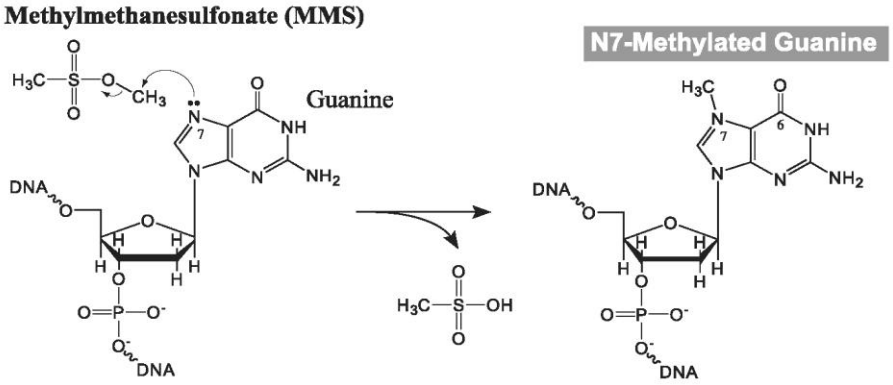


Figure 1.2: DNA methylation by the alkylating agent methyl methanesulfonate (MMS). The MMS methyl group, CH_3, accepts an electron pair from guanine nitrogen atom 7 (N7). The methyl group becomes attached to the N7 atom and generates a methyl lesion at the guanine base (DePamphilis, 2006).

Alkylating agents are classified as either S_N1 or S_N2 agents depending on the chemical mechanism for nucleophilic substitution (S_N). It is found that the former alkylate both N- and O-atoms whereas the latter almost exclusively alkylate N-atoms of DNA bases (Begley and Samson, 2003; Drabløs et al., 2007). The type of alkyl lesion and its specificity depend on the base position at which the nucleotide is modified, the type of alkyl group that is added and the alkylating agent. *O*-alkylations such as O^6-alkylguanine (O^6alkylG) and O^4-alkylthymine (O^4alkylT) are highly mutagenic and genotoxic, whereas *N*-alkylations such as 3-alkyladenine (3alkylA) and 1-alkyladenine (1alkylA) are cytotoxic, but less mutagenic (Drabløs et al., 2007). Alkyl lesions can obstruct Watson-Crick complementary base pairing and lead to blocking of both replication and transcription. Blockage of these processes will lead to cell death, thus making alkyl lesions cytotoxic. Some alkyl lesions might also be bypassed by more error-prone DNA polymerases, thus making them mutagenic (Falnes et al., 2007). This has implicated them in carcinogenesis, neurodegenerative disease and aging (Begley and Samson, 2003; Drabløs et al., 2007).

As mentioned, alkylating agents can also react with RNA molecules. Whether they become functionally inactive depends on the type of RNA. For instance, transfer RNAs (tRNAs) are more tolerant to methylations compared to mRNAs. This is because many positions on a

tRNA can be mutated without interfering with its function. Alkyl lesions within the open reading frame of an mRNA molecule can interrupt correct base-pairing with associated tRNAs during translation and thus be deleterious. In addition mRNA lesions might also cause the ribosome to stall during translation and consequently block translation of the mRNA into a protein. One can also consider the scenario that lesions in mRNAs can cause the formation of mutant proteins (Falnes et al., 2007).

1.3 DNA Repair

In order to protect the genome integrity and promote cell survival, cells have developed several responses to DNA damage. Cells possess several different repair systems where specialized repair systems have been developed for the various types of damages. Different repair systems include: direct repair where a damaged base is directly converted back to its original structure; excision repair where multiple types of DNA damages are excised from the DNA molecule; mismatch repair where errors of DNA replication are corrected; and recombination repair, where gaps are repaired by strand exchange (Freidberg, et al., 2006; Brown, 2007; Freidberg, 2008).

Direct repair can for instance involve the direct reversal of various types of alkylation damage. Based on the process in which the damage base is removed, excision repair can be classified as nucleotide excision repair (NER) and base excision repair (BER). NER is used to describe excision of a longer stretch of nucleotides containing a helix-distortive bulky compound such as pyrimidine (C and T) dimers. BER refers to the removal of damaged nucleotide bases and other inappropriate bases in DNA such as uracil (U) (Freidberg, 2008; Visnes et al., 2009). U replaces T during DNA transcription into mRNA, whereas U is an inappropriate base in DNA (Visnes et al., 2009). Mismatch repair depends on the cell's ability to distinguish between the newly synthesized stand containing the mis-incorporated nucleotide from the parental strand containing the correct nucleotide (Freidberg, 2008).

1.3.1 Repair of Alkyl Lesions

At least three of the previously mentioned repair mechanisms are known to be involved in repair of alkyl lesions in both *Escherichia coli* (*E. coli*) and in eukaryotic cells. These include direct base repair by either methyltransferases or oxidative demethylases, BER initiated by DNA glycosylases, and NER (Drabløs et al., 2007). The proteins involved in these essential

repair systems were initially discovered in *E. coli* mutants that showed increased sensitivity towards alkylation agents (Falnes et al., 2007).

In *E. coli* three different proteins, AlkA, Ada and AlkB, work together in the adaptive response to alkylation damage (Sedgwick and Lindahl, 2002). AlkA is a DNA glycosylase that catalyzes the excision of certain methylated bases such as 3-methyladenine (3MeA) from DNA. Ada is a transcription factor involved in governing the adaptive response as well as being a DNA methyltransferase repairing *O*-methylated bases such as O^6-methylguanine (O^6MeG). AlkB reverse methyl lesions in DNA by oxidative demethylation (Falnes et al., 2007). The knowledge of alkylation repair in *E. coli* has led to the discovery of similar repair processes in eukaryotes. AlkA, Ada and AlkB all have functional homologues in various organisms including humans. This emphasizes the fact that DNA repair of alkyl lesions is an important and fundamental cellular function (Drabløs et al., 2007; Falnes et al., 2007).

1.4 Repair by AlkB

The *E. coli* AlkB protein is involved in direct repair of N1-metyladenine (1MeA) and N3-methylcytosine (3MeC) lesions in DNA and RNA by oxidative demethylation (Drabløs et al., 2007). The gene, a*lkB*, was first described in *E. coli* by Kataoka et al. (1983) when it was found to provide resistance to the S_N2-methylating agent, MMS. Indications that it could be involved in DNA repair came from their studies with alkylated λ phage that survived slightly better when infecting wild-type *E. coli* cells versus *alkB* mutant *E. coli* cells (Kataoka et al., 1983). Research following these initial discoveries indicated that AlkB had a different activity than DNA-methyltransferases, -glycosylases or –nucleases (reviewed by Begley and Samson, 2003). Dinglay et al. (1998) demonstrated that AlkB preferentially binds S_N2-alkylated single stranded (ss) DNA versus S_N2-alkylated double stranded (ds) DNA and therefore proposed that AlkB worked on ss DNA at the replication fork. At this time it was also known that S_N2-methylating agents produce 1MeA and 3MeC more efficiently in ss DNA compared to ds DNA. Dinglay et al. (1998) therefore proposed that AlkB somehow repairs 1MeA and 3MeC DNA lesion in either ss or ds DNA.

Although AlkB was known to be involved in DNA repair for a long time, the actual biochemical function was first demonstrated in 2002 (Falnes et al., 2002; Trewick et al., 2002). Unlike most other DNA repair proteins, AlkB is dependent on specific cofactors and this made it difficult to demonstrate its function. A very important discovery was brought

forward in 2001 when Aravind and Koonin used a bioinformatics approach showing that AlkB belongs to the 2-oxoglutarate (2-OG) and Fe(II)-dependent oxygenase superfamily (Aravin et al., 2001). Proteins belonging to this superfamily use Fe^{2+} ions and 2-OG, also known as 2-ketoglutarate, to oxidize organic substrates with molecular oxygen (O_2) (Drabløs et al., 2007).

2-OG and Fe(II)-dependent oxygenases share a double stranded β-helix (DSBH) protein fold with two conserved histidines and an aspartic acid. These three residues are involved in the co-ordination of Fe^{2+} into the tertiary structure of the protein, whereas an arginine residue within the same fold is involved in co-ordination of 2-OG alongside Fe^{2+}. In almost every case oxidation of the enzyme's substrate is coupled to the conversion of 2-OG into succinate and carbon dioxide (CO_2) (Begley and Samson, 2003; Drabløs et al., 2007). Aravind and Koonin (2001) put this information together with that of Dinglay et al. (1998) and hypothesized that the enzymatic function of AlkB was oxidative demethylation of 1MeA and 3MeC lesions where the oxidized methyl group was removed spontaneously or via a secondary step (Begley and Samson, 2003).

This structural information led to the demonstration of the enzymatic activity of AlkB by two independent groups in 2002 (Falnes, et al., 2002; Trewick et al., 2002). Both groups showed that AlkB could repair 1MeA and 3MeC in ss and ds DNA by using O_2 as an oxidizing agent, 2-OG as a co-substrate and Fe^{2+} as a co-factor (figure 1.3). In the repair process the methyl lesion is oxidized to a hydroxymethyl, which is spontaneously released as formaldehyde with regeneration of the normal base. The co-substrate 2-OG is decarboxylated to yield succinate and CO_2. Although the lesions repaired by AlkB are preferentially introduced in ss DNA they are also found in ds DNA when ss DNA has returned to ds DNA conformation after transcription and replication. AlkB is therefore involved in both repairing ss and ds DNA methyl lesions although the latter at a slightly lower frequency (Falnes et al., 2007).

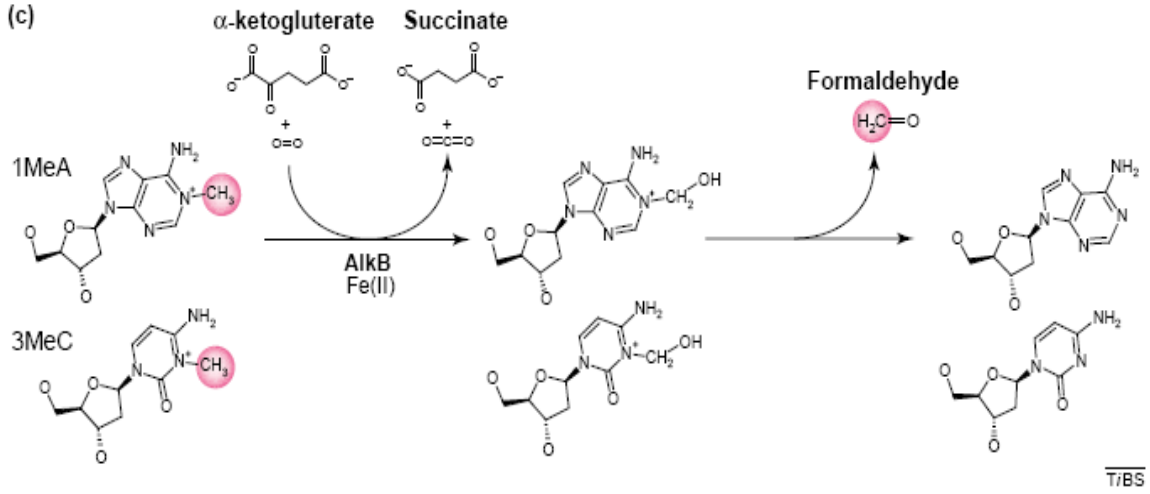


Figure 1.3: The enzymatic activity of AlkB: repair of 1MeA and 3MeC lesions in DNA by oxidative demethylation. Methyl groups are indicated by pink. 2-OG (α-ketogluterate) is converted to succinate and CO_2 using Fe(II) as a cofactor. Removal of methyl groups generates formaldehyde with the regeneration of the normal DNA bases: adenine and guanine top and bottom, respectively (Begley and Samson., 2003).

Following these initial studies *E. coli* AlkB and two human AlkB homologues (ALKBH2 and 3) and *E. coli* AlkB proteins have been shown to work on additional substrates such as the structurally similar 1-methylguanine (1MeG) and 3-methylthymine (3MeT) lesions (figure 1.4) (Falnes, 2004; Koivisto et al., 2004). In addition to these methyl lesions bulkier alkyl lesions such as ethyl, propyl, hydroxyethyl, and hydroxypropyl are also substrates for AlkB but with a much lower efficiency (Duncan et al., 2002; Koivisto et al., 2003).

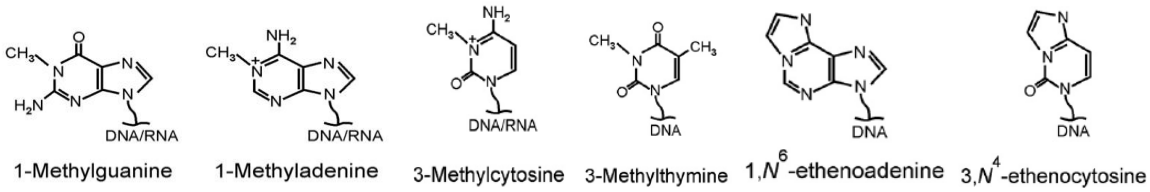


Figure 1.4: Methyl and etheno lesions in DNA and RNA demonstrated to be repaired by AlkB proteins (Falnes et al., 2007).

1.5 AlkB Homologues

There are eight different AlkB homologues found in mammals, which in humans are denoted ALKBH1-8 and Alkbh1 in mouse (Falnes et al., 2007). A bioinformatics approach was used to map these homologues in the human genome. They were found by searching with the *E. coli* sequence against a library of human EST sequences. ESTs are short for expressed sequence tags that represent transcribed gene sequences. The analysis also showed that AlkB

homologues are found in most bacteria and eukaryotes, often with several homologues, as well as in some viruses (Kurowski et al., 2003).

Bacteria, in contrast to eukaryotes, have just one or at the most two *alkB*-like genes, which suggest that eukaryotes are much more complex with increased compartmentalization and differences in sub-cellular localization. An increased number of homologues also suggest that some *alkB*-genes may have taken on new roles in eukaryotes. Furthermore, homologues in RNA viruses suggested that AlkB and its homologues could be involved in RNA processing (Drabløs et al., 2007).

1.5.1 Human AlkB Homologue 1

The first AlkB human homologue, ALKBH1, located at chromosome 14q24 has the highest sequence similarity with *E. coli* AlkB, however it is 173 residues longer (71 at the amino terminus, 42 at the carboxyl terminus and the remainder within the sequence). The full-length sequence of ALKBH1 is highly conserved in humans, mouse and chicken with 70-83% identity (Kurowski et al., 2003). Although its predictions as a functional AlkB homologue have not been fully confirmed, the ALKBH1 protein has been shown to partially rescue an *E. coli alkB* mutant from MMS-induced cell death (Wei et al., 1996). Two groups have however been able to demonstrate two different biochemical actions of ALKBH1. Westbye et al. (2008) described ALKBH1 as a mitochondrial demethylase repairing 3MeC, but not 1MeA lesions in both DNA and RNA. Two years later Müller et al. (2010) reported an additional enzymatic activity – that ALKBH1 could cleave DNA at abasic sites independent of Fe^{2+} and 2-OG.

1.5.2 Human AlkB Homologues 2 and 3

Shortly after the first human homologue had been identified two additional human AlkB homologues, ALKBH2 at chromosomal position 12q24.1 and ALKBH3 at position 11q11, were reported along with their enzymatic activities (Duncan et al., 2002; Aas et al., 2003). Both were shown to remove methyl groups from 1MeA and 3MeC in methylated polynucleotides in a 2-OG dependent manner. This was slightly stimulated by ascorbate and inhibited by Ethylenedinitrilo-tetraacetic acid (EDTA), which is characteristic of 2-OG and Fe(II)-dependent oxygenase activity (Duncan et al., 2002; Aas et al., 2003). Ascorbate can be used as a reducing agent whereas EDTA binds Fe^{2+} and the ions exhibit diminished reactivity.

ALKBH2 and 3 have, however, been shown to demethylate 1MeA and 3MeC lesions in DNA with slightly different preferences. Whereas ALKBH2 has a nuclear localization with higher activity towards ds DNA substrates, ALKBH3 is found in both the nucleus and the cytosol with higher preference for ss DNA substrates (Aas et al., 2003; Falnes et al., 2004).

In addition, ALKBH3 as well as *E. coli* AlkB, have been found to remove 1MeA, 3MeC and 1MeG lesions from RNA *in vitro*, which propose a possible function in RNA repair (Aas et al., 2003; Falnes, 2004). RNA repair has also been demonstrated *in vivo* using *E. coli* mutants infected with chemically methylated ss RNA phage. It was found that mutants produced fewer progeny phage than *E. coli* cells expressing AlkB or ALKBH3 (Aas et al., 2003).

Mouse models have been used to elucidate the mechanisms and *in vivo* roles of Alkbh2 and 3 (Ringvoll et al., 2006). Alkbh2 has been shown to be required for efficient removal of 1MeA and 3MeC from synthetic DNA oligonucleotides *in vitro* and removal of 1MeA from genomic DNA *in vivo.* Alkbh3 has also been demonstrated to remove such lesions *in vitro*, but no detectable repair defect has been observed in cells lacking Alkbh3. In Alkbh2 defective mice 1MeA lesions accumulates during aging and Alkbh2 is able to prevent accumulation of these lesions in the genome (Ringvoll et al., 2006).

1.5.3 Human AlkB Homologues 4 – 7

The functions of ALKBH4-7 have not yet been described. Although displaying high sequence homology with *E. coli* AlkB these homologues may not have a role in repair of methyl lesions in DNA. Since there are several human AlkB homologues these proteins might have taken on other specialized functions within the cell. Some of the proposed roles of ALKBH4-7 are the involvement of neutralizing damaged precursors of DNA synthesis, remove large alkyl groups from DNA or repair minor alkylated DNA lesions such as N2-methylguaninine. They could also be involved in demethylation of methylated proteins such as histones or in reversal of epigenetic silencing by demethylation of 5-methylcytosine (5meC) residues (Sedgwick et al., 2007).

1.5.4 Human AlkB Homologue 8

The human ALKBH8 protein contains a central 2OG and Fe2(II) oxygenase domain and a short N-terminal region containing a RNA recognition motif (RRM). This motif has also been found in some ss DNA binding proteins. A weak similarity to a DNA-binding helix-turn-helix

motif and distinct carboxyl- (C) terminal domain that strongly resembles an *S*-adenosylmethionine (SAM)-dependent methylase has also been identified (Drabløs et al., 2004; Sedgwick et al., 2007).

Two recently published studies have reported that ALKBH8 contains, in addition to its conserved AlkB like domain, a tRNA methyltransferase (Dragony et al., 2010; Sogne-Møller et al., 2010). Dragnony et al. (2010) showed that ALKBH8 catalyzes tRNA methylation with the generation of 5-methylcarboxymethyl uridine (mcm^5U) at the wobble position of certain tRNAs. They also demonstrated that this tRNA modification is critical to DNA damage survival, i.e. that deletion of ALKBH8 in human cells reduced mcm^5U levels in tRNA that in turn increased cellular sensitivity to DNA damaging agents (Dragony et al., 2010).

Songe-Møller et al. (2010) demonstrated the same activity through analysis of tRNAs from *Alkbh8* deficient mice. They demonstrated that Alkbh8 is required for the final step in biogenesis of mcm^5U and that interaction with a small accessory protein, TRM112, is required in order for Alkbh8 to form a functional tRNA methyltransferase (Songe-Møller et al., 2010). These two studies indicate a role of ALKBH8 to be more than just direct DNA repair, but an involvement in the DNA damage response pathway involving tRNA modification.

1.5.5 Sequence Conservation among AlkB Homologues

When aligning the protein sequences of AlkB homologues there are five completely conserved positions: an HXD motif (residues 462-464); a single H (residue 557); and a RXXXXXR motif (residues 610-616) (figure 1.5). The latter, which can also be written RhphphR, is consistent with a β-strand structure where the h and p represent hydrophopic and polar residues, respectively. One side of the strand faces a polar substrate binding cleft and the other faces the hydrophobic core of the protein. The H and D residues of the HXD motif have been associated with the co-ordination of the Fe^{2+} iron, whereas the first R most likely binds 2-oxoglutarate. It is also very likely that the last R may be involved in AlkB-specific substrate binding (Drabløs et al., 2007).

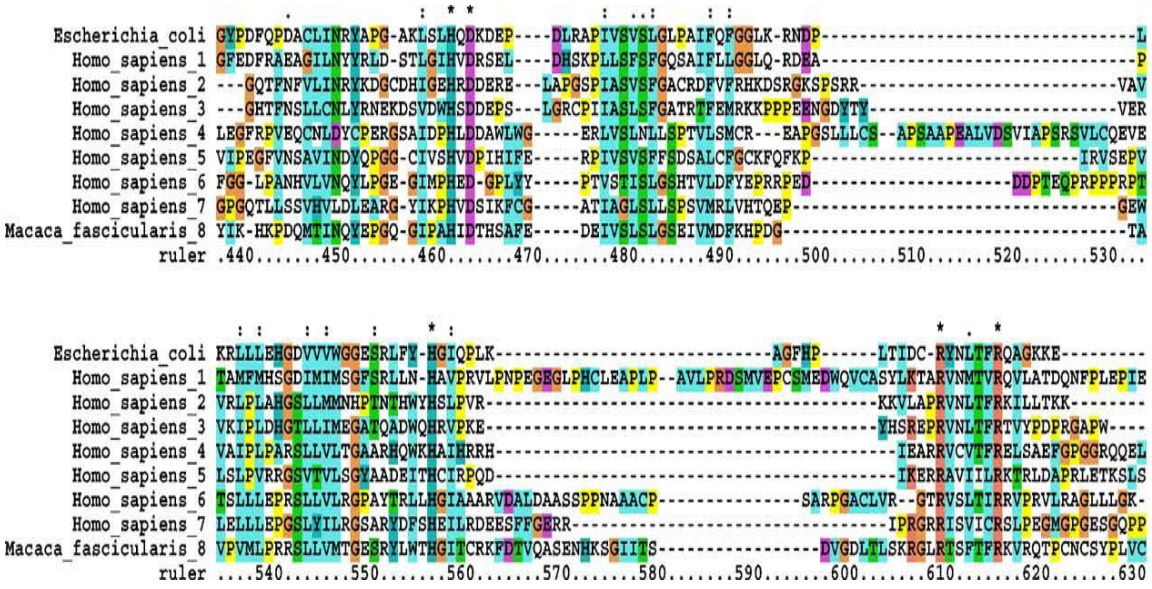


Figure.1.5: Multiple sequence alignment of *E. coli* AlkB and human AlkB homologues. ALKBH8 is represented by the *Macaca fascicularis* sequence, which is 98% identical to the corresponding full-length human protein sequence. (*) indicate positions of the conserved motifs (Drabløs et al., 2007).

1.5.6 Phylogenetic Relationship among AlkB Homologues

Another extensive bioinformatics mapping of AlkB proteins showed that AlkB homologues can be classified into subfamilies based on phylogenetic properties (figure 1.6). Phylogenetics is the study of evolutionary relationship among species and it can be used to study homology and conservation amongst species by aligning DNA and/or protein sequences. When constructing a phylogenetic tree based on an alignment of non-viral AlkB sequences, the lineages can be divided into two sub-trees: one bacterial/eukaryotic lineage (indicated by blue line) including ALKBH1-3 and *E. coli* AlkB, and one eukaryotic lineage (indicated by yellow line) including ALKBH4-8. The former sub-tree can be further divided into two branches with *E. coli* AlkB and ALKBH1 in one branch (indicated by red line), and ALKBH2 and 3 in the other (indicated by purple line) (Drabløs et al., 2007).

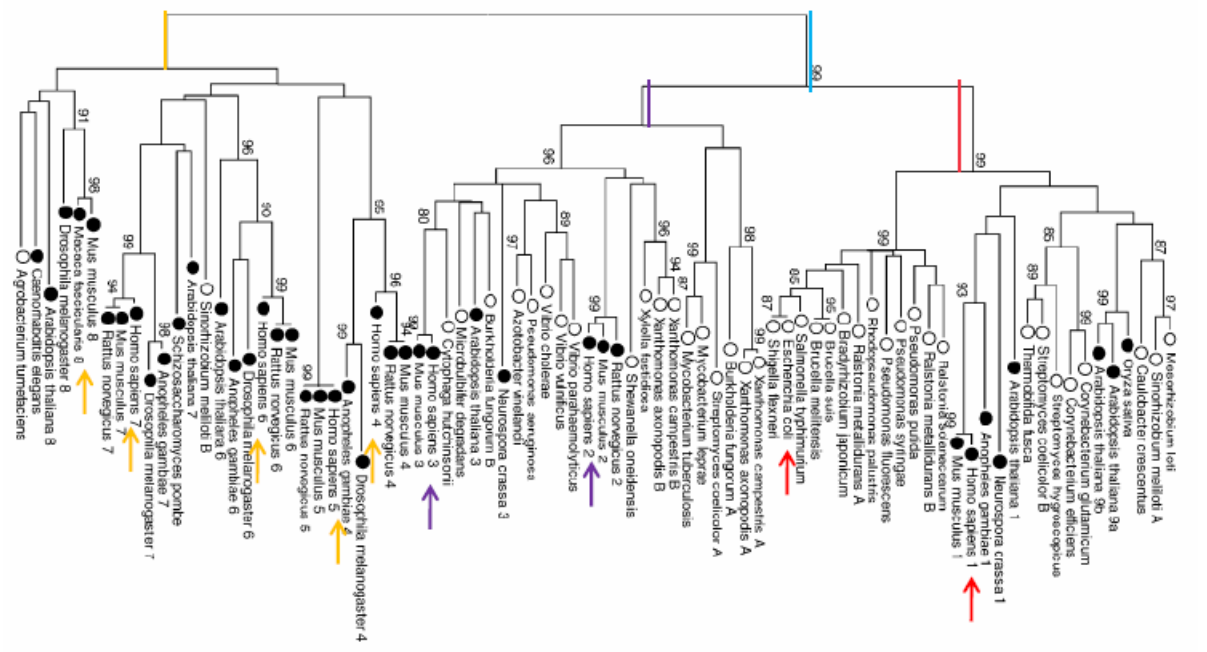


Figure 1.6: Phylogenetic tree of AlkB homologues based on conserved regions found by ClustalX alignment of AlkB sequence. ClustalX is a multiple sequence alignment tool. The phylogeny was computed with Mega2: a molecular evolutionary genetics analysis software. Bacteria and eukaryotes are indicated with open and filled circles, respectively. Numbers on branches represent confidence probability (%), whereas the letters (A, B) or numbers (1, 2, 3, etc.) indicate species with more than one *ALKBH*-type gene. Arrows point out ALKBH1-8 plus *E. coli* AlkB. ALKBH8 is represented by the *Macaca fascicularis* sequence, which is 98% identical to the corresponding full-length human protein sequence (Drabløs et al., 2007).

Information that can be drawn from such a tree is that ALKBH1-3 are of relatively ancient origin being placed in the bacterial/eukaryotic sub-tree, whereas ALKBH4-8 seem to have a more recent origin indicating specific requirement of eukaryotic organisms (Drabløs et al., 2007).

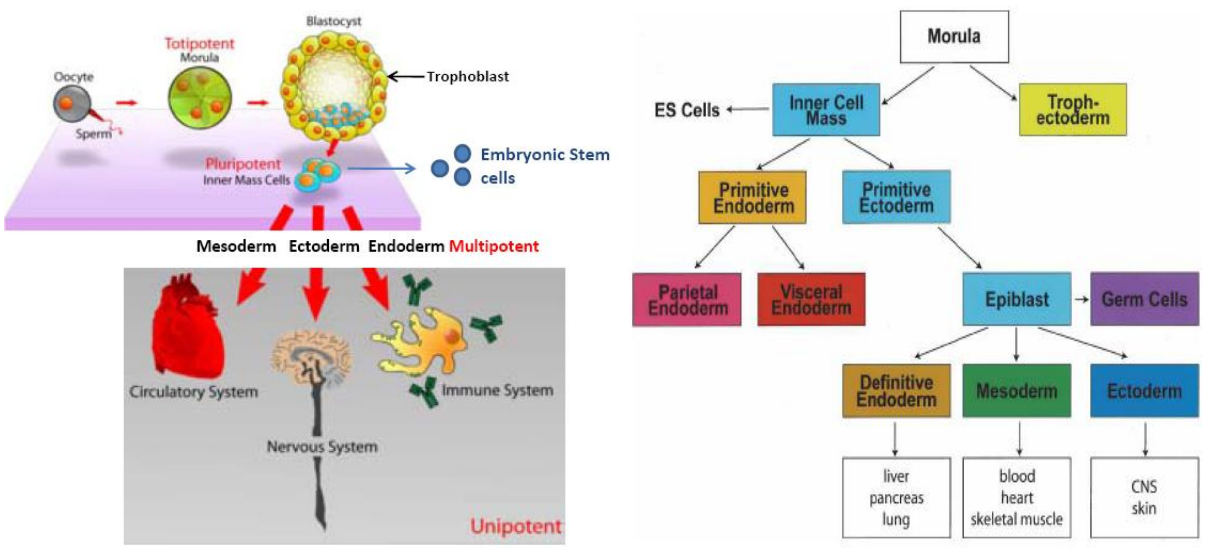
1.6 Embryonic Stem Cells

Embryonic stem (ES) cells are cells that have potential of self-renewal and the ability to develop into any differentiated cell type (Smith, 2001). This is reflected when a ES cell divides – each new copy has the potential to either remain a ES cell or differentiate into another type of cell such as a muscle cell, a red blood cells or a brain cells (Watt and Driskell, 2010).

Stem cells are classified according to their potency (figure 1.7 A). During mammalian fertilization, the zygote and sperm fuse to one cell, the zygote. This single cell is totipotent and has the potential to give rise to both the embryo and the placenta (Ratajczak et al., 2007).

Totipotent stem cells therefore have the ability to give rise to a whole organism. When the zygote divides it forms blastomeres followed by the morula and finally the blastocyst (figure 1.7 B). A fully developed blastocyst consists of an outer layer of trophoblast cells (which gives rise to the placenta) and the inner cell mass (ICM). The cells of the ICM are defined as pluripotent and can give rise to all three germ layers of the developing embryo (Pan et al., 2002; Ratajczak et al., 2007). As the blastocyst develops, the ICM segregates into two lineages: the epiblast and primitive endoderm (PE). The epiblast will give rise to the embryo, whereas the PE will give rise to the yolk sac (Yamanaka et al., 2006).

While pluripotent stem cells can give rise to all the cells in the body, multipotent stem cells, also called somatic stem cells, are capable of self-renewal and have the potential to differentiate into multiple types of cells although within a specific germ layer (Friel et al., 2005; Ratajczak et al., 2007). Multipotent stem cells further give rise to monopotent stem cells (also referred to as unipotent stem cells) that can only differentiate into cells of a specific lineage. For instance multipotent mesodermal stem cells give rise to monopotent skeletal muscle, heart and endothelial cells, while multipotent endodermal stem cells give rise to monopotent liver, pancreas, and gut epithelial cells, and multipotent ectodermal stem cells give rise to monopotent brain cells and nerves, as well as eye, epidermal, and skin tissues (Ratajczak et al., 2007).



A *B*

Figure 1.7: (A) Stem cell pluripotency and (B) embryonic development. Figure A is adapted from Wikipedia (2010b) and figure B is taken from Keller (2005).

1.7 ES Cell Research

In 1981 the first ES cells from early mouse embryos could be harvested and cultured *in vitro* (Evans and Kaufman, 1981; Martin, 1981). Research on mouse ES cells led to the development of methods for harvesting human ES cells from human embryos in 1998 (Thomson et al., 1998). More precisely, ES cells are derived from the ICM of the blastocyst and are therefore classified as pluripotent stem cells. From the ICM the ES cells can be plated onto a feeder layer consisting of non-dividing embryonic fibroblast cells. These feeder cells release nutrients, which inhibit differentiation, into the medium. One of the factors was identified in the late 1980 as the IL-6 cytokine, leukemia inhibitory factor (LIF) (Smith et al., 1988; Williams et al., 1988). This discovery enabled proliferation of certain mouse ES cell lines without the need for feeder cells when recombinant LIF is added to the culture medium (Friel et al., 2005).

1.7.1 LIF Signaling Pathway

The receptor for LIF is a heteromeric complex consisting of glycoprotein 130 and the LIF receptor (LIFR, also referred to as LIFRβ) (Davis et al., 1993). When LIF binds to its receptor it results in a major signaling cascade, involving activation of the Janus kinase (JAK) tyrosine kinase family members leading to the activation of a family of transcriptions factors known as signal transducers and activators of transcription (STAT). First, tyrosine residues of both LIF receptors are phosphorylated by JAK. STAT1 and STAT3 are then recruited, which form active dimers upon phosphoryation by JAK. STAT1 and STAT3 dimers are able to translocate into the nucleus where they function as transcriptional factors (Okita and Yamanaka, 2006).

STAT3 can regulate mouse ES cell pluripotency by various mechanisms. It can maintain the expression of specific genes that are involved in pluripotency or inhibit signal transducers that normally will promote differentiation of ES cells. Such a signal transducer is the extracellular signal-regulated kinase (ERK) and STAT3 can prevent ERK from being activated. STAT3 can also maintain the transcriptional level of *Myc*, which is key to mouse ES cell self-renewal at stable transcription levels (YuXiao et al., 2007).

Although LIF is very important for maintaining self-renewal of mouse ES cells through the activation of STAT3, it is only sufficient when these cells are grown in media containing fetal bovine serum. Furthermore, LIF cannot promote self-renewal of human ES cells and it seems

like human ES cells maintain pluripotency in a LIF/STAT3 independent manner (Okita and Yamanaka, 2006).

Other factors have also shown to support self-renewal of mouse ES cell. One of them is bone morphogenetic protein 4 (BMP4), which cooperates with LIF to maintain self-renewal and pluripotency of mouse ES cells. This is achieved by activating members of the inhibitor of differentiation (id) gene family (Ying et al., 2003). In human ES cells, however, BMPs can cause rapid differentiation (Pan and Thomson, 2007). BMP4, for instance, causes differentiation into mesoderm and ectoderm while BMP2 promotes extra-embryonic endoderm differentiation (Okita and Yamanaka, 2006). BMPs can also cause down-regulation of NANOG and OCT4 in human ES cells, which are core regulators of both mouse and human ES cell pluripotency (YuXiao et al., 2007).

1.7.2 Core Regulators of ES Cell Pluripotency

The transcription factors OCT4, NANOG, and SOX2 play essential roles in the maintenance of pluripotency in both human and mouse ES cells (Niwa et al., 2000; Chambers et al., 2003; Masui et al., 2007). These core factors contributes to the hallmark characteristics of ES cells by activating target genes that encodes pluripotency and self-renewal mechanisms and repress signaling pathways that promote differentiation (Orkin, 2005).

1.7.2.1 OCT4

OCT4 is a transcription factor of the POU family and is also known as OCT3 or OCT3/4 (Friel et al, 2005). Studies have shown that OCT4 is almost exclusively expressed in ES cells and plays a key role in the development of the embryo. OCT4 is initially expressed in all blastomeres, and as development of the embryo progresses OCT4 expression becomes restricted to the ICM. When the embryo matures OCT4 expression is exclusive to the developing germ cells (Pan et al., 2002).

Null mutation of the *oct4* gene results in early embryonic lethality (Nichols et al., 1998). In 2000, Niwa et al. showed that small changes in the level of OCT4 promote differentiation of ES cell. Even a twofold increase of the OCT4 protein induces differentiation toward primitive endoderm and mesoderm, whereas a 50% decrease causes differentiation into trophectoderm (Niwa et al., 2000). Studies on both mouse and human ES cells have indicated that OCT4 maintains ES cell pluripotency by working together with other transcriptional co-factors.

OCT4 directly activates or suppresses the expression of these co-factors and cooperatively maintains pluripotency (YuXiao et al., 2007). The Sry-related factor SOX2 is such a co-factor that forms a complex with OCT4 that stimulates the induction of most pluripotency-associated genes, including *nanog* (Masui et al., 2007).

1.7.2.2 SOX2

Sox2 belongs to the High Mobility Group (HMG) protein superfamily that all possess a HMG box DNA-binding domain (Gubbay et al., 1990). It is also known as a Sry-related factor as the SOX subfamily is defined by its relationship to the testes determining factor Sry (Bowles et al., 2000). Like OCT4 and NANOG, SOX2 is required for maintaining pluripotency and self-renewal in ES cells, and SOX2 and OCT4 are known to bind DNA cooperatively (Ambrosetti et al., 1997; Ambrosetti et al., 2000). Being a co-factor of OCT4, *sox2* mutants might be expected to behave as *oct4* mutants. This is however not completely true as *sox2* mutants are lethal at a slightly later stage. Mutants also exhibit defects in the epiblast also at a later stage than *oct4* mutants (Avilion et al., 2003). When SOX2 levels are reduced in ES cells the cells will start to form trophoblast-like cells (Li et al., 2007; Masui et al., 2007). This is also seen in ES cells with reduced OCT4 levels (Niwa et al., 2000). This supports the idea of SOX2 and OCT4 interacting to repress trophoblast differentiation. In addition, over-expression of SOX2 also leads to ES cell differentiation as seen with elevated OCT4 levels (Kopp et al., 2008).

When studying the expression of many OCT4/SOX2 target genes, *sox2* depletion did not affect the expression of these genes to a major extent (Masui et al., 2007). This can be explained by the expression of additional SOX family protein members such as SOX4, SOX11 and SOX15. The essential function of SOX2 for ES cells appears to be maintaining the correct level of OCT4 by regulating multiple transcription factors that affects OCT4 expression (Masui et al., 2007; Chambers and Tomlinson, 2009).

1.7.2.3 NANOG

NANOG is the most recently described core regulator of ES cell pluripotency. It was initially described by Wang et al. (2003) as an ENK gene (early embryonic specific NK) that was specifically expressed in ES cells. The gene was later re-cloned and re-named as *nanog* by two independent groups that analyzed its function. Chambers et al. (2003) and Mitsui et al. (2003) described NANOG as a divergent homeobox transcription factor that promotes ES cell self-renewal, pluripotency and epiblast formation.

During embryonic development NANOG is first expressed in the interior cells of the morula, and then confined to the ICM while disappearing in the trophectoderm at the blastocyst stage. As the blastocyst develops expression of NANOG is restricted to the epiblast and excluded from the PE. mRNA expression studies of NANOG have shown increased expression in pluripotent cell lines such as ES, embryonic germ and embryonic carcinoma cells. When these cells differentiate NANOG expression is down-regulated (Chambers et al., 2003; Mitsui et al., 2003).

In culture of mouse ES cells the most important role of NANOG is to maintain pluripotency in the absence of LIF (Mitsui et al., 2003). In human ES cells high levels of NANOG enable growth in feeder-free conditions (Darr et al., 2006). Down-regulation of Nanog induces differentiation to extra-embryonic lineages in both mouse and human ES cells (Hyslop et al., 2005). A central role of NANOG therefore seems to be conserved in both mouse and human ES cells (Pan and Thomson, 2007).

As mentioned above, LIF is not needed to maintain pluripotency of mouse ES cells when NANOG is over-expressed. Furthermore, the level of phosphorylated STAT3 does not change when NANOG is over-expressed and elevated STAT3 signaling does not affect NANOG expression (Chambers et al., 2003). This suggests that NANOG is not a direct transcriptional target of STAT3 and NANOG does not seem to regulate STAT3 either (Pan and Thomson, 2007). However, the presence of LIF and over-expression of Nanog increase cell proliferation (Chambers et al., 2003). This suggests that although NANOG maintains self-renewal in a STAT3-independent manner, together they might have additive effects (Friel et al., 2005). More studies, however, are needed to confirm a direct link between NANOG and LIF-STAT3 signaling (Pan and Thomson, 2007).

1.8 Transcriptional Networks

Genome-wide studies have shown that OCT4, SOX2 and/or NANOG bind to promoters of a large number of both transcribed and inactive genes in ES cells (Boyer et al., 2005; Loh et al., 2006). This was achieved using a CHIP-Chip assay that combines chromatin immunoprecipitation coupled with DNA microarrays. One of most exciting findings was the high frequency of ES cells maintenance genes that are co-occupied by at least two of the three factors, including the OCT4, NANOG, and SOX2 themselves forming interconnected regulatory loops (figure 1.18) (Boyer et al., 2005; Loh et al., 2006; Zhou et al., 2007).

This was confirmed in consequent study by Pan et al. (2006) that described a negative feedback loop formed by NANOG, OCT4 and another pluripotent factor, FOXD3. In this and other studies, it has been shown that all three transcription factors can act as activators or repressors, depending on the circumstances (Liu and Labosky, 2008). FOXD3 activates both OCT4 and NANOG in human ES cells (Pan et al., 2006). At a steady state level of OCT4, OCT4 also activates NANOG by binding to the *nanog* promoter. However, when OCT4 expression is elevated, OCT4 repress both NANOG and its own promoter, thus achieving a negative feedback regulation to limit its own expression when in excess. In addition, NANOG activates OCT4 expression, and this complex regulation keeps the level of OCT4 in a very narrow window to maintain pluripotency in ES cells (Pan and Thomson, 2007).

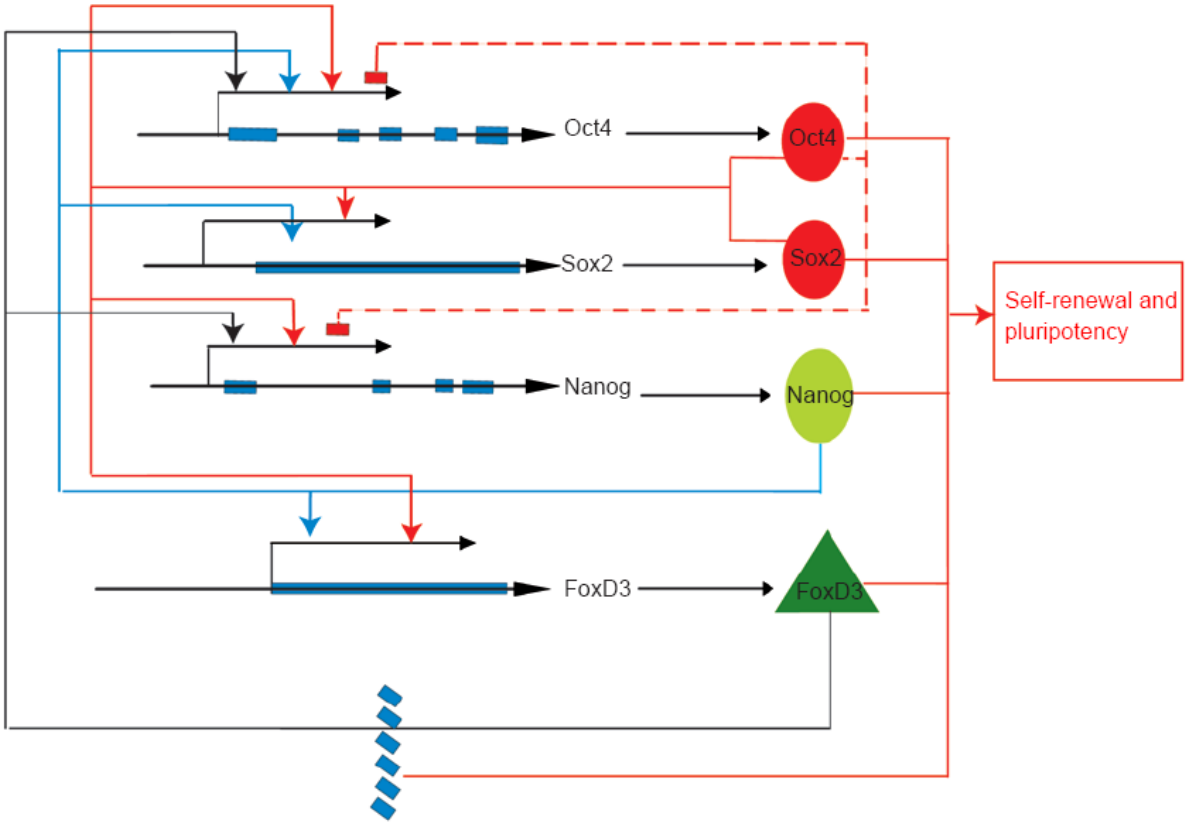


Figure 1.18: Regulation of ES cell self-renewal and pluripotency by a regulatory network of key transcriptional factors. OCT4, NANOG, SOX2 and FOXD3 being to each other's promoters to support or limit each other's expression. Arrows connected to factors by solid lines indicate positive regulation whereas broken lines linking to OCT4 indicate negative regulation (Pan and Thomson, 2007)

Furthermore, Wang et al. (2006) identified proteins that physically associate with NANOG in mouse ES cells including OCT4. These NANOG-associated proteins were used to screen for further partner proteins. They were then able to describe a protein interaction network in maintaining ES cell pluripotency. All of these studies indicate that the key factors of ES cell

pluripotency always work together rather than individually to control target genes as well as each other (Pan and Thomson, 2007).

1.9 Differentiation

Given the pluripotent nature of ES cells, they have the capacity to differentiate into every cell type in the body. ES cells are also capable of forming differentiated cell types in culture. Development of appropriate culture conditions and protocols has provided means of generating a broad spectrum of lineages *in vitro*. This is important for studying factors that regulate the earliest stages of lineage induction and specification (Keller, 2005).

When removing factors that maintain ES cells in their undifferentiated state, they will start to differentiate spontaneously. Using appropriate culture conditions, differentiation towards specialized cells of the three embryonic germ layers: mesoderm, endoderm, and ectoderm can be achieved. Three general approaches can be used to initiate ES cell differentiation. When ES cells are allowed to aggregate in culture they will form three-dimensional colonies known as embryonic bodies (EBs). When ES cells are cultured directly onto stromal (connective tissue) cells, differentiation will occur when in contact with these cells. ES cells can also differentiate when grown in a monolayer on extracellular matrix proteins (Keller, 2005).

EBs are formed when individual cell self-assemble via cell-cell adhesion receptors (Dang et al., 2004). EBs can be made by simply adding a suspension of ES cells in special dishes that inhibits cell adhesion to the dish. In such conditions the cells spontaneously aggregate via cell-cell adhesions. Following cell aggregation, differentiation can be seen by the formation of a PE layer on the exterior surface of the EBs (Maurer et al., 2008). The specific factors that stimulate PE differentiation remain unknown, but PE differentiation seems to be dependent on fibroblast growth factor signaling. As development of EBs progresses differentiated cell phenotypes of all three germ layers begin to arise (Keller, 1995). Expression of phenotypic markers of endoderm (such as GATA4/6), mesoderm (such as BRACHYURY), and ectoderm (such as SOX1) have been revealed in global DNA microarray analysis (Itskovitz-Eldor et al., 2000). This demonstrates that EBs can generate cells from all three germ layers (Bratt-Leal et al., 2009).

Differentiation towards ectoderm lineages have been well established in mouse ES cells. Several protocols have been developed to promote neuroectoderm differentiation (Bain et al., 1995; Okabe et al., 1996; Kawasaki et al., 2000; Tropepe et al., 2001) and include treatment

with retinoic acid (RA) (Guan et al., 2001; Keller, 2005). RA acid is a derivative of vitamin A and used *in vitro* to induce neuronal differentiation of ES cells. High concentrations of RA results in efficient neuronal differentiation and is seen together with expression of tissue-specific genes, proteins, ion channels and receptors in a developmentally controlled manner (reviewed by Guan et al., 2001).

The mechanisms in which RA induces neural differentiation of ES cells *in vitro* are not fully understood, however a complex signaling model has been proposed for the molecular mechanisms of RA action during embryogenesis *in vivo* (reviewed by Gambon 1996). In this model RA binds to cellular RA-binding proteins (CRABP) that interact with the nuclear RA and retinoid X receptors (RAR, RXR). Homodimers of RXR and heterodimers of RAR/RXR bind specific target genes known as RA response elements (RARE). It is plausible that RA act in the same way *in vitro*. This is supported by significant up-regulation of RA receptor-α (RARα) and RA receptor-β (RARβ) mRNA during RA-induced differentiation in mouse ES cell. However a rapid down-regulation of RA receptor-γ (RARγ) and retinoid X receptor-γ (RXRγ) mRNA is also observed (Jonk et al. 1992; Yokota and Ohkubo 1996). This suggests a role for RARα, -β, -γ and RXRγ during neuroectodermal differentiation. Neural-specific genes are also induced by RA (reviewed by Guan et al., 2001) whereas the expression of mesodermal genes such as *brachyury* is down-regulated (Bain et al., 1996). This suggests that RA differentiation also represses mesodermal differentiation. Although these findings provide clues about how RA induces neuroectodermal differentiation, the exact mechanisms by which RA does so need to be further analyzed (Guan et al., 2001).

1.10 Aim of Study

The biochemical function of *E. coli* AlkB homologue 1 (ALKBH1) is under debate as several groups have been unable to demonstrate a DNA demethylation activity similar to that of AlkB, ALKBH2 and ALKBH3 (Duncan et al., 2002; Aas et al., 2003). A role in epigenetic regulation have been suggested as the mechanism used by JmjC-domain containing histone demethylases is identical to that of AlkB, i.e. 2-OG and Fe(II) dependent oxidative demethylation, where 2-OG is decarboxylated to succinate (Trewick et al., 2005; Tsukada et al., 2006; Whetstine et al., 2006). The first aim of this study was therefore to purify human ALKBH1 and characterize the biochemical function.

The second aim of this study was to elucidate the putative role of ALKBH1 in maintaining self-renewal and pluripotency in ES cells. Two different protocols for inducing ES cell differentiation were also tested to evaluate expression of Alkbh1 along with pluripotency markers and specific markers of the three different germ layers. NANOG and OCT4 binds the *ALKBH1* promoter in ES cells (Boyer et al., 2005) implying that ALKBH1 might be important in ES cells. This will be studied in further details by DNA:protein and protein-protein interactions with ALKBH1 and core regulators of ES cell pluripotency.

2 Materials

Standard chemicals that were used were produced by Sigma-Aldrich®, VWR International, Invitrogen and Fluka Analytical.

Milli-Q H_2O (mQ- H_2O) is purified and deionized water by Millipore Milli-Q Integral system.

Materials that have been produced at the Laboratory for genome repair and regulation or at the Laboratory of embryonic stem cell research at Rikshospitalet are referred to as Rikshospitalet only.

2.1 Bacterial strains

Strain	Genotype	Produced by
BL21-CodonPlus® (DE3)-RIL	E. coli B F^- ompT hsdS(r_B^- m_B^-) dcm^+ Tet^r gal λ(DE3) endA Hte [argU ileY leuW Cam^r]	Stratagene

2.2 Plasmids

Plasmid	Produced by
pET28a-ALKBH1	Rikshospitalet

2.3 Glycerol stocks

Stock	Produced by
pTYB12-ALKBH1 BL21 (DE3)- RIL	Rikshospitalet
pTYB12-ALKBH1 DH5 α	Rikshospitalet

2.4 Proteins and Enzymes

2.4.1 Proteins used in reaction and interaction studies

Protein	Description	Produced by
S. pombe AlkB homologue 2 ALKBH2	Recombinant *Schizosaccharomyces pombe* produced in *E.coli*	Rikshospitalet
AlkB	*E. coli* AlkB	Rikshospitalet
NANOG	Recombinant human produced in *E. coli*	Peprotech
OCT4/POU5F1	Recombinant human produced in *E. coli*	Nordic BioSite
SOX2	Recombinant human produced in *E. coli*	Peprotech
Histone H1	from calf thymus	Roche Diagnostics
Histone H2A	Recombinant human produced in *E. coli*	New England Biolabs
Histone H2A	from calf thymus	Roche Diagnostics
Histone H2B	Recombinant human produced in *E. coli*	New England Biolabs
Histone H3	from calf thymus	Roche Diagnostics
Histone H4	Recombinant human produced in *E. coli*	New England Biolabs
Bovine Serum Albumin (BSA)	Molecular biology grade, protease and nuclease-free, faction V starting material	New England Biolabs
Lysozyme	Heat inactivated	Rikshospitalet

2.4.2 Enzymes

Enzyme	Buffer/Reagents	Produced by
Pfu Turbo® DNA Polymerase	10x Cloned Pfu Reaction Buffer	Stratagene
AmpliTaq Gold™ DNA polymerase	10x PCR Buffer $MgCl_2$ Solution	Applied Biosystems
T4 Polynucleotide Kinase	10x T4 PNK Reaction Buffer A	Fermentas

2.5 Antibodies

Antibody	Description	Produced by
ALKBH1 antibody (ab18525)	Polyclonal primary antibody raised in rabbit	Abcam
Anti-rabbit IgG (A3812)	Secondary antibody conjugated with Alkaline phosphatase, produced in goat	Sigma-Aldrich®

2.6 Isotopes

Isotope	Specific activity	Produced by
Adenosine 5'-triphosphate, [γ-^{32}P]	30 Ci/mmol	PerkinElmer
2-Ketoglutaric acid [5-14C]	50 mCi/mmol	Moravek Biochemicals

2.7 Alkaline phosphatase substrates

Substrate	Buffer	Produced by
Western Blue® Stabilized Substrate for Alkaline phosphatase	-	Promega
PPNP (*p*-nitrophenyl phosphate alkaline phosphatase) Tablets	Diethanolamine Substrate Buffer 5x concentrate	Thermo Scientific

2.8 Molecular weight standards

2.8.1 Protein standards

Standard	Produced by
ClearPAGE™ 2-Color SDS Marker	C.B.S. Scientific
NuPAGE® SeeBlue Plus 2 Prestained Standard	Invitrogen

2.8.2 DNA standards

Standard	Produced by
GeneRuler™ DNA Ladder Mix	Fermentas

2.9 Protein dye reagent

Reagent	Produced by
Dye Reagent Concentrate 5x	Bio-Rad

2.10 DNA loading dye

Dye	Produced by
6x DNA Loading Dye	Fermentas

2.11 Liquid Scintillation Cocktail (LSC)

LSC	Produced by
Ultima Gold™ MW LSC	PerkinElmer

2.12 Primers

All primers are supplied by Eurofins MWG Operon. "F" or "forw" (forward) in the 5'→3' direction and "R" or "rev" (reverse) also in the 5'→3' direction on opposite strand

Primer	Used in	Sequence (5'→3')
GAforw781	PCR	GCGCCGCATCCCTGTCAGTT
GArev781	PCR	CTCGACGCCGCAGTCCAAGG
GAforw434	PCR	GCGGGGCAAAGCGATGGAGA
GAforw353	PCR	GGGCCGCCGAGAAGTCGATG
GAforwoct4	PCR/sequencing	CTCTGTACCTTTTGGGCACC
GArevoct4	PCR/sequencing	TCTCGAGCCGAGGTCGCGGC
GAforwnanog	PCR/sequencing	TCTGCGGTCCCGGGCCGGCT
GArevnanog	PCR/sequencing	TTCTACGCAGCGCCGCAGCA
GAoctF	RT-PCR	AACCTTCAGGAGATATGCAAATCG
GAoctR	PT-PCR	TTCTCAATGCTAGTTCGCTTTCTCT
GAnanogF	RT-PCR	TCAGAAGGGCTCAGCACCA
GAnanogR	RT-PCR	GCGTTCACCAGATAGCCCTG
GAsox2F	RT-PCR	CTGCAGTACAACTCCATGACCAG
GAsox2R	RT-PCR	GGACTTGACCACAGAGCCCAT
ABH1_F	RT-PCR	TCAGCTTTGGACAGTCTGCCATCT
ABH1_R	RT-PCR	ACCCGACATTACCATGATGTCACC
GAPDH_F	RT-PCR	TCGTCCCGTAGACAAAATGGT
GAPDH_R	RT-PCR	CGCCCAATACGGCCAAA
GAcalbF	RT-PCR	GCTTCTATCTGGCGGAAGG
GAcalbR	RT-PCR	TGTCATCTGGCTACCTTCCC
GAfgf5F	RT-PCR	TTGCGACCCAGGAGCTTAAT
GAfgf5R	RT-PCR	CTACGCCTCTTTATTGCAGC
GAsox18F	RT-PCR	AACAAAATCCGGATCTGCAC
GAsox18R	RT-PCR	CGAGGCCGGTACTTGTAGTT
GAbracF	RT-PCR	CCAAGGACAGAGAGACGGCT
GAbracR	RT-PCR	AGTAGGCATGTTCCAAGGGC
GAgata6F	RT-PCR	CCCACTTCTGTGTTCCCAATTG
GAgata6R	RT-PCR	TTGGTCACGTGGTACAGGCG
GAgata4F	RT-PCR	GCTCCTTCAGGCAGTGAGAG
GAgata4R	RT-PCR	CTGTGCCCGTAGTGAGATGA

2.13 Chromatography materials

2.13.1 Affinity chromatography matrix

Matrix	Produced by
HIS-Select® Cobalt Affinity Gel	Sigma-Aldrich®
Chitin Beads	New England BioLabs

2.13.2 FPLC Columns

Column	Produced by
HiTrap™ SP HP 1 ml column	GE Healthcare
RESOURCE™ S 1 ml column	GE Healthcare
RESOURCE™ Q 1 ml column	GE Healthcare

2.13.3 Glass chromatography columns

Column	Produced by
Glass Econo-Column Chromatography column	Bio-Rad

2.14 Dialysis columns

Column	Produced by
4 ml Amicon® Ultra-4 Centrifugal Filter Device	Millipore
Vivaspin 500 µl Centrifugal Filter Tubes	Sigma-Aldrich®

2.15 Gel electrophoresis material

Product	Produced by
Horizon® 58 Horizontal Gel Electrophoresis System	Biometra
HY13 Midi Horizontal Gel Electrophoresis Unit	Scie-Plas
XCell *SureLock*™ Mini-Cell	Invitrogen
Gel Cassettes, 1.0 mm	Invitrogen
30% Acrylamide/Bis Solution 29:1	Bio-Rad

2.16 Embryonic stem cell material

2.16.1 Mouse Embryonic Cells

Cells	Produced by
CF-1 MEF Mouse embryonic fibroblast cells, Irradiated	Globalstem
Wild-type mouse embryonic stem cell	Rikshospitalet

2.16.2 Media components

Component	Produced by
GIBCO® Dulbecco's Modified Eagle Medium (DMEM)	Invitrogen
GIBCO® Knockout™ DMEM, Optimized for ES cells 1x	Invitrogen
GIBCO® MEM NEAA 100x	Invitrogen
GIBCO® GlutaMAX™-I 100x	Invitrogen
BioWhittaker® FBS, ES Qualified	LONZO
BioWhittaker® Pen-Strep, 10.000 U Penicillin ml^{-1}, 10.000 U Streptromycin ml^{-1}	LONZO
β-Mercaptoethanol, >= 99%	Sigma-Aldrich®
ESGRO® LIF, 10^7 U ml^{-1}	Millipore
Retinoic Acid	Sigma Aldrich®

2.16.3 Culture plate/dish

Plate/Dish	Produced by
FALCON® MULTIWELL™ Tissue Culture Plate, 6 well	Becton Dickinson
FALCON® Tissue Culture Dish	Becton Dickinson
CORNING® Ultra-Low Culture Dish	Corning

2.16.4 Other related products

Product	Produced by
ESGRO® COMPLETE™ 0.1% Gelatin	Millipore
GIBCO® DPBS Dulbecco's Phosphate Buffered Saline (DPBS) 1x	Invitrogen
GIBCO® 0.05% Trypsin-EDTA	Invitrogen

2.17 Kits

SDS-PAGE Kits	Produced by
ClearPAGE™ Precast Gels & Accessories	C.B.S. Scientific
ClearPAGE™ SDS 12% 12 wells Gel	
ClearPAGE™ SDS Run buffer	
NuPAGE® Novex Bis-Tris Gels & Accessories	Invitrogen
NuPAGE® 10% 1.0 mm*10 well Gel	
NuPAGE® 12% 1.0 mm*10 well Gel	
NuPAGE® 10% 1.0 mm*15 well Gel	
NuPAGE® LDS Sample Buffer	
NuPAGE® MOPS SDS Running Buffer	

Real-Time PCR Kit	Produced by
Power SYBR® Green PCR Master Mix	Applied Biosystems
MicroA® Fast 96-Well Reaction Plate	Applied Biosystems
MicroAmp™ Optical Adhesive Film	Applied Biosystems

High-Capacity cDNA Reverse Transcription (RT) Kit	Produced by
MultiScribe™ Reverse Transcriptase	Applied Biosystems
10x RT Buffer	Applied Biosystems
10x RT Random Primers	Applied Biosystems
25x dNTPs	Applied Biosystems

Other Kits	Produced by
miRNeasy® Mini Kit	QIAGEN
QIAquick® Nucleotide Removal Kit	QIAGEN
QIAquick® PCR Purification Kit	QIAGEN
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems

2.18 Other products

Product	Produced by
CovaLink plate	Nunc
General-purpose (GP) Storage Phosphor Screen	Amersham Biosciences
Protran® Nitrocellulose Transfer Membrane	Whatman

3 Methods

3.1 Bacteria Related Methods

3.1.1 Transformation of Expression Vectors into BL21-CodonPlus® Competent Cells

Introduction of plasmid DNA into competent cells is called transformation. After transformation bacteria cells are plated onto selective lysogeny broth (LB) plates where only cell lines carrying the antibiotic resistance genes on their introduced plasmid DNA will grow. The Stratagene protocol (pp. 9 – 10) was used for transformation of pET28a-ALKBH1 plasmid DNA into BL21-CodonPlus® (DE3)-RIL competent host strains (Stratagene, 2005). Transformed cells were spread onto LB kanamycin plates (appendix 2) as the pET28a vector carries the *kan* gene that infer resistance kanamycin. Transformed colonies appeared following incubation ON at 37ºC and were used in ALKBH1 protein production and purification

3.2 Protein Related Methods

3.2.1 Protein Production

The human *ALKBH1* gene had previously been cloned into two separate expression vectors, namely pTYB12 and pET28a. Production of the ALKBH1 recombinant protein was performed by over-expressing it in a competent *E. coli* host strains – BL21-CodonPlus® (DE3)-RIL.

Both pET and pTYB expression vectors use a T7 promoter-driven system to achieve high levels of expression of a target gene (figure 3.1). This means that expression is under control of the bacteriophage T7 promoter. This promoter is induced by providing a source of T7 RNA polymerase in the host strain. BL21 (DE3)-RIL host strains carry the T7 RNA polymerase gene (*λDE3*) and are referred to as λDE3 lysogens. In λDE3 lysogens the T7 RNA polymerase gene is under the control of the *lacUV5* promoter and expression is induced by the addition of iso-propyl-β-D-thiogalactoside (IPTG). When IPTG is added to the growing *E. coli* culture it induces T7 RNA polymerase production, which then transcribes a target gene in the expression vector (Novagen, 2006; NEB, 2006).

However, there is some expression of T7 RNA polymerase from the *lacUV5* promoter in the absence of IPTG, thus a basal expression of the target protein will occur. This may interfere in

normal functioning of the *E. coli* host strain and may also be toxic to the cell. pET and pTYB vectors therefore contain a T7*lac* promoter in order to control basal expression. Both plasmids contain a *lac* operator sequence downstream of the T7*lac* promoter as well as the *lacI* gene encoding the lac repressor. In λDE3 lysogens the lac repressor acts on the *lacUV5* promoter in the host chromosome in order to repress transcription of the T7 RNA polymerase gene by the host polymerase. The lac repressor also acts on the T7*lac* promoter in the vector to block transcription of the target gene by any T7 RNA polymerase that might be present (Novagen, 2006).

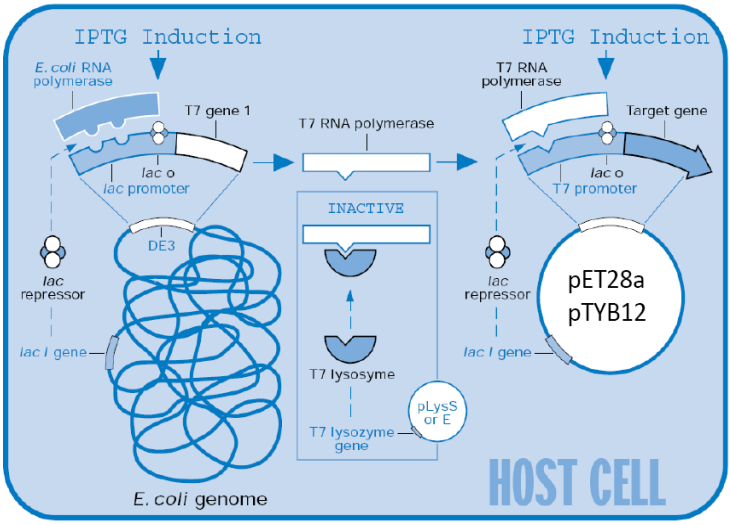


Figure 3.1: Schematic diagram of host and vector elements that control T7 RNA polymerase levels and subsequent transcription of a target gene in pET28a and pTYB12 vectors. The illustration in the middle of T7 lysosyme is only applicable when using pLysS and pLysE host strains. Figure is adapted from Mierendorf et al., (1994).

The pTYB12 vector contains a cleavable intein tag which can be fused to NH_2- (N) terminus of a target protein. The pET28a vector contains a histidine fusion tag (His•Tag) that can either be fused to the N- or Carboxyl- (C) terminal of the target protein. Both expression vectors contain antibiotic resistance genes to allow for selection of transformed host cell colonies. pET28a carry the *kan* gene and pTYB12 carry *amp* gene that infer resistance to ampicillin and kanamycin, respectively (Novagen, 2006; NEB, 2006).

BL21-CodonPlus® competent cells contain extra copies tRNA genes. Production of a recombinant protein in *E. coli* can be limited by the absence of certain tRNAs that are usually found in the organisms from which the recombinant proteins are derived. Over-expression of recombinant proteins can also exhaust existing tRNAs found in the host strain. This will stall protein translation. However, availability of tRNAs in BL21-CodonPlus® cells allow for high-level expression of recombinant genes. The BL2 (DE3)-RIL cells contain extra copies of the

argU, ileY, and *leuW* tRNA genes. These tRNA genes are carried by the pACYC-based, which also carries the gene for chloramphenicol resistance (Stratagene, 2005).

3.2.2 Protein Expression and Cell Lysis

1. LB ampicillin and kanamycin plates (appendix 2) were inoculated with pTYB12-ALKBH1 BL21-(DE3)-RIL and pET28a-ALKBH1 BL21 (DE3)-RIL glycerol stocks, respectively, and incubated over night (ON) at 37˚C. Alternatively, LB kanamycin plates with BL21-(DE3)-RIL colonies containing the pET28a-ALKBH1 plasmid (section 3.1.1) were used before a glycerol stock had been made.
2. One ON colony was transferred to 100 ml LB medium (appendix 2) containing the appropriate antibiotic (ampicillin or kanamycin to a final concentration of 100 µg ml^{-1} and chloramphenicol at final concentration of 34 µg ml^{-1}).
3. The culture was incubated ON at 37˚C with shaking at 225 revolutions per minute (rpm). Of remaining ON culture glycerol stocks were made by mixing 800 µl culture with 200 µl 60% glycerol. These were stored at 70˚C.
4. 1 liter (L) LB medium containing the appropriate antibiotic (same concentrations as above) was inoculated with 10 ml ON culture. 6, 12 or 24 L cultures were incubated at 37˚C at 180 rpm until OD_{600} reached 0.6 – 0.8.
5. IPTG (to a final concentration of 0.5 mM) was added to each L culture to allow for expression of ALKBH1.
6. The culture was further incubated at 37˚C for 1 hour or at 18˚C ON.
7. Cells were harvested by centrifugation at 5500 rpm for 20 min at 4˚C.
8. Cell pellet was resuspended in ice cold Cell Lysis Buffer (approximately 50 ml L^{-1} culture [appendix 1]).
9. Cells were disrupted by sonication on ice for 3x 30 seconds with 1 minute (min) intervals.
10. The clarified cell lysate was obtained by centrifugation for 30 min at 13000 rpm at 4˚C. The recombinant ALKBH1 protein was extracted and purified further from this cell lysate by affinity and ion exchange chromatography.

3.2.3 Affinity Chromatography

Affinity chromatography is based on a reversible interaction between a target protein and a specific ligand attached to a chromatographic matrix. A sample containing the target protein

is applied under conditions that favor specific binding to the ligand and unbound material is washed away. The bound target protein is then recovered by changing the conditions that favor desorption from the matrix (Amersham Biosciences, 2002).

3.2.3.1 Protein Purification with HIS-Select® Cobalt Affinity Gel

HIS-Select® Cobalt Affinity Gel is an Immobilized Metalion Affinity Chromatography (IMAC) product. The cobalt affinity gel contains agarose charged with cobalt (Co^{2+}) and was used here to purify recombinant histidine-tagged ALKBH1 encoded by the pET28a vector. After binding of histidine-tagged ALKBH1 to the Co^{2+} charged agarose gel, imidazole is used to elute ALKBH1 as imidazole has an extremely high affinity towards the Co^{2+} charged gel, thus competing with ALKBH1.

Protocol

1. 3 ml HIS-Select® Cobalt Affinity Gel was added to an Econo-column chromatography column and washed with several volumes of mQ-H_2O.
2. The column was equilibrated with 2 – 3x column volume with Washing Buffer 1 (10 mM imidazole [appendix 1]).
3. Cell lysate was added to the column by filtration and mixed with the gel by gentle head-over tail rotation at 4ºC for 1 hour.
4. Unbound proteins were eluted by gravity flow and a 1 ml sample was collected from flow-through.
5. The column was washed twice with 100 ml Washing Buffer 1. A 1 ml sample was collected – wash 1 sample.
6. The column was further washed with Washing Buffer 2 (30 mM imidazole [appendix 1]) prior to protein elution. A 1 ml sample was collected – wash 2 sample.
7. 10 ml of Elution Buffer (250 mM imidazole [appendix 1]) was added in three consecutive elutions by collection of eluate and reapplication to the column. Protein samples were collected in 1 ml fractions from the third elution.
8. Collected samples and protein fractions were analyzed by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE).

3.2.3.2 Affinity chromatography by the IMPACT™ system

The Intein Mediated Purification with an Affinity Chitin-binding Tag (IMPACT™) system was used here to purify recombinant intein-tagged ALKBH1 encoded by the pTYB12 vector. The intein tag has a chitin binding domain (CBD) with extremely high affinity towards a chitin column. When thiols such as Dithiothreitol (DTT) are added to the column, the intein tag undergoes self cleavage and releases ALKBH1 (NEB, 2006).

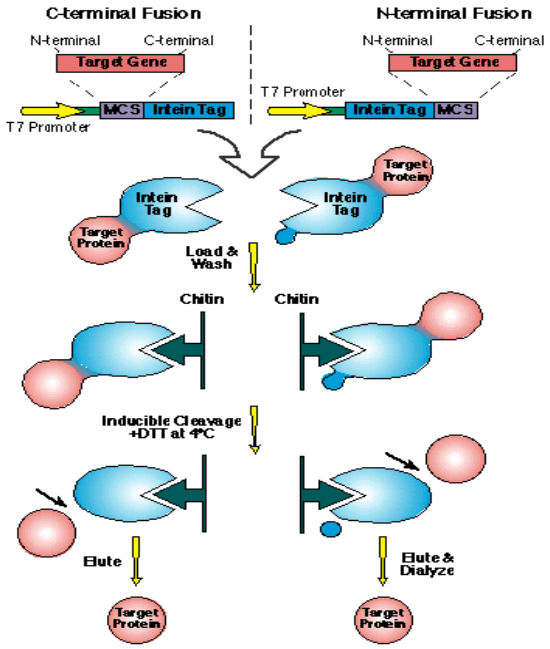


Figure 3.2: Schematic illustration of intein-mediated protein purification with a self-cleavable affinity tag. The target gene is cloned into the multiple cloning site (MCS) of the pTYB12 vector where the N-terminal is fused to the intein tag. When using other pTYB vectors the target protein can be fused to the intein tag by its C-terminal. The intein tag binds to the chitin column through its CBD, which has an extremely high affinity towards the chitin beads. In the presence of thiols the intein tag mediates self-cleavage from the target protein, which is then eluted from the column (NEB, 2006).

The following purification procedure was adapted from the IMPACT™-CN Instructions Manual (NEB, 2006).

1. 20 ml chitin resin was added to an Econo-column chromatography column and washed with several volumes of mQ-H_2O.
2. The column was equilibrated with 2 – 3x column volume with Cell Lysis Buffer.
9. Cell lysate was added to the column by filtration and mixed with the chitin resin by gentle head-over tail rotation at 4ºC for 1 hour.
3. Unbound proteins were eluted by gravity flow. 1 ml sample was collected from the flow-through.
4. The column was washed with at least 20x column volumes of Column Buffer (appendix 1). A 1 ml sample was collected from the wash.

5. On-column cleavage was induced by quickly flushing the column with 35 ml Cleavage Buffer (50 mM DTT [appendix 1]). A 1 ml flush sample was collected.
6. The column flow was stopped and another 10 ml Cleavage Buffer was added followed by incubation ON or for 40 hours at 4˚C.
7. Protein samples were eluted in 1 ml fractions and analyzed by SDS-PAGE.

3.2.4 Dialysis and Concentration of Protein Samples

Dialysis of protein samples is used to eliminate small molecular weight substances that might interfere with downstream experiments. It is also used to change the buffer system of a protein sample for downstream application such as ion exchange chromatography (Thermo Scientific, 2010a). By the use of Amicon® Ultra-4 Centrifugal Filter Devices protein samples were dialyzed and concentrated following affinity chromatography and prior to ion exchange chromatography. By using a Low Salt Buffer (pH 6 or pH 8, 50 mM NaCl [appendix 1]) protein samples were in the appropriate buffer system prior to ion exchange chromatography. Vivaspin 500 µl Centrifugal Filter Tubes were used for concentration of individual protein samples without the use of buffer. Both centrifugal filter devices are equipped with a semi-permeable membrane that allows for protein samples to be retained on one side of the filter membrane whereas smaller molecules diffuse through membrane.

3.2.4.1 Dialysis with 4 ml Amicon® Ultra-4 Centrifugal Filter Devices

1. An appropriate number of centrifugal filter tubes were washed with 4 ml mQ-H_2O by centrifugation at 3900 rpm for 8 min at 4˚C.
2. 1 ml protein sample together with 3 ml Low Salt Buffer was added to each tube and tubes were centrifuged at 3900 rpm for 12 min at 4˚C.
3. Liquid collected in bottom centrifugal tube was discarded and the dialyzed and concentrated protein sample was retained in the filter tube.
4. Steps 2 and 3 were repeated until entire protein sample had been dialyzed and concentrated.
5. Protein samples were analyzed by SDS-PAGE and/or subjected directly to IEX.

3.2.4.2 Concentration with Vivaspin 500 µl Centrifugal Filter Tubes

1. An appropriate number of tubes were washed with 0.6 ml mQ-H_2O by centrifugation at 13000 rpm for 10 min at 4˚C.

2. 0.5 ml protein sample was added to individual tubes and centrifuged at 13000 rpm for 10 min at 4˚C.
3. Liquid collected in centrifugal tube was discarded and concentrated protein sample was retained in the filter tube.
4. Steps 2 and 3 were repeated until the entire protein sample had been concentrated.
5. Protein samples were analyzed by SDS-PAGE.

3.2.5 Ion Exchange Chromatography

Separation of proteins by ion exchange chromatography is based on the reversible interaction between a charged protein and an oppositely charged chromatography matrix. IEX can therefore separate proteins based on their net surface charge, which is dependent on pH. Proteins have no net charge when the pH of the surrounding environment (in this case a buffer) is equivalent to its isoelectric point (pI). When the pH is raised above the pI value of a protein the protein will have a negative net surface charge and will therefore bind to a positively charged matrix – an anion exchange column. At a pH below its pI value the protein will become positively charged and can therefore bind to a negatively charged matrix – a cation exchange column (Amersham Biosciences, 2004).

The ion exchange chromatography matrix is first equilibrated with buffer of a certain pH and ionic strength to ensure that the protein of interest binds to the matrix while unwanted proteins pass through the column. The column is then washed to ensure that all non-binding proteins pass through the column. The target protein can then be eluted by increasing the ionic strength (salt concentration) of the buffer. As the ionic strength increases, salt ions such as Na^+ and Cl^- compete with the target protein for the charged matrix (Amersham Biosciences, 2004).

Ion exchange chromatography was used in further purification of affinity purified ALKBH1 protein samples. It was performed at 4˚C using pre-packed RESOURCE™ Q anion or S cation exchange columns, or a HiTrap™ SP HP cation exchange column. All three columns are strong ion exchangers designed for separation of proteins in operation with ÄKTA™ Explorer Fast Performance Liquid Chromatography (FPLC) system or ÄKTA™ Purifier FPLC system. ÄKTA FPLC systems are computer controlled equipment that allow for automated separation of proteins. The system consists of various components that control the flow rate and ionic gradients, monitor UV adsorption and pH, and allow for sample loading

and sample collection. A chromatogram showing the UV absorption at 280 nm of the eluted protein samples indicates which of the collected fractions contain protein.

Depending on the ion exchange column used (anion or cation exchanger) the ALKBH1 protein sample was dialyzed prior to loading onto the column as described in section 3.2.4.1. ALKBH1 has a pI of 7.0857, thus when dialyzed in a Low Salt Buffer at pH 6 or pH 8 it will be positively or negatively charged, respectively.

Protocol

1. The FPLC system and column was washed with Low Salt Buffer (50 mM NaCl [appendix 1]).
2. The column was equilibrated with High (1 M NaCl [appendix 1]) and Low Salt Buffers, finishing with Low Salt Buffer.
3. The protein sample was loaded onto the column through the injection valve
4. Unbound proteins were collected in an equal flow-through volume as the volume added to the column.
5. The column was washed by a continuous flow of Low Salt Buffer in order to elute all non-binding proteins.
6. ALKBH1 was eluted by applying a linear gradient of increasing salt concentration. This was achieved as the FPLC system applied more and more High Salt Buffer to the column. 1 ml fractions were collected during the elution.
7. Eluted fractions, flow-through and wash samples were analyzed by SDS-PAGE.

3.2.6 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

Gel electrophoresis is a widely used method in which charged molecules are separated according to charge or mass when forced through a gel matrix by an electrical current. Proteins are commonly separated using polyacrylamide gel electrophoresis (PAGE). Polyacrylamide gels are made by mixing acrylamide with bisacrylamide forming a crosslinked polymer network when the polymerizing agent, ammonium persulfate (APS), is added. *N,N,N',N'*-tetramethylethylenediamine (TEMED) is also added to catalyze the polymerization reaction. This crosslinking creates pores in the polyacrylamide gel and the pore size is inversely related to the amount of acrylamide used (Sambrook and Russel, 2001a).

Non-denaturing PAGE or native PAGE separates proteins according to their mass-charge ratio, whereas denaturing SDS-PAGE (figure 3.3) separates proteins based on their molecular mass (Daltons [Da]). When proteins are heated with the ionic detergent SDS, they become denatures and SDS binds tightly to the proteins making them evenly negatively charged. Reducing agents such as DTT or β-mercaptoethanol are also added to cleave proteins disulfide bonds to ensure that no quaternary or tertiary protein structures remain. When an electrical current is applied proteins will migrate towards the positively charged electrode – the anode. Proteins with small molecular masses will travel more quickly through the gel as proteins with increasing molecular masses will be retained by the gel matrix. A molecular weight standard is run alongside the proteins samples and is used as a reference for protein molecular masses (Sambrook and Russel, 2001a).

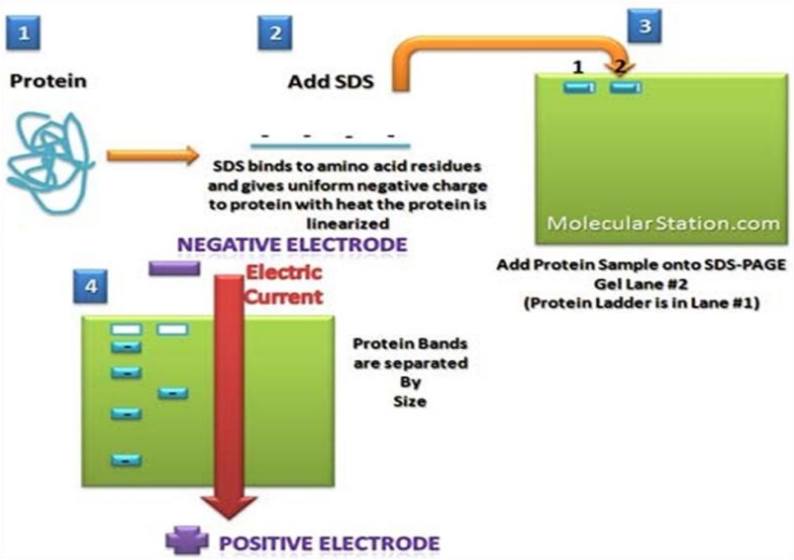


Figure 3.3: Schematic illustration of SDS-PAGE analysis. Figure is adapted from Brebeuf Jesuit Biotechnology (2010)

After separation in the gel the proteins can be visualized by gel staining. Coomassie dye in an acidic buffer was used here to stain proteins bands. The dye will bind to basic and hydrophobic residues of proteins (Sambrook and Russel, 2001a). SDS-PAGE was performed using precast gels, which are commercially available in a variety of percentages. ClearPAGE^{TM} and NuPAGE$^{®}$ precast gels and accessories were used in this study. SDS-PAGE was used here to analyze samples collected during affinity and ion exchange chromatography, including eluted fractions as well as other samples collected during these purification procedures.

Protocol

1. 10 µl sample was mixed with 10 µl Gel Loading Buffer (GLB [appendix 1]). Cell pellet was mixed with 200 – 300 µl of GLB.
2. Samples were heated for 10 min at 70ºC.
3. 20 µl of all samples were loaded to individual wells in the precast gel, which had been assembled into an XCell *SureLock*™ Mini-Cell gel chamber.
4. 10 µl NuPAGE® MOPS molecular weight standard or ClearPAGE™ 2-Color SDS Marker was added to one well depending on the type of precast gel used.
5. The gel chamber was filled up with either ClearPAGE™ SDS running buffer (appendix 1) or NuPAGE® MOPS buffer (appendix 1) depending on the precast gel used.
6. An electrical current was connected to the gel chamber and electrophoresis was performed at 200 Volts (V) for 50 min.
7. Gels were stained for about 15 min in staining solution (appendix 1) followed by washing in destaining solution (appendix 1) until proteins bands could be clearly visualized.

3.2.7 Bio-Rad Protein Assay

The Bio-Rad Protein Assay is based on the Bradford method for determining protein concentrations. It involves the addition of an acidic dye to protein solutions and measurement of absorbance at 595 nm using a spectrophotometer. Coomassie Brilliant Blue G-250 is such an acidic dye where the absorbance maximum shifts from 465 nm to 595 nm when binding to proteins. By comparison to a standard curve of known protein concentrations, unknown protein concentrations can be determined. The Bio-Rad Dye Reagent binds to proteins and a color change of the dye occurs in response to various concentrations of proteins.

The Bio-Rad Protein Assay was used here to determine the concentration of the purified recombinant ALKBH1 protein. It was carried out as described by Bio-Rad (1994, pp. 3 – 4) using the Bio-Rad Dye Reagent and BSA for constructing a standard curve. ALKBH1 protein samples were diluted 1:10.

3.2.8 Protein-protein Interaction by Dot-Blot Immunobinding Assay

Dot-Blot immunobinding assay is an adapted method for assessing protein-protein interactions by immobilizing primary target proteins onto a nitrocellulose membrane followed by incubation with a secondary protein (ALKBH1). By the use of primary and secondary antibodies a color reaction can be visualized as a result of protein-protein interactions (figure 3.4). A primary antibody raised in rabbit (anti-ALKBH1) was used here to detect whether the secondary protein had interacted with the immobilized primary target proteins. An anti-rabbit secondary antibody conjugated with alkaline phosphatase was used to detect primary antibody binding. Alkaline phosphatase is an enzyme that removes phosphate groups from molecules. When Western Blue® Stabilized Substrate was added a dark purple stain could visualized on proteins spots that carried the alkaline phosphatase conjugated secondary antibody.

Phosphate Buffered Saline (PBS) supplemented with Tween100 and BSA was used in this assay. PBS immobilizes proteins onto the nitrocellulose membrane, at the same time preventing denaturation or other conformational changes to the proteins. Tween100 prevents non-specific antibody binding and removes unbound antibodies in order to reduce unspecific background. When used in combination with BSA, Tween100 saturates the binding sites on the membrane, thereby blocking unspecific sites.

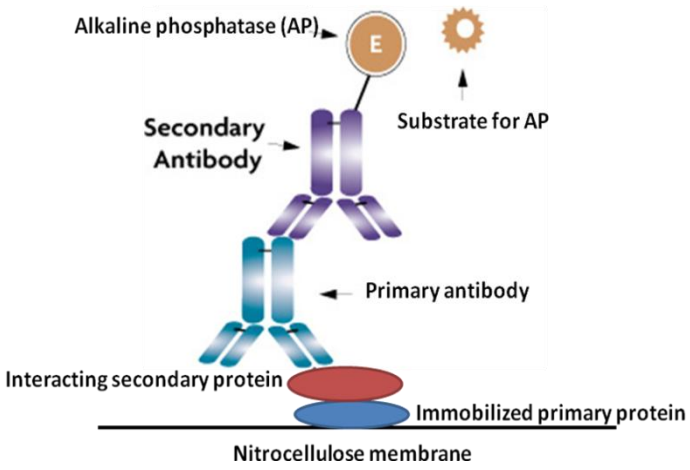


Figure 3.4: Schematic illustration of Dot-Blot immunobinding assay. Figure is adapted from Active Motif (2010)

Protocol

1. 0.5 µg of primary target proteins, NANOG, OCT4, SOX2, histones H1, H2A, H2B, H3 and H4, were spotted onto two separate nitrocellulose membranes together with

ALKBH1 and two negative control proteins – heat-denatured lysozyme and BSA (fraction V).

2. Membranes were dried before blocking of non-specific sites by soaking the membranes in 3%BSA/PBST (appendix 1) for 2 hours at room temperature (RT).
3. One membrane was incubated with ALKBH1 in 3%BSA/PBST (3.325 µg ml^{-1}), whereas the second membrane was incubated with just 3%BSA/PBST.
4. Both membranes were incubated ON at 4ºC by head-over-tail rotation.
5. Membranes were then washed 4 times, 15 min each, in PBST (appendix 1) at 4ºC.
6. Both membranes were incubated with anti-ALKBH1 primary antibody (ab18525 [0.4 mg μl^{-1}]) in 3%BSA/PBST (1:1000 dilution) for 2 hours at 4ºC.
7. Membranes were then washed as in step 5 before being incubated with an anti-rabbit secondary antibody conjugated with alkaline phosphatase (A3812) in 3%BSA/PBST (1:5000 dilution) at 4˚C for 30 min.
8. Membranes were washed as in step 5 and protein-protein interactions were visualized by adding the Western Blue® Stabilized Substrate for alkaline phosphatase (1 ml per membrane).

3.2.9 Protein-protein Interaction by Enzyme-Linked Immunosorbent Assay

An adapted Enzyme-linked immunosorbent assay (ELISA) was performed to assess protein-protein interaction using a CovaLink NH plate. Primary target proteins were attached to the plate followed by incubation with a secondary protein (ALKBH1). Detection of protein-protein interactions is by the same principle as the Dot-Blot immunobinding assay with the use of primary and secondary antibodies. However, this assay allows for detection of protein-protein interactions using increasing concentrations of the ALKBH1.

The CovaLink NH plate (figure 3.5) can be used to bind proteins covalently to a polystyrene surface to which secondary amino groups (NH) have been attached by the use of a spacer arm (Nunc, 2003).

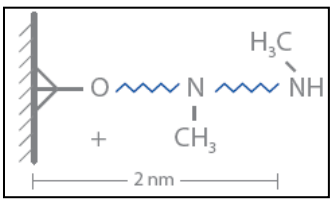


Figure 3.5: Chemical and physical configuration of the CovaLink NH surface. The NH groups are distanced from the polystyrene surface by approx. 2 nm long, chemically defined, spacer arms covalently anchored to the polystyrene surface (Nunc, 2003).

The coupling of primary target proteins to the CovaLink plate is by the formation of amide bonds between carboxylic acids and the CovaLink NH groups. *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC) is a crosslinking agent used to couple carboxyl groups of proteins to NH groups. EDC reacts with carboxyl to form an amine-reactive *O*-acylisourea intermediate. In the presence of *N*-hydroxysulfosuccinimide (Sulfo-NHS) EDC can be used to convert carboxyl groups to amine-reactive Sulfo-NHS esters. This is accomplished by mixing the EDC with a carboxyl containing molecule such as proteins and adding Sulfo-NHS. Figure 3.6 shows an overview of the crosslinking reaction that was used in this assay (Thermo Scientific, 2010b).

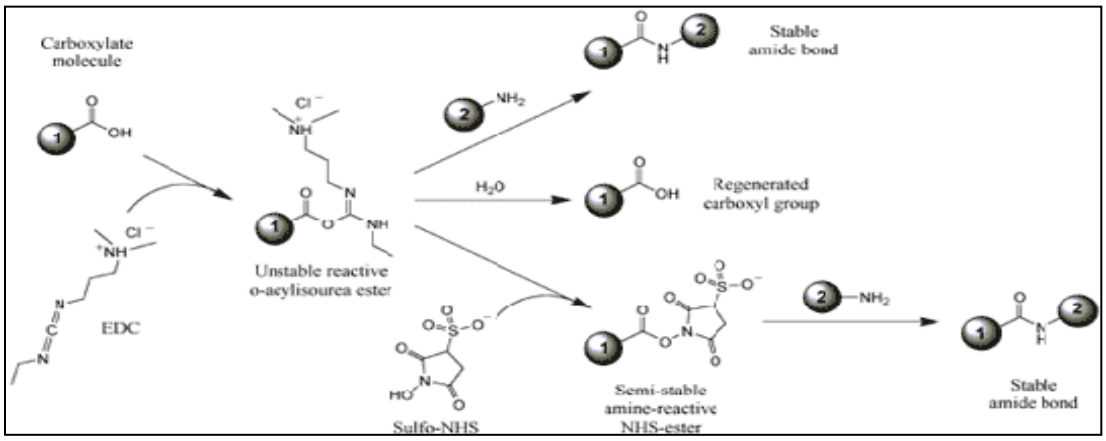


Figure 3.6: Schematic illustration of the crosslinking of primary target proteins to the CovaLink plate. EDC reacts with a carboxyl group on molecule #1 (primary target protein), forming an amine-reactive O-acylisourea intermediate. This intermediate can react with the NH group on molecule #2 (CovaLink NH group) joining the two by an amide bond. This intermediate is also susceptible to hydrolysis, which makes it unstable in an aqueous solution. Addition of Sulfo-NHS stabilizes the intermediate by converting it to an amine-reactive Sulfo-NHS ester, thus increasing the efficiency of EDC crosslinking (Thermo Scientific, 2010b).

Protocol

1. 62.5 fmol of primary target proteins NANOG, OCT4, SOX2, histones H1, H2A, H2B, H3, H4 and BSA in Binding Buffer (appendix 1) containing 1.6 mM Sulfo-NHS were added to each well (100 µl well^{-1}) on a CovaLink plate.
2. Binding of primary proteins to the plate was initiated by adding 6.5 mM EDC in Binding buffer (50 µl well^{-1}).
3. Binding reactions were incubated for 2 hours at RT.
4. Wells were washed 4 times with 300 µl Washing Buffer (appendix 1).
5. Wells were blocked with 350 µl Blocking Buffer (appendix 1) ON at 4°C.
6. Wells were washed as in step 4.

7. A titration of ALKBH1 (0 – 2500 nM) in Washing buffer was added to the wells (100 µl well^{-1}) and incubated for 2 hours at RT.
8. Wells were washed as in step 4 and anti-ALKBH1 primary antibody (ab18525 [0.4 mg µl^{-1}]) in Washing Buffer (1:10,000 dilution) was added to the wells (100 µl well^{-1}).
9. The plate was incubated for 2 hours at RT and wells were washed as in step 4
10. The secondary antibody (A3812) in Washing buffer (1:10,000 dilution) was added to the wells (100 µl well-1).
11. The plate was incubated for 30 min at RT.
12. 100 µl of *p*-nitrophenyl phosphate (PNPP) was added to each well and the reaction was incubated for 30 min at RT.
    - PNPP is a widely used substrate for detecting alkaline phosphatase in ELISA applications. When alkaline phosphatase reacts with PNPP a yellow water-soluble reaction product is formed.
13. Reactions were stopped by adding 50 µl of 2 M NaOH to each well.
14. The color reaction was read at 405 nm using a VICTOR^2 1420 Multilabel counter.
15. The data was analyzed by subtracting the background ALKBH1 bound to BSA from the ALKBH1 bound to primary target proteins.

3.2.10 Succinate Formation Assay

The succinate formation assay can be used to determine biochemical mechanisms of AlkB and AlkB homologues. AlkB uses O_2 in oxidative demethylation of its substrates (figure 3.7). Fe^{2+} is required as cofactor, while 2-OG is a required co-substrate, which is decarboxylated to yield succinate and CO_2 (Falnes, et al., 2003). As ALKBH1 is a homologue of AlkB it is very likely to depend on the same specific cofactors as AlkB to exert its enzymatic function.

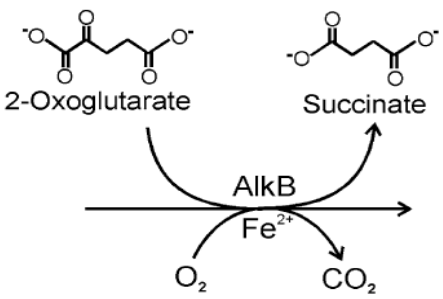


Figure 3.7: Decarboxylation of co-substrate 2-OG to succinate during oxidative demethylation by AlkB. Figure adapted from Falnes et al. (2003)

If the purified recombinant ALKBH1 protein is active and depends on the same cofactors as AlkB, 2-OG will be decarboxylated to yield succinate in the succinate formation assay. The

assay relies on 2,4-Dinitrophenylhydrazine (DNPH) to quantify the formation of succinate as DNPH forms a precipitate with 2-OG, but not with succinate (Kaule and Giinzler, 1990) .

The assay also relies on the use of a radioactive isotope, carbon-14 or ^{14}C, where 2-ketoglutaric acid [5-^{14}C] (carbon 5 is substituted with ^{14}C) can be used as a co-substrate. The decarboxylation activity of ALKBH1 can therefore be determined by measuring the formation of radioactive succinate. In radioactive decay of ^{14}C, β-particles are emitted and succinate quantification can be achieved by measuring the intensity of radiation emitted. This is normally expressed in disintegrations per minute (dpm).

A Tri-Carb 2900 TR Liquid Scintillation Analyzer was used to measure dmp values when the radioactive succinate was added to an Ultima Gold^{TM} MW Liquid Scintillation Cocktail (LSC). The LSC absorbs the energy emitted by radioisotopes and re-emits the energy in flashes of light, which is then read by the Scintillation analyzer.

The succinate formation assay was carried out to analyze the enzymatic activity of ALKBH1 with and without a histone H2A substrate.

3.2.10.1 Succinate Formation Assay by **ALKBH1** *without H2A Substrate*

1. 50 µl reaction mixtures of 50 mM Tris-HCl (pH 7.5), 2 mM Ascorbic acid, 80 µM $FeSO_4$, 0.04 mg ml^{-1} BSA, 0.1 mg ml^{-1} Catalase, 2 mM DTT and 160 µM 2-OG (10 % 2-Ketoglutaric acid [5-^{14}C], 50 mCi/mmol) were added to an appropriate number of 1.5 ml eppendorf tubes.
2. Increasing concentrations (3, 6, 12 µM) of the purified ALKBH1 protein was added to separate reaction tubes. Non-enzymatic decarboxylation of 2-OG was determined by replacing ALKBH1 with mQ-H_2O.
3. 3 µM AlkB was added to a separate reaction mixture and used a control protein.
4. Reactions were incubated for 30 min at 37˚C and terminated by incubation on ice.
5. 25 µl carrier solution (1:1 mixture of 40 mM succinic acid and 40 mM 2-OG) was added to each reaction tube on ice.
6. 2-OG was precipitated at RT for 30 min by adding 25 µl of 0.16 M 2,4 DNPH in 30% $HClO_4$ to each reaction tube.
7. 50 µl 1M 2-OG clearing solution was added to each reaction tube prior to centrifugation at 13000 rpm for 30 min.

8. 40 µl supernatant containing the radioactive succinate was transferred from each reaction tube to separate liquid scintillation vials containing 2 ml Ultima Gold^{TM} MW LSC.
9. Reaction tubes were centrifuged for another 15 min at 13000 rpm and another 40 µl of the supernatant was transferred to the liquid scintillation vials.
10. The radioactivity (dpm) was counted (3x 5 min) for each scintillation vial in a Tri-Carb 2900 TR Liquid Scintillation Analyzer.

3.2.10.2 Succinate Formation Assay by ALKBH1 *with H2A Substrate*

The assay was carried out as in section 3.2.10.1, but with the addition of increasing concentration (2, 4, 8 or 10 µg) of calf thymus histone H2A (cH2A) to separate reaction mixtures. 6 µM purified ALKBH1 was added to each reaction mixture. *Schizosaccharomyces pombe* AlkB homologue 2 was used as a control protein. It has been demonstrated that this enzyme shows increased decarboxylation of 2-OG to succinate by the addition of cH2A (Bjørås, M., unpublished data). Its exact biochemical function, however, has not been verified. 6 µM of *S. pombe* AlkB homologue 2 was used together with 2 or 4 µg cH2A substrate.

3.3 DNA Related Methods

3.3.1 Agarose Gel Electrophoresis

Agarose gel electrophoresis is used to separate DNA fragments by size as they migrate through the gel in an electric field. DNA is negatively charged due to the negatively charged phosphate groups along the backbone of the DNA molecule. DNA will therefore migrate towards the positive electrode. The pores within the agarose gel matrix allow short DNA fragments to migrate more easily and cause retention of longer DNA fragments. The separated DNA fragments are visualized in the gel by addition of ethidium bromide (EtBr). EtBr will intercalate into the structure of DNA molecules resulting in a complex that fluoresces brightly under UV light. Agarose gels with a low agarose percentage (0.7%) will give good separation of large DNA fragments (5 – 10 kilobases [kb]), whereas high percentage gels (2%) will give good resolution of small DNA fragments (0.2 – 1 kb) (Sambrook and Russel, 2001a).

A DNA molecular weight standard is used to size the separated DNA fragments. A DNA loading dye is added to DNA samples in order to increase their specific weight so that they fall to the bottom of the wells in the gel. The progress of the electrophoresis can also be assessed by following the movement of the dye through the gel (Sambrook and Russel, 2001a).

Agarose gel electrophoresis was used here to analyze amplified DNA fragments from Polymerase Chain Reaction (PCR).

Protocol

1. A 1.2% agarose gels was prepared by mixing 0.48 g in 40 ml 1x TAE electrophoresis buffer (appendix 1).
2. The mixture was boiled (in a microwave) until all agarose particles were dissolved
3. The gel was cooled before adding EtBr (2.5 µl 40 ml^{-1} agarose solution).
4. The gel was poured into a gel tray that was placed in an electrophoresis chamber with an assembled comb towards to negative electrode in order to create sample wells.
5. After the gel had solidified the comb was removed and the electrophoresis chamber was filled with 1xTEA buffer so that gel was submerged in buffer
6. 8 µl DNA sample was mixed with 2 µl 6x Gel DNA loading dye and the total volume was added to separate wells. 4 µl of the GeneRuler^{TM} DNA Ladder Mix was added to one well.
7. An electrophoresis power supply was connected with the cathode (negative electrode) at the side of the loaded samples and the anode (positive electrode) at the opposite side.
8. Electrophoresis was run at 85 V until the dye was near the bottom of the gel.
9. The separated DNA fragments were visualized under UV light in a UV-1601 Visible Spectrophotometer.

3.3.2 QIAquick PCR Purification Kit

The QIAquick PCR purification kit is designed to purify ss or ds DNA fragments from PCR reactions as enzymes and other contaminants can interfere with subsequent analysis of the DNA fragments. QIAquick spin columns contain a silica membrane to which DNA is absorbed while contaminants pass through the column. Purified DNA is subsequently eluted from the column (QIAGEN, 2008).

The QIAquick PCR purification kit was used to purify PCR amplified DNA fragments. These were used for subsequent radioactive labeling prior to electrophoretic mobility shift assay (EMSA). The QIAquick PCR Purification Kit Protocol using a microcentrifuge was applied (QIAGEN, 2008, pp. 19 – 20).

3.3.3 5'-labeling of DNA with [γ-^{32}P]ATP

DNA fragments can be labeled with the radioactive [γ-^{32}P] Adenosine triphosphate (ATP) isotope using the T4 Polynucleotide Kinase (PNK). T4 PNK catalyzes the transfer of the γ-phosphate from ATP to the 5'-OH group of ss and ds DNA fragments (Fermentas, 2004). The T4 PNK enzyme was used for [γ-^{32}P]ATP radioactive labeling of DNA fragments that were used in EMSA.

1. The following reaction mixture was incubated for 30 min at 37ºC.
    - 5 pmol DNA
    - 3 µl γ-[^{32}P] ATP (250 µCi)
    - 1 µl T4 PNK (10 Units [U] µl^{-1})
    - 5 µl 10x T4 PNK Reaction Buffer A
    - Nuclease free water to a final volume of 50 µl
2. 0.5 µl of 100 µM non-radioactive ATP was added and the reaction was further incubated at 37°C for 10 min.

3.3.4 QIAquick Nucleotide Removal Kit

The QIAquick Nucleotide Removal Kit was used for cleanup of radioactive [γ-^{32}P]ATP labeled DNA fragments. Primers, enzymes, salts and unincorporated nucleotides are removed by the same principle as the QIAquick PCR Purification Kit. The QIAquick Nucleotide Removal Kit protocol using a microcentrifuge was applied (QIAGEN, 2008, pp. 23 – 24). The DNA fragments were then used in EMSA.

3.3.5 NanoDrop® Nucleic Acid Quantification

Quantitative analysis of nucleic acids (DNA and RNA) can be determined by spectroscopy measuring absorption in the UV range at 260 nm. DNA/RNA concentrations are evaluated against a blank measurement (usually water) at 260 nm and then via an evaluating factor. The evaluation factors for ds DNA, ss DNA and RNA in a UV light path length of 1 cm are: 1 OD

(Absorbance [A]) unit corresponds to approximately 50 µg ml^{-1}, 33 µg ml^{-1} and 40 µg ml^{-1} for ds DNA, ss DNA and RNA, respectively. The interference of protein contaminant is evaluated by calculation of the A_{260}/A_{280} ratio as proteins absorb at 280 nm. Pure DNA should have a ratio of 1.8, whereas RNA should have a ratio of 2.0 (Brinkmann Instruments, 2003).

NanoDrop® is a microvolume spectrophotometer used for measuring DNA/RNA concentration without the use of cuvettes. By the use of fiber optic technology and surface tension the sample is held between two optical surfaces that define the path length. Concentrations ranging from 2 ng $µl^{-1}$ to 3700 ng $µl^{-1}$ can be measured as samples are assessed in both 0.2 mm and 1.0 mm path lengths (Gallagher and Desjardins, 2006).

The following procedure was used to measure RNA and DNA concentration of various samples:

1. The appropriate program was selected in the NanoDrop® software (e.g. ds DNA or RNA).
2. Upper and lower optical surfaces were cleaned with mQ-H_2O.
3. The spectrophotometer was initialized by adding 2 µl mQ- H_2O onto the lower optical surface, lowering the upper arm and selecting the "initialize" option in the NanoDrop® software.
4. A blank measurement was performed by adding 2 µl nuclease free water and selecting the "blank" option in the NanoDrop® software.
5. Both optical surfaces were cleaned.
6. Measurements were made adding 2 µl DNA/RNA sample to the lower optical surface, lowering the upper arm and selecting the "measure" option in the NanoDrop® software.

3.4 Bioinformatics Analysis of DNA regions bound by NANOG and OCT4

In the study by Boyer et al. (2005) target gene sequences bound by NANOG and OCT4 were identified in human ES cells. This was achieved using chromatin immunoprecipitation coupled to DNA microarrays. They found that NANOG binds at position 77243985 to 77244251 at chromosome 14 (C14), overlapping *ALKBH1* exon 1 and open reading frame 156 (C14orf156) upstream of *ALKBH1*. OCT4 binds at position 77243910 to 77244190 at C14 also overlapping *ALKBH1* exon 1 and C14orf156. All sequences and coordinates used by

Boyer et al. (2005) were retrieved from the May 2004 build of the human genome (NCBI Build 35).

The human genome assembly found at National Center for Biotechnology Information (NCBI, 2010a) web site is based on the human genome assembly from the "Genome Reference Consortium" (GRC, 2009). From here different assemblies of the human genome can be accessed (Build 35, 36 and the newest version, Build 37). The nucleotide sequence from position 77243001 to 77244620 on C14 was retrieved from three builds (35, 36 and 37). All three sequences were submitted to the web based Multiple Sequence Alignment (MSA) tool ClustalW2 (EBI, 2010) to see if they were identical for this region of the genome.

A NCBI nucleotide Basic Local Alignment Search Tool (nBlast) search was performed against the "Human genomic + transcripts" database using the sequence from Build 36 and the sequence from Build 37 as a query. nBlast is a bioinformatics tool that finds nucleotide regions of high similarity (NCBI, 2010b). Results from this analysis are presented in results section 4.5. The sequence from Build 36 was used in subsequent primer design for PCR amplification of the DNA regions bound by NANOG and OCT4, as explained in results section 4.5.

3.5 Polymerase Chain Reaction

The Polymerase Chain Reaction (PCR) is a very powerful and versatile technique used to amplify ds DNA regions more than 1 billion-fold. It is based on the use of the DNA polymerases, which in cells are used for DNA replication and synthesis of complementary ss DNA. Primers with complementary sequences to the regions flanking the ds template DNA serve as starting points for DNA synthesis by the DNA polymerase. After heat-denaturing of the ds DNA template, the primers are annealed to the ss DNA and extended by the DNA polymerase in 5'→3' direction. This is achieved in the presence of all four deoxy-trinucleotide phosphate (dNTP), which are incorporated into the newly synthesized DNA strand (Sambrook and Russel, 2001b).

By using a thermostable form of DNA polymerase, Taq polymerase, no new enzyme has to be added to reaction as Taq polymerase does not become inactive during the heat-denaturing step. The reaction is carried out for usually 30 cycles or more with each cycle increasing the amount of target DNA as the number of duplex DNA molecules doubles in each successive cycle. Because the borders of the newly synthesized DNA strands are defined by the primers,

only ds DNA sequences within the primer-binding sites will accumulate with increasing number of cycles and form the vast majority of the PCR products (Sambrook and Russel, 2001b).

3.5.1 Primer Design

The nucleotide sequence (position 77243001 to 77244620 on C14) from the NCBI Build 36 of the human genome was used to design primers for consecutive PCR amplification of the DNA regions bound by NANOG and OCT4 as described by Boyer et al. (2005).

The sequence was submitted to the NCBI primer-Blast tool in order to find good starting primers (NCBI, 2010c). The first primer pair (GAforw781 and GArev781) enclosed a 781 bp region within the original nucleotide sequence. This 781 bp sequence was subsequently submitted to the primer-Blast tool and a second primer pair (GAforw434 and GArev781) was found. This primer pair enclosed a 434 bp region within the 781 bp sequence. From this 434 bp sequence, specific primers for the DNA regions bound by NANOG and OCT4 were designed manually. The GAforwnanog and GArevnanog primers pair was used for PCR amplification of the DNA region bound by NANOG. The GAforwoct4 and GArevoct4 primer pair was used for PCR amplification of the DNA region bound by OCT4. Complete nucleotide sequences including binding regions and primers are listed in appendix 3.

3.5.2 PCR Amplification of DNA regions bound by NANOG and OCT4

PCR amplification of the DNA regions bound by NANOG and OCT4 was carried out using the Techne TC-512 Thermal Cycler. Two different PCR DNA polymerases were used, the Pfu Turbo® DNA Polymerase or AmpliTaq Gold™ DNA Polymerase. Different reaction mixtures were prepared depending on the DNA polymerase used. Gradient PCR reactions were carried out to determine optimum annealing temperatures for primers. Different concentrations of the DNA template were also used. This optimization of annealing temperature and DNA template concentration was done in order to minimize amplification of unspecific DNA fragments. The different PCR reactions and cycling parameters applied are described below.

3.5.2.1 Gradient PCR Amplification

Reaction mixtures for three annealing temperatures, 50ºC, 55ºC or 60ºC, containing different concentrations of DNA template were prepared. Forward primers are abbreviated to → and reverse primers are abbreviated to ←

Table 3.1: DNA templates, respective primers and PCR products

DNA template	**DNA concentration/volume**	**→ primer**	**← primer**	**PCR product**
Genomic DNA (10 ng μl^{-1})	3 ng or 5 ng	GAforw781	GArev781	781 bp sequence
PCR amplified 781 bp sequence	Pipette dip or 0.5 µl	GAforw434	GArev781	434 bp sequence
PCR amplified 434 bp sequence	Pipette dip or 0.5 µl	GAforwnanog	GArevnanog	NANOG DNA binding sequence
PCR amplified 434 bp sequence	Pipette dip or 0.5 µl	GAforwoct4	GArevoct4	OCT4 DNA binding sequence

When PCR amplified DNA sequences were used as DNA templates the exact DNA concentration is not listed as the concentration of the DNA templates were not determined prior to amplification. However, knowing that the template was a PCR product the DNA concentration is very high, thus a very small volume was used.

Table 3.2: Reaction mixture for Pfu Turbo® DNA Polymerase

Reagent	**Volume**
Pfu Turbo® (25 U μl^{-1})	0.5 µl
10x Cloned Pfu Reaction Buffer	3 µl
dNTP	3 µl
→ primer (3.5 pmol μl^{-1})	3 µl
← primer (3.5 pmol μl^{-1})	3 µl
DNA template	Pipette dip, 0.5 µl, 3 µl or 5 µl
Nuclease free water	To a final volume of 30 µl

Table 3.3: Reaction mixture for AmpliTaq Gold™ DNA Polymerase

Reagent	Volume
AmpliTaq Gold™ (5 U μl^{-1})	0.3 µl
10x PCR Buffer	2 µl
$MgCl_2$ solution (25 mM)	1 µl
dNTP	1.6 µl
→ primer (10 pmol μl^{-1})	1 µl
← primer (10 pmol μl^{-1})	1 µl
DNA template	Pipette dip or 0.5 µl
Nuclease free water	To a total volume of 20 µl

Table 3.4: Gradient PCR cycling parameters

Step	Temperature	Time	Number of cycles
Denaturing	95ºC	5 min	1
Denaturing	95ºC	30 sec	30
Annealing	50ºC/55ºC/60ºC	30 sec	
Extension	72ºC	1 min	
Extension	72ºC	7 min	1

The amplified DNA samples were analyzed by agarose gel electrophoresis and samples with the least unspecific DNA bands were used as DNA templates in subsequent PCR amplifications. In further amplification DNA regions bound by NANOG and OCT4, the AmpliTaq Gold™ reaction mixture (table 3.3) was used together with the same PCR cycling parameters as in table 3.4, although the annealing step was set to 60ºC only. A pipette dip of the 434 bp DNA template was used.

3.6 DNA Sequencing

DNA sequencing was performed by the sequence facility at "Mikrobiologisk Institutt, Rikshospitalet", using the DNA Analyzer 3730. DNA regions bound by NANOG and OCT4 had been amplified by PCR (section 3.5.2) and were used as DNA templates. The DNA region bound by NANOG was sequenced using either the GAforwnanog or GArevnanog primer, whereas the DNA region bound by OCT4 was sequenced using either the GAforwoct4 or GArevoct4 primer. Sequencing mixtures contained 5 ng DNA template, 3.2 pmol primer and nuclease free water to a total volume of 15 µl.

3.6.1 Bioinformatics Analysis of Sequencing Results

Obtained sequencing results of the DNA regions bound by NANOG and OCT4 were formatted into the fasta format. When DNA templates had been sequenced using reverse primers the obtained nucleotide sequence were converted to its reverse complement in order to get the correct orientation of the DNA sequence. Formatting was done using the web based program, Nucleic Acid Sequence Massager (Attotron Biosensor Corporation, 1998).

The nucleotide sequences were submitted to an nBlast (NCBI, 2010b) search against the "Human Genomic + transcripts" database. It was applied here to see whether the two sequences overlapped the *ALKBH1* promoter sequence as demonstrated by Boyer et al. (2005)

3.7 Electrophoretic Mobility Shift Assay

EMSA can be used to determine protein:DNA interactions. The techniques takes advantage of the observation that protein:DNA complexes migrate more slowly than free linear DNA fragments in non-denaturing PAGE. 5'-end labeling of DNA templates with the radioactive isotope ^{32}P using [γ-^{32}P]ATP is performed for visualization of unbound DNA fragments and protein:DNA complexes. Unlabeled nucleic acids can be used to minimize unspecific binding to the labeled target DNA and Poly (dI•dC) was used here as a nonspecific competitor (Thermo Scientific, 2010c).

Following electrophoresis the gel can be exposed to a phosphor imaging screen that detects ^{32}P radioisotopes. The screen is composed of crystals of $BaFBr:EU^{+2}$ and when the screen is exposed to γ radiation, the electrons from Eu^{+2} are excited and then trapped in an "F-center" of the $BaFBr^{-}$ complex. This results in the oxidation of Eu^{+2} to Eu^{+3}, which forms an image on the screen. When the screen is exposed to a scanner Eu^{+3} is reverted back to Eu^{+2} releasing a photon at 390 nm. This is collected by the scanner, which provides a representation of the image on screen (Voytas and Ke, 1999).

EMSA was used here to detect protein:DNA interactions between NANOG and OCT4 and their respective DNA binding sequences. These sequences had been prepared by PCR and subsequent labeling with [γ-^{32}P]ATP (sections 3.5.2 and 3.3.4, respectively).

Protocol

1. A 5% native PAGE gel (appendix 2) was prepared using 1.0 mm Gel Cassettes and left to solidify for 1.5 – 2 hours.
2. 20 µl reaction mixtures containing 10 mM HEPES (pH 7.5), 10 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 1 mM EDTA, 10 % glycerol, 1 µg poly(dI•dC) and 0.5 ng labeled DNA template were prepared.
3. Increasing concentration of NANOG or OCT4 (0 nM, 8 nM, 40 nM, 200 nM, 1 µM and 2 µM) were added to separate reaction mixtures that were incubated at RT for 30 min.
4. Reactions were stopped by adding 5 µl 75% glycerol.
5. Binding reactions were resolved in the 5% PAGE gel by electrophoresis run at 60 V for 4-6 hours using 0.5x TBE running buffer (appendix 1).
6. 5 µl GeneRuler^{TM} DNA Ladder Mix was added to one well as an indicator for migration of unbound DNA templates.
7. Following electrophoresis the gel was placed on a sheet of Whatman 3MM filter paper and covered with plastic wrap.
8. The gel was exposed to a general purpose storage phosphor screen at RT ON and the screen image was scanned using the Typhoon 9410 Variable Mode Imager Scanner the following day.

3.8 Culture of Mouse Embryonic Stem Cells

In order for mouse ES (mES) cells to maintain their self-renewal capacity and remain undifferentiated, they have to be cultured under specific conditions. The mES cells thereby retain the potential to differentiate into all three germ layers. In order to maintain these characteristics of the wild-type mES cell line used in this study, they were cultured on a murine embryonic fibroblast (MEF) monolayer using a medium supplemented with LIF and fetal bovine serum (FBS). Induced differentiation was performed by the formation of embryonic bodies (EBs) or by the addition of Retinoic Acid (RA) to the culture medium. LIF was excluded from the medium when mES cells were differentiated. Samples were collected on Day 0, 6, 9 and 12 from both EBs and RA differentiated mES cells.

A laminar flow hood was used in all protocols and all media solutions, GIBCO® DPBS and GIBCO® 0.05% Trypsin-EDTA were preheated in a 37˚C water bath prior to use. Aseptic working techniques were conducted with the used of sterile solutions and materials (pipettes,

tubes, etc.). 70% ethanol and sterile wipes were used to clean the hood area and other materials prior to use. All cells were incubated in a 37˚C, 5% CO_2 humidified incubator.

3.8.1 Preparation of MEF Plates

1. Gelatin plates were prepared by coating FALCON® MULTIWELL™ Tissue Culture Plates (6 wells) or FALCON® Tissue Culture Dishes with ESGRO® COMPLETE™ 0.1% Gelatin with a volume that covered the plate/dish surface[1].
2. Plates were incubated for at least 30 min at RT.
3. A vial of MEF cells ($4.5x10^5$ cells per vial) were rapidly thawed in a 37˚C water bath just until a bit of ice was left.
4. Cells were aseptically transferred to a 15 ml tube containing 9 ml prewarmed MEF medium (appendix 2).
5. The tube was centrifuged for 5 min at 800 rpm.
6. The medium was aspirated and cell pellet resuspended in an appropriate volume of fresh MEF medium.
7. The gelatin solution was removed from the plates and cells were immediately added without letting the gelatin dry.
8. MEF cells were incubated ON and a confluent (entire plastic area covered) settled feeder monolayer was observed the following day.

A total volume of 10 ml or 2 ml MEF medium were used per dish or per well, respectively.

3.8.2 Thawing of mES cells and Transfer to MEF plates

1. 9 ml mES cell medium (appendix 2) was preheated in a 15 ml tube.
2. A vial with mES cells were thawed rapidly in a 37˚C water bath just until a bit of ice was left.
3. Cells were aseptically transferred to the 9 ml prewarmed mES cell medium and centrifuged for 5 min at 800 rpm.
4. Medium was aspirated and cell pellet resuspended in an appropriate volume of fresh mES cell medium.
5. MEF medium was aspired from ready prepared MEF plates and mES cells were aseptically transferred to the plates.

[1] For simplicity dishes and plates will be referred to as plates only

6. mES cells were incubated ON and small bright colonies were observed the following day.

The mES cell medium was supplemented with 61.8 µl LIF (10^7 U ml^{-1}). A total volume of 8 ml or 2 ml medium were used per dish or per well, respectively.

3.8.3 Feeding and Passage of Undifferentiated mES cells

Undifferentiated mES cells were given fresh mES cell medium every day. These cells were also split/passed every second day when colonies were big and individual cells were difficult to see by the following procedure.

1. Medium was aspirated from plates and mES cells were washed with GIBCO® DPBS.
2. DPBS was aspired and GIBCO® 0.05% Trypsin-EDTA was added to cover the cells (a volume just enough to cover the surface).
3. Plates were left in the 37ºC incubator until cells started to disrupt from the bottom of the plate.
4. mES cell medium was added to neutralize the trypsin (twice the volume as Trypsin-EDTA) and cells were disrupted further into single cell colonies by pipetting up and down.
5. Dissociated cells were aseptically transferred to a 15 ml tube and centrifuged for 5 min at 800 rpm.
6. MEF medium was removed from ready prepared MEF plates and an appropriate volume of fresh mES cell medium was added to the plates.
7. The mES cell medium was aspirated from the centrifuged tube and cell pellet resuspended in an appropriate volume of fresh mES cell medium before being aseptically transferred onto MEF plates.

The mES cell medium was supplemented with 61.8 µl LIF (10^7 U ml^{-1}). A total volume of 8 ml or 2 ml medium were used per dish or per well, respectively.

3.8.4 Induced Differentiation of mES cells with RA

3.8.4.1 Passage of Undifferentiated mES cells from MEF Plates onto Gelatin Plates

1. Medium was aspirated from plates with undifferentiated mES cells growing on a MEF monolayer.
2. mES cells were dissociated from the plates as described in section 3.8.3, steps 1 – 5.

3. The medium was aspirated from the 15 ml centrifuged tube and cell pellet resuspended in an appropriate volume of fresh mES cell medium.
4. Cells were aseptically transferred to a ready prepared gelatin plate
5. Plates were incubated for at least 20 min in order for MEF cells to settle onto the gelatin plate leaving mES cells in suspension in the medium.
6. An appropriate volume of fresh mES cell medium was added to a new ready prepared gelatin plate and mES cells were aseptically transferred to the plate.
7. The plates were incubated ON before differentiation with RA was induced.

The mES cell medium was supplemented with 61.8 µl LIF (10^7 U ml^{-1}). A total volume of 8 ml or 2 ml medium were used per dish or per well, respectively.

3.8.4.2 Differentiation with RA

1. An appropriate volume of mES cell medium was prewarmed in a 37˚C water bath
2. A vial of 100x RA was defrosted and added to the prewarmed mES cell medium to a final concentration of 1x RA.
3. Medium was aspired from mES cells growing on a gelatin plate and fresh mES cells medium with 1xRA was added to the plates.

RA acid differentiated mES cells were given fresh mES cell medium with 1x RA acid every day.

The mES cell medium was not supplemented with LIF. A total volume of 7 ml or 2 ml medium were used per dish or per well, respectively.

3.8.4.3 Collection of Samples from RA Differentiating mES cells

Samples were collected from RA differentiated mES cells on Day 6, 9 and 12 by the following procedure.

1. Medium was aspirated from plates with RA differentiated mES cells
2. Cells were dissociated from the plate as described in section 3.8.3, steps 1 – 5.
3. The medium was aspired from the 15 ml centrifuged tube and cell pellet was resuspended in 1.5 ml fresh mES cell medium and added to a 1.5 ml eppendorf tube.
4. The sample was spun down for 10 min at 6000 rpm. The supernatant was aspired and cell pellet was stored at -70˚C.

3.8.5 Differentiation of mES cells by Formation of EBs

3.8.5.1 Preparation of EBs

1. The same procedure as in section 3.8.4.1, steps 1 – 5, was used to separate mES cells from MEF cells.
2. mES cells were aseptically transferred to CORNING® Ultra-Low Culture Dish instead of a gelatin plate.
    - By allowing mES cells to grow in suspension instead of being attached to the plate they clump together to form EBs.
3. A 1.5 ml Day 0 sample of undifferentiated mES cells was collected at the same time.
4. The sample was spun down for 10 min at 6000 rpm. The supernatant was aspired and cell pellet was stored at -70˚C.

The mES cell medium used was not supplemented with LIF. A total volume of 7 ml or 2 ml medium were used per dish or per well, respectively.

3.8.5.2 Feeding EBs before Day 4

1. EBs growing in suspension were aseptically transferred from plates to a 15 ml tube.
2. Tubes were incubated until EBs had settled to the bottom of the tube (5 – 15 min depending on the size of the EBs).
3. Top part of the medium was aspirated leaving the settles EBs.
4. EBs were resuspended in an appropriate volume of fresh mES cell medium and aseptically transferred back to the plates.

The mES cell medium was not supplemented with LIF. A total volume of 8 ml or 2 ml medium were used per dish or per well, respectively.

3.8.5.3 Attaching EBs to Plates on Day 4

After EBs have been grown in suspension for four days they become so large that they start to settle down onto the plate. This can also be done manually.

1. Same procedure as in section 3.8.5.2, steps 1 – 3.
2. EBs were resuspended in an appropriate volume of mES cell medium and aseptically transferred onto ready prepared gelatin plates allowing EBs to attach to the plate.

The attached EBs were given fresh mES cell medium every day.

The mES cell medium used was not supplemented with LIF. A total volume of 7 ml or 2 ml medium were used per dish or per well, respectively.

3.8.5.4 Collection of Samples from Settled EBs

Samples were collected from EBs on Day 6, 9 and 12 by the same procedure in section 3.8.4.3.

3.9 Purification of Total RNA from Animal Cells

The QIAGEN® miRNeasy Mini Kit (QIAGEN, 2007) can be used for extraction and purification of RNA from cultured animal cells. The principle for RNA extraction and purification using this kit is based on phenol/guanidine lysis of samples together with a silica membrane for RNA purification. The kit is supplied with a QIAzol Lysis Reagent that facilitates cell lysis, inhibits RNases and removes most of the cellular DNA and proteins from the lysate by organic extraction (QIAGEN, 2007). The reagent contains phenol and guanidine thiocyanate. The latter is a chaotropic agent and can therefore be used as a general protein denaturant as it disrupts the three dimensional structure of proteins, including RNase enzymes (Mason, et al., 2003).

After cell homogenization with the QIAzol Lysis Reagent chloroform is added to allow for separation into aqueous and organic phases by centrifugation (QIAGEN, 2007). Phenol-chloroform extraction is a widely used liquid-liquid extraction technique in molecular biology for isolating DNA, RNA or proteins. It relies on phase separation during centrifugation which results in an upper aqueous phase and a lower organic phase (mainly chloroform). Nearly all of the RNA is present in the aqueous phase, while DNA and protein partition in the organic phase and interphase, respectively (Chomczynski and Sacchi, 1987). RNA is recovered from the aqueous phase and ethanol is added such that the RNA molecules will have good binding conditions to the RNeasy Mini spin column. In the column RNA binds to the membrane while phenol and other contaminants are washed off. RNA is then eluted from the column using RNase-free water (QIAGEN, 2007).

The QIAGEN protocol for Purification of total RNA, including small RNAs, from Animals Cells (QIAGEN, 2007, pp. 19 – 22) using the QIAGEN® miRNeasy Mini Kit was applied here to extract RNA from cell samples collected from both EBs and RA differentiated mES cells at Day 0, 6, 9 and 12. RNA extraction was started at step 2 in the protocol and after

completion the RNA concentration was measured with NanoDrop®. The RNA samples were used for subsequent reverse transcription dependent conversion into complementary DNA (cDNA).

3.10 Real-Time PCR

Real-time PCR (RT-PCR) is used to detect PCR products as they accumulate. This allows for quantification of gene expression at mRNA levels by first converting mRNA into cDNA by reverse transcription. The cDNA is then amplified by PCR and detection of the PCR product relies on the use of a fluorescent reporter dye. The amount of DNA produced is proportional to the increase in fluorescent signal. SYBR Green is a fluorescent reporter dye that binds ds DNA and emits light upon excitation. As the PCR product accumulate, the fluorescence increases. As SYBR Green will bind to any ds DNA in the reaction, including primer-dimers and other non-specific reaction product well designed primers have to be used. This limits non-specific reaction products. A PCR amplification plot of the cDNA target contains the information needed for quantitative measurements. A threshold line is the point at which the reaction reaches a fluorescent intensity above the background. The cycle at which the sample reaches this level is called the cycle threshold (Ct). The more cDNA present in the starting material the lower the Ct value. (Nolan, et al., 2006; Applied Biosystems, 2010).

3.10.1 Reverse Transcription of mRNA into cDNA

Reverse transcriptase in an enzyme derived from retroviruses that use mRNA as a template to synthesize ss cDNA. This ss cDNA is used as a template for ds DNA synthesis (Griffith, et al., 1999). RNA extracted from all mES cell samples were transcribed into cDNA using the High-Capacity cDNA Reverse Transcription kit and protocol (Applied Biosystems, 2006, pp. 7 – 8). After reverse transcription the amount of cDNA is assumed to be equal to the starting RNA concentration. In this protocol 10 µl RNA (500 ng μl^{-1}) is mixed with 10 µl 2x Reverse Transcriptase Mastermix (total volume of 20 µl), which means that the amount of cDNA produced was half of the amount of RNA, i.e. 250 ng μl^{-1}.

3.10.2 Real-Time PCR with SYBR green

The StepOnePlus RT-PCR instrument was used together with the StepOnePlus 2.1 software to perform RT-PCR experiments of cDNA samples prepared and collected the cultured mES

cells. Experiments were designed using the software prior to preparation of RT-PCR plates. Comparative Ct experiments and Standard Curve experiments were designed for each target (gene). The non-regulated house-keeping gene glyceraldehydes-3-phosphate dehydrogenase (*gapdh*) was used as an endogenous control gene to normalize the amplification levels of the other targets. The target genes that were analyzed included *Alkbh1*, *nanog*, *oct4*, *sox2*, *fgf5* (ectoderm specific), *brachyury* (mesoderm specific), and *gata6* (endoderm specific).

Gapdh and *Alkbh1* primers had previously been designed by colleagues at Rikshospitalet. The *Alkbh1* primers had also been designed by colleagues at Rikshospitalet using the web-based program PrimerQuest^{SM} (Integrated DNA Technologies, 2010). Primers for *nanog*, *oct4* and *sox2* were the same as those used by Aloia et al. (2010), whereas primer sequences for *fgf5*, *brachyury* and *gata6* were the same as those used by Singh et al. (2008).

All cDNA samples were diluted to 5 ng μl^{-1} or 20 ng μl^{-1} prior to Comparative Ct experiments. Standard Curve experiments were designed for each target with decreasing concentrations (250 ng μl^{-1}, 25 ng μl^{-1}, 2.5 ng μl^{-1} or 0.25 ng μl^{-1}) of cDNA template. All primers were diluted to 900 nM. SYBR green RT-PCR reaction mixtures were prepared as a mastermix in order to keep pipetting errors to the minimal. One separate mastermix was prepared for *gapdh*, and another separate mastermix was prepared for the target gene that was analyzed.

Table 3.8: SYBR green RT-PCR reaction mixture

Reagent	**Volume (1x)**
Power SYBR® Green PCR mastermix	10 µl
Forward primer (900 mM)	1 µl
Reverse primer (900 nM)	1 µl
Nuclease free water	7 µl
cDNA template (0.25, 2.5 5, 20, 25 or 250 ng μl^{-1})	1 µl
Total volume	20 µl

The SYBR green RT-PCR mastermix was prepared containing all reagents listed in table 3.5 except for the cDNA template. 19 µl mastermix was added to individual wells in a MicroAmp® Fast 96-Well Reaction Plate. 1 µl cDNA from each sample was then added to six individual wells, three wells that contained the *gapdh* mastermix and three wells that contained the target mastermix. In this way all samples were prepared in triplicates for both *gapdh* and the target in question. The plate was sealed with a MicroAmp^{TM} Optical Adhesive

Film. The plate was centrifuged by quick spin at 800 rpm at 4˚C and then inserted into the instrument.

3.10.3 Quantitative Analysis of RT-PCR Experiments

Relative quantification of the RT-PCR results was applied to analyze changes in gene expression in a given sample relative to a reference sample in comparison to an endogenous control gene. The sample collected at Day 0 when mES cells were undifferentiated was used as a reference sample, whereas *gapdh* was used as an endogenous control gene.

The mathematical model proposed by Pfaffl (2001) was used to determine changes in the relative expression levels of all target genes in comparison to *gapdh*. This model can be used when PCR efficiencies (E) of different targets are different. The relative expression ratio (R) of a target gene is calculated based on E and the Ct deviation of an unknown sample versus a control sample, and expressed in comparison to a reference gene.

$$R = \frac{(E_{target})^{\Delta Ct_{target}\ (control\ -sample)}}{(E_{ref})^{\Delta Ct_{ref}\ (control\ -sample)}}$$

- E_{target} is the PCR efficiency of the gene target being analyzed. The PCR efficiency for each target was determined by setting up standard curve experiments for each target and *gapdh*.
- E_{ref} is the PCR efficiency of *gapdh*.
- The Ct value is the point at which fluorescence rises above the background fluorescence.
- ΔCt_{target}(control-sample) is the deviation between the Ct value of the sample you are analyzing and Ct value of the control sample (day 0) for each specific target.
- ΔCt_{ref}(control-sample) is the deviation between the Ct value of the sample you are analyzing and Ct value of the control sample (day 0) for *gapdh*.

4 Results

4.1 Purification of ALKBH1

Later experiments in this thesis required purified ALKBH1 protein and this work therefore begun with purification of recombinant ALKBH1. Recombinant ALKBH1 protein was over-expressed in BL21-CodonPlus® (DE3)-RIL cells with the use of either pET28a or pTYB12 expression vectors followed by affinity chromatography with either HIS-Select® Cobalt Affinity Gel or the IMPACT™ system, respectively. The protocols for protein purification are outlined in the Method section. In brief, HIS-Select® Cobalt Affinity Gel contains beaded agarose charged with cobalt (Co^{2+}) was used to purify recombinant histidine-tagged ALKBH1. Elution of ALKBH1 from the column is achieved when imidazole is added to the column. Additionally, the IMPACT™ system was used to purify recombinant intein-tagged ALKBH1 as the intein tag contains a chitin binding domain with high affinity towards a chitin column. The intein tag undergoes self-cleavage from ALKBH1 when DTT is added to the column.

Different culture volumes and duration times for IPTG induction, at 18˚C and 37˚C, were testes in order to determine the optimal purification procedure. Affinity chromatography eluted protein samples were dialyzed and concentrated prior to ion exchange chromatography and both cation and anion exchange columns were tested. The presentation of the results in 4.1.1-3 follows the experimental progress made during the thesis work.

4.1.1 IMPACT™ Purification and RESOURCE™ Q Anion Exchange Chromatography

The first purification of ALKBH1 was performed using a pTYB12-ALKBH1 BL21 (DE3)-RIL glycerol stock. ALKBH1 had previously been cloned into the pTYB12 vector by fusion of the N-terminal to the intein tag to create pTYB12-ALKBH1. A 6 L culture was prepared and IPTG (0.5 mM) induction was performed at 37˚C for 1 hour as described in Methods. On-column cleavage of the intein tag was performed ON at 4˚C and samples were eluted in nine 1 ml fractions. SDS-PAGE analysis of samples collected during purification and eluted fractions, are presented in figure 4.1. However, this strategy did not lead to visible amounts of ALKBH1, which is expected to appear at around 44 kDa (ALKBH1 MW is 43.8 kDa).

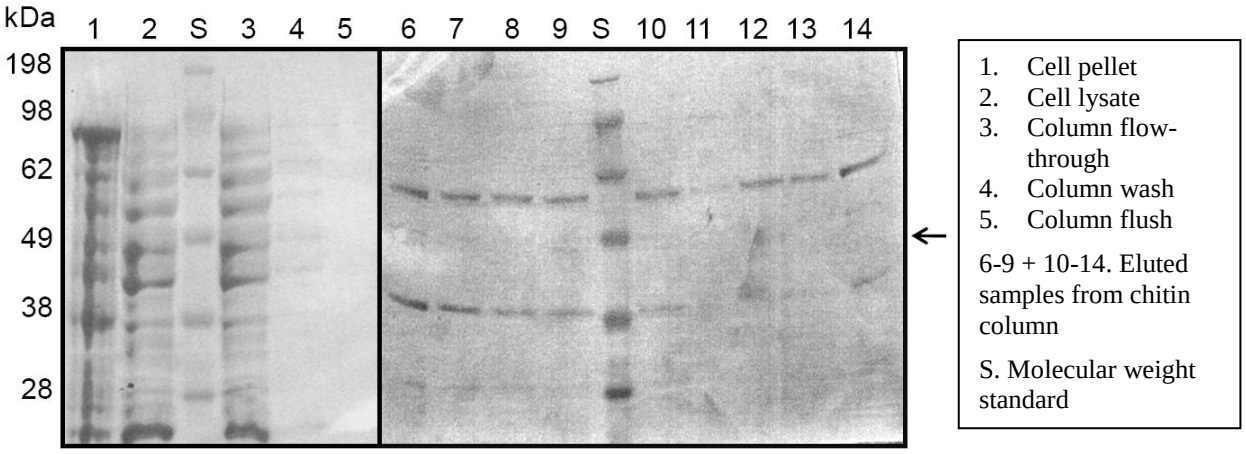


Figure 4.1: SDS-PAGE analysis of intein-tagged ALKBH1 expressed from pTYB12 and purified by the IMACT™ system. Expression of ALKBH1 in *E. coli* (6 L culture) was induced by 0.5 mM IPTG for 1 hour at 37˚C. On-column cleavage of intein tag and elution of ALKBH1 was performed ON at 4˚C with 50 mM DTT. Arrow indicates the expected position of purified ALKBH1 and the content of each lane is indicated on the right hand side**.**

In order to possible identify ALKBH1 in the nine eluted fractions in figure 4.1 (lanes 6-9 and 10-14) these were dialyzed in Low Salt Buffer, pH 8, and concentrated. If samples need to be purified further by ion exchange chromatography they are usually loaded onto an ion exchange column under condition of low ionic strength, thus a low salt buffer was used in dialysis. Following this scheme a band possible representing ALKBH1 could be observed after upconcentration (figure 4.2, right panel). Thus proteins upconcentrated by this protocol were fractionated further by RESOURCE™ Q anion exchange chromatography on the ÄKTA explorer FPLC system. SDS-PAGE analysis of the eluted fractions did not reveal visible amounts of the putative ALKBH1 protein (data not shown).

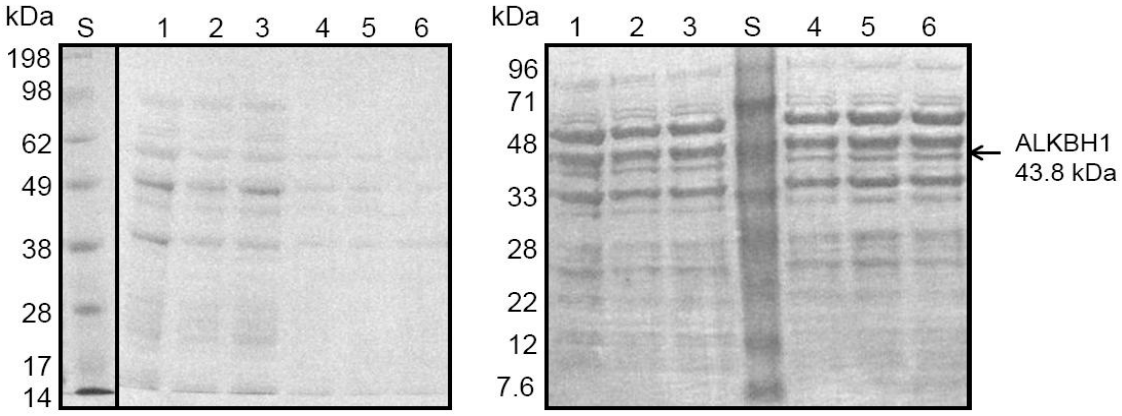


Figure 4.2: Dialyzed and upconcetrated protein fractions. SDS-PAGE analysis of dialyzed protein samples in lanes 1 – 6 (left) and concentrated protein samples in lanes 1 – 6 (right). These samples had been purified by the IMPACT™ system prior to dialysis and concentration. Lanes 1-6 (both panels) represents fractions 1-9 (lanes 6-9 and 10-14 in figure 4.1) that had been pooled, dialyzed and upconcentrated using six Amicon® Ultra-4 centrifugal filter devices as described in Methods. S: molecular weight standard.

We were unable to achieve successful purification of ALKBH1 using the above purification procedure based on the IMPACT™ system. An alternative purification procedure using the HIS-Select® Cobalt Affinity Gel was therefore performed.

4.1.2 HIS-Select® Purification and RESOURCE™ S Cation Exchange Chromatography

A pET28a-ALKBH1 plasmid DNA was first transformed into BL21-(DE3)-RIL competent cells as no pET28a-ALKBH1 BL21-(DE3)-RIL glycerol stock was available. Plasmid pET28a-ALKBH1 contains the *ALKBH1* gene in frame with a histidine tag at the N-terminal. A 12 L culture was prepared and IPTG (0.5 mM) induction was performed ON at 18˚C as described in Methods. ALKBH1 was eluted from the HIS-column with 250 mM imidazole and samples collected in ten 1 ml fractions analyzed by SDS-PAGE, figure 4.3. A band representing ALKBH1 was clearly observed around 44 kDa indicating successful expression and purification.

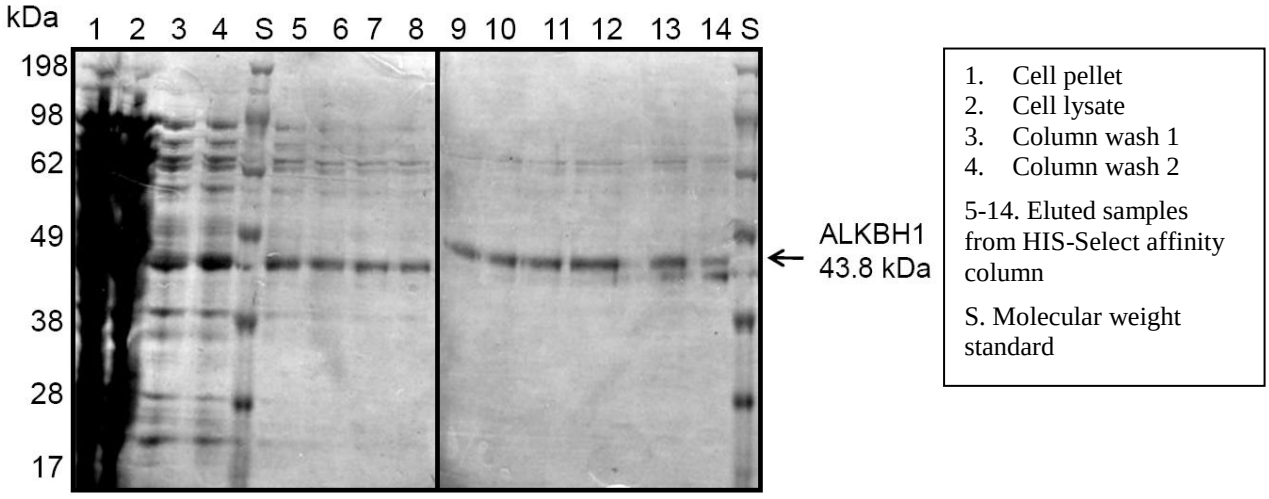


Figure 4.3: SDS-PAGE analysis of histidine-tagged ALKBH1 expressed from pET28a and purified by HIS-Select® cobalt affinity gel. Expression of ALKBH1 in *E. coli* (12 L culture) was induced by 0.5 mM IPTG ON at 18˚C. ALKBH1 was eluted from the column with 250 mM imidazole. The content of each lane is indicated on the right hand.

The 10 fractions shown in figure 4.3 (lanes 5-14) were pooled, dialyzed in Low Salt Buffer, pH 6, and concentrated prior to further purification by RESOURCE™ S cation exchange chromatography on the ÄKTA explorer FPLC system. SDS-PAGE analysis of the eluted fractions did not reveal visible amounts of the ALKBH1 protein (data not shown). It was suspected that the column used was either contaminated or exhausted as the sample contained ALKBH1 prior to loading.

An up-scaling of the above HIS-Select® Cobalt Affinity Gel purification procedure was then performed using a 24 L culture. Eight 1 ml fractions eluted from the HIS-column were individually concentrated immediately after elution from the column using 500 µl Vivaspin Centrifugal Filter Tubes as described in Methods. SDS-PAGE analyses of upconcentrated fractions are presented in figure 4.4. High yields of ALKBH1 were apparent in all fractions. All eight fractions were pooled and dialyzed in Low Salt Buffer, pH 6, using six Amicon® Ultra-4 centrifugal filter devices as described in Methods. SDS-PAGE analysis of dialyzed fractions is presented in figure 4.5. Although the purity of ALKBH1 was sufficient for subsequent experiments, an optimized purification procedure using the IMPACT™ system was performed. Expression of ALKBH1 proved to be more efficient by IPTG induction ON at 18˚C as observed from SDS-PAGE analysis of the HIS-Select® Affinity Cobalt Gel purification procedure. This was therefore applied in the following optimized purification procedure.

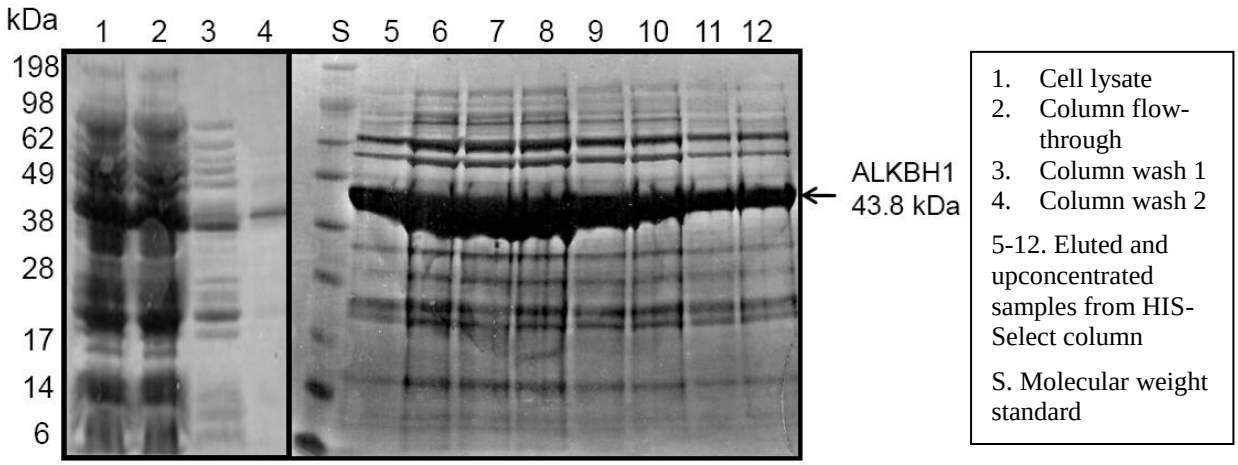


Figure 4.4: SDS-PAGE analysis of concentrated histidine-tagged ALKBH1 samples expressed from pET28a and purified by HIS-Select® cobalt affinity gel. Expression in *E. coli* (24 L culture) was induced by 0.5 mM IPTG ON at 18˚C. ALKBH1 was eluted from column with 250 mM imidazole. The content of each lane is indicated on the right hand side.

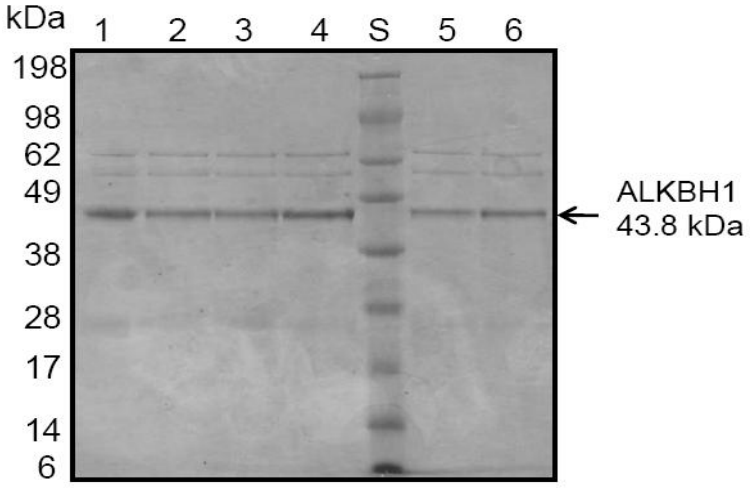


Figure 4.5: SDS-PAGE analysis of dialyzed concentrated protein samples purified by HIS-Select® cobalt affinity gel. Lanes 1-4 and 5-6 represent fractions 1-8 (lanes 5-12 in figure 4.4) that were pooled and dialyzed using six Amicon® Ultra-4 centrifugal filter devices. S: molecular weight standard

4.1.3 Optimized IMPACT™ Purification and HiTrap™ SP HP Cation Exchange Chromatography

Taking advantage of the protocol used in 4.1.2, an optimized IMPACT™ purification procedure was prepared in which a 6 L culture was induced with IPTG (0.5 mM) ON at 18˚C as described in Methods. On-column cleavage of the intein tag was performed for 48 hours at 4˚C and samples were eluted in ten 1 ml fractions. A sample of the chitin resin was also taken after elution of ALKBH1. SDS-PAGE analysis of samples collected during purification, eluted fractions are presented in figure 4.6. High yields of ALKBH1 of good purity were observed in bands around 44 kDa indicating that this was a successful purification procedure.

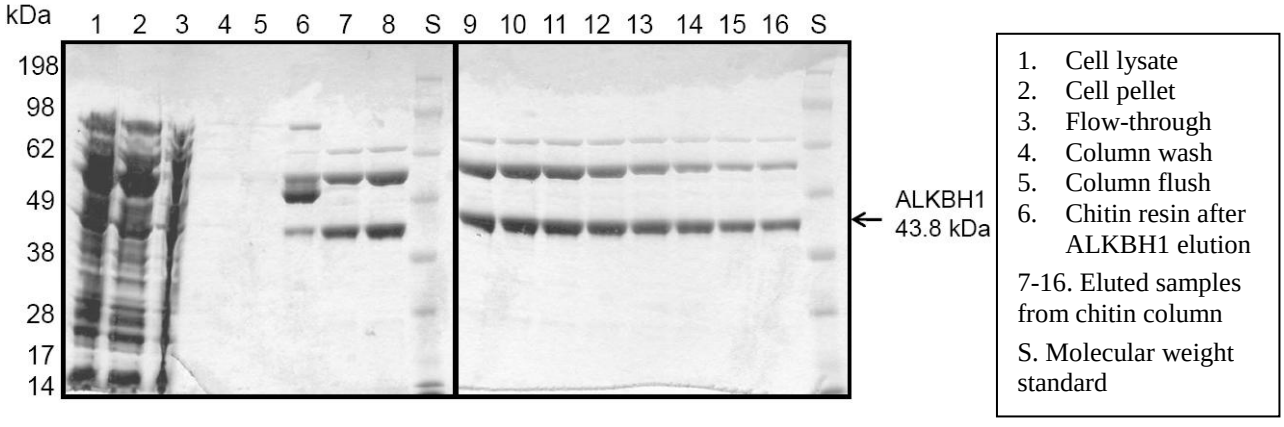


Figure 4.6: SDS-PAGE analysis of intein-tagged ALKBH1 expressed from pTYB12 and purified by the IMACT™ system. Expression in *E. coli* (6 L culture) was induced by 0.5 mM IPTG ON at 18˚C. On-column cleavage of intein tag was performed for 48 hours at 4˚C with 50 mM DTT. The content of each lane is indicated on the right hand side.

Protein samples in lanes 7 – 16 (figure 4.6) were pooled and dialyzed in Low Salt Buffer, pH 6, prior to HiTrap™ SP HP cation exchange chromatography using the ÄKTA purifier FPLC system. A fresh HiTrap™ SP HP column was used to avoid possible contamination or exhaustion of the column. Cation exchange chromatography was performed in two separate trials to ensure successful purification before using the entire protein sample. ALKBH1 was eluted from the column using a linear gradient of increasing salt concentrations. Column flow-through and column wash samples were analyzed by SDS-PAGE together with eluted fractions. Results presented in figure 4.7 show superior separation of ALKBH1 with impurities eluted in the wash and flow-through samples.

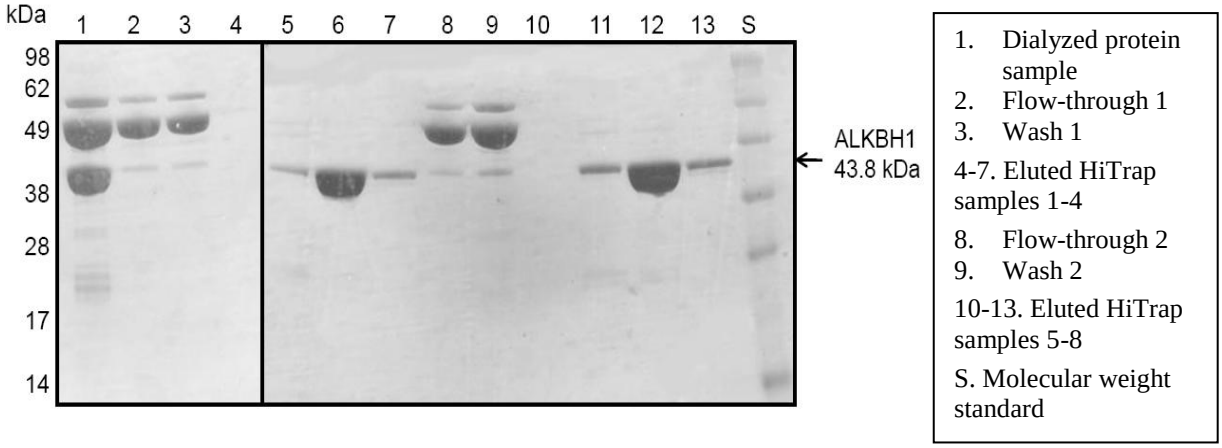


Figure 4.7: SDS-PAGE analysis of protein samples purified by HiTrap™ SP HP cation exchange chromatography. Loaded protein sample had been purified by the IMPACT™ system and dialyzed in low salt buffer. Elution was achieved by applying a linear gradient of increasing ionic strength. The content of each lane is indicated on the right hand side. Flow-through 1 and wash 1 samples are from one HiTrap separation with elution of ALKBH1 in samples 1-4 (lanes 4-7), whereas flow-through 2 and wash 2 samples are from a second HiTrap separation with elution of ALKBH1 in samples 5-8 (lanes 10-13).

4.2 Protein Concentration by Bio-Rad Protein Assay

The protein concentration of purified protein samples in lanes 5-7 and 11-13 (figure 4.7) was determined by the Bio-Rad Protein Assay. Table 4.1 shows the mean measured absorbance values of the respective samples (prepared in triplicates) together with standard deviation (SD) and the calculated protein concentration (mg ml^{-1}). Figure 4.8 shows the standard curve that was made from prepared dilutions of BSA.

Table 4.1: Absorbance at 595nm of ALKBH1 protein samples

	Sample 5	Sample 6	Sample 7	Sample 11	Sample 12	Sample 13
A_{595} replicate 1	0.069	0.188	0.072	0.071	0.206	0.103
A_{595} replicate 2	0.065	0.186	0.072	0.081	0.164	0.069
A_{595} replicate 3	0.065	0.189	0.070	0.085	0.189	0.065
Average A_{595}	0.0663	0.187	0.071	0.079	0.186	0.079
SD	0.002	0.002	0.001	0.007	0.021	0.021
Concentration (mg ml^{-1})	0.333	1.158	0.367	0.419	1.149	0.4187

The concentration of protein samples 5-7 and 11-13 was calculated from the equation of the standard curve: y = 1,4697x + 0,1747.

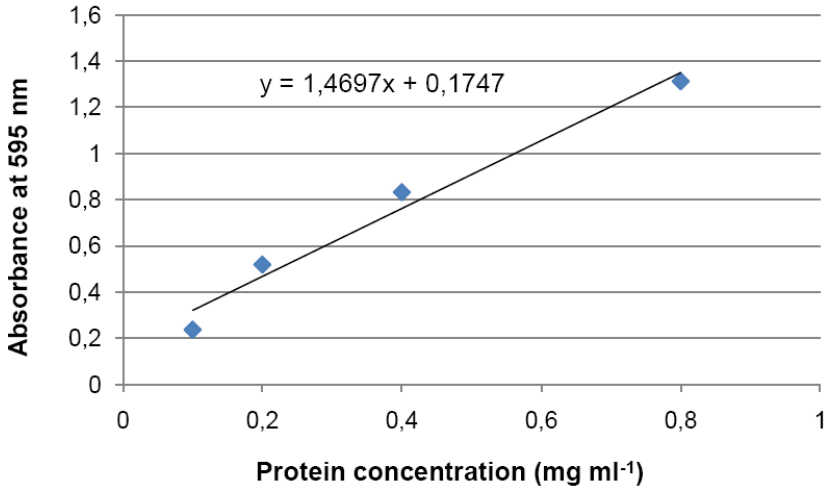


Figure 4.8: Standard curve for Bio-Rad protein assay using bovine serum albumin. Concentration of protein samples of unknown concentration is calculated from the equation. Example of protein sample 5 in table 4.1: Protein concentration (mg ml^{-1}) = ((Abs595*dilution)-0.1747)/1.4697 = ((0.066333*10)-0.1747)/1.4697 = 0.332471 mg ml^{-1}

4.3 Biochemical Analysis of ALKBH1 by Succinate Formation

The putative enzymatic activity of purified recombinant ALKBH1 protein was addressed using two different approaches. First, increasing concentrations of ALKBH1 (3 µM – 12 µM) was incubated with 2-ketoglutarate [5-^{14}C] and $FeSO_4$. If ALKBH1 is a decarboxylase, it should decarboxylate 2-ketoglurate [5-^{14}C] and form radioactive succinate in the presence of Fe^{2+}. This is called the un-coupled reaction, since no biological substrate is present. The *E. coli* AlkB protein (3 µM) was used as a positive control. Formation of succinate was determined by measuring the intensity of radiation emitted, expressed in dpm (Methods). Succinate formation by ALKBH1 in the uncoupled reaction was performed up to eight times,

with results from one of the trials presented in figure 4.9. Reaction mixtures were prepared in triplicates and the mean dpm value ± SD (n = 3) is presented in figure 4.9, which represents one of the trials. DMP values were corrected against the mean dpm background value (reaction mixture without any enzyme). An increase in dpm values indicate increased decarboxylation of 2-OG, whereas negative dmp values indicate no catalytic activity. High variance between replicates is indicated by large SD bars. The positive control enzyme, AlkB, showed successful decarboxylation of 2-ketoglutarate to succinate

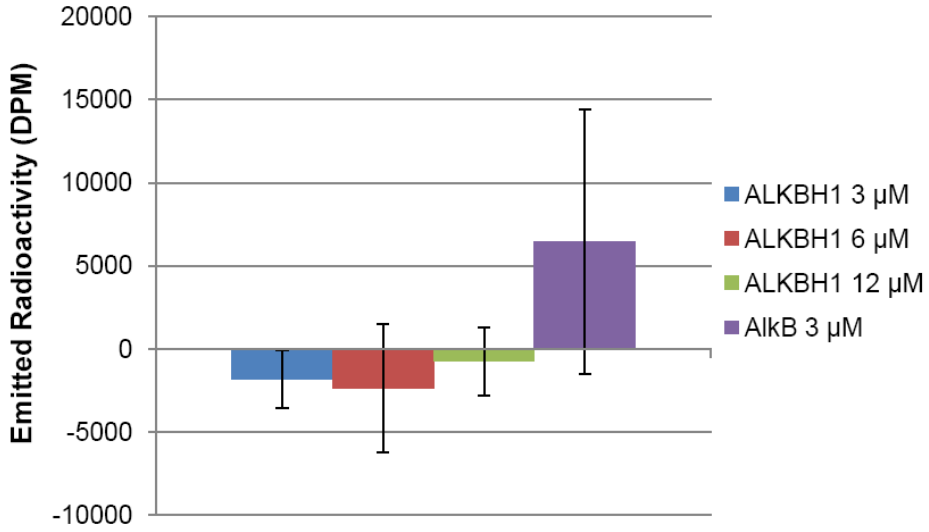


Figure 4.9: Succinate formation. Decarboxylation catalyzed by ALKBH1 and AlkB in the presence of 80 µM $FeSO_4$ and 160 µM 2-OG (10 % 2-Ketoglutaric acid [5-^{14}C]) is shown. 2-OG was precipitated with 2,4 DPNH and radioactive succinate was measured by scintillation counting. **by** Error bars represent mean ± SD (n = 3).

In the second approach it was tested whether cH2A (calf histone H2A), a putative substrate of ALKBH1 (results from another research group and therefore not shown) would increase decarboxylation of 2-ketoglutarate above the rate of the uncoupled reaction. AlkB homologue 2 from *S. pombe* was used as a positive control upon addition of cH2A to the reaction. 6 µM ALKBH1 was incubated with 2, 4, 8 or 10 µg of substrate cH2A and 6 µM AlkB homologue 2 from *S. pombe* was incubated with 2 or 4 µg of substrate cH2A. Formation of succinate was determined by measuring the intensity of radiation emitted, expressed in dpm (Methods). Reaction mixtures were prepared in triplets and the mean dpm value ± SD (n = 3) is presented in figure 4.10. DMP values were corrected against the mean dpm background value (reaction mixture without any enzyme). An increase in dpm values indicate increased decarboxylation upon addition of cH2A, whereas negative dmp values indicate no catalytic activity. High variance between replicates is indicated by large SD bars and makes interpretation of results difficult.

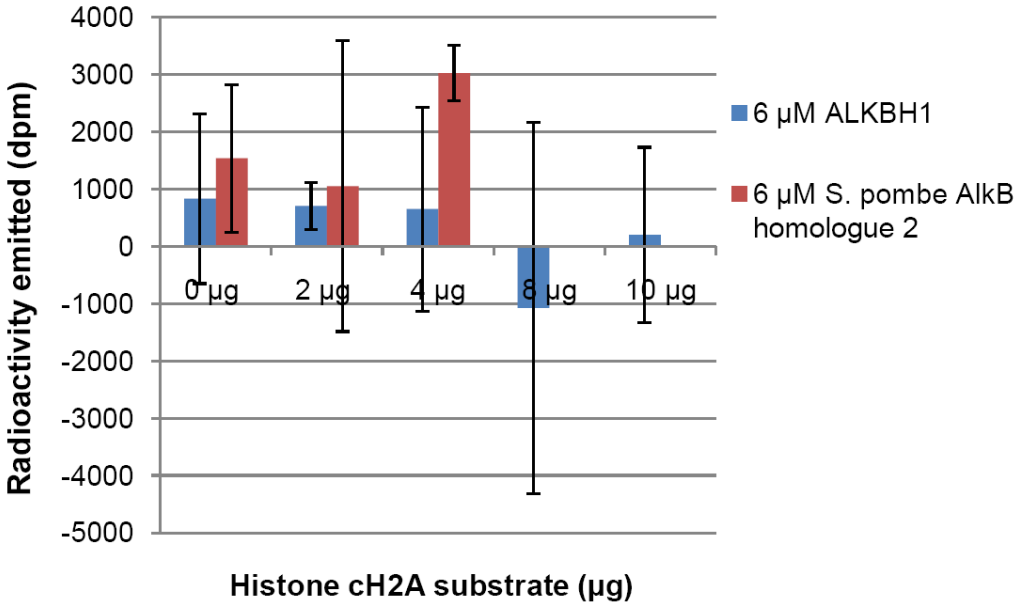


Figure 4.10: Succinate formation by ALKBH1 and *S. pombe* AlkB homologue 2. Succinate formation in the presence of 80 µM $FeSO_4$, 160 µM 2-OG (10 % 2-Ketoglutaric acid [5-^{14}C]) was tested with increasing concentrations of cH2A. 2-OG was precipitated with 2,4 DPNH and radioactive succinate was measured by scintillation counting. Error bars represent mean ± SD (n = 3).

4.4 Protein-Protein Interaction

Protein-protein interactions between ALKBH1 and core regulators of ES cell pluripotency: NANOG, OCT4, and SOX2, in addition to histones H1, H2A, H2B, H3 and H4 were tested by Dot-Blot immunobinding assay and an adapted ELISA (Methods). Dot-Blot immunobinding assay was used as a rough estimate of protein-protein interaction whereas ELISA was used for further quantification using an increasing concentration of ALKBH1.

4.4.1 Dot-Blot Immunobinding Assay

All proteins indicated above were individually spotted onto nitrocellulose membranes that were incubated with or without ALKBH1 to investigate protein-protein interactions. For each protein tested two membranes were spotted with 5 ng of four different proteins: the test protein, two negative control proteins (BSA and lysozyme) and ALKBH1. Protein-protein interactions were determined by the use of primary and secondary antibodies as described (Methods). This assay was performed up to three times for each protein and results presented in figure 4.11 represents one representative assay. Upper panels show membranes that were incubated with both ALKBH1 protein and a primary antibody (ab18525) directed against ALKBH1. Lower panels show membranes that were only incubated with the primary antibody. Positive results produced only at the ALKBH1 spots in lower panels indicate that

there was no background binding between the primary ALKBH1 antibody and the other protein spots. Negative results between ALKBH1 and control proteins (BSA and lysozyme) in upper panels indicate that the positive results between ALKBH1 and NANOG, OCT4, SOX2 and histones H2A and H4 are true protein-protein interactions. Interactions between ALKBH1 and histones H1, H2B and H3 did not show any positive results.

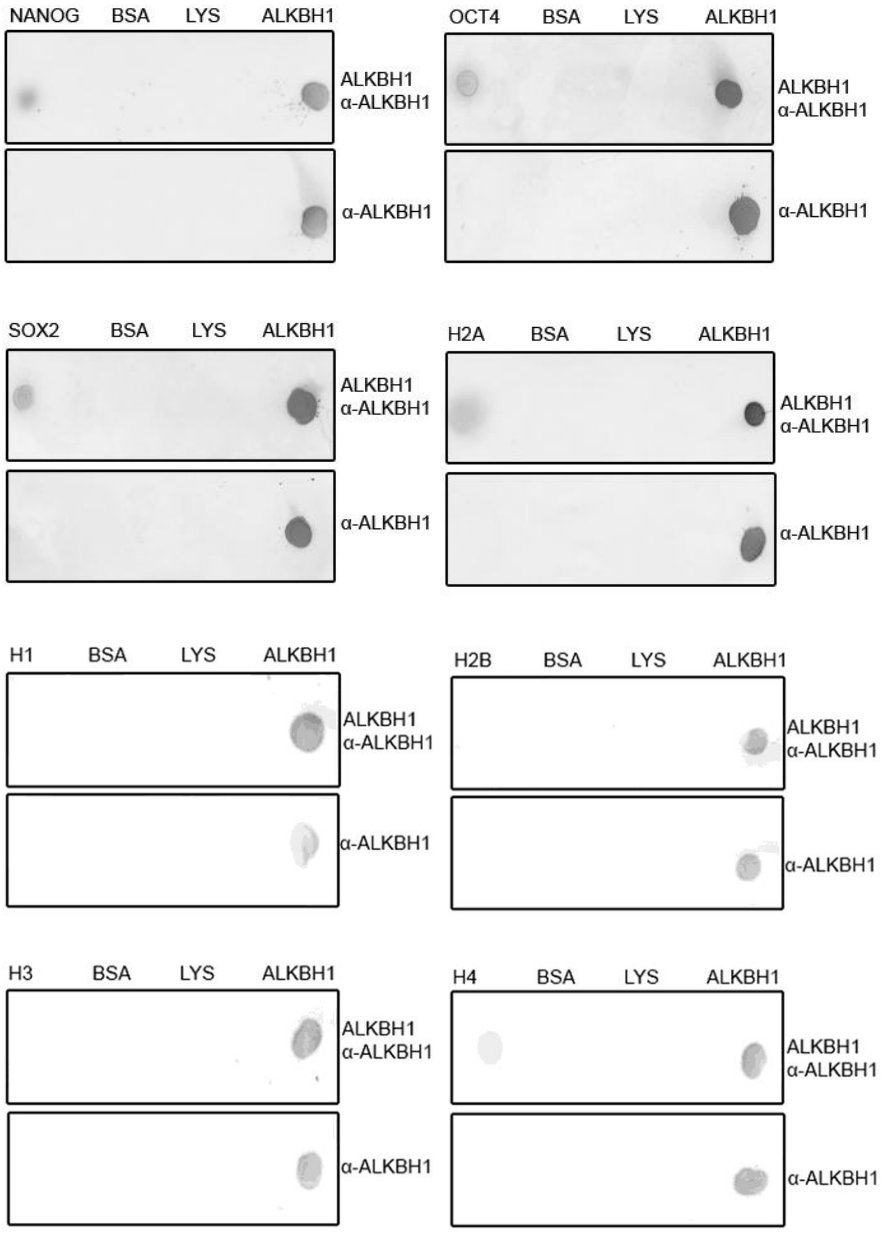


Figure 4.11: Protein- protein interactions by Dot Blot immunobinding assay. Proteins were spotted onto a nitrocellulose membrane that were incubated with ALKBH1 and primary antibody against ALKBH1 (α-ALKBH1) in upper panels for each primary target protein (NANOG, OCT4, SOX2, H2A, H1, H2B, H3 or H4) or with just α-ALKBH1 in lower panels for each primary target proteins. Protein-protein interactions were determined by a secondary antibody conjugated with alkaline phosphatase as described (Methods).

4.4.2 Enzyme-linked Immunosorbent Assay

In this assay, the same proteins that were analyzed in the Dot-Blot assay above were investigated further. Proteins (62.5 fmol) were bound to a Covalink plate as described in Methods and incubated with increasing concentrations of ALKBH1 (0 – 2.5 M). Protein-protein interactions were determined by the use of primary and secondary antibodies as described in Methods. The assay was performed two times for proteins that had produced positive interaction in the Dot-Blot assay (NANOG, OCT4, SOX2, H2A and H4). Results are presented in figure 4.12 with average absorbance values from the two trials. Proteins that showed no interaction with ALKBH1 in the Dot-Blot assay (H1, H2B and H3) were also tested, however these interaction were only tested by one ELISA with results presented in figure 4.13. Positive interactions were observed between ALKBH1 and NANOG, OCT4, SOX2, H2A, H2B and H3 (figure 4.12 and 4.13), whereas a weak interaction was observed between ALKBH1 and H4 (figure 4.12), and no interaction was observed between ALKBH1 and H1 (figure 4.13). As the concentration of ALKBH1 increases, so does the absorbance value. The absorbance value relates to the color reaction produced by alkaline phosphatase as described in Methods.

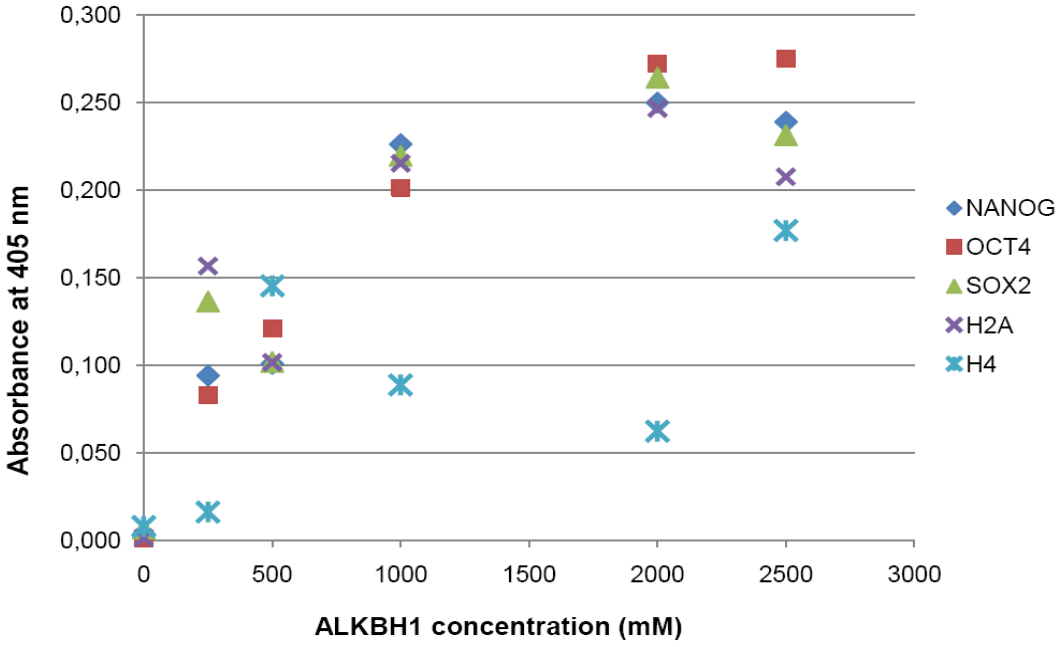


Figure 4.12: Enzyme-linked immunosorbent assay measuring the affinity of ALKBH1 towards core regulators of embryonic stem cell pluripotency (NANOG, OCT4 and SOX2) and histones H2A and H4. 62.5 fmol of NANOG, OCT4, SOX2, H2A and H4 was bound to a Covalink plate and incubated with increasing concentrations, as indicated, of ALKBH1. Protein-protein interactions were determined with the use a primary and a secondary antibody conjugated with alkaline phosphatase. The absorbance values relate to the color reaction produced by alkaline phosphatase and is an indirect measurement of protein interactions.

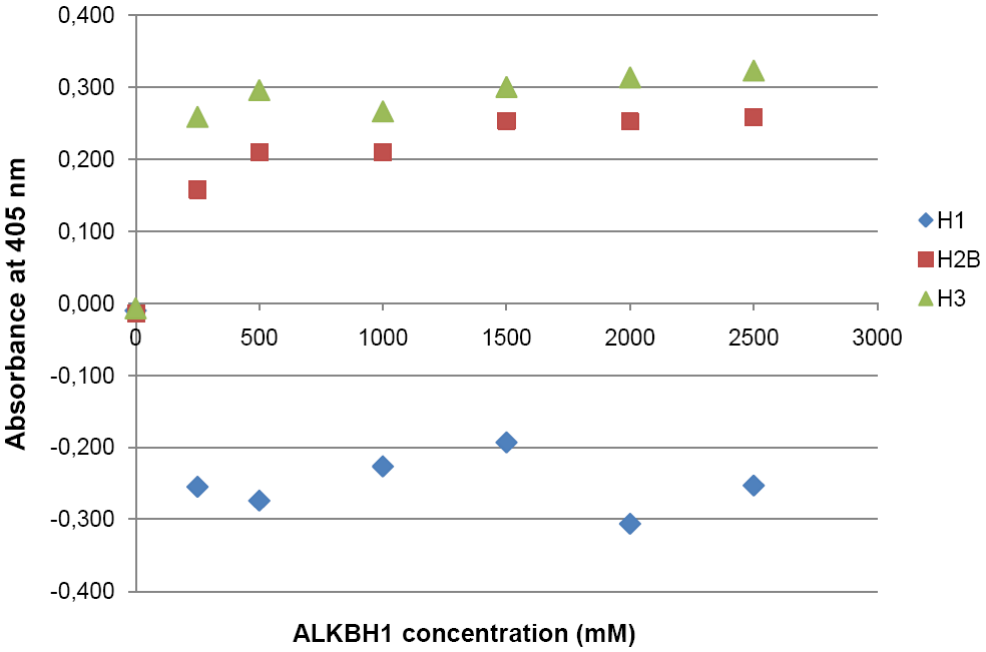


Figure 4.13: Enzyme-linked immunosorbent assay measuring the affinity of ALKBH1 towards and histones H1, H2B and H3. 62.5 fmol of H1, H2B and H3 was bound to a Covalink plate and incubated with increasing concentrations, as indicated, of ALKBH1. Protein-protein interactions were determined with the use a primary and a secondary antibody conjugated with alkaline phosphatase. The absorbance values relate to the color reaction produced by alkaline phosphatase and is an indirect measurement of protein interactions.

4.5 Bioinformatics Analysis of DNA regions bound by NANOG and OCT4

In the study by Boyer et al. (2005) target gene sequences bound by NANOG and OCT4 were identified in human ES cells. They found that both NANOG and OCT bound to a region overlapping *ALKBH1* exon 1 and the orf156, both located on chromosome 14 (C14) (figure 4.14). In order to retrieve these DNA regions some basic bioinformatics tools and databases were used. The nucleotide sequence from position 77243001 to 77244620 of C14 was retrieved from Build 35, 36 and 37 of the human genome as described in Methods. Nucleotide sequences were submitted to the Multiple Sequence Alignment tool ClustalW2 to verify that they were identical. The MSA output presented in appendix 4 indicated that the three builds are not identical for this sequence. However, Build 35 and 36 are identical for this sequence as shown in the MSA output in appendix 4.

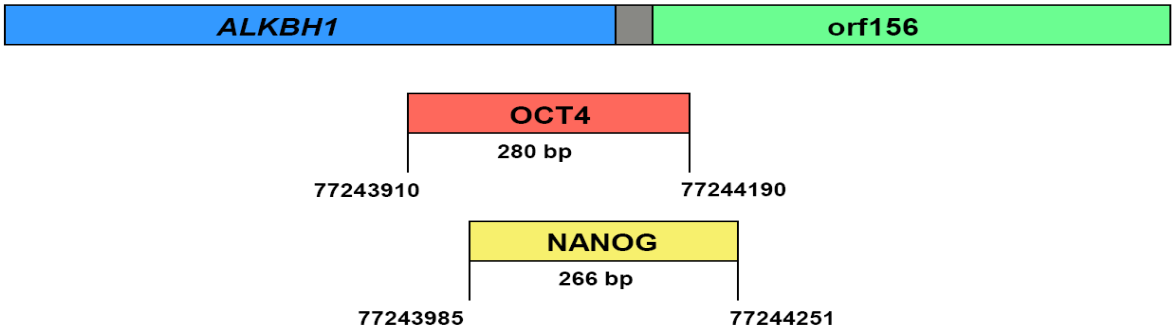


Figure 4.14: DNA regions bound by OCT4 and NANOG with indicated nucleotide position and the length of the DNA region (bp). Illustration shows genome regions of *ALKBH1* found in a study by Boyer et al. (2005) to be bound by OCT4 and NANOG using chromatin immunoprecipitation coupled to DNA microarrays. Both proteins bind to a region overlapping *ALKBH1* exon 1 and orf156 on chromosome 14.

The sequences from Build 36 or 37 were then used as a query sequences in two separate nBlast search against the "Human genomic + transcripts" database. In figure 4.15 the nBlast results clearly show that the sequence from Build 36 overlaps with the genomic regions, ALKBH1 and C14orf156, found by Boyer et al. (2005). This sequence was therefore used for designing primers for PCR amplification of the DNA regions bound by NANOG and OCT4 as described in Methods. Nucleotide sequences and primers can be found in appendix 3. Results from the nBlast search using the sequence from Build 37 are presented in appendix 5. This sequence did not overlap the regions found by Boyer et al. (2005).

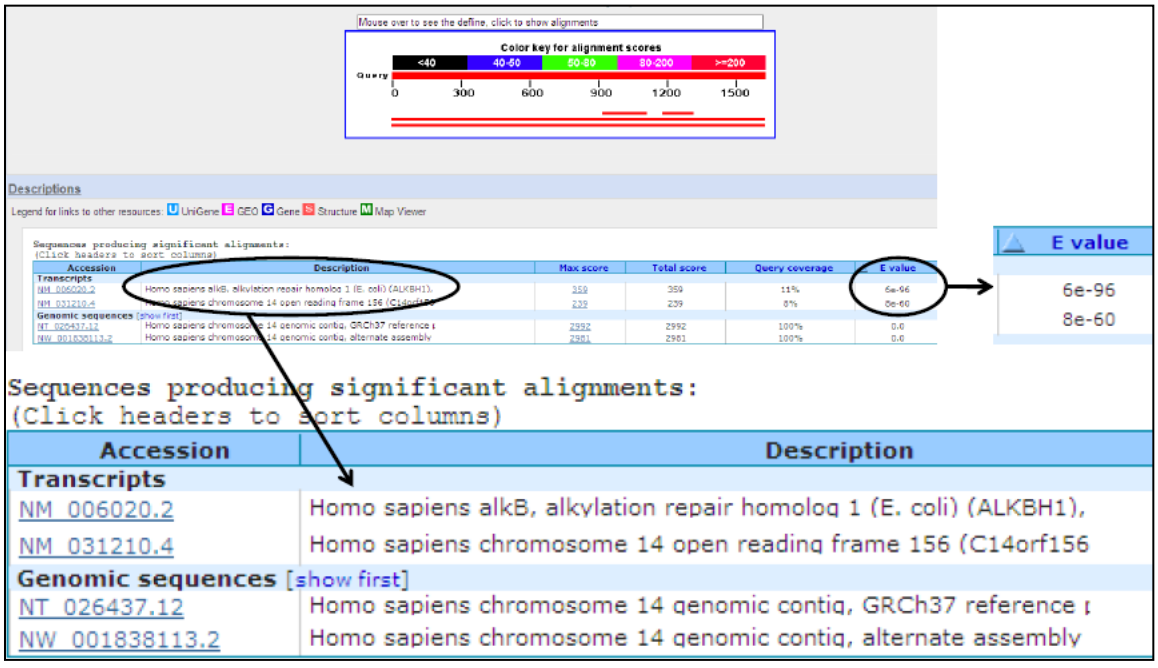


Figure 4.15: nBlast results showing hits towards ALKBH1 and C14orf156 when the nucleotide sequence from Build 36 of the human genome (position 77243001 to 77244620) on chromosome 14) was used as the query sequence. E-values of 6e-96 and 8e-60 indicate very significant hits. The E-value is the expected value in which the query sequence can produce this alignment by change.

4.6 PCR amplification of DNA regions bound by NANOG and OCT4

Four different primer pairs were used in the PCR amplification of the DNA regions bound by NANOG and OCT4. This was done in order to achieve a more specific amplification of desired DNA regions using primer pairs within previously amplified DNA regions. This procedure is described in more detail in Methods.

The first set of primers was designed to produce a 781 bp PCR product. A gradient PCR was used to determine optimum annealing temperature for primers and optimal DNA template concentrations. The DNA template used was genomic DNA isolated from human blood. The amplified PCR products were analyzed by agarose gel electrophoresis and presented in figure 4.16.

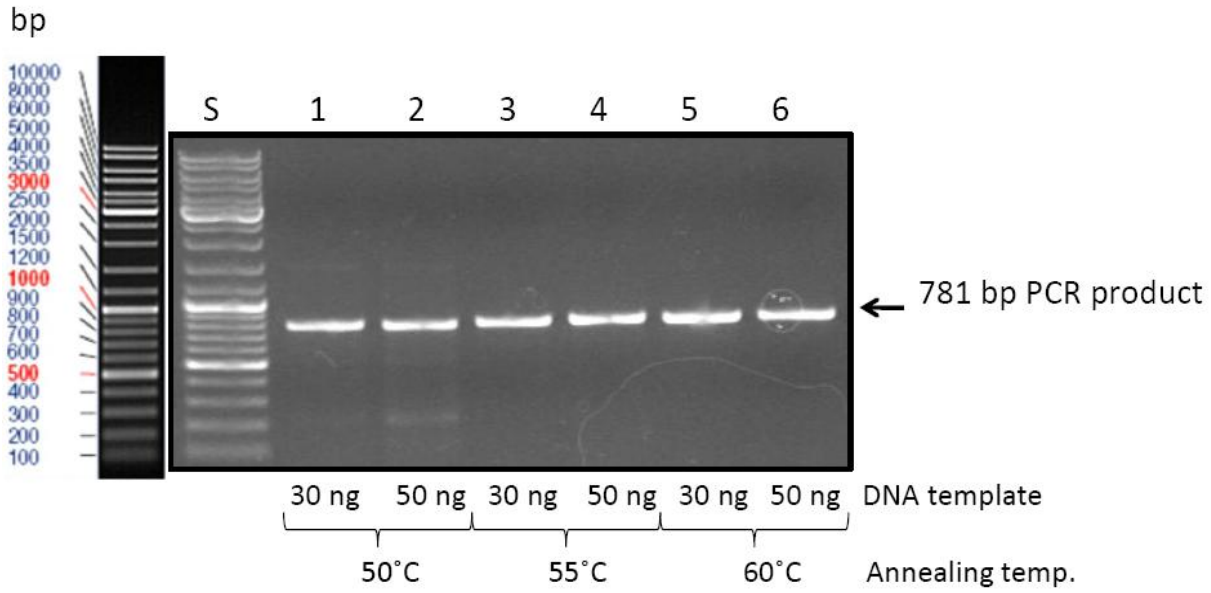


Figure 4.16: Amplification of relevant ALKBH1 promoter sequence. A 781 bp DNA product was produced by gradient PCR using different annealing temperatures and varying concentrations of DNA template, as indicated (S: DNA standard).

The PCR product in lane six (figure 4.16) was used as the DNA template for subsequent gradient PCR reactions using a second primer pair. This second primer pair was designed to produce a 434 bp PCR product within the 781 bp region. Amplified PCR products were analyzed by agarose gel electrophoresis and presented in figure 4.17.

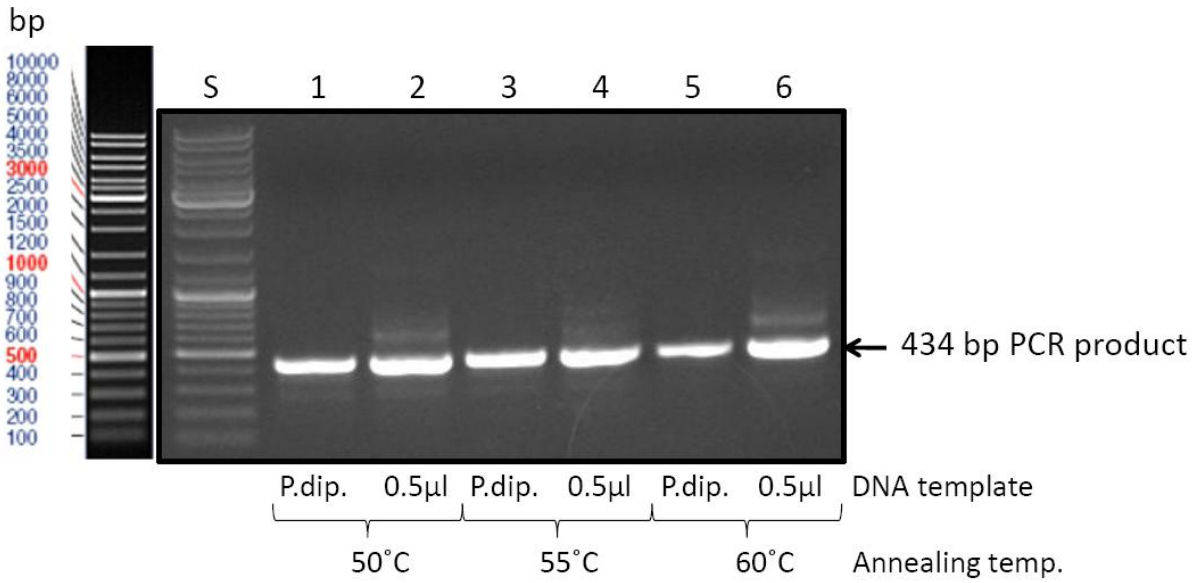


Figure 4.17: Amplification of relevant ALKBH1 promoter sequence. A 434 bp DNA product was produced by gradient PCR using different annealing temperatures and varying concentrations of DNA template, as indicated (S: DNA standard, P.dip: pipette dip.).

The PCR product in lane 5 was used as the DNA template in gradient PCR reactions with primer pairs specifically designed to amplify the precise DNA regions to which NANOG and OCT4 were reported to bind. Results from both gradient PCRs are presented in figure 4.18 with 266 bp and 280 bp PCR products for DNA regions bound by NANOG and OCT4, respectively.

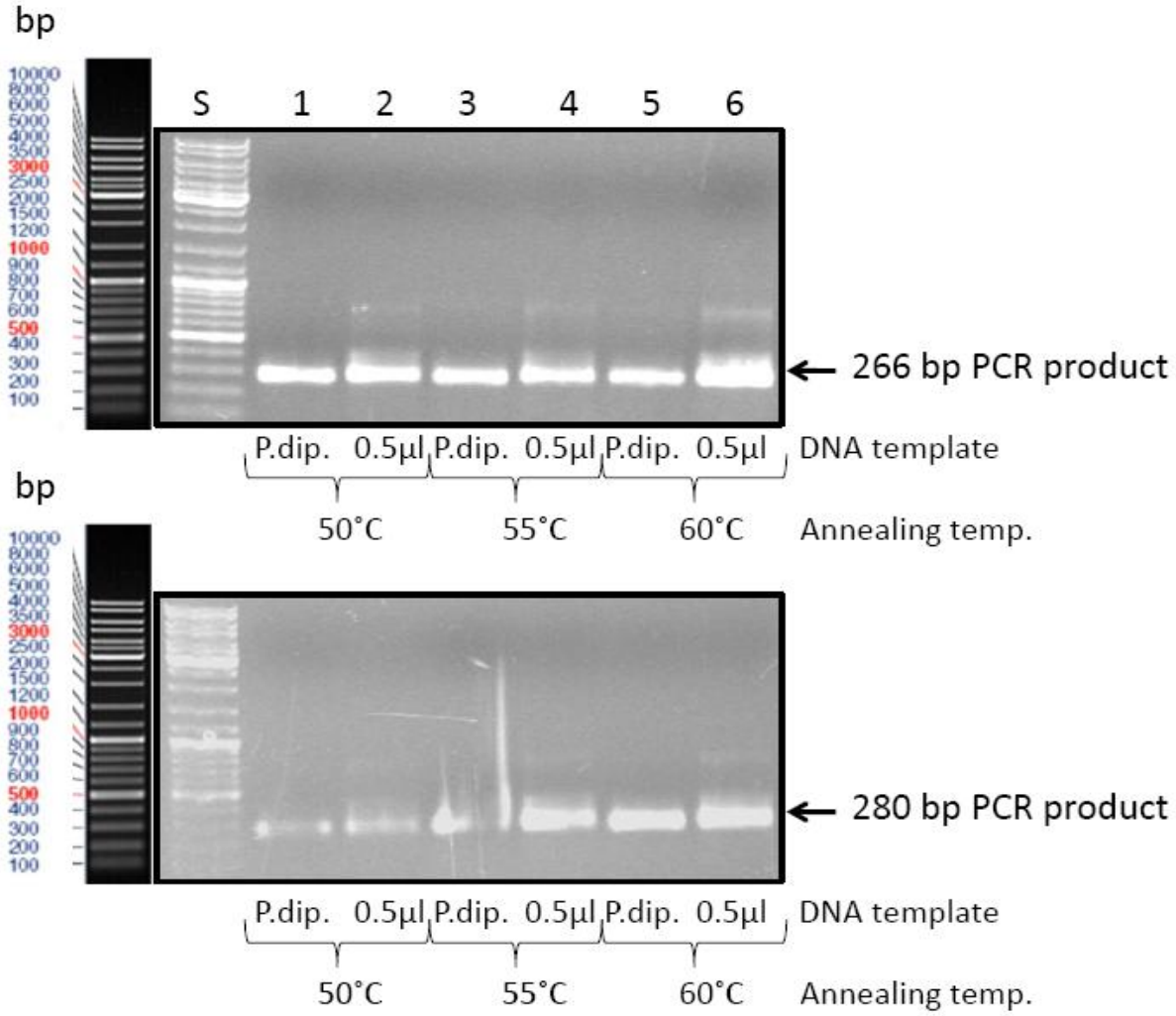


Figure 4.18: Amplification of precise *ALKBH1* DNA sequence for subsequent OCT4 and NANOG binding assays. A 266 bp and a 280 bp DNA products from gradient PCR using different annealing temperatures and varying amounts of DNA template, as indicated. The 266 bp DNA product is the amplified DNA region bound by NANOG and the 280 bp DNA product is the amplified DNA region bound by OCT4 (S: DNA standard, P.dip: Pipette dip).

The PCR products in lanes 5 from both agarose gels in figure 4.18 indicated that 60˚C was the optimal annealing temperature for both NANOG and OCT4 primer pairs. A pipette dip of DNA template was the optimal starting amount for successful amplification of these two DNA products. Thus a second traditional PCR was used to amplify the NANOG and OCT4 DNA binding sequences further using an annealing temperature of 60˚C and a pipette dip of the 434 bp DNA template. The amplified PCR products were analyzed by gel electrophoresis and results presented in figure 4.19 show successful amplification of NANOG and OCT4 DNA binding sequences with minimal, if any, amplification of unspecific PCR products.

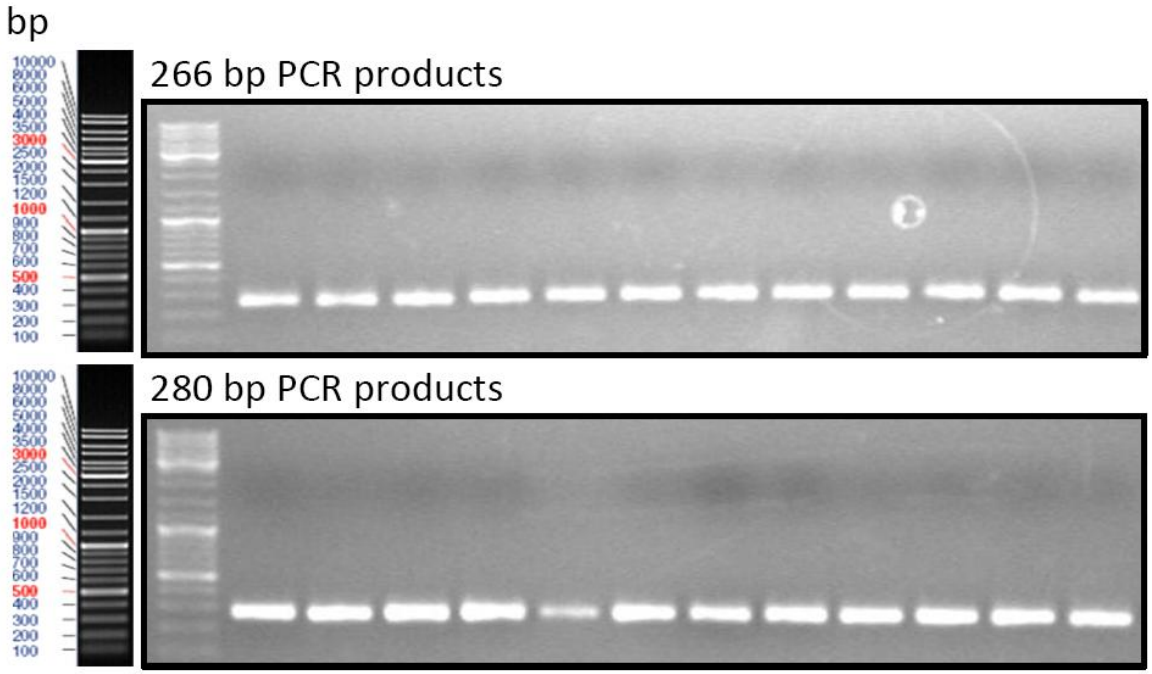


Figure 4.19: Amplification of precise *ALKBH1* DNA sequence for subsequent OCT4 and NANOG binding assays. Amplified DNA product of 266 bp and 280 bp for DNA regions bound by NANOG and OCT4, respectively, using an annealing temperature of 60˚C and a pipette dip of DNA template. Each individual PCR reaction produced PCR products of equal lengths – a 266 bp PCR product in each lane in upper panel and a 280 bp PCR product in each lane in lower panel.

The DNA sequence of PCR products in figure 4.19 were analyzed and used in an nBlast search against the "Human genomic + transcripts" database in order to validate if they were the DNA regions found by Boyer et al. (2005) to be bound by NANOG and OCT4. The results presented in figure 4.20 show that the amplified PCR products contained the sequences that overlap ALKBH1 and C14orf156.

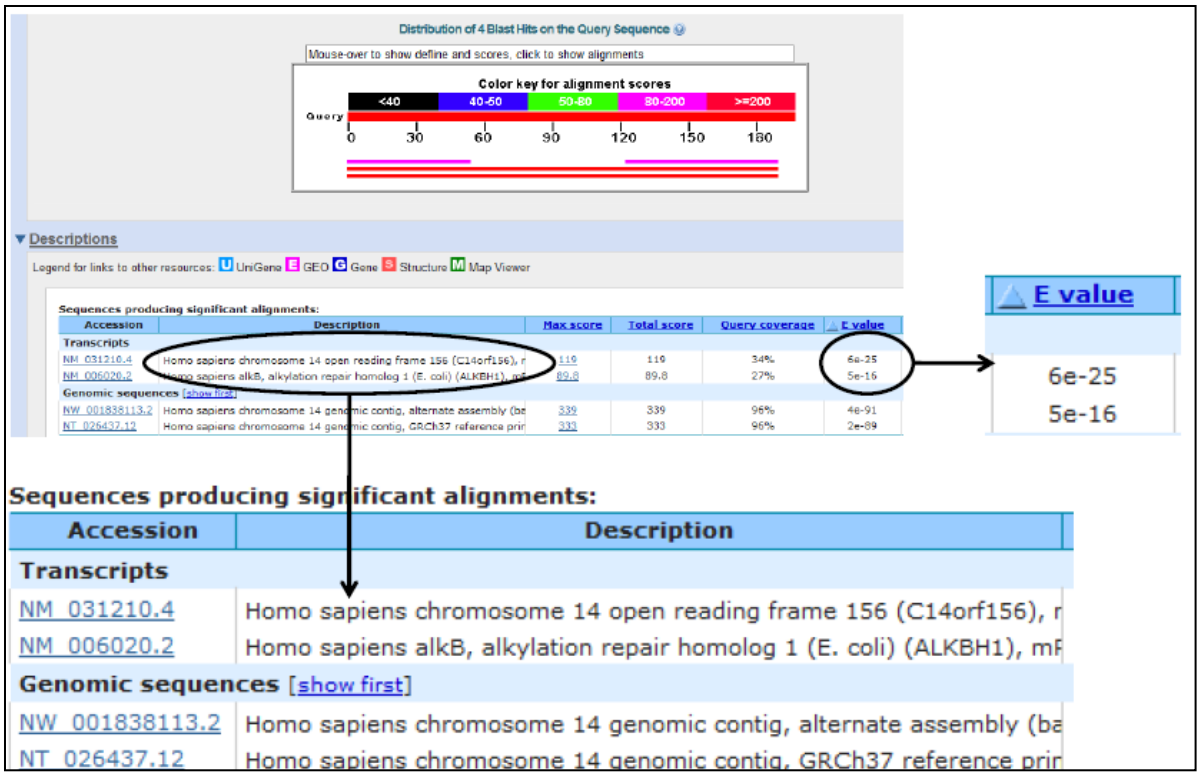


Figure 4.20: nBlast results of nucleotide sequences bound by OCT4/NANOG against the "Human and Genomic transcripts" database. Sequences overlap C14orf156 and ALKBH1 with E-values of 6e-25 and 5e-16, respectively. These hits are very significant as E-values are extremely low.

4.7 Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic Mobility Shift Assay (EMSA) was performed to investigate whether NANOG and OCT4 would bind to the DNA regions that had been amplified by PCR. The amplified DNA fragments were labeled with the radioactive isotope [γ-^{32}P]ATP and incubated with increasing concentrations of NANOG and OCT4 (0 nM – 2000 nM). Results, presented in figure 4.21, show that the EMSA procedure applied here was unsuccessful in determining whether NANOG and OCT4 bind to these DNA fragments as no apparent band shift was identified. If NANOG and OCT4 had formed a protein:DNA complex with these DNA fragments, DNA bands would have been shifted upwards with increasing concentrations of protein, as a protein:DNA complex would be retained in the gel compared to unbound DNA.

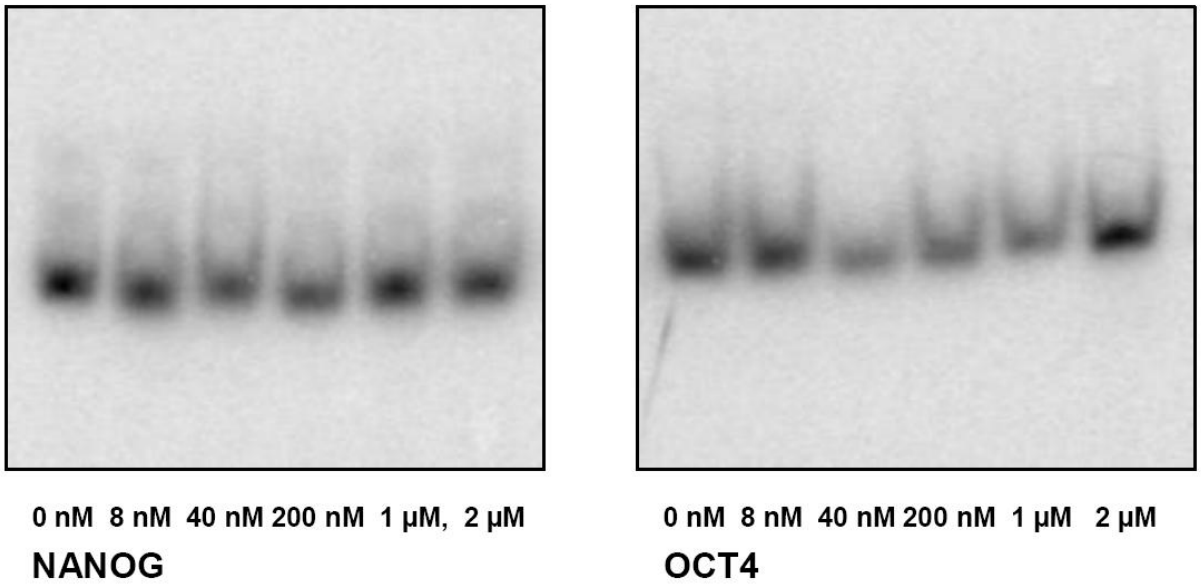


Figure 4.21: Electrophoretic Mobility Shift Assay of OCT4 and NANOG at the ALKBH1 promoter. Radiolabeled DNA probes (0.5 ng) were incubated with increasing concentrations of NANOG or OCT4 as indicated. DNA probes were amplified PCR products that contained regions previously identified as NANOG and OCT recognition sequences. Possible protein:DNA complexes were resolved in a 5 % native PAGE gel, but no apparent band shift is present indicating that proteins did not bind to their respective DNA regions.

4.8 RT-PCR Analysis of Differentiated mES cells

Previous studies have characterized the expression of multiple pluripotency markers in differentiating ES cells. It is crucial that the differentiation can be performed in a reproducible manner, thus, the expression of such markers were characterized in our wild-type ES cell line. Differentiation of mES cells were achieved by removal of LIF and by formation of embryonic bodies (EBs) or by the addition of retinoic acid (RA) to the culture medium in the absence of MEF monolayer (Methods). Samples from Day 0, 6, 9 and 12 (Day 0 is day of LIF removal) were collected from EBs and RA differentiated mES cells. RNA was extracted and reverse transcribed into cDNA. Real-time PCR analysis was performed for all samples using the following target genes:

- *gapdh1* (endogenous control housekeeping gene)
- *nanog, oct4* and *sox2* (genes encoding core transcriptional factors in ES cell pluripotency)
- *brachyury* (mesoderm specific gene)
- *gata6* (endoderm specific gene)
- *fgf5* (ectoderm specific gene)
- *Alkbh1*

The pureness of the RNA samples is listed in table 4.2, as A_{260}/A_{280} ratio, with a ratio of 2 corresponding to complete pureness.

Table 4.2: Calculated A_{260}/A_{280} of RNA samples

Sample	A_{260}/A_{280}
Day 0	2.09
Day 6 EB	2.10
Day 6 RA	2.12
Day 9 EB	2.12
Day 9 RA	2.10
Day 12 EB	2.07
Day 12 RA	2.11

EB: Embryonic bodies, RA: Retinoic acid

A series of dilutions (prepared in duplicates) ranging from 250 ng μl^{-1} to 0.25 ng μl^{-1} cDNA were prepared for standard curves experiments, resulting in a linear regression between template cDNA concentration and the Ct value. cDNA was collected from all samples, pooled and diluted to the concentrations indicated above. The PCR efficiencies (E) of each target were calculated according to the established equation $E = 10^{-1/slope}$ as described by Pfaffl (2001). The slope and calculated E value for each target are listed in table 4.3 together with the correlation coefficient (R^2) of the linear regression. R^2 is a measure of the proximity between the two variables, cDNA concentration and Ct value, with $R^2=1$ indicating perfect correlation. Standard curve of *fgf5* is not included as it gave very poor results, i.e. very high Ct values or no amplification.

Table 4.3: RT-PCR targets, slope of linear regression, E and R^2measurements.

Target gene	Slope	E	R^2
gapdh	-3,365	1,98233396	0,998
nanog	-3,429	1,95717749	0,998
oct4	-4,078	1,75880724	0,997
sox2	-4,05	1,76568643	0,995
brachyury	-3,604	1,89439037	0,987
gata6	-2,917	2,20200329	0,99
Alkbh1	-3,962	1,7881246	0,995

Comparative Ct experiments were performed for all targets listed in table 4.3. All individual samples were diluted such that the starting cDNA template concentration was of either 5 ng μl^{-1} when *gapdh, nanog, oct4* and *Alkbh1* were analyzed or 20 ng μl^{-1} when *sox2, brachyury* and *gata6* were analyzed. The mathematical model proposed by Pfaffl (2001) was used to

calculate the relative expression of all targets in all samples as described in Methods. The range in Ct values between all samples for each target is listed in table 4.4.

Table 4.4: Range in Ct values between all samples for each target

Target gene	Lowest Ct value	Highest Ct value
nanog	27.46	36.95
oct4	23.69	37.06
sox2	27.77	35.88
brachyury	26.41	36.96
gata6	27.85	38.98
Alkbh1	29.94	30.96

Results presented in figure 4.22 show the mean relative expression ± SD (n = 3) of core transcriptional factors of ES cell pluripotency in undifferentiated and differentiated mES cells: NANOG, OCT4 and SOX2[2]. The expression of all three proteins was down-regulated in both methods used to induce differentiation: formation of EBs or addition of RA to the culture medium.

[2] When the expression of targets is in question, target names will be written in capital letters as it refers to the level of protein expression. Alkbh1 is not written in capital letters as this would refer to the human homologue

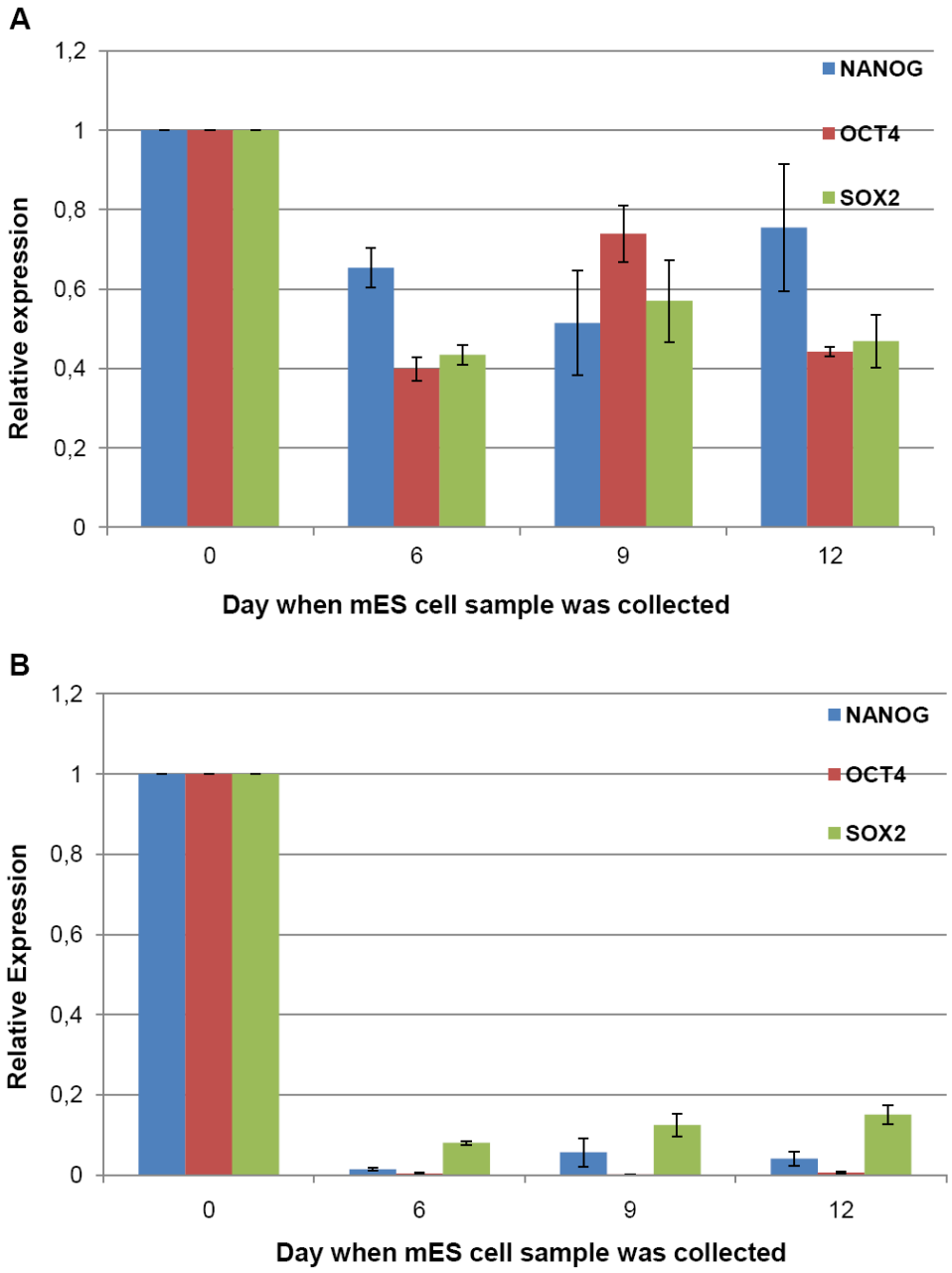


Figure 4.22: Comparative Ct RT-PCR analysis of core transcription factors in ES cell pluripotency during induced differentiation of mouse ES (mES) cells. Relative expression of NANOG, OCT4 and SOX2 at Day 6, 9 or 12 after removal of LIF, versus a control sample (Day 0, undifferentiated) is compared to a reference housekeeping gene, *gapdh*, and calculated by the mathematical model proposed by Pfaffl (2001). A) Relative expression in embryonic bodies (EBs), B) Relative expression in retionic acid (RA) differentiated mES cells. The expression of all targets is down-regulated in both EBs and in RA differentiated mES cells, with a much steeper decline in the RA differentiated mES cells. Error bars represent mean ± SD (n = 3).

The mean relative expression ± SD (n = 3) of mesoderm (BRACHYURY) and endoderm (GATA6) specific markers is presented in figure 4.23. GATA6 is highly up-regulated in both EBs and in RA differentiated mES cells. BRACHYURY is highly up-regulated on Day 6 in

EBs and to a lesser extent on Day 9 and 12 in EBs. In RA differentiated mES cells the expression of BRACHYURY remains very low. The mean relative expression ± SD (*n* = 3) of Alkbh1 is presented in figure 4.24 that shows a slight up-regulation of Alkbh1 in both EBs and in RA differentiated mES cells.

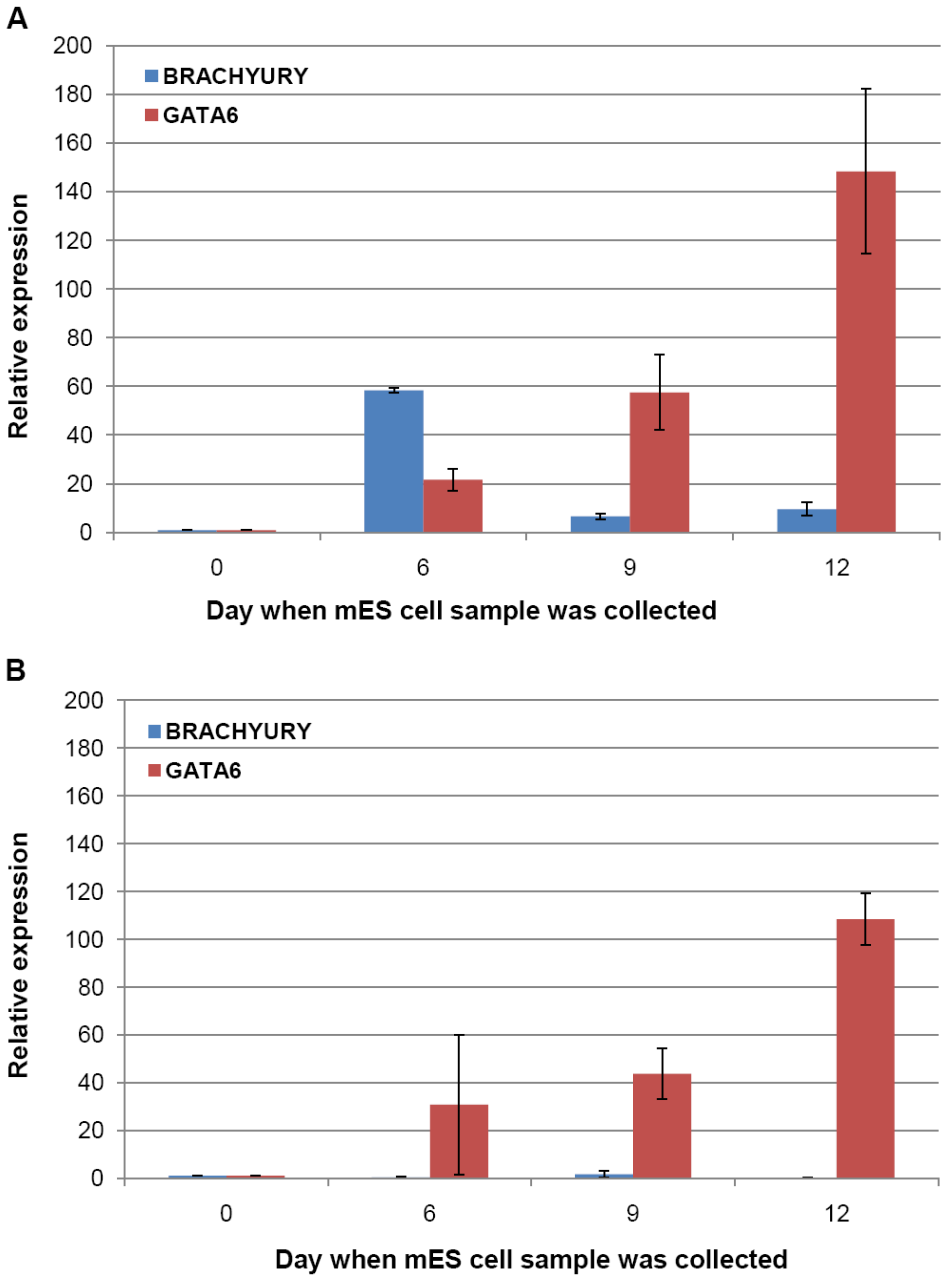


Figure 4.23: Comparative Ct RT-PCR analysis of mesoderm and endoderm specific proteins during induced differentiation of mouse ES (mES) cells. Relative expression of BRACHYURY and GATA6 targets at Day 6, 9 or 12 after removal of LIF, versus a control sample (Day 0, undifferentiated) in comparison to a reference gene, *gapdh*, is calculated by the mathematical model proposed by Pfaffl (2001). A) Relative expression in embryonic bodies (EBs), B) Relative expression in retinoic acid (RA) differentiated mES cells. Expression of BRACHYURY is slightly up-regulated on days 9 and 12 and especially on day 6 in EBs. Expression of GATA6 is also up-regulated in EBs, reaching a peak on day 12. In RA differententiated mES cells expression of GATA6 is up-regulated, whereas expression of BRACHYURY remains very low. Error bars represent mean ± SD (*n* = 3).

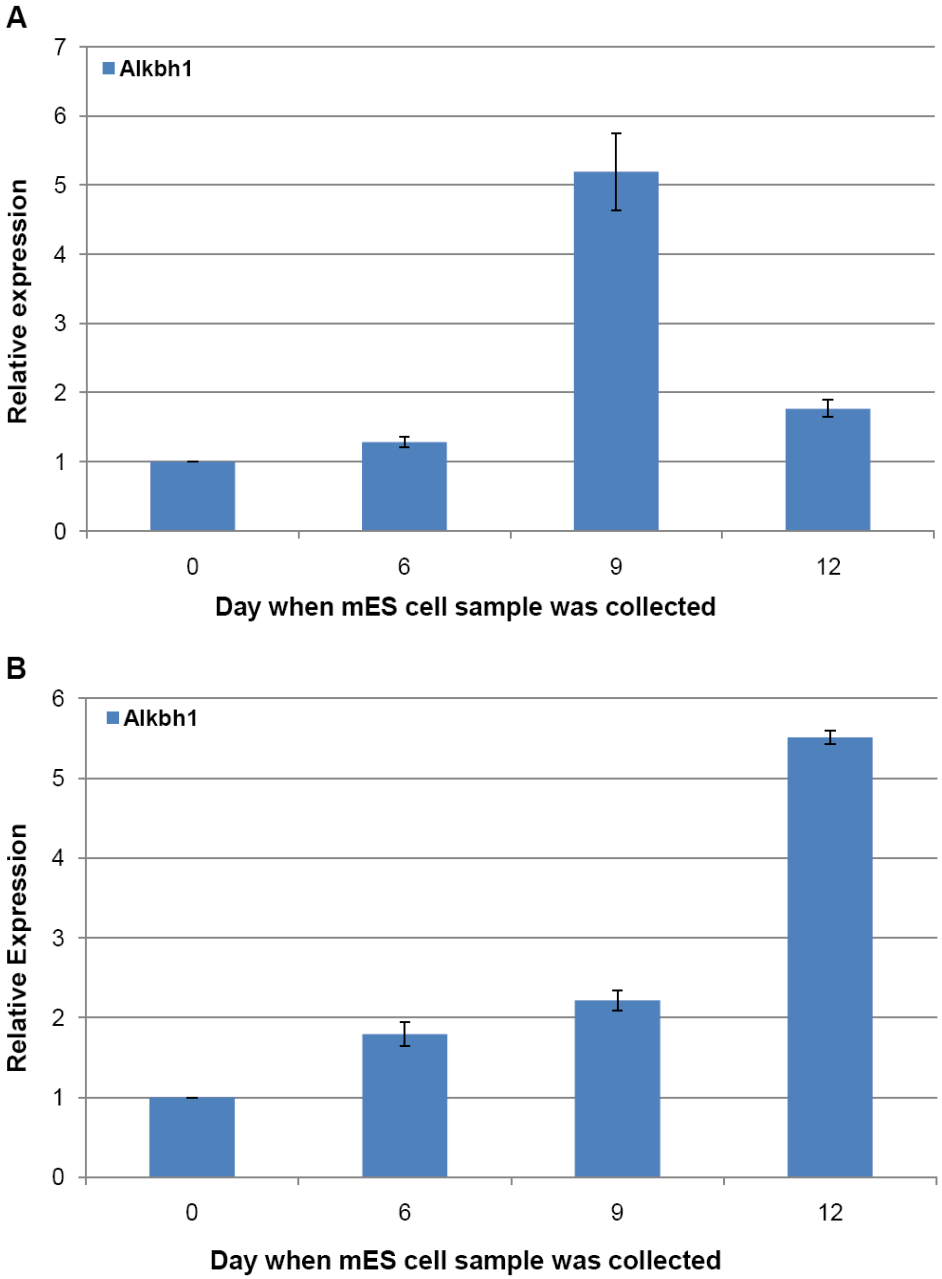


Figure 4.24: Comparative Ct RT-PCR analysis of Alkbh1 during induced differentiation of mouse ES (mES) cells. Relative expression of Alkbh1 at Day 6, 9 or 12 after removal of LIF, versus a control sample (Day 0, undifferentitated) in comparison to a reference gene, *gapdh*, is calculated by the mathematical model proposed by Pfaffl (2001). A) Relative expression of Alkbh1 in embryonic bodies (EBs), B) Relative expression of Alkbh1 in retinoic acid (RA) differentiated mES cells. Alkbh1 is slightly up-regulated, especially on day 9 and 12 in EBs and in RA differentiated mES cells, respectively. Error bars represent mean ± SD (n = 3).

Differentiation of mES cells and subsequent RT-PCR analysis were performed twize in this study and the same findings were observed in each experiment. Results presented in figures 4.22 – 4.24 therefore represents only one of the experiments.

5 Discussion

5.1 Purification of ALKBH1

Purification of ALKBH1 proved to be a quite challenging and labor-intensive part of this thesis as the methods had to be optimized several times and large volumes of *E. coli* cultures were required in order to achieve good yield. Affinity-purification by the IMPACT™ system, however, proved to be most efficient in the end, requiring only a volume of 6 L compared to 24 L using the HIS-Select® cobalt affinity gel. This was achieved when the purification protocol had been optimized by lowering the induction temperature (18˚C) in combination with ON expression. Also, by allowing on-column cleavage with DTT to take place for 48 hours the total protein yield was quite high.

Reasons to why the IMPACT™ -pTYB12 system worked better than the HIS-Select®-pET28a protocol is that ALKBH1 tagged with intein tag seemed to be more soluble during protein expression, thus allowing for better binding to the chitin column. Histidine-tagged ALKBH1 was observed in cell lysate and flow-through samples (figure 4.4) using the pET28a expression vector indicating that it was less soluble. Some ALKBH1 was also eluted from the cobalt affinity column in wash 1 and 2 (figure 4.3 and 4.4). As washing buffers 1 and 2 contain increasing concentrations of imidazole (10 mM and 30 mM, respectively) this indicates that the cobalt affinity gel did not have optimal binding capacity towards the histidine-tagged ALKBH1. In contrast, no ALKBH1 was eluted from the IMPACT™ column during the wash and flush (figure 4.6), indicating excellent binding of intein-tagged ALKBH1 to the chitin resin.

5.2 Biochemical Function of ALKBH1

Although displaying the highest sequence similarity to *E. coli* AlkB, ALKBH1 has not been confirmed as a true functional homologue of AlkB. However, a truncated recombinant ALKBH1 protein has shown to demethylate 3MeC *in vitro* (Westbye et al., 2008). Methylation of histones is known to be important for transcriptional regulation and predictions that AlkB homologues might play role in epigenetic regulation come from studies on JmjC domain-containing proteins. These proteins display similar features to 2-OG and Fe(II)-dependent oxygenases and it was therefore proposed that JmjC-containing proteins could be histone demethylases acting in an AlkB-like mechanism (Trewick et al., 2005). It

was later demonstrated that the JmjC domain-containing histone demethylases, (JHDM)1A and JHDM2, demethylates mono- and di-methylated H3-K36 (lysine residue 36 of histone 3 tails) and trimethylated H3-K36 and monomethylated H3-K9, respectively, by Fe^{2+}-dependent oxidative demethylation coupled to decarboxylation of 2-OG to succinate (Tsukada et al., 2006; Whetstine et al, 2006). As human AlkB homologues belong to the 2-OG and Fe(II)-dependent oxygenase superfamily, ALKBH1 might have a role in histone demethylation.

5.2.1 Succinate Formation Assay

Succinate formation assays were performed to investigate the ability of purified recombinant ALKBH1 to decarboxylate 2-OG into succinate, with and without a cH2A as substrate. A reliable interpretation of the results could, however, not be made. Obtained results from uncoupled reaction (figure 4.9) showed negative dpm values for ALKBH1 decarboxylation. The control protein, AlkB, however, showed positive dmp values. Replicates displayed high variance as indicated by the SD error bars, also in the case of AlkB. This was most likely due to pipetting errors as the assay relies on 2,4 DNPH to selectively precipitate 2-OG. It is therefore very difficult to obtain a pure supernatant including succinate only. Small particles from the precipitate might follow into the supernatant upon transfer into the scintillation vial. This will result in some very high dpm values, which interferes with the actual dpm value of the formed succinate. The results can therefore not be used to interpret whether the purified ALKBH1 had successfully decarboxylated 2-OG into succinate.

The addition of cH2A substrate was also investigated. The obtained results (figure 4.10) could not be used to conclude whether H2A is a substrate of ALKBH1. One would also expect to see an increase in dpm values when the amount of cH2A increases, but a steady level or decrease was observed instead. Furthermore, reaction replicates displayed high SD values, which most likely are due to the same reasons as stated above. The control enzyme, *S. pombe* AlkB homologue 2, did however show some increased decarboxylation of 2-OG upon addition of cH2A.

Albeit continuous efforts to produce reliable results in the succinate formation assay, the same problems were encountered each time, i.e. high variance between replicates and/or negative results for positive control enzymes and/or ALKBH1. Failure to demonstrate the enzymatic activity of ALKBH1 could also be due to difficulties in purifying the active form of the protein. In addition to this, freezing and thawing of protein samples might denature proteins,

thus making them inactive. Most likely, it is difficult to obtain reliable results due to the practicality of the assay, i.e. difficult to obtain a pure supernatant without interference with the precipitate. A CO_2-capture assay might be more reliable in determining ALKBH1 ability to decarboxylate 2-OG. In such an assay carbon 1 of 2-OG is substituted with ^{14}C, which will be found on CO_2 upon decarboxylation of the co-substrate. The released CO_2 gas can be collected without the need to precipitate 2-OG, thus avoiding radioactive contamination from the reaction mixture.

5.2.2 Enzymatic Activities and Sub-Cellular Localization of ALKBH1

ALKBH1 was reported as a mitochondrial demethylase that demethylates 3MeC but not 1MeA in both DNA and RNA by Westbye et al. (2008). In contrast to our study, Westbye and co-workers used a truncated recombinant version of ALKBH1(Δ3), where three amino acids had been deleted from the N-terminal region. They also used considerably higher concentration of the recombinant protein (up to a 100 pmol) when demonstrating successful decarboxylation of 2-OG in a CO_2 capture assay. The reported repair activity was also much lower than that of AlkB, ALKBH2 and ALKBH3 activities (Westbye et al., 2008). By northern blot analysis, Westbye et al (2008) showed that ALKBH1 mRNA was widely expressed in different human tissues with the highest expression in skeletal muscle and heart that are both rich in mitochondria. The indication of a mitochondrial localization of ALKBH1 was supported by evidence from fluorescence microscopy, immunocytochemical studies, and treatment with Proteinase K or digitonin on isolated mitochondria (Westbye et al., 2008).

However, expression and sub-cellular localization of ALKBH1 is a matter of debate as other studies using real-time PCR indicate that ALKBH1 is most highly expressed in spleen (Tsujikawa et al., 2007). Furthermore, one cannot ignore the observation that ALKBH1 mRNA is expressed in all human tissues, which indicates that ALKBH1 could be a housekeeping gene playing a fundamental role in most tissues (Falnes et al., 2007; Tsujikawa et al., 2007). An elevated level of ALKBH1 expression has also been observed in the pancreas and testis (Tsujikawa et al., 2007). By fusing EmGFP (Emerald Green Fluorescent Protein) to the N-terminus of all ALKBH coding regions Tsujikawa et al. (2007) examined the sub-cellular localization of all ALKBH proteins in HeLa cells. These localization studies showed ALKBH1 throughout the cells. These studies oppose the theory that ALKBH1 is a mitochondrial protein.

5.3 Possible Roles of ALKBH1 in ES cell Pluripotency and Epigenetic Regulation

In the study by Boyer et al. (2005) NANOG and OCT4 was found to bind to the promoter region of *ALKBH1* in human ES cells, implying that ALKBH1 might be important in ES cell self-renewal and pluripotency. *Alkbh1* deficient mice have also been created (Pan et al., 2008, Nordstand et al., unpublished data). Although being viable these mouse models show developmental defects and defects in placental trophoblast lineage differentiation (Pan et al., 2008; Nordstrand et al., unpublished data). It was proposed that ALKBH1 could participate in epigenetic gene regulation when it was found to bind to the co-chaperone protein Mrj (Pan et al., 2008). Mrj is known to modulate the conformation and activity of multiple proteins including class II histone deacetylases (HDACs) (Dai et al., 2005). By binding to Mrj, ALKBH1 could prevent interaction with class II HDACs, thus relieving HDAC mediated gene repression (Pan et al., 2008). Histone demethylation by JMJD6 (a JmjC-containing 2-OG and Fe(II)-dependent demethylase) could also explain developmental defects seen in *Jmjd6* deficient mice (Chang et al., 2007). Most interestingly, two 2-OG and FE(II)-dependent histone demethylases, JMJD1A and JMJD2C, have been shown to regulate self-renewal of ES cells (Loh et al., 2010). Depletion of the two proteins led to ES cell differentiation and NANOG expression was reduced upon JMJD2C depletion. OCT4 was shown to up-regulate JMJD1A and JMJD2C, which in turn maintained a low level of lysine 9 histone 3 (H3K9) methylation patterns at promoter regions of genes critical for self-renewal of ES cells. In addition, reduced NANOG expression suggested that JMJD2C was required to reverse H3K9 methylation at the *nanog* promoter region (Loh et al., 2010).

5.3.1 ALKBH1 Interacts with Core Regulators of ES cell Pluripotency and Histones

In this study, positive protein-protein interactions between recombinant ALKBH1 and core regulators of ES pluripotency, including NANOG, OCT4 and SOX2, were demonstrated in two *in vitro* assays. Positive interactions were also seen between ALKBH1 and core histones H2A and H4 in both assays, and with H2B and H3 in one of the assays. The Dot-Blot immunobinding assay, (figure 4.11) demonstrated positive interactions between ALKBH1 and NANOG, OCT4, SOX2, H2A and H4. These results were supported by negative interactions between ALKBH1 and the two control proteins, BSA and lysozyme. No background interference between the primary ALKBH1 antibody and non-ALKBH1 protein also validates these findings. No interaction was detected between ALKBH1 and histones H1, H2B or H3 in

the Dot-Blot assay (figure 4.11). All protein interactions were tested further in a second a second *in vitro* ELISA. Again, positive interactions were seen between ALKBH1 and the proteins described above (figure 4.12), in addition to H2B, H3 (figure 4.13). Interactions with histone H4 was however more weak than the other interactions. The reason why all the four core histones and not histone H1 showed positive interactions with ALKBH1 can be explained by the high level of evolutionary conservation among core histones. In chromatin structure these form the core particles of nucleosomes, whereas H1, however, acts as a linker histone between nucleosomes (reviewed by Verdone et al., 2006).

One might regard these positive interactions as an effect of protein aggregation. However, from the observation of non-positive reactions with two control proteins in the Dot-Blot assay, and negative absorbance values of histone H1, this seems unlikely. In addition, background BSA absorbance values were subtracted from all absorbance values of target proteins, thereby separating results caused by non-specific protein aggregation from more likely proteins interactions.

In conclusion, these results indicate that core histone are possible substrates of ALKBH1 acting as a histone demethylase in regulation of ES cell pluripotency. Histone H2A is the most likely substrate as it displayed positive results in both assays. Supporting evidence from the succinate formation assay would have helped to justify this theory. In light of these findings it is also worth mentioning that the pI value of ALKBH1 (7.08) also points towards a protein being the substrate of ALKBH1, and thus possible a histone. Histone methylation occurs at arginine and lysine residues, thus at a physiological pH of 7.4 ALKBH1 will have a negative net surface charge and an interaction with positively charged arginine and lysine residues is therefore plausible. In contrast, E. coli AlkB, ALKBH2 and ALKBH3 have relatively high pI values (7.9, 10.0 and 8.5, respectively) and positively charged surfaces, which seems to be important characteristics for associating DNA as DNA is negatively charged. An association between ALKBH1 and DNA cannot be predicted based on the protein's pI value (Sedwick et al., 2007). It would also be interesting to study if ALKBH1 depends on the interaction with partner proteins, such as NANOG, OCT4 and SOX2, to express its function.

5.4 NANOG and OCT4 Binds to the Promoter Sequence of *ALKBH1*

The possibility that NANOG and OCT4 regulate *ALKBH1* gene expression by binding to its promoter is even more likely as both proteins have been found to bind to the *ALKBH1* promoter in CHIP-chip assays (Boyer et al., 2005). This interaction was investigated further by EMSA. Although, bioinformatics analysis of the DNA fragments used in this assay (figure 4.21) showed that they were the actual sequences found by Boyer et al. (2005), a positive results of a protein:DNA complex between NANOG and OCT4 and their respective DNA binding sequences could not be demonstrated. Reasons for this could be explained by the EMSA procedure itself. Several factors such as the pH and ionic strength of the binding buffer have to be optimized further to strengthen the specific protein:DNA interactions. The temperature and time of the binding reaction are also important in order to allow for stable interaction complexes to form. Thus a lower temperature and a longer incubation time would have to be tested as well. It might also be more suitable to use a larger gel format as the DNA sequences were quite long (266 and 280 bp) in order to resolve any positive gel shift and to run the gel in a cold room.

In order to validate protein:DNA interactions between NANOG and OCT4 and the *ALKBH1* promoter, several of the factors mentioned above have to be optimized. Another issue is the fact that naked DNA was used here to investigate these interactions, whereas Boyer et al. (2005) relied on a method where chromatin DNA and proteins were cross-linked *in vivo.* Many additional factors could therefore be involved in the NANOG and OCT4 interaction with the *ALKBH1* promoter, including associated proteins, the local chromatin structure and *cis* elements located far away from the promoter region, all of which would be absent in an EMSA experiment.

5.5 Differentiation of mES Cells

To further investigate the role of ALKBH1 in ES cell self-renewal and pluripotency wild-type mES cells were cultured and differentiated. Differentiation of mES cells by either RA induced differentiation or by the formation of EBs were successful as demonstrated by the reduced level of expression of all three core regulators in ES cell pluripotency – NANOG, OCT4 and SOX2 (figure 4.22). A much more rapid decline in expression levels were observed in RA differentiated cells compared to EBs, which would be expected as commitment to a neuroectoderm differentiation by the addition of RA have been demonstrated to be rapid and

efficient (reviewed by Keller, 2005). Differentiation by the formation of EBs would be slower as aggregates of ES cells would have to be formed before spontaneous differentiation will occur. In the first few days following removal of LIF (day 0 – 3) EBs will consist of mES cell aggregates, followed by EBs surrounded by a layer of endodermal like cells, whereas mature EBs containing many differentiated cell types will occur after several days of culture (typically 14 days in the case of human ES cells) (Itskovit-Eldor et al., 2000).

The expression of mesoderm and endoderm markers (BRACHYURY and GATA6, respectively) was up-regulated in EBs with a continuous up-regulation of GATA6 expression levels reaching its peak on day 12 (figure 4.23). A rapid increase in BRACHYURY expression levels was seen on day 6 followed by a decline in expression levels. BRACHYURY is a marker of the primitive streak which forms upon gastrulation – the process in which the mouse embryo establishes mesoderm and definitive endoderm (reviewed by Gadue et al., 2005). The elevated level of BRACHYURY expression is therefore consistent with early embryonic development. BRACHYURY is not only expressed in the primitive streak but also in the developing mesoderm during ES cell differentiation as demonstrated by Fehling et al. (2003) using EBs. This is also consistent with the findings in this study as BRACHYURY was expressed during maturation of EBs.

GATA6 has been indicated as an essential factor in the formation of the primitive endoderm (Koutsourakis et al, 1999; Li et al., 2004), where *gata6* mutants are incapable of endoderm differentiation (Cai et al., 2008). The primitive endoderm does not express BRACHYURY (Kubo et al., 2004), which is also consistent with these findings, i.e. the expression level of BACHYURY declines as the expression level of GATA6 increases. In conclusion, a mixture of endoderm and mesoderm lineages was observed at different stages in EB development.

When mES cells were differentiated by the addition of RA to the culture medium (figure 4.23), very low levels of BRACHYURY was observed, whereas the expression levels of GATA6 increased as observed in EBs. The low expression levels of BRACHYURY are again consistent with other studies where mesodermal genes including BRACHYURY become down-regulated during RA differentiation (Bain et al., 1996). This suggests that neuronal differentiation is seen together with repression of mesodermal differentiation. Studies have shown that treatment with RA also stimulates differentiation towards endoderm lineages (Wendling et al., 2000; Matt et al., 2003), which is consistent with the expression of GATA6 in RA differentiated mES cell. However, expression of FGF5, a marker of the primitive

ectoderm could not be observed. This could be explained by the fact that the primitive endoderm starts to develop prior to primitive ectoderm (reviewed by Gadue et al., 2005), thus FGF5 expression might appear at a later stage. It might be more efficient to allow EBs to form before RA induced neuroectodermal differentiation, as this has been a successful method for induction of neuronal differentiation (reviewed by Guan et al., 2001).

The levels of Alkbh1 expression was slightly up-regulated when mES cells differentiated, especially on day 9 in EBs and on day 12 in RA differentiated mES cells. As the Ct values seemed to be fairly consistent in all samples (table 4.4), this elevated level of expression was suspected to be misleading. Inspection of the GAPDH Ct value in these two samples showed that they deviated by +2 Ct values compared to the GAPDH Ct values in the other samples. It is advisable that Ct values of the endogenous control gene don't vary more than ±1 mean Ct value of all samples, to ensure equal cDNA amplification in all samples. One might therefore want to disregard these two samples, as the higher GAPDH Ct value will result in a misleadingly higher expression of Alkbh1 when relative expression is calculated.

In conclusion, no extreme changes in Alkbh1 expression were observed when mES cells differentiated, which doesn't lead to any specific interpretation of its role in ES cell pluripotency. However, as it is expressed both in the undifferentiated state as well as during differentiations, this indicates a possible role ES cell pluripotency or a role in embryonic development. It would also be valuable to have an *Alkbh1* deficient cell line to compare with to see if any changes in NANOG, OCT4 or SOX2 would be affected. However, as successful methods of induced differentiation have been developed, these can be used in further studies using an *Alkbh1* deficient mES cell line. Some fine-tuning of the RA induced neuroectodermal differentiation procedure might be necessary as mES cells didn't express the FGF5 primitive ectoderm marker.

6 References

Aas, P. A., Otterlei, M., Falnes, P. O., Vagbo, C. B., Skorpen, F., Akbari, M., Sundheim, O., Bjøras, M., Slupphaug, G., Seeberg, E. and Krokan, H. E., 2003. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature*, 421, pp. 859-863.

Active Motif, 2010. *FunctionELISA™ IκBα* [Online] Available at: http://www.activemotif.com/catalog/165/functionelisa-i%CE%BAb%CE%B1.html [Accessed 8 April 2010]

Aloia, L., Parisi, S., Fusco, L., Pastore, L., and Russo, T., 2010. Differentiation of ESCs 1 (Dies1) is a component of Bone Morphogenetic Protein 4 (BMP4) signaling pathway required for proper differentiation of mouse embryonic stem cells. *The Journal of Biological Chemistry*, 285(10), pp. 7776-7783.

Ambrosetti, D. C., Basilico, C. and Dailey, L., 1997. Synergistic activation of the fibroblast growth factor 4 enhancer by sox2 and oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Molecular and Cellular Biology.* 17, 6321-6329.

Ambrosetti, D. C., Scholer, H. R., Dailey, L. and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *The Journal of Biological Chemistry*, 275, pp. 23387-23397.

Amersham Biosciences, 2002. *Affinity Chromatography – Principles and Methods* [Handbook]. 18-1022-29. AD ed., pp. 9.

Amersham Biosciences, 2004. *Ion Exchange Chromatography and Chromatofocusing – Principles and Methods* [Handbook]. 11-0004-21, AA ed., pp. 11-12.

Applied Biosystems, 2010. *Essentials of Real-Time* [Online] Available at: http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_042485.pdf [Accessed 14 April 2010]

Attotron Biosensor Corporation, 1998. *Nucleic Acid Sequence Massager* [Online] Available at: http://www.attotron.com/cybertory/analysis/seqMassager.htm [Accessed 20 September 2009 and 15 February 2010]

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes and Development*, 17, pp. 126-140.

Bain, G., Kitchens, D., Yao, M., Huettner, J. E., and Gottlieb, D. I., 1995. Embryonic stem cells express neuronal properties in vitro. *Developmental Biology,* 168, pp. 342-357.

Bain, G., Ray, W. J., Yao, M. and Gottlieb, D.I., 1996. Retinoic acid promotes neural and represses mesodermal gene expression in mouse embryonic stem cells in culture. *Biochemical and Biophysical Research Communications*, 223(3), pp. 691-694.

Begley, T. J. and Samson, L. D., 2003. AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. *Trends in Biochemical Science*, 28 (1), pp. 2-5

Bowles, J., Schepers, G. and Koopman, P., 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Developmental Biology*, 227, pp. 239-255.

Boyer, L. A, Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R. and Young, R. A., 2005. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*, 122, pp. 947-956

Bratt-Leal, A. M., Carpenedo, R. L. and McDevitt, T. C., 2009. Engineering the Embryoid Body Microenvironment to Direct Embryonic Stem Cell Differentiation. *Biotechnology Progress*, 25(1), pp. 43-51

Brebeuf Jesuit Biotechnology, 2010. *SDS-PAGE* [Online] Available at: http://bjpsbiotech.edublogs.org/content-review/sds-page/ [Accessed 8 April 2010]

Brinkmann Instruments, 2003. *Quantification made easy* [Online] Available at: http://www.protocol-online.org/cgi-bin/prot/view_cache.cgi?ID=1059 [Accessed 8 April 2010]

Brown, T. A., 2007. *Genomes*. 3rd ed. New York: Garland Science Publishing

Cai, K. Q., Capo-Chichi, C. D., Rula, M. E., Yang, D-H., and Xu, X-X., 2008. Dynamic GATA6 Expression in Primitive Endoderm Formation and Maturation in Early Mouse Embryogenesis. *Developmental Dynamics*, 237, pp. 2820-2829

Cambon, P., 1996. A decade of molecular biology of retinoic acid receptors. *The FASEB Journal*, 10(9), pp. 940-954.

Chambers, I., Colby, D., Roberston, 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*, 113, pp. 643-655.

Chambers, I. and Tomlinson, S. R., 2009. The transcriptional foundation of pluripotency. *Development*, 136, pp. 2311-2322.

Chan, B., Chen, Y., Zhao, Y. and Bruick, R. K., 2007. JMJD6 is a histone arginine demethylase. *Science*, 318(5849), pp. 444-447.

Chomczynski, P. and Sacchi, N., 1987. Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Analytical Biochemistry*, 162, pp. 156-159

Dang, S. M., Gerecht-Nir, S., Chen, J., Itskovitz-Eldor, J. and Zandstra, P. W., 2004. Controlled, scalable embryonic stem cell differentiation culture. *Stem Cells*, 22, pp. 275-282.

Darr, H., Mayshar, Y. and Benvenisty, N., 2006. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development*, 133, pp. 1193-1201.

Davis, S., Aldrich, T. H., Stahl, N., Pan, L., Taga, T., Kishimoto, T., Ip, N. Y. and Yancopoulos, G. D., 1993. LIFRβ and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor, *Science,* 260, pp.1805-1808.

DePamphilis, M. L., 2006. *DNA Replication and Human Disease*. New York: Cold Spring Harbor Laboratory Press.

Dingley, S., Gold, B. and Sedgwick, B., 1998. Repair in *Escherichia coli alkB* mutants of abasic sites and 3-methyladenine residues in DNA. *DNA Repair,* 407(2), pp. 109-116.

Drabløs, F., Feyzi, E., Aas, P. A., Vaagbø, Kavli, B., Bratlie, M. S., Pena-Diaz, J., Otterlei, M., Slupphaug, G. and Krokan, H. E., 2007. Alkylation damage in DNA and RNA—repair mechanisms and medical significance. *DNA Repair*, 3, pp. 1289-1407.

Dragony, F., Brophy, A. N., Clement, T. Y. Chan, K. A., Atmore, K. A., Begley, U., Paules, R. S., Dedon, P. C., Begley, T. J. and Samson, L. D., 2010. Human AlkB homolog ABH8 is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. *Molecular and Cell Biology,* doi:10.1128/MCB.01604-09 [Online]

Duncan, T., Trewick, S. C., Koivisto, P., Bates, P.A., Lindahl, T. and Sedgwick, B., 2002. Reversal of DNA alkylation damage by two human dioxygenases. *PNAS,* 99, pp. 16660-16665.

Ensembl, 2010. *Gene: ALKBH1 (ENSG00000100601)* [Online] Available at: http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000100601;r=14:78138749-78174356;t=ENST00000216489 [Accessed 20 September 2009]

European Bioinformatics Institute (EBI), 2010. *ClustalW2* [Online] Available at: http://www.ebi.ac.uk/Tools/clustalw2/index.html [Accessed 20 September 2009]

Evans, M. J. and Kaufman, M. H., 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292, pp. 154-156.

Falnes, P. Ø., Johansen, R. F. and Seeberg, E., 2002. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli. Nature,* 12(419), pp. 178-182.

Falnes, P. Ø., 2004. Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Research*, 32, pp. 6260-6267.

Falnes, P. Ø., Bjørås, M, Aas, P. A., Sundheim, O. and Seeberg, O., 2004. Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Research,* 32, pp. 3456-3461.

Falnes, P. Ø., Klungland, A. and Alseth, I., 2007. Repair of Methyl Lesions in DNA and RNA by oxidative demethylation. *Neuroscience*, 145, pp. 1222-1232.

Fehling, H. J., Lacaud, G., Kubo, A. Kennedy, M., Robertson, S., Keller, G. and Kouskoff, V., 2003. Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development*, 130, pp.4217-4227.

Fermentas, 2004. *T4 Polynucleotide Kinase* [Certificate of Analysis] Revised 7.11.2004, pp. 1-4.

Freidberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A. and Ellenberger, T., 2006. *DNA Repair and Mutagenesis.* 2nd ed. Washington D. C.: ASM Press.

Freidberg, E. C., 2008. A brief history of the DNA repair field. *Cell Research,* 18, pp. 3-7

Friel, R., Sar, S. and Mee, P. J., 2005. Embryonic stem cells: Understanding their history, cell biology and signaling. *Advanced Drug Delivery Reviews*, 57, pp. 1894-1903.

Gadue, P., Huber, T. L., Nostro, M. C., Kattman, S. and Keller, G. M., 2005. Germ layer induction from embryonic stem cells. *Experimental Hematology,* 33, pp. 955-964.

Gallagher, S. and Desjardins, P. R., 2006. Quantitation of DNA and RNA with Absorption and Fluorescence Spectroscopy. In: F. M. Ausubel et al., ed 2006. *Current Protocols in Molecular Biology*, John Viley & Sons, Inc. Appendix 3D.

Genome Reference Consortium, 2009. *Human Genome Assembly Information* [Online] Available at: http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/data/index.shtml [Accessed 21 November 2009]

Guan, K., Chang, H., Rolletschek, A. and Wobus, A. M., 2001. Embryonic stem cell-derived neurogenesis - Retinoic acid induction and lineage selection of neuronal cells. *Cell Tissue Research,* 305, pp. 171-176.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P. and Lovell-Badge, R., 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, 346, pp. 245-250.

Hyslop, L., Stojkovic, M, Armstrong, L., Walter, T., Stojkovic, P., Przyborski, S., Herbert, M., Murdoch, A., Strachan, T., and Lakoa, M., 2005. Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem Cells,* 23, pp. 1035-1043.

Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H. and Benvenisty, N., 2000. Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Molecular Medicine,* 6, pp. 88-95.

Jonk, L. J., de Jonge, M. E., Kruyt, F. A., Mummery, C. L., van der Saag, P. T. and Kruijer, W., 1992. Aggregation and cell cycle dependent retinoic acid receptor mRNA expression in P19 embryonal carcinoma cells. *Mechanisms of Development,* 36(3), pp. 165-172.

Kataoka, H., Yamamoto, Y. and Skiguchi, M., 1983. A New Gene (alkB) of Escherichia coli That Controls Sensitivity to Methyl Methane Sulfonate. *Journal of Bacteriology,* 153(3), pp. 1301-1307.

Kaule, G. and Giinzler, V., 1990. Assay for 2-Oxoglutarate Decarboxylating Enzymes Based on the Determination of [l-^{14}C]Succinate: Application to Prolyl4-Hydroxylase. *Analytical Biochemistry,* 184, pp. 291-297

Kawasaki, H., Mizuseki, K., Nishikawa, S., Kaneko, S., Kuwana, Y., Nakanishi, S., Nishikawa, S. I., and Sasai, Y. 2000. Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. *Neuron,* 28, pp. 31-40.

Keller, G. M., 1995. In vitro differentiation of embryonic stem cells. *Current Opinion in Cell Biology,* 7, pp. 862–869.

Keller, G., 2005. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes and Development,* 19, pp. 1129-1155

Koivisto, P., Duncan, T., Lindahl, T. and Sedgwick, B., 2003. Minimal methylated substrate and extended substrate range of Escherichia coli AlkB protein, a 1-methyladenine-DNA dioxygenase. *Journal of Biological Chemistry,* 278, pp. 44348-44354.

Koivisto, P., Robins, P., Lindahl, T. and Sedgwick, B., 2004. Demethylation of 3-Methylthymine in DNA by Bacterial and Human DNA Dioxygenases. *Journal of Biological Chemistry,* 279, pp. 40470-40474.

Kopp, J. L., Ormsbee, B. D., Desler, M. and Rizzino, A., 2008. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells,* 26, pp. 903-911.

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R. and Grosveld, F., 1999. The transcription factor GATA6 is essential for early extraembryonic development. *Development,* 126, pp. 723–732.

Kubo, A., Shinozaki, K., Shannon, J. M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H. J., and Gordon Keller, 2004. Development of definitive endoderm from embryonic stem cells in Culture. *Development and Disease,* 131, pp. 1651-1662.

Kurowski, M. A., Bhagwat, A. S., Papaj, G. and Bujnicki, J. M., 2003. Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB, *BMC Genomics,* 4(48).

Li, J., Pan, G., Cui, K., Liu, Y., Xu, S. and Pei, D., 2007. A dominant-negative form of mouse SOX2 induces trophectoderm differentiation and progressive polyploidy in mouse embryonic stem cells. *Journal of Biological Chemistry,* 282, pp. 19481–19492.

Li, L., Arman, E., Ekblom, P., Edgar, D., Murray, P. and Lonai, P., 2004. Distinct GATA6- and laminin-dependent mechanisms regulate endodermal and ectodermal embryonic stem cell fates. *Development,* 131, pp. 5277-5286.

Lindahl, T. and Wood, R. D., 1999. Quality Control by DNA Repair. *Science,* 286, pp. 1897-1905.

Loh, Y-H., Wu, Q., Chew, J-L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K-Y., Sung, K. W., Lee, C. W. H., Zhao, X-D., Chiu, K-P., Lipovich, L., Kuznetsov, V. A., Robson, P., Stanton, L. W., Wei, C-L., Lim, B. and Ng, H-H., 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics,* 38, pp. 431-440

Loh, Y-H., Zang, W., Chen, X., George, J. and Ng, H-H., 2010. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes and Development,* 21, pp. 2545-2557.

Liu, Y and Labosky, P. A., 2008. Regulation of embryonic stem cell self-renewal and pluripotency by Foxd3. *Stem cells,* 10, pp. 2475-2484

Martin, G., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *PNAS*, 78, pp. 7634- 7638.

Mason, P. E., Neilson, G. W., Dempsey, C. E., Barnes, A. C. and Cruickshank, J. M., 2003. The hydration structure of guanidinium and thiocyanate ions: Implications for protein stability in aqueous solution. *PNAS*, 100(8), pp. 4557-4561.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., Ko, M. S. H. and Niwa, H., 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology*, 9, pp. 625–635.

Matt, N., Ghyselinck, N. B., Wendling, O., Chambon, P. and Mark, M., 2003. Retinoic acid-induced developmental defects are mediated by RARbeta/RXR heterodimers in the pharyngeal endoderm. *Development*, 130, pp. 2083-2093.

Maurer, J., Nelson, B., Cecena, G., Bajpai, R., Mercola, M., Terskikh, A. and Oshima, R. G., 2008. Contrasting expression of keratins in mouse and human embryonic stem cells. *PLoS ONE*, 3(10), pp. e3451

Mierendorf, R., Yeager, K. and Novy, R., 1994.The pET System:Your Choice for Expression. *Inovations*, 1(1), pp. 1-3.

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*, 113, pp. 631– 642.

Muller, T. A., Meek, K. and Hausinger, R. P., 2010. Human AlkB homologue 1 (ABH1) exhibits DNA lyase activity at abasic sites. *DNA Repair*, 9, pp. 58-65.

Nakamura, Y. and Miyoshi, N., 2010. Electrophiles in Foods: The Current Status of Isothiocyanates and Their Chemical Biology. *Bioscience, Biotechnology, and Biochemistry*, 74 (2), pp. 242-255.

NCBI, 2010a. *National Centre of Biotechnology Information* [Online] Available at: http://www.ncbi.nlm.nih.gov/ [Accessed 1 December 2009]

NCBI, 2010b. *Basic Local Alignment Search Tool* [Online] Available at: http://blast.ncbi.nlm.nih.gov/Blast.cgi [Accessed 15 February 2010]

NCBI, 2010c. *Primer-Blast, a tool for finding specific primers* [Online] Available at: http://www.ncbi.nlm.nih.gov/tools/primer-blast/ [Accessed 2 December 2009]

New England Biolabs (NEB), 2006. IMPACT^{TM}-CN Instruction Manual, 2.1 version, pp. 1-48.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H. and Smith, A., 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, 95, pp.379-391.

Niwa, H., Miyazaki, J. and Smith, A. G., 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, 24, pp. 372-376.

Nolan, T., Hands, R. E., and Bustin, S. A., 2006. Quantification of mRNA using real-time RT-PCR. *Nature Protocols,* 1(3), pp. 1559-1582.

Novagen, 2006. Pet System Manual, 11th ed., pp. 1-80.

Nunc, 2003. CovaLink NH, 2nd ed., pp. 3 – 15.

Okabe, S., Forsberg-Nilsson, K., Spiro, A. C., Segal, M., and McKay, R. D., 1996. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mechanisms of Development,* 58, pp. 89-102.

Okita, K. and Yamanaka, S., 2006. Intracellular Signaling Pathways Regulating Pluripotency of Embryonic Stem cells. *Current Stem Cell Research and Therapy,* 1, pp. 103-111

Orkin, S. H., 2005. Chipping away at the embryonic stem cell network. *Cell,* 122(6), pp. 947-956.

Pan, G., Chang, Z. Y., Schöler, H. R. and Pei, D., 2002. Stem cell pluripotency and transcription factor Oct4. *Cell Research*, 12(5-6), pp. 321-329.

Pan, G., Li, J., Zhou, Y., Zheng, H. and Pei, D., 2006. A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *The FASEB Journal,* 20, pp. 1730-1732.

Pan, G. and Thomson, J. A., 2007. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Research*, 17, pp. 42-49

Pfaffl, M. A., 2001. A new mathematical model for relative quantification in real time RT-PCR. *Nucleic Acid Research*, 29(9), pp. 2002-2007.

Ringvoll, J., Nordstrand, L. M., Vagbo, C. B., Talstad, V., Reite, K., Aas, P. A., Lauritzen, K. H., Liabakk, N. B., Bjork, A., Doughty, R. W., Falnes, P. Ø., Krokan, H. E. and Klungland, A., 2006. Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *The EMBO Journal*, 25, pp. 2189-2198.

Ratajczak, M. Z., Zuba-Surma, E. K., Machalinski, B. and Kucia, M., 2007. Bone-marrow-derived stem cells – our key to longevity? *Journal of Applied Genetics*, 48(4), pp. 307-319.

QIAGEN, 2007. miRNA Mini Handbook, pp. 1-50.

QIAGEN, 2008. QIAquick® Spin Handbook, pp. 1-44.

Sambrook, J. and Russel, D. W., 2001a. *Molecular Cloning – A Laboratory Manual, Volume 1*. 3rd ed. New York: Cold Spring Laboratory Press

Sambrook, J. and Russel, D. W., 2001a. *Molecular Cloning – A Laboratory Manual, Volume 2*. 3rd ed. New York: Cold Spring Laboratory Press

Sedgwick, B. and Lindahl, T., 2002. Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene,* 21, pp. 8886-8894

Sedgwick, B., Bates, P. A., Paik, J., Jacobs, S. C., Lindahl, T., 2007. Repair of alkylated DNA: Recent advances. *DNA repair,* 6, pp. 429-442

Singh, S. K., Kagalwala, M. N., Parker-Thornburg, J., Adams, H. and Majumder, S., 2008. REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature,* 453, pp. 223-229.

Smith, A., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D., 1988. Inhibition of pluripotent embryonic stem cell differentiation by purified polypeptides. *Nature,* 336, pp. 688-690.

Smith, A. G., 2001. Embryo derived stem cells: of mice and men. *Annual Review of Cell and Developmental Biology,* 17, pp. 435-462.

Stratagene, 2005. BL21-CodonPlus® Competent Cells Instructions Manual, pp. 1-16.

Thermo Scientific, 2010a. *Dialysis Methods for Protein Research* [Online] Available at: http://www.piercenet.com/browse.cfm?fldID=5753AFD9-5056-8A76-4E13-5F9E9B4324DA [Accessed 19 April 2010]

Thermo Scientific, 2010b. *EDC (1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide Hydrochloride)* [Online] Available at: http://www.piercenet.com/products/browse.cfm?fldID=02030312 [Accessed 8 April 2010]

Thermo Scientific, 2010c. *Gel Shift Assay (EMSA)* [Online] Available at: http://www.piercenet.com/browse.cfm?fldID=7EDAC33E-3981-4E58-8A57-057A8280C68F [Accessed 15 April 2010]

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. and Jones, J. M., 1998. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science,* 282, pp. 1145-1147.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P. and Zhang, Y., 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature,* 439, pp. 811-816.

Trewick, S. C., Henshaw, T. F., Husinger, R. P., Lindahl, T. and Sedgwick, B., 2002. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature,* 12(419), pp. 174-178

Tropepe, V., Hitoshi, S., Sirard, C., Mak, T.W., Rossant, J., and van der Kooy, D., 2001. Direct neural fate specification from embryonic stem cells: A primitive mammalian neural stem cell stage acquired through a default mechanism. *Neuron,* 30, pp. 65-78.

Tsujikawa, K., Koike, K., Kitae, K., Shinkawa, A., Arima, H., Suzuki, T., Tsuchiya, M., Makino, Y., Furukawa, T., Konishi, N., and Yamamoto, H., 2007. Expression and sub-cellular localization of human ABH family molecules. *Journal of Cellular and Molecular Medicine,* 11 (5), pp. 1105-1116

Visnes, T., Doseth, B., Pettersen, H. S., Hagen, L., Sousa, M. M. L., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G. and Krokan, H. E., 2009. Uracil in DNA and its processing by different DNA glycosylases. *Philosophical Transactions of The Royal Society B,* 364, pp. 563–568.

Yokota, Y. and Ohkubo, H., 1996. 9-cis-retinoic acid induces neuronal differentiation of retinoic acid-nonresponsive embryonal carcinoma cells. *Experimental Cell Research,* 228(1), pp. 1-7.

Voytas, D. and Ke, N., 1999. Detection and Quantitation of Radiolabeled Proteins and DNA in Gels and Blots. In: In: F. M. Ausubel et al., ed 1999. *Current Protocols in Molecular Biology*, John Viley & Sons, Inc. Appendix 3.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W. and Orkin, S. H., 2006. A protein interaction network for pluripotency of embryonic stem cells. *Nature*, 444, pp. 364-368

Wang, S. H., Tsai, M. S., Chiang, M. F., and Li, H., 2003. A novel NK-type homeobox gene, ENK (early embryo specific NK), preferentially expressed in embryonic stem cells. *Gene Expression Patterns*, 3, pp. 99-103

Watson, J. D. and Crick, F. H. C., 1953. Molecular structure of nucleic acids. *Nature*, 4356(171), pp. 737-738

Watt, F. M. and Driskell, R. R., 2010. The therapeutic potential of stem cells. *Philosophical Transactions of The Royal Society B,*365, pp. 155-163

Wei, Y. F., Carter, K. C., Wang, R. P. and Shell, B. K., 1996. Molecular cloning and functional analysis of a human cDNA encoding an Escherichia coli AlkB homolog, a protein involved in DNA alkylation damage repair. *Nucleic Acids Research*, 24, pp. 931-937.

Wendling, O., Dennefeld, C., Chambon, P. and Mark, M., 2000. Retinoid signaling is essential for patterning the endoderm of the third and fourth pharyngeal arches. *Development*, 127, pp. 1553-1562.

Westbye, M. P, Feyzi, E., Aas, P. A., Vågbø, C. B., Talstad, V. A., Kavli, B., Hagen, L., Sundheim, O., Akbari, M., Liabakk, N.-B., Slupphaug, G., Otterlei, M. and Krokan, H. E., 2008. Human AlkB Homolog 1 Is a Mitochondrial Protein That Demethylates 3-Methylcytosine in DNA and RNA. *The Journal of Biological Chemistry*, 283 (36), pp. 25046-25056

Wikipedia, 2010a. *DNA* [Online] (Updated 10 April 2010) Available at: http://en.wikipedia.org/wiki/DNA [Accessed 12 April 2010]

Wikipedia, 2010b. *Stem Cell* [Online] (Updated 27 April 2010) Available at: http://en.wikipedia.org/wiki/Stem_cell [Accessed 27 April 2010]

Williams, R. L., Hilton, D. J. Pease, S., Wilson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A., and Gough, M. M., 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells, *Nature*, 336, pp. 684-687.

Yamanaka, Y., Ralston, A., Stephenson, R. O. and Rossant, J., 2006. Cell and Molecular Regulation of the Mouse Blastocyst. *Developmental Dynamics*, 235, pp. 2301-2314

Ying, Q-L., Nichols, J., Chambers, I. and Smith, A., 2003. BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. *Cell*, 115, pp. 281-292.

YuXiao, L., Lei, J., Yue, T., YunFang, W. and XueTao, P., 2007. The molecular mechanism of embryonic stem cell pluripotency and self-renewal. *Science in China Series C-Life Sciences*, 50 (5), pp. 619-623

Zhou, Q, Chipperfield, H, Melton, D. A. and Wong W. H., 2007. A gene regulatory network in mouse embryonic stem cells. *PNAS*, 104(42), pp. 16438-16443.

Appendix 1: Buffers and solutions

Buffers used in IMPACT™ protein purification

Buffer	Reagents	Concentration/Amount
Cell Lysis Buffer	Tris-HCl, pH 8	50 mM
	NaCl	500 mM
	EDTA	1 mM
	Triton®X-100	0.1% (v/v)
	20 µM PMSF	20 µM
Column Buffer	Tris-HCl, pH 8	50 mM
	NaCl	500 mM
	EDTA	1 mM
	Triton®X-100	0.1 % (v/v)
Cleavage Buffer	Tris-HCl, pH 8	50 mM
	NaCl	500 mM
	EDTA	1 mM
	Triton®X-100	0,1 % (v/v)
	DTT	50 mM

Buffers used in HIS-Select® protein purification

Buffer	Reagent	Concentration/Amount
Cell Lysis Buffer	NaP, pH 8	50 mM
	NaCl	500 mM
	Imidazole	10 mM
	β-merceptoethanol	5 mM
	Complete	2 tablets per 50 ml
Washing Buffer 1	NaP, pH 8	50 mM
	NaCl	500 mM
	Imidazole	10 mM
	β-merceptoethanol	5 mM
Washing Buffer 2	NaP, pH 8	50 mM
	NaCl	500 mM
	Imidazole	30 mM
	β-merceptoethanol	5 mM
Elution Buffer	NaP, pH 8	50 mM
	NaCl	500 mM
	Imidazole	250 mM
	β-merceptoethanol	5 mM

Buffers used in Dialysis and FPLC

Buffer	Reagent	Concentration/Amount
Resource Q Low Salt/Dialysis Buffer, pH 8	Tris-HCl, pH 8	50 mM
	NaCl	50 mM
	DTT	5 mM
Resource Q High Salt Buffer, pH 8	Tris-HCl, pH 8	50 mM
	NaCl	1 M
	DTT	5 mM

Resource S and HiTrap SP HP Low Salt/Dialysis Buffer, pH 6	MES, pH 6	50 mM
	NaCl	50 mM
	β-mercaptoethanol	5 mM
Resource S and HiTrap SP HP High Salt Buffer, pH 6	MES, pH 6	50 mM
	NaCl	1 M
	β-mercaptoethanol	5 mM

Protein Gel Loading Buffer (GLB)

Buffer	Reagent	Concentration/Amount
Reducing GLB	NuPAGE® LDS Sample Buffer	250 µl
	β-merceptoethanol	50 µl
	mQ-H_2O	300 µl

Electrophoresis Buffers

Buffer	Reagent	Concentration/Amount
10x TBE buffer	Tris Base	108 g
	Boric Acid	55 g
	0.5 M EDTA	40 ml
	mQ-H_2O	To a final volume of 1 L
0.5x TBE running buffer	10x TBE buffer	45 ml
	mQ-H_2O	To a final volume of 0.9 L
50x TAE	Tris Base	242 g
	Glacial acid	57.1 ml
	0.5 M EDTA	100 ml
	mQ-H_2O	To a final volume of 1 L
1x TAE electrophoresis buffer	50x TAE	200 ml
	mQ-H_2O	To a final volume of 10 L
1x NuPAGE® MOPS SDS Running Buffer	NuPAGE® MOPS SDS Running buffer (20x)	50 ml
	mQ-H_2O	To a final volume of 1 L
1x ClearPAGE™ SDS running buffer	ClearPAGE™ SDS running buffer (5x)	200 ml
	mQ-H_2O	To a final volume of 1 L

Buffers used in protein-protein interaction studies

Buffer	Reagent	Concentration/Amount
1xPBS, pH 7.4	NaCl	8 g
	KCl	0.2 g
	Sodium phosphate, dibasic. Na_2HPO_4	1.44 g
	Potassium phosphate, monobasic. KH_2PO_4	0.24 g
	mQ-H_2O	To a final volume of 1 L
PBST	PBS	
	Tween100	0.5% (v/v)
3%BSA/PBST	PBST	
	BSA (electrophoresis grade)	3% (w/v)
ELISA Binding Buffer	HEPES, pH 7.5	25 mM
	NaCl	50 mM
ELISA Washing Buffer	HEPES, pH 7.5	25 mM
	NaCl	50 mM
	Tween20	0.1% (v/v)
	BSA (electrophoresis grade)	2% (w/v)

ELISA Blocking Buffer	HEPES, pH 7.5	25 mM
	NaCl	50 mM
	Tween20	0.1% (v/v)
	BSA (electrophoresis grade)	3% (w/v)

Solutions

Solution	**Recipe**
0.1 M NaP, pH 8	26.5 ml monobasic (16.6 g NaH_2PO_4 + 500 ml mQ-H_2O) 473.5 ml dibasic (53.65 g Na_2HPO_4 + 1 L mQ-H_2O) mQ-H_2O to a total volume of 1 L
SDS-PAGE Staining Solution (2 L)	40 % (800 ml) Methanol 10 % (200 ml) Acetic Acid 0.1 % Coomassie Brilliant Blue R-250 1 L mQ-H_2O
SDS-PAGE Destaining Solution (5 L)	40 % (2 L) Methanol 10 % (500 ml) Acetic Acid 200 ml Glycerol 2.3 L mQ-H_2O

Appendix 2: Gels, Media and Plates

5% non-denaturing PAGE gel (50 ml solution)

Stock solution	5% PAGE
30% Acrylamide/bis	8.25 ml
10 x TBE buffer	12.5 ml
10% ammonium persulfate	250 µl
TEMED	25 µl
mQ-H_2O	~29 ml

Media

Medium	Reagent	Concentration/Amount
LB Medium (1L) Autoclave to sterilize	Bacto Tryptone Bacto Yeast extract NaCl pH 7.5 with NaCl mQ-H_2O	10 g 5 g 10 g To a final volume of 1 L
LB Medium (>1 L) Autoclave to sterilize	Difco™ LB Broth (Becton, Dickson and Company)	25 g per L mQ-H_2O
SOC medium Autoclave to sterilize	Bacto Tryptone Bactro Yeast Extract NaCl KCl pH 7.0 with NaOH 1M Glucose $MgCl_2$ mQ-H_2O	20 g 5 g 0.5 g 2.5 mM 5 ml 10 mM To a final volume of 1 L
MEF medium Filter sterilize	GIBCO® DMEM BioWhittaker® Pen-Strep GIBCO® GlutaMAX™-I BioWhittaker® FBS, ES qualified	500 ml 5.5 ml 5.5 ml 55 ml
mES cell medium Filter sterilize	GIBCO® KO DMEM BioWhittaker® Pen-Strep GIBCO® GlutaMAX™-I GIBCO® MEM NEAA BioWhittaker® FBS, ES qualified β-mercaptoethanol ESGRO® LIF	500 ml 6.2 ml 6.2 ml 6.2 ml 99 ml 4.4 µl 61.8 µl

Plates

Plate, protocol	LB agar + (antibiotic)	Concentration/Amount
LB plate	Bacto-Tryptone	10 g
Prepare LB agar,	Bactro-yeast extract	5 g
Autoclave to sterilize	NaCl	10 g
Pour 25 ml per 100 mm-plate	Bacto-agar	15 g
	pH 7.0 with NaOH	
	mQ-H_2O	To a final volume of 1 L
LB ampicillin plate	LB agar	1 L
Preapare LB agar	Ampicillin (10 mg/ml stock)	10 ml
Autoclave to sterilize		
Add ampicillin		
Pour 25 ml per 100 mm-plate		
LB kanamycin plate	LB agar	1 L
Preapare LB agar	Kanamycin (10 mg/ml stock)	
Autoclave to sterilize		
Add kanamycin		
Pour 25 ml per 100 mm-plate		

Appendix 3: NANOG and OCT4 DNA binding sequence and primers

Nucleotide sequence with regions bound by NANOG and OCT4

>Build 36 Chr14 77243001 - 77244620
AGAATGCCTCATCTAGACATCTGTGGACAACACCAAGAGCAAATTAAGGGACTGCATATCCTCAACGTTA
AAGTGGATTGTAAAGAAAGAAAATATAAATTGGACTACAGTTTTGGTTCCCCGTTTTCCCTCTTGCCTGT
TATGTTCCCTTAAAAAAATAATTTTTAAAAAATCTATTTACTTGAGAACAGGTTATGAGACTGGCTAATT
TTTTGTATTTTTGGCGGAAACGGGGGTCTTGCCATGCCCCAGGCTGGTCTCGAACCCCTGGGTTCAAGCG
ATCCACCCGCCTCGGCCTCCGAAGTGCGGGATTACAGGCTTGAGCCACCGCTCCCGGACTGTGTTCCCCT
TCTTAATCTCCATGGGGAAAAGCTGAGAGAGAGAACGGTGCCTGTATTCCATATATACTTGTTTGCCATT
TATTTTCAACACGGAGGGATTCTCAGAATCCTTCAATGACCTCCTTCTAAAAGAGCCAGAGGAGAAGACA
GGGACTAGTTTTACGCACAGTGGCTTAACAAACA**GCGCCGCATCCCTGTCAGTT**CTGGGATTCCTCGGGA
GTGGTTTTTTTTAAGCTGCAGAAGTCTACCCTACCGTCTACAAAGACGTTCTATGGGGTTTTACAAAAAC
CCAAACAGAAATGTGAGGAGACAGTTACAGTAAGAAATAAACACGGATTCCCAGGCTCTTTTGACTGAGC
GCCCCCGCCAGGAATGGCCGCCTAAAGAAACAGGCATCTCAACGCTTTATTGCTAAGTGTGACGGTTAGT
GGGGAGTTCGACTCTGCAGCCAAAGGAGGTCAGGAATCGTTCAATAACAATCGGTCTGGAGGTGAAAAGA
GGCGAAGAAGACACGTAAGTCCCTCCAAGACGCGGGGCAAAGCGATGGAGAGACGCGCGCCACTCCTCTC
TCTGTACCTTTTGGGCACCAGGACCCTTGCCACGGGCTGCGTGGGCCGCCGAGAAGTCGATGACCCCTTC
CAGGTCTGCGGTCCCGGGCCGGCTCTGACGGTAGAAGCGGAAAAGTTTCCGAAAGGCGTCCTCCCCGGGC
TCAGTCGCCAGAGTCGCCACAGAGCCCACGGCCGCTGCCATCTTCCCCATCTCGCGGCCTATACCCTCTG
ATCCGGAAGCAGATTCTCTCGTGTTCCGGATCCGGAAATTTTTTCCGGGGCCGCGACCTCGGCTCGAGAA
GGTGCTTTAGTCTGAAGATGGCGGCCTCAGCAGCGAGAGGTGCTGCGGCGCTGCGTAGAAGTATCAATCA
GCCGGTTGCTTTTGTGAGAAGAATT**CCTTGGACTGCGGCGTCGAG**TGAGTGATGGAGATGTGGGAGTAGT
GGAATTTTTAGGTCCAAGTGATCCTCAAAGCTTCTGCTTTTTGCAGGGTACTTAAGTCAGCGTGGTGGCT
GAATTAGGAGACTGCTCCTGAGCATGCAAGTGAGCAGAAAGAAATTTTGTTTGCAGTTAACCGTTTAGGG
ATGTCCCTTGAGCTTTTGGGGTCACGAATGCACCGGCTCTTGCCTTTAGGAGTGATTTGGTGTAAAGATT
TATTACTTGGTAACCAAGATTTACAATTTAGAAGGGATAATTTTGACTGTAGTGTGGAGGATGGATTCAA
GGTGGCAAGA

Nanog binds to the region from position 77243985 to 77244251
Oct4 binds to regions from position 77243910 to 77244190.
Both proteins bind to this region

First primer pair produces a 781 bp product

	Sequence 5'-> 3'	Length	Tm (ºC)	GC%
Forward	GCGCCGCATCCCTGTCAGTT (GAforw781)	20 bp	60.04	65%
Reverse	CTCGACGCCGCAGTCCAAGG (GArev781)	20 bp	60.11	70%

> 781 bp product produced from first primer pair >
GCGCCGCATCCCTGTCAGTTCTGGGATTCCTCGGGAGTGGTTTTTTTTAAGCTGCAGAAGTCTACCCTAC
CGTCTACAAAGACGTTCTATGGGGTTTTACAAAAACCCAAACAGAAATGTGAGGAGACAGTTACAGTAAG
AAATAAACACGGATTCCCAGGCTCTTTTGACTGAGCGCCCCCGCCAGGAATGGCCGCCTAAAGAAACAGG
CATCTCAACGCTTTATTGCTAAGTGTGACGGTTAGTGGGGAGTTCGACTCTGCAGCCAAAGGAGGTCAGG
AATCGTTCAATAACAATCGGTCTGGAGGTGAAAAGAGGCGAAGAAGACACGTAAGTCCCTCCAAGAC**GCG**
GGGCAAAGCGATGGAGAGACGCGCGCCACTCCTCT**CTCTGTACCTTTTGGGCACCAGGACCCTTGCCACG**
GGCTGCGTGGGCCGCCGAGAAGTCGATGACCCCTTCCAGGTCTGCGGTCCCGGGCCGGCTCTGACGGTAG
AAGCGGAAAAGTTTCCGAAAGGCGTCCTCCCCGGGCTCAGTCGCCAGAGTCGCCACAGAGCCCACGGCCG
CTGCCATCTTCCCCATCTCGCGGCCTATACCCTCTGATCCGGAAGCAGATTCTCTCGTGTTCCGGATCCG
GAAATTTTTTCCGGGGCCGCGACCTCGGCTCGAGAAGGTGCTTTAGTCTGAAGATGGCGGCCTCAGCAGC
GAGAGGTGCTGCGGCGCTGCGTAGAAGTATCAATCAGCCGGTTGCTTTTGTGAGAAGAATT**CCTTGGACT**
GCGGCGTCGAG

Second primer pair produces a 434 bp product

	Sequence 5'-> 3'	Length	Tm (ºC)	GC%
Forward	GCGGGGCAAAGCGATGGAGA (GAforw434)	20 bp	59.76	65%
Reverse	CTCGACGCCGCAGTCCAAGG (same sequence as GArev781)	20 bp	60.11	70%

> 434 bp product produced from second primer pair
GCGGGGCAAAGCGATGGAGAGACGCGCGCCACTCCTCT**CTCTGTACCTTTTGGGCACCAGGACCCTTGCC**
ACGGGCTGCGTGGGCCGCCGAGAAGTCGATGACCCCTTCCAGGTCTGCGGTCCCGGGCCGGCTCTGACGG
TAGAAGCGGAAAAGTTTCCGAAAGGCGTCCTCCCCGGGCTCAGTCGCCAGAGTCGCCACAGAGCCCACGG
CCGCTGCCATCTTCCCCATCTCGCGGCCTATACCCTCTGATCCGGAAGCAGATTCTCTCGTGTTCCGGAT
CCGGAAATTTTTTCCGGGGCCGCGACCTCGGCTCGAGAAGGTGCTTTAGTCTGAAGATGGCGGCCTCAGC
AGCGAGAGGTGCTGCGGCGCTGCGTAGAAGTATCAATCAGCCGGTTGCTTTTGTGAGAAGAATT**CCTTGG**
ACTGCGGCGTCGAG

Primer pair for Oct4 binding site produces a 280 bp product

	Sequence 5'->3'	Length	Tm (ºC)	GC%
Forward	CTCTGTACCTTTTGGGCACC (GAforwoct4)	20	54	55
Reverse	TCTCGAGCCGAGGTCGCGGC (GArevoct4)	20	62	75

> Oct4 DNA binding site, 280 bp

CTCTGTACCTTTTGGGCACCAGGACCCTTGCCACGGGCTGCGTGGGCCGCCGAGAAGTCGATGACCCCTT
CCAGGTCTGCGGTCCCGGGCCGGCTCTGACGGTAGAAGCGGAAAAGTTTCCGAAAGGCGTCCTCCCCGGG
CTCAGTCGCCAGAGTCGCCACAGAGCCCACGGCCGCTGCCATCTTCCCCATCTCGCGGCCTATACCCTCT
GATCCGGAAGCAGATTCTCTCGTGTTCCGGATCCGGAAATTTTTTCCGGGGCCGCGACCTCGGCTCGAGA

Primer pair for Nanog binding site produces a 266 bp product

	Sequence 5'->3'	Length	Tm (ºC)	GC%
Forward	TCTGCGGTCCCGGGCCGGCT (GAforwnanog)	20	64	80
Reverse	TTCTACGCAGCGCCGCAGCA (GArevnanog)	20	58	65

> Nanog DNA binding site, 266 bp

TCTGCGGTCCCGGGCCGGCTCTGACGGTAGAAGCGGAAAAGTTTCCGAAAGGCGTCCTCCCCGGGCTCAG
TCGCCAGAGTCGCCACAGAGCCCACGGCCGCTGCCATCTTCCCCATCTCGCGGCCTATACCCTCTGATCC
GGAAGCAGATTCTCTCGTGTTCCGGATCCGGAAATTTTTTCCGGGGCCGCGACCTCGGCTCGAGAAGGTG
CTTTAGTCTGAAGATGGCGGCCTCAGCAGCGAGAGGTGCTGCGGCGCTGCGTAGAA

Melting temperature and GC% content calculated using BioMath calculators at http://www.promega.com/biomath/calc11.htm

Appendix 4: ClustalW2 Multiple sequence alignment (MSA) results

MSA of all three builds (35, 36 and 37) of the human genome

Only a section of the MSA output file is shown as the entire file is very big. * indicate nucleotide positions that are identical in all three query sequences.

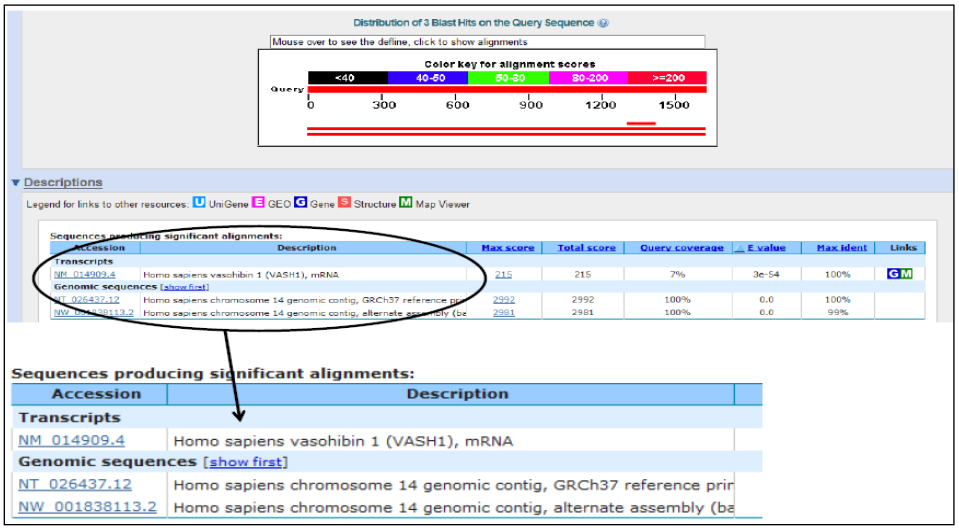
```
Build 35        AGAATGCCTCATCTAGACATCTGTGGACAACACCAAGAGCAAATTAAGGGACTGCATATC 60
Build 36        AGAATGCCTCATCTAGACATCTGTGGACAACACCAAGAGCAAATTAAGGGACTGCATATC 60
Build 37        ------------TCAGAC-CATGTGGAAGACAATGAGGGCAGTGTTCCTTCTTCCTCTTC 47
                            ****   ******  ***   ** ***   *       * *   **

Build 35        CT--CAACGTTAAAGTGGATTGTAA-AGAAAGAAAATATAAATTGGACTACAGTTTTGGT 117
Build 36        CT--CAACGTTAAAGTGGATTGTAA-AGAAAGAAAATATAAATTGGACTACAGTTTTGGT 117
Build 37        CCACCCACACTTGAGCAGGAGGTGATAGGGAGTTGGGCTGGGGGCTGTTGTACCAAAAAC 107
                *   * **  *  **  *   ** * **  **      *         *  *

Build 35        TCCCCGTTTTCCCTCTTGCCTGTTATGTTCCCTTAAAAAAATAATTTTTAAAAAATCTAT 177
Build 36        TCCCCGTTTTCCCTCTTGCCTGTTATGTTCCCTTAAAAAAATAATTTTTAAAAAATCTAT 177
Build 37        AGCTGGGCTGAGCTGTTGAGCCCTAGAGCCTTTCAGCAGGCAGAATTCTGACATCTGCCT 167
                  *  *  *   ** ***     **    *  * *  *     * ** * * *  *   *

Build 35        ---TTACTTGAGAACAGGTTATGAGACTGGCTAATTTTTTGTATT-TTTGGCGGAAACGG 233
Build 36        ---TTACTTGAGAACAGGTTATGAGACTGGCTAATTTTTTGTATT-TTTGGCGGAAACGG 233
Build 37        GGGTTCTCCTGGGGCTGCCAGGGCCCCCTGCTGGGTATCTCCATAGCTCAGTCACTGCTG 227
                   **      *  * *     *   *  ***   * * *  **   *  *      * *

Build 35        GGGTCTTGCCATGCCCCAGGCTGGTCTCGAACCCCTGGGTTCAAGCGATCCACCCGCCTC 293
Build 36        GGGTCTTGCCATGCCCCAGGCTGGTCTCGAACCCCTGGGTTCAAGCGATCCACCCGCCTC 293
Build 37        CACTGTGACTGCCCCTCCAGGCAGCCTGGAGCCTCCGC-CTCAA-----CCACCGACCTC 281
                   * *  *    ** *  *   * ** ** ** * *   ****     *****  ****
```

MSA of builds 35 and 36 of the human genome

Only a section of the MSA output file is shown as the entire file is very big. * indicate nucleotide positions that are identical in both query sequences, which include all positions.

```
Build 35        AGAATGCCTCATCTAGACATCTGTGGACAACACCAAGAGCAAATTAAGGGACTGCATATC 60
Build 36        AGAATGCCTCATCTAGACATCTGTGGACAACACCAAGAGCAAATTAAGGGACTGCATATC 60
                ************************************************************

Build 35        CTCAACGTTAAAGTGGATTGTAAAGAAAGAAAATATAAATTGGACTACAGTTTTGGTTCC 120
Build 36        CTCAACGTTAAAGTGGATTGTAAAGAAAGAAAATATAAATTGGACTACAGTTTTGGTTCC 120
                ************************************************************

Build 35        CCGTTTTCCCTCTTGCCTGTTATGTTCCCTTAAAAAAATAATTTTTAAAAAATCTATTTA 180
Build 36        CCGTTTTCCCTCTTGCCTGTTATGTTCCCTTAAAAAAATAATTTTTAAAAAATCTATTTA 180
                ************************************************************

Build 35        CTTGAGAACAGGTTATGAGACTGGCTAATTTTTTGTATTTTTGGCGGAAACGGGGGTCTT 240
Build 36        CTTGAGAACAGGTTATGAGACTGGCTAATTTTTTGTATTTTTGGCGGAAACGGGGGTCTT 240
                ************************************************************

Build 35        GCCATGCCCCAGGCTGGTCTCGAACCCCTGGGTTCAAGCGATCCACCCGCCTCGGCCTCC 300
Build 36        GCCATGCCCCAGGCTGGTCTCGAACCCCTGGGTTCAAGCGATCCACCCGCCTCGGCCTCC 300
                ************************************************************

Build 35        GAAGTGCGGGATTACAGGCTTGAGCCACCGCTCCCGGACTGTGTTCCCCTTCTTAATCTC 360
Build 36        GAAGTGCGGGATTACAGGCTTGAGCCACCGCTCCCGGACTGTGTTCCCCTTCTTAATCTC 360
                ************************************************************

Build 35        CATGGGGAAAAGCTGAGAGAGAGAACGGTGCCTGTATTCCATATATACTTGTTTGCCATT 420
Build 36        CATGGGGAAAAGCTGAGAGAGAGAACGGTGCCTGTATTCCATATATACTTGTTTGCCATT 420
                ************************************************************
```

Appendix 5: Build 37 nBlast hits against Human genomic + transcripts" database



Appendix 6: Instruments

Instrument	Produced by
TC-512, Thermal Cycler	Techne
PTC-200, Peltier Thermal Cycler	MJ Reseach
StepOnePlus Real-Time PCR System	Applied Biosystems
ÄKTA explorer FLPC system	Amersham Pharmacia Biotech
ÄKTA purifier FLPC system	Amersham Pharmacia Biotech
Megafuge 1.0	Heraeus Instruments
Allegra™ X-22R Centrifuge	BECKMAN COULTIER
Avanti™ J-25	BECKMAN COULTIER
Innova™ 4230 Refridgerated Incubator Shaker	New Brunswick Scientific
Innova™ 4000 Incubator Shaker	New Brunswick Scientific
EPS-301, Electrophoresis Power Supply	Amersham Pharmacia Biotech
EPS 2A200, Electrophoresis Power Supply	Hoefer
UV-1601, UV-visible Spectrophotometer	SHIMADZU
VICTOR^2 1420 Multilabel Counter	Wallac
Typhoon 9410 Variable Mode Imager Scanner	Amersham Biosciences
Tri-Carb 2900 TR Liquid Scintillation Analyzer	Packard
NanoDrop®	Saveen Werner
DNA Analyzer 3730	International Biotechnologies Inc

Appendix 7: Abbreviations and Units

Abbreviations

‘	Prime
1alkylA	1-alkyladenine
1MeA	N1-methyladenine
1MeG	1-methylguanine
^{14}C	Carbon 14
2-OG	2-oxoglutarate
3alkylA	3-alkyladenine
3MeA	3MeA
3MeC	N3-methylcytosine
3MeT	3-methylthymine
^{32}P	Phosphate 32
5MeC	5-methylcytosine
A	Adenine
ALKBH	AlkB homologue (in humans)
Alkbh	AlkB homologue (in mice)
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BER	Base excision repair
Blast	Basic Local Alignment Search Tool
BMP	Bone morphogenic protein
BP	Base Pairs
BSA	Bovine serum albumin
C	Cytosine
CBD	Chitin binding domain
cDNA	Complementary DNA
CH_3	Methyl
CHIP	Chromatin Immunoprecipitation
CNS	Central Nervous system
CO_2	Carbon dioxide
Cobalt	Co^{2+}
CRABP	Cellular RA-binding protein
Ct	Cycle treshold
C-terminal	Carboxyl-terminal
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DNPH	2,4-Dinitrophenylhydrazine
dNTP	Deoxy-trinucleotide phosphate
DPBS	Dulbecco’s Phosphate Buffered Saline
ds DNA	Double stranded DNA
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EBs	Embryonic Bodies
EDC	*N*-(3-dimethylaminopropyl)-*N*’-ethylcarbodiimide hydrochloride
EDTA	Ethylenedinitrilo-tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EmGFP	Emeral Green Fluorescent Protein
EMSA	Electrophoretic mobility shift assay
ENK	Early embryonic specific NK
EPI	Epiblast

ERK	Extracellular signal-regulated kinase
ES cell	Embryonic stem cell
EST	Expressed sequence tag
EtBr	Ethidium Bromide
FBS	Fetal bovine serum
FPLC	Fast Performance Liquid Chromatography
G	Guanine
GAPDH	Glyceraldehydes-3-phosphate dehydrogenase
GLB	Gel Loading Buffer
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H4	Histone 4
H-bonds	Hydrogen bonds
HCl	Hydrogen chloride
HDACs	Histone deacetylases
HEPES	*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid
HMG	High mobility group
ICM	Inner Cell Mass
Id	Inhibitor of differentiation
IEX	Ion exchange chromatography
IMAC	Immobilized metalion affinity chromatography
IMPACT	Intein Mediated Purification with an Affinity Chitin-binding Tag
IPTG	Iso-propyl-β-D-thiogalactoside
JAK	Janus Kinase
JHDM	JmjC domain containing histone demethylase
K	Lysine
KCl	Potassium chloride
LB	Lysogeny broth
LIF	Leukemia inhibitory factor
LIFR/LIFRβ	LIF receptor
LSC	Liquid Scintillation Cocktail
Mcm^5U	5-methylcarboxymethyl uridine
MEF	Murine embryonic fibroblast
MEM NEAA	Minimal essential medium, non-essentail amino acid solution
MES	2-(*N*-morpholino)ethanesulfonic acid
mES cells	Mouse Embryonic Stem cells
Min	Minute
MMS	Methylmethanesulfonate
mRNA	Messenger RNA
MSA	Multiple Sequence Alignment
MSC	Multiple cloning site
N	Nitrogen
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NaP	Sodium phosphate
nBlast	Nucleotide Blast
NER	Nucleotide excision repair
NH	Amine
N-terminal	NH_2-terminal
O	Oxygen
O_2	Dioxygen/molecular oxygen

O^4alkylT	O^4-alkylthymine
O^6alkylG	O^6-alkylguanine
O^6MeG	O^6-methylguanine
ON	Overnight
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PE	Primitive endoderm
pI	Isoelectric point
PMSF	Phenylmethylsulfonyl fluoride
PNPP	*p*-nitrophenyl phosphate
RA	Retinoic Acid
RAR	RA receptor
RARE	RA response elements
RARα/β/γ	RA receptor-α/β/γ
RNA	Ribonucleic acid
RRM	RNA recognition motif
RT	Room temperature
RT-PCR	Real-time PCR
RXR	Retinoid X receptor
RXRγ	Retinoid X receptor γ
S. pombe	*Schizosaccharomyces pombe*
SAM	*S*-adenosylmethionine
SD	Standard deviation
SDS-PAGE	Sodium dodecyl Sulfate Polyacryamide Gel Electrophoresis
S_N	Nucleophilic substitution
ss DNA	Single stranded DNA
STAT	Signal transducer and activator of transcription
Sulfo-NHS	*N*-hydroxysulfosuccinimide
T	Thymine
T4 PNK	T4 polynucleotide kinase
TEA	Tris acetate (buffer)
TBE	Tris borate (buffer)
TEMED	*N*,*N*,*N'*,*N'*-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
U	Uracil
UV	Ultraviolet

Units

˚C	Degrees Celsius
g	Grams
mg	Milligram (10^{-3} g)
µg	Microgram (10^{-6} g)
ng	Nanogram (10^{-9} g)
L	Liter
ml	Milliliter (10^{-3} L)
µl	Microliter (10^{-6} L)
M	Molar
mM	Millimolar (10^{-3} M)
µM	Micromolar (10^{-6} M)
pmol	Picomole (10^{-12} mole)
fmol	Femtomole (10^{-15} mole)

A	Absorbance
OD	Optical density
DPM	disintegrations per minute
nm	Nanometer (wavelength of light)
kb	Kilobases (10^3 b)
kDa	Kilo Daltons (10^3 Daltons)
RPM	Revolutions per minute
U	Units
V	Volts