DNA Damage and Base Excision Repair; An Important Role during Zebrafish Embryogenesis.

By Stephen Paul Gifford Moore

B.S. in Biology, University of Auckland
P.G.Dip.Sci, in Marine Science, University of Auckland
M.S. in Marine Science, University of Auckland

A dissertation submitted to

The Faculty of
the College of Science of
Northeastern University
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

December 1st, 2017

Dissertation directed by

Phyllis R. Strauss
Dedication


Gone, but not forgotten.
Acknowledgements

This dissertation represents over five years of extremely hard work, some of which I did.

First and foremost, I wish to extend my deepest gratitude to my advisor, Dr Phyllis Strauss, for giving me the opportunity to join her lab. I learnt so much during my dissertation, and have become a more educated, confident, and responsible scientist and person. I owe this to Dr. Strauss’ unwavering patience, support, and mentorship over the years.

I would also like to thank Dr Erin Cram, Dr Veronica Godoy, and Dr James Monaghan, from the Northeastern University Department of Biology, who were members of my dissertation committee. I am especially grateful for all the help and advice I received on my research goals, and help keeping me focused on end goals. I would especially like to thank the fifth member of my committee, Dr Leona Samson of the Center for Environmental Health Sciences at MIT, for her mentorship and advice over the years.

I would like to save a very big and special thank you to all the past members of the Strauss lab for their companionship and help. I would especially like to thank Ben Snow for the research that he did as part of his M.S. His findings were instrumental in guiding the direction that I took my own research in. Also, thank you to Kim Toomire and Josh Kruchten, who both did research projects on my behalf that we managed to publish. Thanks to Brian Dobosh too, who also did an amazing job with some qPCR work, that unfortunately did not end up being used. I also want to say thank you to everyone in the Cram lab, where I spent a great deal of time in my final year and half at Northeastern. All your humor, friendship and helpful advice inspired me to continue in my work, despite frustrations.

Thank you to the Biology Department of Northeastern University, particularly Dr Wendy Smith for all the work in the background that she did and for providing me with bench space to work at once our lab had closed. Thanks to Adrian Gilbert who made stuff happen, and Frauke Argyros who made teaching Biochemistry labs much more tolerable. A big thank you to all my fellow graduate students in the Biology department over the years who made my time at Northeastern University all that more enjoyable.
And finally, thank you to all my friends and family who supported me unwaveringly during a very busy and difficult time. A very special chur bro to my wife Maria who had to make so many sacrifices over the years while I chased rainbows and hunted unicorns, thank you! The final thank you goes to my daughter Ada, thank you for being you.
Abstract of Dissertation

Damage to DNA is an unavoidable consequence of life. Unrepaired DNA damage is mutagenic, promotes genomic instability, and leads to the onset of numerous serious diseases. Multiple pathways have evolved to repair specific types of DNA damage. Of these, the multi-enzyme base excision repair (BER) pathway is considered the most active, because BER repairable damage is ongoing. Embryogenesis is a highly complex and regulated process. Since developing embryonic cells divide rapidly, unrepaired DNA damage leads to mutation and cell death. Many BER enzymes are embryonic lethal as knockouts, but not in adult cell cultures. This indicates an important role during embryonic development. Active DNA demethylation of 5-methylcytosine (5mC) in CpG islands during embryogenesis may result in DNA damage requiring BER. AP endonuclease (Apex1) is an essential BER enzyme. It incises abasic (AP) sites resulting from removal of DNA lesions by DNA glycosylases, allowing subsequent repair. Accumulation of AP sites is toxic. Knockdown (K/D) of Apex1 in zebrafish embryos results in a consistent and concurrent loss of the critical transcription factor (TF) Creb1. Many Creb1 dependent genes are subsequently perturbed resulting in abnormal embryo development. Death occurs at ~7 days post fertilization (7dpf). Research in this dissertation shows that DNA damage increases in zebrafish embryos at the mid-blastula transition (MBT) when zygotic genome activation (ZGA) occurs. Damage is further elevated with Apex1 K/D, thereby illustrating the importance of fully functional BER during embryonic development. Furthermore, increased DNA damage inversely correlates with decreased levels of 5mC and vice versa, providing indirect evidence that active DNA demethylation is at least one source of elevated embryonic DNA damage. The Creb1 binding site (CRE site [TGACGTCA]) is present within the promoter CpG island of the creb1 gene and some 5000 other genes. It is likely that active DNA demethylation of the Creb1 promoter may also affect the CRE site CpG dinucleotide. My published work shows that binding of recombinant CREB1 to the CRE site is modulated by DNA damage, and abolished by 5mC in the CpG. We also find that both recombinant CREB1 and BER glycosylases compete when DNA damage occurs within a TFs binding sequence, and demonstrably affects embryo development. These novel findings provide insight into the presence, timing, and potential source of DNA damage during zebrafish embryogenesis. They also show how DNA damage may act in an epigenetic fashion when it occurs in a TF binding site, and that these results are valid both in vitro and in vivo. We speculated that BER substrates act in an epigenetic fashion.
Table of Contents

Dedication ................................................................................................................. 2
Acknowledgements .................................................................................................... 3
Abstract of Dissertation............................................................................................ 5
Table of Contents ..................................................................................................... 6
List of Figures .......................................................................................................... 8
List of Tables .......................................................................................................... 9

Chapter 1. Introduction ............................................................................................ 10
  1.1. DNA damage and repair systems ................................................................. 10
      1.1.1. Nucleotide excision repair ................................................................. 11
      1.1.2. Mismatch repair .............................................................................. 12
      1.1.3. Homologous recombination (HR) and non-homologous end joining (NHEJ) .................................................................................. 12
      1.1.4. Base excision repair ........................................................................ 13
  1.2. AP endonuclease ............................................................................................. 14
      1.2.1. Non-endonuclease function of Apex1 ................................................. 16
  1.3. Zebrafish as a model system .......................................................................... 17
  1.4. Base excision repair and embryonic development ....................................... 17
      1.4.1. Apex1 studies and zebrafish embryogenesis ....................................... 18
  1.5. DNA methylation ........................................................................................... 19
      1.5.1. DNA methylation ............................................................................. 20
      1.5.2. DNA demethylation ...................................................................... 20
      1.5.3. DNA methylation and embryogenesis ............................................. 21

Chapter 2. DNA damage and cytosine methylation status are temporally modulated
  during zebrafish embryogenesis. ............................................................................ 26
  2.1. Introduction .................................................................................................... 27
  2.2. Experimental procedures .............................................................................. 29
  2.3. Results ............................................................................................................ 32
  2.4. Discussion ..................................................................................................... 35

Chapter 3. DNA lesions can modulate transcription factor binding. ....................... 44
  3.1. DNA modifications repaired by base excision repair are epigenetic. .......... 45
  3.2. Abstract .......................................................................................................... 45
  3.3. Abbreviations ............................................................................................... 46
  3.4. Introduction ................................................................................................... 47
3.4.1. Oxidative damage to DNA is ongoing in all cells and not considered to be epigenetic.  
3.4.2. The BER pathway, which repairs most oxidative damage and removes uracil in DNA, is a  
cascade of critical enzymes.                                                                                    47
3.4.3. The oxidative state of DNA could be epigenetic and effect important transcription binding  
factors like CREB1.                                                                                             48

3.5. Methods..................................................................................................................................................... 49
3.5.1. Oligonucleotides ................................................................................................................................. 49
3.5.2. Protein.................................................................................................................................................... 49
3.5.3. Electrophoretic Mobility Shift Assay (EMSA).................................................................................... 49
3.5.4. Binding Kinetics ............................................................................................................................... 49
3.5.5. Statistical analyses ............................................................................................................................ 50

3.6. Results.......................................................................................................................................................... 50
3.6.1. Purified CREB1 recognizes and binds the CRE-site cleanly............................................................. 50
3.6.2. CREB1’s ability to bind to its consensus sequence is affected both positively and negatively  
by specific modifications and their locations.                                                              51
3.6.3. DNA modification and spatial position within the CRE-site affects CREB1’s ability to form  
a dimer.                                                                                                   51

3.7. Discussion .................................................................................................................................................... 52
3.7.1. Major Conclusions ............................................................................................................................ 52
3.7.2. Response to oxidative DNA damage is pleomorphic and probably extends to other  
transcription factors.                                                                                        52
3.7.3. DNA damage repaired by BER is likely to be epigenetic.                                           54

3.8. Acknowledgments ...................................................................................................................................... 54
3.9. Conflict of Interest .................................................................................................................................... 54

Chapter 4. DNA glycosylases and transcription factors compete for damaged  
transcription factor binding sequences........................................................................................................... 62
4.1. Damaged Promoter Sites............................................................................................................................ 63
4.2. Abstract...................................................................................................................................................... 63
4.3. Introduction................................................................................................................................................. 64
4.4. Experimental procedures .......................................................................................................................... 65
4.5. Results ......................................................................................................................................................... 67
4.6. Discussion ................................................................................................................................................... 70
4.7. Acknowledgements .................................................................................................................................... 72
4.8. Conflict of interest ..................................................................................................................................... 72

Chapter 5. General discussion and future directions of study. ................................................................. 81

References.......................................................................................................................................................... 87
List of Figures

Chapter 1
Figure 1 ........................................................................................................................................ 15
Figure 2 ........................................................................................................................................ 16

Chapter 2
Figure 1 ........................................................................................................................................ 30
Figure 2 ........................................................................................................................................ 31
Figure 3 ........................................................................................................................................ 32
Figure 4 ........................................................................................................................................ 33
Figure 5 ........................................................................................................................................ 34

Chapter 3
Figure 1 ........................................................................................................................................ 47
Figure 2 ........................................................................................................................................ 48
Figure 3 ........................................................................................................................................ 49
Figure 4 ........................................................................................................................................ 50
Figure 5 ........................................................................................................................................ 51
Figure 6 ........................................................................................................................................ 52

Chapter 4
Figure 1 ........................................................................................................................................ 64
Figure 2 ........................................................................................................................................ 65
Figure 3 ........................................................................................................................................ 66
Figure 4 ........................................................................................................................................ 67
Figure 5 ........................................................................................................................................ 68
Figure 6 ........................................................................................................................................ 69
Figure 7 ........................................................................................................................................ 70
Figure 8 ........................................................................................................................................ 71
List of Tables

Chapter 1
Table 1 .................................................................................................................. 29
Table 2 .................................................................................................................. 29

Chapter 2
Table 1 .................................................................................................................. 46

Chapter 3
Table 1 .................................................................................................................. 64
Chapter 1. Introduction

1.1. DNA damage and repair systems

All life requires a mechanism to store, retrieve, and translate the information essential to make and maintain itself. In the 1950s it was determined that the primary molecule responsible for this was DNA (Watson and Crick 1953). Initially thought of as an extremely stable molecule, DNA was subsequently found to be highly dynamic and susceptible to damage. In fact, it has been estimated that approximately $10^5$ DNA lesions occur per day in mammalian genomes (Iyama and Wilson 2013). DNA damage can occur through the intrinsic instability of specific chemical bonds including tautomeric shifts, deamination of bases, and loss of bases, or through interaction with various physical and chemical agents (Friedberg et al. 2006; Sinden 1994). Sources of DNA damage agents can be exogenous, such as ultraviolet light generating bulky cyclobutane dimers from adjacent pyrimidines and ionizing radiation that damages bases and breaks phosphodiester bonds (Friedberg et al. 2006). Alternatively, damage agents can be endogenous, for example, uracil as dUTP, an analogue of thymine, is routinely incorporated into DNA during DNA synthesis or arises through spontaneous deamination of cytosine (Zharkov, Mechetin, and Nevinsky 2010), while spontaneous hydrolysis of nucleotides generates highly deleterious apurinic/apyrimidinic or abasic (AP) sites (Sczepanski et al. 2010). However, arguably the most important endogenous source of damage, both quantitatively and qualitatively, is the ongoing generation of reactive oxygen species (ROS).

ROS are ubiquitous in aerobic organisms and mainly arise from leakage of electrons during mitochondrial respiration when oxygen is reduced to water through the electron transport chain (Barzilai and Yamamoto 2004). Of the approximately $10^5$ DNA lesions occurring per day, ~$10^4$ are due to oxidization and/or single stand breaks (Lindahl 1993). ROS consist of, among others, the electrophilic hydroxyl radical (HO•), superoxide radical (O$_2$•$^-$), and hydrogen peroxide (H$_2$O$_2$). It is likely that H$_2$O$_2$ is the primary culprit (Marnett 2000). Most endogenous ROS are generated by mitochondrial respiration in the form of O$_2$•$, this is converted to H$_2$O$_2$ by superoxide dismutase (De Bont and van Larebeke 2004). If H$_2$O$_2$ molecules reach the nucleus, they can react with iron-complexed DNA via Fenton Reactions forming HO•, which will then react with the DNA. Since HO• is highly reactive, it can only travel one or two molecule lengths before reacting with cellular components (Imlay, Chin, and Linn 1988; Izatt, Christensen, and Rytting 1971; Marnett 2000; Pryor 1986). This assault results in either hydrogen abstraction or addition to double bonds resulting in two main modes of DNA attack: (1) addition to the double bonds of DNA bases and (2) hydrogen removal from the deoxyribose sugar units of DNA (Cadenas 1989). These modes of attack result
in the production of over 80 products, as well as fragmentation of DNA, base loss and strand breaks when the sugar residue is attacked. The most common base lesion caused by ROS is 7,8-dihydro-8-oxoguanine (8-oxoG, $^{\text{o}}$G), a modification of guanine where the oxygen radical breaks the normal double bond between a nitrogen and carbon atom and attaches itself to that same carbon atom. $^{\text{o}}$G base-pairs with adenine and can lead to transversion mutations after replication (Friedberg et al. 2006).

DNA damage results in genomic instability, considered the principal cause of most cancers, and promotes cell responses that play an important role in neurodegenerative disease and aging (Collins 1999). Since genomic maintenance is critically important, it comes as no surprise that multiple DNA repair pathways have evolved. The major DNA repair pathways are: nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination (HR), non-homologous end joining (NHEJ), and base excision repair (BER). While each enzymatic pathway is distinct, a number of auxiliary components such as PARP1 participate in cross-talk among multiple pathways (Fishel, Vascotto, and Kelley 2013). Although I will briefly review the first four, I will emphasize BER, as it repairs DNA damage that is ongoing.

1.1.1. Nucleotide excision repair

NER is an extremely versatile DNA repair pathway that recognizes the distorted DNA structure produced by the presence of bulky DNA lesions, typically cyclobutane pyrimidine-dimers (CPD) or 6-4 photoproducts (6-4PP) from UV light. NER has two modes of repair: (1) global genome repair (GGR), where detection and repair of lesions is independent of transcription, and (2) transcription-coupled repair (TCR), in regions of the genome undergoing active transcription. In mammals a complex consisting of three subunits, xeroderma pigmentosum complementation group C (XPC), HR23B, and centrin-2 (CETN2), senses the distorted DNA, though it is XPC that actually recognizes the damage and initiates repair (Sugasawa et al. 1998). In contrast, TCR occurs when actively transcribing RNA polymerase II (RNAPII) stalls at the site of a lesion, serving as a damage recognition signal. Recognition is followed by recruitment of various factors, most notably cockayne syndrome B (CSB) (Hanawalt and Spivak 2008). Both modes of repair then recruit transcription factor II H (TFIIH), subunits of which act as a helicase and ATPase, to unwind the DNA at the lesion site. The endonuclease functions of xeroderma pigmentosum complementation groups G and F (XPG and XPF) cleave the damaged strand 3′ and 5′ respectively to the lesion, resulting in removal of a single-stranded 24-32 nucleotide fragment. Replicative DNA polymerase $\delta$ or $\epsilon$ can then fill the resultant gap by using the undamaged complementary strand as a template. The ensuing nicks are subsequently sealed by DNA ligase III (Hanawalt and Spivak 2008; Kamileri, Karakasilioti, and Garinis 2012; Nouspikel 2009; Sugasawa et al. 1998).
1.1.2. Mismatch repair

Nucleotide mismatches occur at a rate of one every 10^4 – 10^5 nucleotides during normal DNA synthesis in *E. coli*, though DNA polymerase proofreading functions reduce the error rate to approximately one in every 10^7 nucleotides during normal DNA synthesis in *E. coli* (Iyer et al. 2006). These rates are estimated to be the same in mammals (Kunkel 2004). Nucleotide mismatches that escape editing by DNA polymerase proofreading, except for U/G mispairs, are repaired through MMR, a highly conserved repair pathway. Noncanonical base-pairings are recognized by a MutSa heterodimer (formed by MutS homologue MSH2 and MSH6), or a MutSβ heterodimer (formed by MSH2 and MSH3). MutSa preferentially recognizes mispairs of 1-2 base and single base mismatches, while MutSβ recognizes mispairs of 2 bases and larger. Once recognition is achieved, MutL homolog 1 (MLH1) forms a heterodimer with postmeiotic segregation increased 2 (PMS2) to form MutLα. MutLα cleaves the daughter strand 5' to the lesion and activates an excision system that includes PCNA and replication factor C (RFC). Endonuclease 1 (EXO1) in conjunction with MutLα then cleaves DNA at the 3' side of the lesion, resulting in removal of a single-stranded DNA fragment. DNA polymerase δ fills the resultant gap, which is then sealed by DNA ligase I (Iyer et al. 2006; Jiricny 2006; Li 2008).

1.1.3. Homologous recombination (HR) and non-homologous end joining (NHEJ)

Any reagent capable of severing the DNA backbone can introduce double strand breaks (DSBs). Such reagents include ROS, though ionizing radiation is particularly important. Note also that DSBs are the ultimate result of accumulation of single strand breaks occurring when other repair pathways are initiated but fail to proceed to completion. DSBs have serious consequences if unrepaired such as cell death, or genome instability through chromosomal translocations when misrepaired. Either HR or NHEJ pathways can repair DSBs. HR occurs mainly during S and G2 phases of the cell cycle and relies upon the forced invasion of homologous sequences by the strand being repaired. In most eukaryotes, HR is initiated by the binding of the MRN complex (RAD50-MRE11-NBS1) at the site of DSBs. This initiates 5'-3' resection at the DSB generating the 3' ssDNA overhang required for strand invasion. ssDNA is stabilized by binding of replication protein A (RPA), which in turn activates ATR via ATR activating protein (ATRIP); this cascade results in replacement of RPA with the recombinase RAD51. The RAD51-DNA complex invades an intact homologous region, the invading strand is then extended by DNA polymerase δ, and the DNA is sealed by a DNA ligase (Krejci et al. 2012; Li and Heyer 2011; San Filippo, Sung, and Klein 2008).

NHEJ is the major DSB repair pathway in eukaryotes, especially during cell cycle phases where no homologous sister chromatid is present. It is initiated when a Ku70/80 heterodimer captures both ends of the broken DNA molecule. This protein-DNA complex in turn recruits the DNA-dependent protein
kinase catalytic subunit (DNA-PKcs) bringing the DNA ends together. Processing of the DNA ends occurs through the enzymes Artemis, DNA polymerase μ and λ, tyrosyl-DNA phosphodiesterase (TDP1), or polynucleotide kinase (PNK) depending upon the required processing. This is followed by ligation of the processed DSBs by a ligase IV-XRCC4 complex (Davis and Chen 2013; Hefferin and Tomkinson 2005; Mahaney, Meek, and Lees-Miller 2009).

1.1.4. Base excision repair

BER is responsible for the repair of small, non-bulky base damage, such as 6-OH and G/U and T/G mismatches (Zharkov 2008). It may also play a role in the demethylation of 5-methylcytosine at CpG islands (Wu and Zhang 2010). Since oxidation, deamination, and spontaneous hydrolysis occur so frequently, BER occurs at all cell cycle stages and is required for cell viability (Zharkov 2008).

Initiation of BER occurs when a damaged nucleobase is recognized and excised by a lesion specific mono- or bi-functional DNA glycosylase (Figure 1). For example, 8-oxoguanine glycosylase (OGG1) recognizes 6-OH, while uracil-DNA glycosylase (UDG) recognizes uracil mismatched with any canonical base and thymine-DNA glycosylase (TDG) recognizes thymine or uracil mismatched with guanine. Excision occurs by the hydrolysis of the N-glycosidic bond attaching the base to the deoxyribose backbone and generates a toxic AP site. The resulting AP sites are incised by apurinic/apyrimidinic endonuclease 1 (Apex1), which cleaves the DNA backbone immediately 5’ to the AP site. Monofunctional DNA glycosylases possess only glycosylase activity so that cleavage of the resulting AP site by Apex1 leaves a 3’-hydroxyl (3’-OH) and a 5’-deoxyribose phosphate (5’-dRP) group at the nick. Conversely, bifunctional DNA glycosylases possess both glycosylase activity and 3’ AP lyase activity (David, O’Shea, and Kundu 2007; Robertson et al. 2009; Zharkov 2008). After removal of the base a bifunctional glycosylase will cleave the backbone immediately 3’ to the AP site producing a 3’-phospho-unsaturated aldehyde (3’-PUA) that is subsequently removed by Apex1 creating a 3’-OH group (Iyama and Wilson 2013).

Once the necessary 3’-OH and 5’-dRP groups have been generated, completion of BER occurs via one of two pathways. The majority of repair is achieved through short patch (SP) repair, though long patch (LP) repair can occur when ATP concentrations are low, during S-phase (Gary 1999; Iyama and Wilson 2013), or when the downstream 5’ end is blocked. During SP repair DNA polymerase β (Polβ) replaces the missing nucleotide and the resultant nick is sealed by the XRCC1–LIG3α complex. In contrast, LP repair utilizes DNA polymerase δ/ε in conjunction with clamp loading factor RFC and processivity factor PCNA to synthesize a new 2-13 nucleotide sequence. Synthesis of the new sequence results in displacement of the original strand, which is removed by flap endonuclease 1 (FEN1), and the resultant nick is sealed by a DNA ligase (Dianov and Hübscher 2013; Iyama and Wilson 2013; Robertson et al. 2009; Zharkov 2008).
1.2. AP endonuclease

Apurinic/apyrimidinic endonuclease, also known by many aliases including AP endonuclease, HAP1, Ape1, Ref-1, or Apex1, is a ubiquitous, multifunctional, 35.5 kDa nuclear enzyme that catalyzes the incision of DNA at AP sites (Kane and Linn 1981; Robson and Hickson 1991; Shaper, Grafstrom, and Grossman 1982). Found across all prokaryote and eukaryote phyla, it is classified as either a class I or class II AP endonuclease based upon its incision pattern. Class I AP endonucleases are also known as AP lyases, while class II AP endonucleases are hydrolytic endonucleases; the latter form the majority of AP endonucleases. Class II AP endonucleases are further classified into two families (exonuclease III [Xth] and endonuclease IV [nfo]) based upon homology to *Escherichia coli* endonucleases. Apex1 in all metazoan eukaryotes belongs to the exonuclease III [Xth] family (Bapat, Fishel, and Kelley 2009; Mol, Hosfield, and Tainer 2000).

Human Apex1 has been mapped to a 2.6 kB gene on chromosome 14q11.2 with a structure comprising four introns and five exons. The start codon is in the second exon and most cross-species homology is in the fifth exon (Robson et al. 1992). The Apex1 protein is comprised of a highly conserved C-terminal nuclease domain, responsible for the DNA repair function, attached to an N-terminus that is not conserved across phyla (Tsutakawa et al. 2013; Wilson and Simeonov 2010). Structurally Apex1 is a globular α/β protein comprising a two-layered six-stranded β-sheet core flanked by α-helices, while the N-terminal domain, comprised of the first 45 amino acids, is disordered. Additionally, Apex1 is Mg$^{2+}$-dependent (Mol et al. 1995; Strauss and Holt 1998; Tsutakawa et al. 2013).

Apex1 is capable of repairing AP sites generated by both mono- and bifunctional DNA glycosylases. When an AP site is generated by a monofunctional DNA glycosylase, Apex1 hydrolyses the phosphodiester bond 5’ to the AP site leaving standard 3’-OH and 5’-dRP termini. It also possesses a 3’-repair diesterase ability that removes 3’ DNA blocking groups such as 3’ phosphates to produce the 3’-OH group required for replacement synthesis (Chen, Herman, and Demple 1991; Iyama and Wilson 2013; Mol, Hosfield, et al. 2000; Strauss et al. 1997). Apex1 possesses a 3’-5’ exonuclease function, although this is >100-fold lower than its AP endonuclease activity (Wilson et al. 1995; Wilson and Simeonov 2010). Furthermore, it appears to have the capacity to make an incision 5’ to certain oxidized bases, such as 5,6-dihydro-2’-deoxyuridine and 5-hydroxy-2’-deoxyuridine, in a process termed nucleotide incision repair (NIR) (Gros et al. 2004). Apex1 may also be capable of cleaving AP sites when they are present in RNA molecules (Berquist, McNeill, and Wilson 2008; Vascotto et al. 2009).

Apex1 must be able to recognize AP sites in order to repair them. Based on Apex1 mutational analyses residues Tyr$^{128}$, Tyr$^{171}$, and Tyr$^{269}$ are involved in the recognition and subsequent binding of Apex1
to an AP site. Tyr$^{171}$, which is positioned within the active site, is necessary to discriminate between DNA that contains an AP site and DNA that does not (Melo et al. 2007; Mundle et al. 2004, 2009).

Apex1 flips the AP site $\sim$180° into its active site for catalysis to occur. This is mediated by domain I residues interacting with the DNA strand opposite the AP site. Gly$^{127}$, Tyr$^{127}$, and Met$^{270}$ extend the minor groove, while Arg$^{177}$ opens the major groove (Mol, Izumi, et al. 2000; Mol, Hosfield, et al. 2000). Mutational analyses and crystal structures show that when a tetrahydrofuran substrate is used, Apex1 generates an incision in a single-step mechanism coordinated by residues Glu$^{96}$, Tyr$^{171}$, and Asp$^{283}$, with His$^{309}$ catalyzing the hydrolysis reaction. His$^{309}$ abstracts a proton from a water molecule to generate an active site nucleophile. His$^{309}$ is stabilized by a hydrogen bond with Asp$^{283}$, while Glu$^{96}$ binds a divalent metal cation (generally Mg$^{2+}$), thus exposing the negatively-charged phosphate group facilitating nucleophilic attack by the hydroxyl (Melo et al. 2007; Mundle et al. 2004, 2009; Wilson and Simeonov 2010). Recently, a 2.4Å co-crystal structure of Apex1-DNA with a tetrahydrofuran was published. The authors claimed that, based on the distances and angles between individual residues and DNA atoms at the AP site, many of the actions outlined above, particularly those of the tyrosine residues, are not consistent with their structure. However, the authors do not propose any new models or insight into the functionality of Apex1 (Tsutakawa et al. 2013).

Prior research suggests that the absence of Apex1 is incompatible with life (Fung and Demple 2005; Izumi et al. 2005). Recently, the Yu lab at Michigan State University claimed to have generated an APE1 null line in a mouse B cell line (CH12F3) (Masani, Han, and Yu 2013). The authors utilized a gene targeting method to replace exon 4 of the APE1 gene with a puromycin selection cassette producing an APE1$^{+/+}$ CH12F3 cell line. The authors reported that this cell line had no detectable APE1 protein, but was still viable with no apparent changes to cell proliferation, though it was hypersensitive to methyl methanesulphonate, a DNA-alkylating agent. However, it is quite possible the authors generated an APE1 K/D cell line rather than a knockout one. For instance, the authors claim that they targeted exon 4, “the last exon of the APE1 gene, containing $\sim$52% of the coding sequence” (Masani et al. 2013). However, APE1 has five exons; exon 5 forms the bulk of the coding sequence, though exon 1 is non-coding (Harrison et al. 1992). It is unclear from the work of Masani, et al. (Masani et al. 2013) whether they correctly targeted exon 5 by omitting the noncoding exon from their map, or incorrectly targeted exon 4. Furthermore, Masani et al. claim that they completely deleted the entire APE1 gene, whereas they merely deleted an exon from the APE1 gene. The authors also did not provide any epitope information for the anti-APE1 antibody they utilized to determine that they successfully knocked out APE1. It is very possible that they deleted the epitope site, but not the entire enzyme. Finally, they did not provide any kinetics data, whether generated from cell lysate or purified mutant APE1, to prove that their mutation was 100% inactivated. In fact, their
methyl methanesulfonate sensitivity assay would indicate that some ability to process AP sites remained intact. Nevertheless, even if Masani et al. had successfully generated a viable APE1 null cell line, the fact that it is the only one highlights the importance of APE1 to normal cellular growth and development.

1.2.1. Non-endonuclease function of Apex1

In 1992, Steven Xanthoudakis and Tim Curran identified a 37 kDa protein that could stimulate the DNA-binding activity of the heterodimerizing transcription factors Fos and Jun in vitro. The stimulation of transcription factors was based on reduction/oxidation (redox) of cysteine residues in the transcription factors. In light of this observation the authors named the protein redox factor 1 (Ref-1), but they later identified it as Apex1 (Xanthoudakis et al. 1996; Xanthoudakis, Miao, and Curran 1994; Xanthoudakis and Curran 1992). While the exact redox mechanism of Apex1 remains uncertain, several cysteine residues in human Apex1 are known to play a role. Furthermore, one, Cys\(^{65}\), has been identified as being central to Apex1 redox activity. While five of the seven cysteine residues are conserved across all vertebrates, two including Cys\(^{65}\) are only found in mammals. Apex1 in non-mammalian vertebrates, including zebrafish, does not have redox activity (Georgiadis et al. 2008). This implies that the redox function of Apex1 is an evolutionarily modern acquirement and is not essential for its function. The AP endonuclease and redox functions are located separately from each other and deletion of either one does not affect the function of the other in vitro. Since its initial discovery several other transcription factors have been reported to be regulated in vitro by Apex1 using the same redox mechanism, including p53, nuclear factor kappa-light-chain-enhancer of activated B cells (NFkB), Myb, ETS-related gene-1 (Erg-1), and cAMP response element-binding protein-1 (Creb1) (reviewed in (Tell et al. 2009)).

While much work has been done to elucidate the function and mechanism of redox activity in Apex1, it remains somewhat controversial. For example, all crystal structures to date for Apex1 show that Cys\(^{65}\) is buried deep within the protein; this observation is supported by proteolysis experiments (Strauss and Holt 1998). Access to Cys\(^{65}\) would require extensive remodelling, evidence for which is currently lacking, nor has any mechanism to make the various cysteines accessible been proposed (Georgiadis et al. 2008; Tsutakawa et al. 2013). Although Apex1 knockout is embryonic lethal at ~E8.6 in mice, when Cys\(^{64}\) (equivalent to human Cys\(^{65}\)) was mutated to alanine in mouse embryos the embryos develop normally and there is no difference in DNA binding of Fos, Jun or AP-1, indicating that either Cys\(^{64}\) was not involved in the redox activity of Apex1, or redox activity as a whole was not an essential function (Ordway, Eberhart, and Curran 2003).
1.3. Zebrafish as a model system

Zebrafish (Danio rerio) are small tropical shoaling freshwater fish of the Cyprinidae family native to northeastern India, Bangladesh, and Nepal (Spence et al. 2008). They have become an important model organism due to a number of beneficial characteristics including ease of maintenance and breeding, high fecundity (females can lay ~200 eggs a week), early sexual maturity, and ability to spawn year-round. Furthermore, zebrafish embryos display ex utero development, rapid embryonic development times, large size, optical transparency, and ability to tolerate manipulation through microinjection and cell transplantation methods (Dahm and Geisler 2006; Howe et al. 2013; Lawrence 2007; Nusslein-Volhard and Dahm 2002; Spence et al. 2008). The zebrafish genome has also been fully sequenced and has a high (~70%) similarity to that of humans (Howe et al. 2013).

Zebrafish have become a powerful model organism for studies on genetics (Dahm and Geisler 2006; Klee et al. 2012), environmental toxicology (Dai et al. 2013; Peterson and Macrae 2012), development (Page 1990; Veldman and Lin 2008), regeneration (Gemberling et al. 2013; Lin, Chiang, and Tsai 2016; Sehring, Jahn, and Weidinger 2016), and cancer (Amatruda et al. 2002; Feitsma and Cuppen 2008; Leacock et al. 2012; Lin et al. 2016). Additionally, zebrafish make an ideal model organism for studies on DNA damage and repair pathways as they have all the genes involved in the different DNA repair pathways of higher eukaryotes (Pei and Strauss 2013). As can be expected for a popular model organism, many powerful in vivo techniques have been developed for use with zebrafish. They make particularly potent genetic tools, as gene expression can easily be manipulated up or down. Overexpression can be achieved by microinjection of early stage embryos with in vitro transcribed capped mRNA of a specific gene, whereas knockdown (K/D) is easily achieved by microinjection with a morpholino oligonucleotide (MO), a non-degradable oligonucleotide, MOs are designed to target any gene of interest. Of course, appropriate controls are required to exclude off-target effects of MO K/D (Ekker and Larson 2001).

1.4. Base excision repair and embryonic development

BER is undoubtedly extremely important in everyday cell survival. Despite this, little is known of the system in the context of embryonic development. This is not overly surprising considering knockout or K/D of various BER genes is embryonic lethal.

The Strauss lab currently employs the zebrafish model to investigate embryonic BER. Most BER elements are maternally deposited into zebrafish eggs, and cell free extracts from unfertilized and early stage embryos are capable of performing the first three steps of BER. Since Polβ is absent during early
embryogenesis, another aphidicolin-sensitive DNA polymerase executes this gap-filling role. Furthermore, a \( \text{Mg}^{2+} \)-dependent endonuclease, possibly Apex2, appears to compensate for loss of Apex1 in early stage embryos (Fortier et al. 2009).

There is no embryonic lethal phenotype associated with most of the DNA glycosylases suggesting that there may be some redundancy amongst glycosylases or alternative compensatory mechanisms (Engelward et al. 1997; Hang et al. 1997; Karahalil et al. 2003; Nilsen et al. 2000; Parsons and Elder 2003). However, thymine DNA glycosylase (Tdg) is an exception; Tdg is a monofunctional glycosylase responsible for repair of T/G mispairs. Tdg deficient mouse embryos die by embryonic day 11.5 (E11.5) (Cortázar et al. 2011). The role of Tdg in zebrafish embryology has not been fully studied, though it appears that Tdg is not present in early embryos as evidenced by lack of transcript during early development (Wu et al. 2014), and lack of detectable protein until one day post fertilization (1dpf) (Strauss lab, unpublished data).

Apex1 deficient mice die between E4 and E6.5 (Xanthoudakis et al. 1996). Full K/D of Apex1 by MO is lethal to zebrafish embryos at the mid-blastula transition (MBT), which corresponds roughly to the same stage. The MBT corresponds to the developmental stage where the embryo switches from using maternally provisioned transcript and protein to gastrulation and activation of the zygotic genome (Kimmel et al. 1995). Partial Apex1 K/D in zebrafish results in heart and brain abnormalities by 1dpf and death by 7dpf when a fully-functional heart is required. Furthermore, K/D of Apex1 results in a concurrent decrease in the critical transcription factor Creb1 (Snow 2012). Creb1 can be rescued by co-injection with human apex1 message; the human mRNA is used, as the zebrafish Apex1 MO does not bind to the human 5’ end. Nevertheless, rescue only occurs if co-injected apex1 is endonuclease competent (Pei et al. 2011; Wang et al. 2006).

1.4.1. Apex1 studies and zebrafish embryogenesis

Since the Apex1 knockout mutation in mice is embryonic lethal Apex1 may be critical for development. Studies in the Strauss lab identified two gene copies of apex1 (Wang et al. 2006), consistent with a genome duplication event that occurred approximately 350 million years ago in teleost fishes (Postlethwait et al. 1998). One of the apex1 copies has a similar structure to its human ortholog with four introns and five exons, the first of which is not translated. The other copy lacks introns. Furthermore, two apex1 RNA transcripts of different sizes are present in adult zebrafish: one is 2.1 kb in length, while the other is 1.3 kb and lacks the full 3’UTR. Nevertheless, only one 35 kDa protein species has been identified and this displays a 78% homology (64% identity) with the human protein. When a MO directed against the Apex1 translation start site (TS-MO) is used, embryos die at the MBT, yet, embryos subjected to splice site
directed MOs (SS-MO) pass through the MBT but die at ~7 hpf due to heart and brain abnormalities. This implies that both mRNA copies are utilized, and both are important for development (Wang et al. 2006).

Full Apex1 K/D by the TS-MO can be rescued by co-injection of mature human apex1 mRNA. However, no rescue occurred when the mRNA contained a mutation of Tyr^{171} to phenylalanine or alanine. These mutations result in almost complete loss of endonuclease activity to Apex1, without changing its tertiary structure, indicating that the endonuclease function of Apex1 is important for viability and development (Wang et al. 2006).

Partial K/D of Apex1 in zebrafish embryos results in the failure of polβ transcript and protein to appear at the appropriate time (Fortier et al. 2009). Moreover, partial K/D also results in loss of the critical transcription factor, creb1 (Pei et al. 2011), a result that is not overly surprising considering creb1 is known to regulate polβ expression (Narayan, He, and Wilson 1996). These findings also occurred in primary cultures of murine B cells (Pei et al. 2011). Co-injection of creb1 mRNA restored Polβ protein in Apex1 K/D embryos, while co-injection of human apex1 mRNA restored both Creb1 and Polβ protein. Again, restoration of Creb1 and Polβ was only achieved when endonuclease competent Apex1 was used. Apex1 K/D also resulted in the loss of the Creb1 binding partners, Creb1 binding protein (Cbp), cAMP response element modulator (Crem), and Creb1-regulated transcription coactivators 1 and 2 (Crtc1 & 2). Protein levels of the Creb1 binding partners were not tested during the early stages of embryogenesis. These studies show that the multifunctional protein Apex1 regulates the protein levels of Creb1 and its binding partners, and controls the levels of the downstream BER partner polβ. Since Creb1 acts as an anchor protein enabling transcription of ~25% of the eukaryotic genome (Impey et al. 2004; Zhang et al. 2005) and Apex1 regulates Creb1 protein, Apex1 is linked to genomic transcription, but not necessarily via its reported redox capability. Thus, Apex1 has an important, but relatively unexplored, role in development and the maintenance of cellular physiology.

1.5. DNA methylation

DNA is conventionally viewed as being comprised of the four nucleobases C, G, T, and A. However, cytosine can have a methyl group enzymatically attached to the 5-carbon of the cytosine ring to generate 5-methylcytosine (5mC). The methyl group projects into the major groove of DNA and acts to control gene transcription by modulating chromatin structure (Nabel, Manning, and Kohli 2012). The addition of 5mC occurs almost exclusively in the context of CpG (C followed by G) dinucleotides, which often occur together forming genomic regions with a high frequency of CpG dinucleotides referred to as CpG islands (Jones 2012; Smith et al. 2014; Tricarico and Bellacosa 2016). Since 5mC is a heritable DNA
modification it is referred to as being epigenetic; a heritable change in gene expression that does not involve a fundamental change to the underlying DNA structure. DNA demethylation is the reverse of this process, where the methyl group is enzymatically removed from the cytosine ring (Kohli and Zhang 2013; Niehrs 2009).

1.5.1. DNA methylation

Methyl addition to the cytosine ring is primarily conducted by the DNA methyltransferase (Dnmt) family of enzymes (Figure 2). These are subdivided into maintenance (Dnmt1) and de novo (Dnmt3a and Dnmt3b) methyltransferases. Dnmt2 is also present, but recombinant studies and inactivation of Dnmt2 in embryonic stem cells indicate that it possesses no methyltransferase activity (Shimoda et al. 2005), though it possibly functions in methylation of RNA (Goll et al. 2006). The de novo Dnmts, Dnmt3a and Dnmt3b, are responsible for establishing the initial CpG methylation pattern while Dnmt1 maintains this pattern during cell division and repair. The basic mechanism for methyl addition by Dnmts involves the transfer of a methyl group from S-adenyl methionine (SAM) to cytosine. Briefly, Dnmts will bind to the DNA target and “flip” the base out of the double helix. They then covalently modify the C6 of the cytosine ring via a nucleophilic attack by a conserved cysteine residue and transfer the methyl group derived from SAM to C5. The covalent C6 intermediate will then breakdown releasing the Dnmt enzyme (Goll and Bestor 2005).

Zebrafish possess multiple dnmt members, dnmt1 - 8. Both dnmt1 and 2 are highly conserved with those of mammals, while dnmts 3 - dnmt8 are all related to the dnmt3 family (Goll, Mary Grace and Halpern 2011; Wu et al. 2011).

1.5.2. DNA demethylation

Removal of the methyl group from C5 is a much more complicated, and somewhat controversial, process. Methyl group erasure can be classified as passive or active (Niehrs 2009; Tricarico and Bellacosa 2016). Passive demethylation occurs in dividing cells and is simply the failure of newly synthesized DNA strands to adopt the methylation status of the parent strand. With time, the overall methylation levels are diluted by continued cell division (Tricarico and Bellacosa 2016). In contrast, active DNA demethylation involves enzymatic modification of 5mC to revert it back to normal C and occurs in both dividing and non-dividing cells. Though several active DNA demethylation pathways have been proposed, one, the direct substitution of the methyl group with hydrogen, is incompatible with chemical reaction theory (Smith 2000). Of the other pathways, one involves deamination of the amine to a carbonyl group in 5mC by activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme complex (AID/APOBEC),
converting it to thymine (Figure 2). This produces a G/T mismatch which activates BER via TDG, resulting in replacement of T with unmodified cytosine (Bochtler, Kolano, and Xu 2017; Niehrs 2009).

Another pathway involves iterative modification of 5mC by the ten-eleven translocation (TET) enzymes, TET1 – 3 (Figure 2). TET enzymes belong to the Fe(II)/α-ketoglutarate dependent oxygenase family and contain a dioxygenase domain (Wu and Zhang 2017). TET1 adds a hydroxyl group to the methyl group of 5mC, generating 5-hydroxymethyl cytosine (5hmC). Conversion of 5hmC to unmodified cytosine may then occur via two sub-pathways. In the first, AID/APOBEC deaminates 5hmC producing 5-hydroxymethyluracil (5hmU) which is then recognized by TDG which activates the BER pathway. However, in vivo evidence for this sub-pathway is currently lacking. The second sub-pathway relies on iterative oxidation of 5hmC by TET2 and TET3 to produce 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) intermediates. These intermediates are then recognized by TDG, activating the BER pathway resulting in replacement with unmodified cytosine (Kohli and Zhang 2013; Niehrs 2009; Wu and Zhang 2017). The key point is that, regardless of which pathway is utilized, the final steps involve activation of the BER pathway resulting in the generation of an AP site. This requires the presence of Apex1 to cleave the AP site before BER can regenerate the unmodified cytosine.

1.5.3. DNA methylation and embryogenesis

As previously mentioned, DNA methylation occurs primarily at CpG dinucleotides, typically in the context of CpG islands (CGI). CpG islands consist of genomic regions ~250 – 2000bp in length, possess a high density of CpG, typically greater than 55%, and have a ratio of observed-to-expected number of CpG dinucleotides above 0.6 (Siegfried and Simon 2010). They occur throughout the genome in both intra- and inter-genic regions. However, CGIs predominately occur in gene promoters, ~70% of annotated gene promoters contain a CGI, and between 4% and 25% are methylated in normal tissues (Deaton and Bird 2011; Siegfried and Simon 2010). Methylation of promoter CGIs leads to stable silencing of gene expression, mainly by impairing transcription factor binding, and altering chromatin structure to a closed, inhibitory state. DNA demethylation allows chromatin to adopt an open confirmation more permissive for active transcription; thus, the methylation status of a CGI can influence the transcriptional status of a gene and may be an important component of transcriptional regulation (Jones 2012). During embryogenesis, DNA methylation status is highly dynamic (Fang et al. 2013).

At fertilization, embryos are transcriptionally inactive and require reprogramming before embryonic genome activation occurs. Both sperm and oocyte genomes are differentially methylated due to germline genomic imprinting. In mammals, a rapid global loss of both paternal and maternal DNA methylation occurs post-fertilization, followed by remethylation. However, remethylation of the now
diploid genome recapitulates the maternal methylation pattern, the paternal patterning is essentially erased (Smith et al. 2014). This process occurs in the pre-implantation stages, yet not all methylated CGIs are demethylated during the process (Andersen et al. 2012).

Zebrafish also undergoes global DNA demethylation and remethylation post-fertilization, though this process differs dramatically from mammals. Post-fertilization, zebrafish embryos undergo ten rapid, transcription free, cell cycles until they enter the MBT at 3.3 hpf, where activation of zygotic genome occurs (Kane and Kimmel 1993). During the initial ten cell cycles, the paternal and maternal DNA methylation contributions are erased, but not until after the 16-cell stage. However, unlike mammals, the diploid zebrafish genome adopts the paternal DNA methylation pattern globally (Jiang et al. 2013). When DNA methylation is viewed globally, methylation levels tend to increase as embryos develop through the MBT (Andersen et al. 2012; Potok et al. 2013). Yet, despite remethylation typically occurring by the MBT, changes to methylation status at the individual gene level is dynamic, and time-specific. For example, the promotors of many genes crucial to development, including transcriptional regulation, DNA modification and RNA processing, become hypomethylated around the MBT (Jiang et al. 2013), consistent with zygotic genome activation (ZGA) occurring during the MBT (Kimmel et al. 1995). Furthermore, many genes involved with gastrulation, chromatin remodelling, and cell morphogenesis, have their promotors demethylated between the MBT and the germ line gastrula phase at 5.66 hpf (Jiang et al. 2013).

As previously mentioned, Apex1 K/D results in a concurrent loss of the crucial transcription factor Creb1 with Apex1 in zebrafish embryos. Creb1 facilitates the activation of cAMP-responsive genes by binding to the conserved cAMP-responsive element (CRE), TGACGTCA (Zhang et al. 2005), and half CRE, TGACG/CGTCA (Craig et al. 2001). Regardless whether the target CRE sequence is a full or half site, a single CpG dinucleotide (CpG islet) is present. Partial and complete methylation of the CRE sequence CpG have been shown to block Creb1 binding both in vitro (Iguchi-Ariga and Schaffner 1989; Moore, Toomire, and Strauss 2013) and in vivo (Zhang et al. 2005). It is possible that regulation of cAMP-responsive genes is due to methylation/demethylation of the Creb1 consensus sequence itself, a form of regulation apparently common to many transcription factors (Yin et al. 2017). Moreover, the two creb1 genes in zebrafish, creb1a and creb1b, both have CpG islands in their promotors. Both also have half CRE sequences within those CpG island, creb1a has four and creb1b three, indiscriminate methylation/demethylation of the CpG island of the creb1 genes would also affect the CRE sequences which would affect transcription of creb1 itself.

The purpose of this dissertation was to investigate the role that DNA damage and BER, particularly that of Apex1, may play during zebrafish embryogenesis. Loss of endonuclease competent Apex1 in
zebrafish results in a consistent, and concurrent, loss of the transcription factor (TF) Creb1, potentially indicating that non-random DNA damage is a normal occurrence during embryonic development. In light of this I hypothesize that: **During zebrafish embryogenesis, targeted DNA damage occurs in the genome requiring BER to reverse.**

The main **aims** of this dissertation were to:

1) **Quantify DNA damage during embryogenesis.** The effects of Apex1 K/D are due to loss of repair capability; however, we do not know when, or if, DNA damage occurs during embryogenesis. 1b) **If DNA damage is present, identify the mechanism by which it occurs.** Since a consistent loss of Creb1 occurs, DNA damage may be non-random. Is there a particular mechanism by which targeted DNA damage to the genome can occur?

2) **Investigate binding kinetics of Creb1 to the CRE site when it contains DNA damage.** If loss of Creb1 results from loss of functional Apex1, could DNA damage in the binding sequence for Creb1 have an impact on Creb1 binding? Could this be the reason for the observed loss of Creb1?

3) **Investigate competition between BER DNA glycosylases and Creb1 for damage containing CRE sites.** If DNA lesions present in the CRE site can modulate Creb1 binding, would BER repair the lesions before CREb1 could bind or could binding of Creb1 block repair from happening?
**Figure 1. Overview of the basic steps involved in the BER pathway.** Numerals on the left indicate the steps, while Roman numerals on right indicate the different substrates. Step 1 in BER is recognition and excision of a DNA lesion by the appropriate DNA glycosylase leaving an abasic (AP) site. In step 2, the resulting AP site is cleaved by Apex1 to generate the 3’-OH and 5’-dRP required as substrate by the next enzyme in the pathway. Steps 3a, 4a, and 5a illustrate short patch repair, which replaces single nucleotides. Steps 3b, 4b, and 5b illustrates long patch repair, which replaces two to six nucleotides. Image reproduced from (Fortier et al. 2009).
Figure 2. Pathways of active DNA demethylation. Colors denote the enzyme/s responsible for modifications to substrates, except purple, which denotes the entire BER pathway from Apex1 onwards. Cytosine (C) is methylated by DNA methyltransferase 1 (Dnmt1) producing 5-methylcytosine (5mC), 5mC can be deaminated to Thymine (T) by Aid/Apobec (Aid/Ab1-3) family members producing a T/G mispair. BER glycosylases thymine DNA glycosylase or single-strand-selective monofunctional uracil DNA glycosylase (Tdg/Smug1) excise the T leaving an abasic site (AP). The AP site is cleaved by Apex1, and BER inserts a C restoring the C/G. The 5mC can also be hydroxylated by ten-eleven-translocation (Tet1-3) family members, generating 5-hydroxymethylacytosine (5hmC). This can then be either further hydroxylated by Tet members to 5-carboxylycitosine (5caC), or deaminated by Aid/Ab1-3 members to 5-hydroxymethyluracil (5hmU). Both products can then be excised by Tdg/Smug1 to produce an AP site, and BER then restores C.
Chapter 2. DNA damage and cytosine methylation status are temporally modulated during zebrafish embryogenesis.

A major goal of the Strauss lab has been the investigation of the molecular changes that accompany the loss of Apex1. While many BER enzymes, including Apex1, are embryonic lethal, null cell lines exist for all of them except Apex1. Obviously, something important is occurring during zebrafish embryogenesis that requires fully functional BER to resolve, but this requirement is diminished in differentiated, adult cells. Since complete knockout of Apex1 is embryonic lethal at the MBT, the logical inference is that whatever is occurring that requires resolution by BER is transpiring then. By partially knocking down Apex1, the Strauss lab showed that the first detectable loss of Apex1 protein was detectable at the MBT, and was accompanied by a loss of Creb1 (Snow 2012). The concurrent loss of Creb1 indicates that this is a primary effect of Apex1 loss. Creb1 is a pleiotrophic transcription factor that functions to activate or repress ~25% of the genome (Impey et al. 2004; Zhang et al. 2005). Once Creb1 protein is lost, many other genes become dysregulated. Complete Apex1 knockout embryos can transition through the MBT when co-injected with human *Apex1* mRNA, though they develop the same phenotype as Apex1 knockdown embryos. Rescue of knockdown embryos with human *Apex1* message rescued Creb1 protein. Importantly, rescue with enzymatically mutant (Y171F) or (Y171F-P173L-N174K) human *Apex1* message did not rescue knockout embryos. This showed that the effects detected with the loss of Apex1 protein was related to its ability to cleave an AP site and was a function of loss of BER capacity. Therefore, the logical interpretation is that whatever process is occurring during early zebrafish embryogenesis, it is generating DNA damage.

In the following chapter I have examined, by means of the Comet assay, the timing of DNA damage in zebrafish embryos as they develop through the MBT. To my knowledge, this represents the first time that DNA damage has been quantified over the course of early embryogenesis in any organism. I have also examined the generation of AP sites and changes to global 5mC levels over the same developmental period.

Partial Apex1 loss results in many Creb1 partner proteins including Creb-binding protein (Cbp), cAMP response element modulator (Crem), and transducer of regulated Creb-binding proteins 1 and 3 (Torc 1 &

---

1 One Apex1 null cell line does exist, but issues with the study suggest that it may actually be a knockdown line, rather than a knockout; see chapter 1.
3) to be down-regulated, as is the BER enzyme DNA polymerase β (Polβ) (Pei et al. 2011). All of these genes contain at least one Creb1 binding site in their promoters, so are likely to be impacted by the initial loss of Creb1 protein. In the following chapter I have also attempted to provide mechanistic evidence of active DNA demethylation within the CpG islands of creb1 gene promoters.

2.1. Introduction

The process of vertebrate embryogenesis features the generation of a fully functional organism from a single cell. This is a highly complex, tightly regulated, and coordinated process (Capron et al. 2009). During embryogenesis, the essential architectural features of the organism are established. This involves three main features, proliferation of cells, differentiation of cells, and establishment of polarity (Takaoka and Hamada 2012). At the molecular level, both genetic and epigenetic information regulate gene expression resulting in growth, differentiation, and development. Changes to DNA methylation, incorporation of histone variants, post-translation histone modifications, and remodelling of chromatin all help regulate gene expression during embryogenesis (Burton and Torres-Padilla 2014). However, normal cellular process during embryogenesis can generate DNA damage, which may have profound implications to the developing embryo if unrepaired.

The occurrence of DNA damage during embryonic development has received little attention in the past. Most studies have focused upon DNA damage in sperm and how it impacts fertilization rates (Simon et al. 2016), in vitro development of fertilized embryos in IVF (Sturmey et al. 2009), or cryopreservation of gametes and/or embryos (Shi et al. 2017). Yet, the potential for DNA damage, particularly that repaired by BER, to occur during embryonic development must be considered. DNA damage can potentially alter developmental programs as various DNA lesions disrupt important developmental transcription factors from binding their cognate sequences (Hailer-Morrison et al. 2003; Moore et al. 2013). Furthermore, developing embryo cells divide rapidly compared to adult cells, so that failure to repair DNA damage can lead to mutations and cell death (Wells et al. 2010).

An indication that BER repaired DNA damage may be important during embryogenesis is that many components of the BER pathway are embryonic lethal. For example, in mouse, fen1−/− embryos die at embryonic day (E) 3.5 (Larsen et al. 2003), xrcel−/− between E6.5 – E7.5 (Tebbs et al. 1999), apex1−/− between E4 – E6.5 (Xanthoudakis et al. 1996), and lig3−/− halt development at E8.5 and die within 24 hours (Puebla-Osorio et al. 2006). Despite this lethality, knockout cell models for these same components are viable, though they are typically more sensitive to DNA damaging agents. This indicates that DNA damage occurring during the earliest stages of development is more detrimental to the organism than in differentiated, adult cells.
In zebrafish, abrogation of Apex1 results in embryos dying at the MBT (~3.3 hpf), roughly corresponding to E4 – E6.5 in mouse. Partial loss of Apex1 allows embryos to continue development through the MBT, but they display heart and brain abnormalities and die at 7 days post fertilization (Pei et al. 2011; Wang et al. 2006). Furthermore, partial Apex1 loss results in a concurrent and dramatic loss of the developmentally important transcription factor Creb1 (Snow 2012), potentially explaining the manifestation of heart and brain abnormalities. However, not all BER components are embryonic lethal. When most DNA glycosylases are knocked out, embryos are still viable and develop as normal (Engelward et al. 1997; Hang et al. 1997; Jacobs and Schär 2012; Parsons and Elder 2003). An exception is tdg, a monofunctional DNA glycosylase responsible for excising mispaired thymine. Loss of Tdg is embryonic lethal at E11.5 and has been implicated in DNA demethylation (Cortázar et al. 2011).

In vertebrates, the addition of a methyl group to cytosine results in 5-methylcytosine (5mC), a common epigenetic mark (Nabel et al. 2012). 5mC occurs in CpG dinucleotides, and can vary by tissue type, as well as during embryonic development. Removal of the methyl group can be passive, where 5mC is diluted out by replication, or active (Niehrs 2009; Tricarico and Bellacosa 2016). Active DNA demethylation is an iterative process requiring the coordination of a number of enzymes, though the actual pathway is controversial. Three different pathways have been proposed in the past, though one, straight substitution of the methyl group with hydrogen, is inconsistent with chemical reaction theory (Bochtler et al. 2017; Smith 2000). The other pathways all result in the generation of DNA damage requiring BER to resolve (Bochtler et al. 2017).

Activation of BER during active DNA demethylation is thought to occur when tdg recognizes and removes a specific intermediate generated during the stepwise, enzymatic removal of the methyl group from cytosine, creating an AP site (Kohli and Zhang 2013). Subsequent cleavage of the AP site by Apex1 allows Polβ to insert a cytosine into the gap (Niehrs 2009). The role of tdg in active DNA demethylation during embryogenesis could explain why it is embryonic lethal whereas most DNA glycosylases are not.

Since DNA demethylation occurs to the promoters of many zebrafish genes from the MBT onwards (Jiang et al. 2013), DNA damage may increase concomitantly. Furthermore, partial loss of Apex1 could elevate DNA damage due to an inability to process AP sites, and may explain the embryonic lethality of Apex1 knockdown and its timing. In this study, I examine the timing of DNA damage and changes to genomic DNA methylation during zebrafish embryogenesis. I show that DNA damage during embryogenesis is dynamic and correlates well with changes to DNA demethylation. This suggests that active epigenetic reprogramming during embryogenesis results in high levels of DNA damage, highlighting the importance of functional BER during early embryonic development.
2.2. Experimental procedures

**Zebrafish Culture:** Wild-type zebrafish (*Danio rerio*), purchased from Aquatica BioTech (Sun City Center, FL), were maintained and bred using standard protocols in accordance with approved Northeastern University IACUC policies as previously described (Wang et al. 2006).

**Knockdown of Apex1 by morpholino oligonucleotide (MO):** Embryos were microinjected directly into the yolk prior to the 4-cell stage with 0.23 mM MO (GeneTools, Philomath, OR) targeted against the translation start site of *apex1* mRNA as previously described (Fortier et al. 2009; Pei et al. 2011; Wang et al. 2006). The MO was suspended in solution with 1x Danieau buffer and 0.05% phenol red dye to act as an injection indicator. Control microinjections comprised 0.05% phenol red in 1x Danieau buffer. Injected embryos were maintained at 29°C until reaching desired developmental stages.

**Abasic site quantification:** Genomic DNA from control and Apex1 knockdown (K/D) embryos (4.3, 6, and 8 hpf) was prepared using the DNeasy Blood and Tissue Kit (Qiagen, Venlo, Netherlands), with 0.1 mM deferoxamine added to prevent oxidation of DNA (Taghizadeh et al. 2008). Each treatment had ~150 embryos microinjected per experiment. However, since embryos were so early in the developmental process, the extracted DNA concentrations were lower than the minimum amount required for assays. To overcome this, extracted DNA was subjected to ethanol precipitation and concentration in an anaerobic chamber to minimize oxidation of DNA. Due to these issues, we were unable to use 2.5 and 3.3 hpf stage embryos, as pooling of preparations on multiple days would have been required. DNA concentration was quantified with a Nanodrop 1000 (Thermo Scientific, Waltham, MA). AP site quantification was then performed by labelling DNA with an aldehyde reactive probe (ARP), which only binds AP sites, using the DNA Damage Quantification Kit-AP Site Counting (Dojindo Molecular Technologies Inc., Rockville, MD) following the manufacturer's instructions. The OD intensity of the colorimetric reaction was measured using a BioTek Synergy HT plate reader (BioTek, Winooski, VT).

**Global DNA methylation quantification:** Genomic DNA from control and Apex1 K/D embryos (2.5, 3.3, 4.3, 6, and 8 hpf) was prepared using the DNeasy Blood and Tissue Kit (Qiagen, Venlo, Netherlands). For each treatment in an experiment ~150 embryos were microinjected. To account for the high yolk protein content, homogenized embryos from the 2.5 and 3.3 hpf stages were incubated with proteinase K overnight at 56°C and DNA from two separate experiments were pooled. All DNA extractions were subjected to ethanol precipitation and concentration. DNA was quantified with a Nanodrop 1000 (ThermoFisher Scientific, Waltham, MA). Global DNA methylation was measured with the MethylFlash™ Global DNA Methylation Kit (Epigentek group, Farmingdale, NY) following the manufacturer's instructions. The OD intensity of the colorimetric reaction was measured using a BioTek Synergy HT plate reader (BioTek,
Global methylation was calculated from the standard curve and normalized to 100 ng of DNA. To calculate the percentage of global cytosine that was methylated, the calculated value was divided by 19.3%, the cytosine content of zebrafish DNA (Zhou, Bizzaro, and Marx 2004).

**Preparation of comet assay slides:** To prepare comet slides consisting of separated single cells from early stage zebrafish embryos, a mechanical separation technique developed by the Filipič laboratory (Eleršek, Plazar, and Filipič 2013) was employed. Two slightly different methods were utilized based upon embryonic developmental time. For embryos at 4.3, 6, and 8 hpf stages, microscope slides were pre-coated with 1% normal melt point agarose and dried overnight. Microinjected embryos were dechorionated with 2 mg/ml of pronase (Sigma-Aldrich, Darmstadt, Germany) in E2 buffer (15 mM NaCl, 0.5 mM KCl, 2.7 mM CaCl₂, 1 mM MgSO₄, 0.7 mM NaHCO₃, 0.15 mM KH₂PO₄, 0.05 mM Na₂HPO₄) heated to 37°C. Once chorions burst open, embryos were rinsed 3X with E2 buffer and resuspended in E2 buffer on Petri dishes. On each slide, two embryos were placed in 60 µl of 1.5% low melt point (LMP) agarose in phosphate buffered saline (PBS) heated to 37°C, after which a cover slip was applied, and tweezers were used to gently press the cover slip down. The downward pressure squashed the embryos, releasing cells and spreading them throughout the agarose. The slides were incubated at 4°C for five min to solidify the agarose. The cover slip was then removed and another layer of LMP agarose (~80 µl) was applied to the slides which were incubated at 4°C for five min.

Embryos from the 2.5 and 3.3 hpf stages were mainly yolk with a small cell cap on top. The above method resulted in too much yolk protein being present which impacted subsequent imaging and analysis. A slight variation to methodology was therefore developed for early stage embryos. Briefly, after dechorionation and rinsing, embryos were resuspended in deyolking buffer without CaCl₂ (55 mM NaCl, 1.8 mM KCl, 1.25 mM NaHCO₃) and the cell caps physically removed from the yolk using number 5 tweezers. Ten cell caps were then carefully removed with a 200 µl pipette and resuspended in 80 µl of 37°C LMP agarose by pipetting up and down. Then 60 µl of the LMP agarose and cell suspension was applied to each well of a CometSlide (Trevigen, Gaithersburg, MD) and incubated at 4°C.

**Comet assay:** The alkaline comet assay method, which measures combined single and doubled strand breaks in DNA, as well as alkali-labile sites including AP sites, was performed as described by Singh et al. (1988). Slides prepared as above were incubated in lysis buffer (Trevigen, Gaithersburg, MD) at 4°C for 60 min and rinsed 3 times in PBS. The slides were incubated in alkaline electrophoresis buffer (300 mM NaOH pH13, 1 mM EDTA) at 4°C for 20 min to allow DNA unwinding. The slides were electrophoresed at 1V/cm and 300 mA for 22 min at 4°C, followed by neutralization in 0.4 M Tris buffer (pH 7.5) for 5 min. All these steps were performed in the dark. Slides were incubated in 5% ethanol for 5 min, and left to dry overnight.
They were then incubated in 1X Sybr Gold (Invitrogen Molecular Probes, ThermoFisher, Waltham, MA) in PBS for 20 min, and imaged using fluorescence microscopy (Nikon eclipse 80i, Tokyo, Japan). The % Tail DNA and Olive Tail Moment were measured with image analysis software [CASPlab, (Końca et al. 2003)]. At least two embryos, for the 4.3, 6, and 8 hpf stages, and ten embryonic cell caps for the 2.5 and 3.3 hpf stages, per treatment in three independent experiments were examined. Each independent experiment had two technical replicates, and a minimum of 80 nuclei per technical replicate was analyzed.

**PCR:** Genomic DNA previously labelled with ARP for the AP site quantification (4.3, 6, and 8 hpf) was used for PCR analysis. Primers for PCR were designed for the promoter region of six genes, with or without CpG islands, using Primer3 software (Untergasser et al. 2012) (Table 1). PCR was performed using OneTaq® Quick-Load 2x Master Mix (New England Biolabs, Ipswich, MA) with 1 ng of template DNA and 200 nM of each primer in a total reaction volume of 25 µl. The instrument used was an Eppendorf MasterCycler® Gradient 5331 (Eppendorf, Hamburg, Germany) with a profile set of an initial denaturation at 95°C for 30 seconds, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at appropriate temperature (see Table 1) for 30 seconds, and extension at 68°C for one min, followed by final extension at 68°C for five mins. PCR products were resolved on 1.2% agarose gels in TAE buffer, imaged with a BioRad ChemiDoc™ MP Imaging System (BioRad, Hercules, CA) and quantified using ImageJ software (Schneider, Rasband, and Eliceiri 2012).

**Quantitative PCR (qPCR) for DNA damage:** DNA damage to the promoters of Creb1a and Creb1b was quantified by qPCR using the methods outlined by Rothfuss and colleagues (Rothfuss, Gasser, and Patenge 2009). The primers designed above for Creb1a and Creb1b (Table 1) were used for the qPCR. An internal normalization control was also designed using Primer3 software. This consisted of the forward primer listed above and a new reverse primer giving an amplicon of 105 bp (Creb1a) and 121 bp (Creb1b) long (Table 2). qPCR was performed using the DyNAmo Flash SYBR Green qPCR Kit (ThermoFisher Scientific, Waltham, MA) with 1 ng of template DNA and 500 nM of each primer in a total reaction volume of 25 µl. DNA samples came from control or Apex K/D zebrafish embryos (2.5, 3.3, 4.3, 6, and 8 hpf). The instrument used was a StepOnePlus™ Real-Time PCR System (ThermoFisher Scientific, Waltham, MA) with a preincubation phase of 95°C for 10 min followed by 42 cycles of denaturation at 95°C for 10 seconds, annealing at appropriate temperature (see Table 2) for 10 seconds, and extension at 72°C for one min (in large amplicons) or 10 seconds (in short reference amplicons).

DNA damage was quantified as number of lesions per 10000 bp of DNA using the equation:

$$\text{Lesion rate [Lesion per 10 kb DNA]} = (1 - 2^{-\Delta A_{\text{short}}})$$
Where $\Delta_{\text{long}}$ is the threshold cycle (Ct) of the large amplicon and $\Delta_{\text{short}}$ is the smaller internal normalization amplicon’s Ct value. Each sample had four technical replicates, and each was performed in triplicate.

**Statistical analyses:** Data in Figures 1 and 2 were analyzed by one-way ANOVA using GraphPad Prism 7 (GraphPad Software, Inc. La Jolla, CA). Before analysis was conducted, a Shapiro-Wilk’s test for normality and a Brown-Forsythe test for equal variance were performed. Sidak’s *post-hoc* multiple comparisons test was conducted to identify significantly different means. Data in Figure 3 was analyzed by Kruskal-Wallis test using GraphPad Prism 7. Dunn’s *post-hoc* multiple comparisons test was conducted to identify significantly different means. Significance for all statistical tests was set *a priori* at 0.05.

### 2.3. Results

**DNA damage is elevated at the MBT during normal zebrafish development:** To measure the amount of DNA damage occurring during embryogenesis, we conducted comet assays under alkaline conditions to estimate the total amount of damaged DNA in a sample. Prior to the MBT at 2.5 hpf, when the embryo is transcriptionally silent, total DNA damage, quantified as percent of DNA present in the comet tail (Figure 1A), was 15.7% ±3.0% (S.E., 3) in sham injected embryos (Figure 1B). At 3.3 hpf, when the embryo entering the MBT and activating its own genome, total DNA damage increased by 40% to 22.0% ±1.7% (S.E., 3) and remained high at 21.7% ±1.3% (S.E., 3) as the embryo exited the MBT at 4.3 hpf. By 6 hpf total DNA damage reduced by 40% to 12.8% ±0.6% (S.E., 3), and remained stable at 13.0% ±1.5% (S.E., 3) through 8 hpf.

When total DNA damage was quantified by Olive Tail Moment (OTM), a unit-less measure which takes the distance the tail migrated as well as percent of DNA in the comet tail into consideration, the same trend was observable (Figure 1C). Pre-MBT (2.5 hpf), OTM was 7.2 ±1.6 (S.E., 3) in sham injected embryos. This increased by 46% to 10.6 ±1.3 (S.E., 3) during the MBT (3.3 hpf) and remained elevated at 4.3 hpf at 9.8 ±0.7 (S.E., 3). Post-MBT, OMT decreased by 52% compared to the 4.3 hpf value, to 4.6 ±0.4 (S.E., 3), and remained stable at 5.1 ±0.6 (S.E., 3) through 8 hpf. These results indicated that DNA damage was part of a normal physiological process as developing zebrafish embryos enter the MBT.

**DNA damage is more elevated in developing zebrafish embryos with reduced repair capacity:** When the BER repair enzyme Apex1 was knocked down by MO microinjection in zebrafish embryos, reducing BERs
ability to go to completion, a similar trend was observed, but damage was more elevated (Figure 1B). Total DNA damage as percent comet tail DNA pre-MBT was 13.0% ±5.3% (S.E., 3) in Apex1 K/D embryos, which was 17% lower than in control embryos at the same time point, though this difference was not statistically significant. As the embryo entered the MBT (3.3 hpf), loss of Apex1 resulted in an increase in total DNA damage of 118% to 28.3% ±1.3% (S.E., 3); this is a statistically significant 28% increase over control embryos at the same time point. By 4.3 hpf, total DNA damage remained stable at 29.2% ±1.9% (S.E., 3) but began to decrease at 6 hpf to 21.9% ±1.2% (S.E., 3). However, the reduction in total DNA damage in Apex1 K/D embryos lagged that of the control embryos at the same developmental stage. While total DNA damage in control embryos at 6 hpf essentially returned to pre-MBT values, damage in Apex1 K/D embryos remained elevated, 69% higher than its pre-MBT value. By 8 hpf, total DNA damage in Apex1 K/D decreased to 15.4% ±1.1% (S.E., 3), still 19% higher than it was at the pre-MBT, but not significantly different from the 8 hpf control embryos.

Again, when total DNA damage was quantified as OTM in Apex1 K/D embryos, the same trend was evident, but much more elevated (Figure 1C). Pre-MBT, the OTM was 6.6 ±3.0 (S.E., 3) and increased by 150% to 16.6 ±1.9 (S.E., 3) by the MBT (3.3 hpf). By 4.3 hpf, the OTM was still elevated at 17.1 ±2.6 (S.E., 3), but began to decrease at 6 hpf to 10.0 ±1.3 (S.E., 3), still 41% higher than the Apex1 K/D pre-MBT value and significantly higher than the 6 hpf control. By 8 hpf the OTM was 6.2 ±0.4 (S.E., 3), similar to the Apex1 K/D pre-MBT value, and was not significantly different from the 8 hpf control embryos. Taken together these results indicate that DNA damage occurred as part of the normal process of zebrafish embryogenesis, particularly as the embryos transitioned through the MBT. Since the same trend occurred in Apex1 K/D embryos, but was much more pronounced, and Apex1 is required to clear AP sites generated through BER, it would suggest that the small, non-bulky DNA lesions that are BER substrates occur normally as transcription is activated during the MBT and that diminution of Apex1 delays their repair.

Loss of Apex1 results in increased AP sites during zebrafish embryogenesis: As Apex1 is responsible for cleaving AP sites generated during BER initiated DNA repair, I asked whether AP sites would be elevated in Apex1 K/D embryos. Since the AP site quantification kit required a set concentration of DNA, I was unable to obtain enough DNA in a single experiment from 2.5 and 3.3 hpf due to the limited number of cells/embryos and the time required for microinjection. Thus, I was only able to quantify AP sites from the three post-MBT stages (4.3, 6, and 8 hpf). There was a clear increase in the number of AP sites with loss of Apex1 at the first two post-MBT stages, 4.3 and 6 hpf (Figure 2). By 8 hpf, the number of AP sites did not differ between control and Apex1 K/D embryos. These results further strength the proposition that the type of DNA damage occurring during early zebrafish development is repaired by BER and that loss of Apex1 was a limiting factor.
Loss of Apex1 inhibits DNA remethylation post-MBT: Since active DNA demethylation occurs during embryonic development and results in the generation of an AP site via BER, I asked whether the level of global genomic DNA methylation also changed and whether it correlated with observed changes to DNA damage. A clearly discernible, and very similar, trend in DNA methylation was evident in both control and Apex1 K/D embryos (Figure 3). Prior to the MBT (2.5 hpf), 40.9% ± 2.1% (S.E., 6) of cytosine residues were methylated in control embryos, while in Apex K/D embryos 35.1% ± 4.4% (S.E., 6) of cytosine residues were methylated. As embryos entered the MBT at 3.3 hpf, both control and Apex1 K/D embryos experienced a reduction in global cytosine methylation to 23.3% ± 4.9% (S.E., 5) and 28.1% ± 3.1% (S.E., 5) respectively. Post-MBT, global cytosine methylation began increasing to 29.5% ± 2.4% (S.E., 6) in controls, but continued to decrease in Apex1 K/D embryos to 21% ± 4.3% (S.E., 6) by 4.3 hpf. By 6 hpf, global cytosine methylation increased to 36% ± 3.5% (S.E., 6) but only to 27.6% ± 4.8% (S.E., 6) in control and Apex1 K/D embryos respectively. Global cytosine methylation continued to increase at 8 hpf to 41.3% ± 1.9% (S.E., 6) in control embryos, but only 29.8% ± 5.5% (S.E., 6) in Apex1 K/D embryos. By 24 hpf cytosine remethylation diminished dramatically in the Apex1 K/D embryos to 25.9% ± 4.0% (S.E., 6), while methylated cytosine dipped only very slightly to 40.68 ± 3.05% (S.E., 5) in control embryos. The percentage of genomic cytosine present as 5mC in zebrafish embryos at 2.5 hpf and in 24 hpf controls was consistent with HPLC methods on zebrafish livers (Mirbahai et al. 2011) and ELISA based methods on zebrafish embryos (McDougall et al. 2017). Therefore, the loss of Apex1 did not impede global DNA demethylation as zebrafish embryos entered the MBT. However, DNA remethylation post-MBT was reduced compared to control embryos.

Loss of Apex1 does not result in increased DNA damage in Creb1 gene promoters post-MBT: Since a concurrent loss of Creb1 occurred with knockdown of Apex1 at the MBT (Snow 2012), zebrafish creb1 promoters feature a CpG island, and active DNA demethylation occurred during the MBT, I asked whether the CpG island of creb1 might be a target of active DNA demethylation. I utilized a DNA damage qPCR assay to determine if more lesions occurred in the promoters of zebrafish creb1 genes. Zebrafish have two copies of the creb1 gene, creb1a and creb1b, which share ~90% similarity at the mRNA level. There were no differences in the number of DNA lesions present in the gene promoter of creb1a and creb1b in control and Apex K/D embryos (Figure 4 A&B). The number of lesions also did not differ across the three post-MBT stages examined; creb1a had approximately 9.7 DNA lesions per 10 kb DNA, while creb1b had approximately 9 DNA lesions per 10 kb DNA in control and Apex1 K/D embryos, and across all three time points examined.

Loss of Apex1 does not result in detectable changes in amplification of gene promoters using an ARP probe for AP sites: To examine whether DNA demethylation was occurring in promoter CpG islands of particular
genes, I attempted to use genomic DNA labelled with the ARP probe used in the AP site quantification assay coupled with standard PCR. My rationale was that, since the ARP label binds tightly to AP sites and is constructed of an aldehyde group conjugated to biotin, the probe would block polymerases in the PCR reaction through steric hindrance. I would then be able to detect this as reduced PCR reaction product. Of the six genes examined, no difference was detected between control and Apex1 K/D embryos at any of the developmental times (4.3, 6, 8 hpf) examined (Figure 5).

2.4. Discussion

Here I show that total DNA damage occurring during zebrafish embryogenesis in both controls and Apex1 K/Ds is dynamic on a developmental and temporal basis. Increases to total DNA damage correlate well with entry to the MBT, when embryonic genome transcription is activated. While the level of total DNA damage remains high as embryos transition through the MBT to the post-MBT, as embryos transitioned further the amount of total DNA damage returns to similar levels observed prior to the MBT. Thus, elevated DNA damage occurs as the genome of the developing embryo is first being activated. To our knowledge, this represents the first time that DNA damage during embryological development has been quantified.

These results suggest that during zebrafish embryogenesis a normal cellular process occurs that generates DNA damage, either as a component of the process itself or a by-product. Furthermore, when BER capacity is diminished via knockdown of Apex1 the same trend is observed, except the amount of DNA damage increases. This is an important point, as complete loss of Apex1 is embryonic lethal with embryos dying at the MBT (Pei et al. 2011; Wang et al. 2006). Therefore, a logical conclusion is that the main type of DNA damage that occurs during early embryogenesis comprises substrates for BER. This is confirmed, at least in post-MBT stages, by the ~two-fold increase in AP sites observed at 4.3 and 6 hpf stages with partial Apex1 loss. AP sites are a BER intermediate; their presence indicates that BER has been initiated but is limited by the reduced amount of Apex1 enzyme, required to clear the AP sites, resulting in an accumulation of AP sites. While spontaneous depurination of DNA results in ~10,000 AP sites occurring per day per cell (Lindahl and Nyberg 1972), it is likely that the observed increase in AP sites of Apex1 K/D embryos is due to increased DNA damage rather than an increase in spontaneous depurination, since DNA demethylation is a necessary process at this stage.

Changes to global genomic 5mC mimicked the changes in total DNA damage in both control and Apex K/D embryos, though the relationship at each developmental stage was reversed. For example, when DNA damage levels were high, global 5mC were low and vice versa. Since increasing DNA damage correlates well with decreasing levels of 5mC, a logical inference is that the process of DNA demethylation
itself is responsible for the elevated levels of DNA damage observed. The CpG islands of many developmentally important genes become targets for active DNA demethylation at appropriate times in development (Smith and Meissner 2013; Wu and Zhang 2017). When methylated, these genes are typically inactivated; demethylation allows for transcription to occur. During embryogenesis, active demethylation of gene promoters allows controlled development and differentiation of cells (Tricarico and Bellacosa 2016). During zebrafish embryogenesis CpG islands in the promoters of many genes involved in transcriptional regulation, gastrulation, appendage morphogenesis, and cell morphogenesis become hypomethylated as embryos enter the MBT and transition beyond it (Jiang et al. 2013). The process of active DNA demethylation results in activation of BER and generation of an AP site, which requires Apex1 to cleave. The BER pathway will then insert a cytosine base in the gap using the complementary strand as a template and conversion of 5mC to C occurs (Kohli and Zhang 2013; Tricarico and Bellacosa 2016).

Therefore, if large scale DNA demethylation were normally occurring at particular developmental times but repair was reduced, DNA damage levels would be even more elevated due to an accumulation of AP sites.

My results support this interpretation: as embryos enter the MBT at 3.3 hpf total 5mC drops dramatically from ~40% to ~25-30% in both treated and untreated embryos. Concurrently, total DNA damage increases from 15% to 20% in control embryos and to 28% in Apex1 K/D embryos. As embryos begin to transition out of the MBT at 4.3 hpf and total 5mC begins to rebound in the control embryos, 5mC continues to decrease in the Apex1 K/D embryos. Concomitantly, total DNA damage remains high in both control and Apex1 K/D embryos at 4.3 hpf. Interestingly, as embryos from both control and Apex1 K/D treatments transition further into post-MBT stages total DNA damage begins to decrease, though the rate of reduction in the Apex1 K/D considerably lags that of the control embryos. By 8 hpf though, total DNA damage is the same for both treatments indicating that the elevated DNA damage in the Apex1 K/D has been repaired. Quantification of AP sites shows the same results: there are significantly more AP sites in Apex1 K/D embryos at 4.3 and 6 hpf. This result implies that the bulk of the DNA damage is likely to be due to small, non-bulky lesions repaired by BER. By 8 hpf AP sites in the Apex1 K/D embryos return to background levels.

These findings are somewhat surprising when one realizes that at 8 hpf Apex1 levels in K/D embryos are 46% compared to control levels (Snow 2012), theoretically repair of DNA damage at 8 hpf should be impaired compared to earlier developmental times. Abasic sites are chemically unstable and prone to breaking, producing single strand DNA (ssDNA) breaks (Greenberg 2014). Single strand DNA would typically be repaired by the single-strand break repair pathway, which utilizes a number of BER components including Apex1 (Caldecott 2008). However, if active DNA demethylation was occurring in
CpG islands it is possible that ssDNA breaks would occur virtually opposite each other in CpG dinucleotides, resulting in double-strand DNA (dsDNA) breaks. This could lead to activation of one of the two dsDNA repair pathways, homologous recombination (HR) or non-homologous end joining (NHEJ) resulting in the apparent repair of DNA damage seen in the Apex1 K/D embryos. Non-homologous end joining is considered error prone, often causing deletion of several nucleotides at the point of the break (Weterings and Chen 2008). Furthermore, recent findings suggest that HR may be more error prone than originally thought (Rodgers and Mcvey 2016). Activation of these pathways could lead to sequence deletions occurring within the CpG islands, and changes to transcriptional control. Nucleotide loss in CpG islands could erase CpG dinucleotides and could explain why DNA remethylation is lower in Apex1 K/D embryos compared to controls.

Results from the qPCR assay to quantify DNA damage assay indicate that there are no significant differences in the number of lesions present in the CpG island of creb1a and creb1b promoters when Apex1 is knocked-down, at least during the post-MBT stages (4.3, 6, and 8 hpf). This could indicate that the CpG island of Creb1 promoters in zebrafish are not targeted for active DNA methylation at those developmental times. Likewise, an attempt to PCR genomic DNA treated with an ARP probe, produced the same result. The rationale behind this approach was that AP sites are inherently unstable, yet an ARP probe will bind specifically to the AP site and the resultant AP site/ARP probe will be stabilized (Bennett and Kitner 2006). If a DNA sequence includes one or more ARP probes, the Taq polymerase should be stalled through steric hindrance and no exponential amplification of DNA would occur. This would be detectable as a reduction in product compared to non-treated template. All of the genes tested using this method showed no significant difference between the number of AP sites in control and Apex1 K/D embryos of the tested gene promoters, at least during the post-MBT stages (4.3, 6, and 8 hpf). It was not possible to examine earlier times with both qPCR and PCR methods as the amount of template required was greater than could be produced per knock-down experiment.

One assumption made was that more AP sites would be present in the gene promoters of the Apex1 K/D embryos resulting in lower amplification. However, if active DNA demethylation occurred within those promoters as a normal part of cellular physiology, then ARP probes would also be present in the control template. Presumably, the presence of just one ARP probe could block the Taq polymerase, meaning the assay may not be sensitive enough to detect differences between treatments, as amplification may be diminished to a similar degree in both despite the Apex1 K/D template having more AP sites. In addition, the assumption that the ARP probe would block a Taq polymerase is unproven; it is possible a polymerase may be able to simply knock off the ARP or bypass it.
My findings not only show that total DNA damage changes during zebrafish embryogenesis, but that it correlates with changes to DNA methylation status. As DNA damage increases, DNA methylation decreases and vice versa. Partial loss of Apex1 protein results in higher levels of DNA damage, and more AP sites. These results indicate that the majority of the DNA damage occurring during embryogenesis is a substrate for BER. Therefore, it is apparent that a normal cellular process occurring during zebrafish embryogenesis, potentially active DNA demethylation during transcriptional activation, specifically results in the generation of DNA damage as part of the process.

Table 1. Primer sequences used for PCR assay using ARP-labelled DNA. The presence of a CpG island is indicated, as well as size of CpG island in base pairs. The amplicon size in base pairs is also given, as well as the annealing temperature used for each primer pair.

<table>
<thead>
<tr>
<th>Name</th>
<th>Accession #</th>
<th>CpG island</th>
<th>CpG island size</th>
<th>Sequence</th>
<th>Amplicon size</th>
<th>T annealing (°C)</th>
<th>t elongation (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creb1a-F</td>
<td>NC_007112.7</td>
<td>Yes</td>
<td>396 bp</td>
<td>CCGGTGATCGCATTCATAAC GCAATTGAAAGCTATCCAGTCG</td>
<td>1012 bp</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Creb1a-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creb1b-F</td>
<td>NC_007120.7</td>
<td>Yes</td>
<td>249 bp</td>
<td>CAGGGCGTGCTCAAAATATC CTAATTGTCAAGTGCCCCTGC</td>
<td>1004 bp</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Creb1b-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rps21-F</td>
<td>NC_007134.7</td>
<td>No</td>
<td>NA</td>
<td>TGGTTACAGGACTTTGCCGATT TCGTGTTGTTCTTACGATT</td>
<td>968 bp</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Rps21-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun-F</td>
<td>NC_007131.7</td>
<td>Yes</td>
<td>1343 bp</td>
<td>GTATACAGACCTCGCCACAG GCCGATCTGTGTTCATGC</td>
<td>1048 bp</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Jun-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insm1b-F</td>
<td>NC_007128.7</td>
<td>Yes</td>
<td>1571 bp</td>
<td>TTTCTTTGTGTGTCGTCATTTC FATGGCTTGGGTGTTCATTG</td>
<td>975 bp</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Insm1b-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bactin-F</td>
<td>NC_007112.7</td>
<td>No</td>
<td>NA</td>
<td>GCCCACTGGTTCTCATCAAAC TTAGGCGACACGAGACCAAG</td>
<td>966 bp</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Bactin-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Primer sequences for qPCR using genomic DNA. Amplicon size is given as well as annealing temperature and duration of elongation for each primer pair.

<table>
<thead>
<tr>
<th>Name</th>
<th>Accession #</th>
<th>CpG island</th>
<th>CpG island size</th>
<th>Sequence</th>
<th>Amplicon size</th>
<th>T annealing (°C)</th>
<th>t elongation (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creb1a-F</td>
<td>NC_007112.7</td>
<td>Yes</td>
<td>396 bp</td>
<td>CCGGTGATCGCATTCATAAC GCAATTGAAAGCTATCCAGTCG</td>
<td>1012 bp</td>
<td>58</td>
<td>60</td>
</tr>
<tr>
<td>Creb1a-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creb1b-F</td>
<td>NC_007120.7</td>
<td>Yes</td>
<td>249 bp</td>
<td>CAGGGCGTGCTCAAAATATC CTAATTGTCAAGTGCCCCTGC</td>
<td>1004 bp</td>
<td>57</td>
<td>60</td>
</tr>
<tr>
<td>Creb1b-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creb1a-F</td>
<td>NC_007112.7</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creb1a-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creb1b-F</td>
<td>NC_007120.7</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creb1b-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Primer sequences for qPCR using genomic DNA. Amplicon size is given as well as annealing temperature and duration of elongation for each primer pair.
Figure 1. The level of DNA damage occurring during zebrafish embryogenesis is dynamic. A, representative image of comets from 3.3 hpf embryos. The tail is noticeably longer in the Apex K/D sample than the control. B, total DNA damage, as % DNA in the comet tail, rises as embryos enter the MBT before dropping back down by 6 hpf. This effect is more pronounced in embryos with compromised BER. C, total Olive Tail Moment reflects the same trend. Trend-lines added to visualize overall trend. Line marked MBT denotes developmental time point embryos enter the MBT. Error bars represent ± S.E., ** p ≤ 0.005; * p ≤ 0.05. Significance testing was between control and Apex1 K/D at each developmental time point.

Figure 2. Loss of Apex1 protein results in increased AP sites post-MBT. There is approximately a two-fold increase in the number of AP sites in the Apex1 K/D embryos as they transitioned to the post-MBT stages. By 8 hpf, the number of AP sites in Apex1 K/D embryos has returned to control levels. Error bars represent ± S.E.
Figure 3. Genomic cytosine undergoes demethylation/remethylation around the MBT. Prior to the MBT, ~40% of genomic cytosine is methylated, which drops to under 30% as embryos transition to the MBT. As embryos develop, remethylation occurs, resulting in cytosine methylation levels returning to ~40%. The same trend is evident with Apex1 loss, but recovery of cytosine methylation is delayed and diminished compared to control. Trend-lines added to visualize overall trend. Line marked MBT denotes developmental time point embryos enter the MBT. Error bars represent ± S.E., * p ≤ 0.05. Significance testing is between Control and Apex1 K/D at each developmental time point.
Figure 4. Loss of Apex1 protein does not affect the number of DNA lesions found in both Creb1 gene promoters. A, gene promoter of Creb1a. Both control and Apex1 K/D embryos had ~10 DNA lesions per 10 kb at the three post-MBT stages tested. B, gene promoter of Creb1b. As with Creb1a no difference was apparent in the number of DNA lesions in control and Apex1 K/D embryos. There were ~9 DNA lesions per 10 kb at the three post-MBT stages tested. Error bars represent ± S.E.
Figure 5. No difference is detectable between post-MBT control and Apex1 K/D DNA labelled with an ARP probe. An ARP probe binds stably to AP sites. PCR of template could result in reduced amplification of target if an AP site was present. No differences in amplification of the promoter regions of five select genes were detected across all three post-MBT stages examined. Error bars represent ± S.E.
Chapter 3. DNA lesions can modulate transcription factor binding.

As previously discussed the Strauss lab had discovered that loss of Apex1 protein during zebrafish embryonic development resulted in a concurrent, and dramatic, loss of the transcription factor (TF) Creb1. Loss of Creb1 protein during embryogenesis must therefore be considered a primary effect of Apex1 loss. Additionally, many other genes reliant upon Creb1 for activation were also downregulated, though the timing suggested that this was due to the initial loss of Creb1 protein. Thus, it is possible that embryonic lethality through Apex1 knockout is through the loss of Creb1 and the resultant misregulation of the many genes reliant upon Creb1 for transcription. Furthermore, Creb1 protein could be rescued by co-injection with human Apex1 message, but only when it had the ability to process AP sites. The logical inference was that loss of Creb1 protein occurred due to an inability to complete BER. These findings showed that reduced BER capacity resulted in a consistent loss of Creb1. This is interesting as the traditional view of DNA damage is of a random process; yet, for loss of Creb1 to occur when DNA repair capacity was reduced, damage is likely to manifest through the portion of the genome involved in regulating Creb1’s transcription. If damage occurred in the coding portion of the gene, then we would likely still detect normal levels of Creb1 protein, but the protein would carry point mutations or be non-functional. Based on this supposition we speculated that small, non-bulky DNA lesions must be preferentially occurring in the promoter of Creb1. Indeed, oxidative DNA damage generated by lysine demethylase 1 (LSD1) mediated demethylation of dimethyl-Lys⁵ in histone H3 (H3K9me2) drives estrogen-induced transcription (Perillo et al. 2008). This provides a precedent that controlled DNA damage and repair can regulate active transcription. Other studies have also shown that site specific DNA damage can activate gene transcription (Ba et al. 2014; Fleming, Ding, and Burrows 2017; Pan et al. 2016; Pastukh et al. 2015), though the mechanism through which damage occurs to specific points in the genome remains relatively unexplored.

Based on these previous findings, we hypothesized that DNA lesions occurring at specific locations within Creb1’s DNA binding sequence would act to regulate transcription of Creb1 by modulating Creb1 binding. The following research paper was published in the journal DNA Repair in 2013 (Moore et al. 2013) and represents the first steps in addressing this question. In the article, we investigated the binding kinetics of Creb1 when a small number of small, non-bulky DNA lesions and the intermediates that are formed during repair by BER were present in its consensus sequence. We focused upon ⁶O-G, the most common form of oxidative DNA lesion, and the intermediates generated during its repair. Since demethylation of 5-methylcytosine (5mC) can result in the activation of BER and subsequent generation of an AP site, we also examined Creb1 binding in the presence of 5mC and demethylation intermediates including 5-hydroxymethylcytosine (5hmC), thymine (T), and uracil (U). Uracil not only occurs through random deamination of cytosine (C) or misincorporation, but has been shown to be an intermediate generated during
demethylation of 5mC in zebrafish (Wu et al. 2014). By measuring the binding of purified recombinant CREB1 to modified oligonucleotides by electrophoresis mobility shift assay (EMSA), we were able to determine that BER substrates could act in an epigenetic-like fashion.

3.1. DNA modifications repaired by base excision repair are epigenetic.

Stephen P.G. Moore¹, Kimberly J. Toomire¹, and Phyllis R. Strauss¹

¹Department of Biology, Northeastern University, Boston MA02115

This work was supported by the G. Harold and Leila Y. Mathers Foundation; Aid for Cancer Research; and funds from Northeastern University.

3.2. Abstract

CREB1 controls ~25% of the mammalian transcriptome. Small changes in binding to its consensus (CRE) sequence are likely to be amplified many fold in initiating transcription. Here we show that DNA lesions repaired by the base excision repair (BER) pathway modulate CREB1 binding to CRE. We generated $K_d$ values by electrophoretic mobility shift assays using purified human CREB1 and a 39-mer double-stranded oligonucleotide containing modified or wild-type CRE. CRE contains two guanine residues per strand, one in a CpG islet. Alterations in CRE resulted in positive or negative changes in $K_d$ over two orders of magnitude depending on location and modification. Cytosine methylation or oxidation of both guanines greatly diminished binding; a G/U mispair in the CpG context enhanced binding. Intermediates in the BER pathway at one G residue or the other resulted in reduced binding, depending on the specific location, while there was no change in binding when the single G residue outside of the CpG islet was oxidized. CREB1 recruits other partners after dimers form on DNA. Only UpG increased DNA-CREB1 dimer formation. Since oxidation is ongoing and conversion of cytosine to uracil occurs spontaneously or at specific times during differentiation and development, we propose that BER substrates are epigenetic and modulate TF recognition/binding.

Key words: chromatin, CREB1, DNA base excision repair, DNA-protein interaction, epigenetics, transcription factors
3.3. Abbreviations

**AID**  activation induced deaminase

**AP**  abasic

**APEX1**  human AP endonuclease 1

**ATF-1, ATF-2**  activating transcription factor-1, -2

**BER**  base excision repair

**CRE**  cAMP response element

**CREB1**  cyclic AMP response element binding protein 1

**CREM**  cAMP response element modulator

**°G**  8-oxo-7,8-dihydroguanine

**hmC**  hydroxymethylcytosine

**hmU**  5-hydroxymethyluracil

**OGG1**  oxoguanine glycosylase

**RNS**  reactive nitrogen species

**ROS**  reactive oxygen species

**THF**  tetrahydrofuran
3.4. Introduction

3.4.1. Oxidative damage to DNA is ongoing in all cells and not considered to be epigenetic.

Epigenetics encompasses DNA methylation, histone modification, nucleosome localization and synthesis of non-coding RNA, all of which modulate access to the transcription machinery. These modifications are central to differentiation, embryonic development and carcinogenesis (Riddihough and Zahn 2010). Other DNA modifications including tautomeric shifts, deamination, conversion of cytosine to uracil (U), oxidation of bases and spontaneous base loss are ongoing. Most of these lesions are repaired by the base excision DNA repair (BER) pathway and have never been considered epigenetic, because BER efficiently removes them. However, oxidative DNA damage due to reactive oxygen species (ROS) or reactive nitrogen species (RNS) from either intra- or extracellular sources is a continuous process in all cells (David et al. 2007), with oxidation of guanine (G) to 8-oxo-7,8-dihydroguanine ("G) being especially common. Repair is efficient but not perfect and a low level of "G and U can be found in the genome.

3.4.2. The BER pathway, which repairs most oxidative damage and removes uracil in DNA, is a cascade of critical enzymes.

The BER pathway is initiated by specific DNA glycosylases that largely remove the aberrant base without cleaving the DNA backbone and leave an abasic (AP) site, the accumulation of which is toxic (Zharkov and Grollman 2005). The endonuclease function of the enzyme AP endonuclease (APEX1) then recognizes and cleaves this AP site on the 5' side of the phosphodeoxyribose, after which a single nucleotide is inserted by DNA polymerase β (Zharkov 2008). However, in some cases, such as with "G, the lesion is recognized by oxoguanine glycosylase (OGG1) that not only removes the aberrant base but frequently cleaves the AP site on the 3' side requiring the removal of the sugar by APEX1 (Zharkov and Grollman 2005). The loss of APEX1 is embryonic lethal in mice (Xanthoudakis et al. 1996), while knockdown of Apex1 in zebrafish embryos leads to improper development of heart and neural tissue, so that development continues until heart function is required several days later and the embryos die (Wang et al. 2006). Furthermore, knockdown of Apex1 results in a p53-independent loss of cyclic AMP response element binding protein (CREB1) (Pei et al. 2011).

Cytosine can be modified by a number of different mechanisms associated with chromatin modification and gene activation. In lymphocytes, C is converted directly to U by activation induced deaminase (AID) during immunoglobulin processing (Krokan, Drabløs, and Slupphaug 2002; Maul et al. 2011). Meanwhile, for chromatin to be activated, mC residues need to be converted to C through the removal of the methyl group. Demethylation of mC occurs through several enzymatic pathways which generate various
intermediates such as T, $^{hm}$C, 5-carboxylcytosine, 5-formylcytosine or 5-hydroxymethyluracil ($^{hm}$U). BER is then required to convert the resulting intermediates to C (Wu and Zhang 2010).

3.4.3. The oxidative state of DNA could be epigenetic and effect important transcription binding factors like CREB1.

Cellular redox status, reflected in the levels of glutathione and thioredoxin/thioredoxin reductase, is usually construed as the relative ability to make or break disulfide bonds (Manda, Nechifor, and Neagu 2009). Despite the ongoing production of ROS and resultant oxidative damage to DNA, the oxidative state of DNA is overlooked. We have chosen to examine the ability of CREB1 to bind to its CRE consensus sequence containing U or selected oxidative damage, as CREB1 is responsible for the transcriptional activation of ~25% of the genome (Impey et al. 2004; Zhang et al. 2005). CREB1 binds to the eight-base palindromic cAMP-response element (CRE) TGACGTCA, or the non-palindromic half-site motif CGTCA (Mayr and Montminy 2001). The full length CRE-site contains two G residues per strand that could be susceptible to oxidation, one located in a CpG islet that is subject to methylation/demethylation or conversion to UpG (Figure 1). CREB1 commonly functions as a homodimer but is capable of forming heterodimers with other bZIP members including cAMP response element modulator (CREM) and activating transcription factor 1 or 2 (ATF-1 or ATF-2) (Shaywitz and Greenberg 1999). Despite the name “activating transcription factor,” several of the ATF’s including ATF-2 in the heterodimer serve as repressors (Hai and Hartman 2001).

Since CREB1 is the foundation on which many other TFs build the complex that will recruit RNA polymerase and initiate transcription, subtle changes in CREB1 binding could amplify or depress transcription many fold. Since even small changes in gene expression can have an enormous impact on cellular development, changes in oxidative state or the presence of a U residue within promoter sites may be epigenetic (Ghosh and Mitchell 1999; Mitchell and Ghosh 2007). In this study, we investigated the possibility that lesions and modifications in DNA repaired by BER might alter CREB1 binding to its CRE target site. We present quantitative data indicating that the oxidative status of DNA itself and, in particular, lesions repaired by BER, are an important component in regulating recognition of TFs to their cognate sequences and, therefore, are likely to play an epigenetic role in cell physiology.
3.5. Methods

3.5.1. Oligonucleotides

39-mer oligonucleotides containing a CRE-site with or without various DNA modifications (Table 1) were sourced either from Midland Certified Reagent Co., Inc. (Midland, Texas) or from Integrated DNA Technologies (Coralville, Iowa). G1 and G2 AP sites and G1 5' and 3' dRP modifications were formed with tetrahydrofuran (THF). Either upper or lower strands were labelled at the 5' end with \([\gamma^{32}\text{P}]\)ATP (Perkin-Elmer, Waltham, Massachusetts) by polynucleotide kinase (PNK) (New England Biolabs, Ipswich, Massachusetts) as previously described (Fortier et al. 2009; Strauss et al. 1997). After the labelling reaction was complete, PNK was heat-inactivated at 95°C for 5 min and the desired complement was added at a 1:1 molar ratio. The mixture was slow-cooled, separated from unincorporated \(^{32}\text{P}\) using Illustra™ MicroSpin™ G-25 columns (GE Healthcare, Little Chalfont, United Kingdom), and stored at -20°C.

3.5.2. Protein

Purified recombinant human CREB1 protein was the kind gift of Dr. Jennifer Nyborg (Colorado State University, Fort Collins, Colorado). Before use CREB1 was diluted into a buffer containing 50 mM Tris-HCl (pH 7.9), 100 mM KCl, 12.5 mM MgCl\(_2\), 1 mM EDTA (pH 8.0), 20% glycerol, and 2 mM DTT.

3.5.3. Electrophoretic Mobility Shift Assay (EMSA)

EMSAs were performed as described by Lopez et al (Lopez, Mick, and Nyborg 2007). Briefly, varying concentrations (0 to 50 nM) of purified CREB1 protein were incubated in a reaction buffer (25 mM Tris [pH 7.9], 50 mM KCl, 6.25 mM MgCl\(_2\), 0.5 mM EDTA, 10% glycerol, 0.1% Tween-20, and 2 mM DTT) with 5 ng poly(dA-dT)-poly(dA-dT), 0.25 \(\mu\)g/\(\mu\)L BSA, and \(^{32}\text{P}\)-end labelled double stranded oligonucleotide probe (0.5 nM) on ice for 1 h in a final volume of 20 \(\mu\)L. Samples were then loaded into the gel without tracking dye and resolved at room temperature on pre-electrophoresed native 5% (49:1 acrylamide/bisacrylamide) gels in a buffer composed of 40 mM Tris-HCl (pH 8.5), 306 mM Tris-glycine and 0.1% Nonidet P-40. Electrophoresis was performed at 90 V for 2.5 h until a tracking lane containing bromophenol blue dye had reached ~2/3 down the gel.

3.5.4. Binding Kinetics

Vacuum dried gels were exposed to a PhospholImager screen and visualized with a GE Storm™ PhosphorImager™; data were quantified with Molecular Dynamics ImageQuant (GE Healthcare, Little Chalfont, United Kingdom). Experiments were performed at least three independent times, except for two
experiments that were conducted twice. \( K_d \) values were calculated using the equation:

\[
K_d = \frac{[\text{unbound DNA}] \times [\text{unbound CREB1}]}{[\text{bound CREB1}]}
\]

Bound CREB1 was calculated by quantifying the amount of bound DNA; unbound CREB1 was calculated as the difference between initial CREB1 and bound CREB1.

For substrates where saturation binding was not achieved under our assay conditions, we used the initial slope to estimate a concentration of CREB1 where half maximal binding would be expected.

### 3.5.5. Statistical analyses

Data were analyzed using SPSS 19.0 (IBM, Armonk, New York). Independent \( t \)-tests were performed between data for each DNA modification and the unmodified control sequence. Before each \( t \)-test was applied, a Levene's test for equal variance was conducted. If the assumption of equal variances was violated, a \( t \)-test with unequal variances was utilized. Significance for all statistical tests was set \textit{a priori} at 0.05.

### 3.6. Results

#### 3.6.1. Purified CREB1 recognizes and binds the CRE-site cleanly.

DNA-protein complexes (monomer and dimer) formed between purified recombinant CREB1 and the 39-mer sequence 5' - TAC CAT GCC TG1 ACG2 TCA GAG AGC ATT CGT AAT CAT GGT - 3' (CRE sequence shown in bold) resolved cleanly using the experimental protocols employed in this study (Figures 2 and 3). The 5’-most G in the sequence is indicated as G1, while the G residue in the CpG islet is indicated as G2. As the concentration of CREB1 increased, the monomer form (DNA.CREB1) appeared first followed at higher concentrations by the dimer form. The \( K_d \) for total CREB1 binding to control 39-mer was 4.8 \pm 0.2 nM (SE,3) (Figure 3 and 4). After CREB1 initially bound to the CRE-site as a monomer (DNA.CREB1), a second molecule of CREB1 bound to DNA.CREB1 to form a dimer (DNA.CREB1\(_2\)), which is in keeping with published reports that CREB1 forms dimers on CRE only after the DNA.monomer complex has formed (Wu et al. 1998). Kinetics of dimer formation are dissected from overall data below.
3.6.2. CREB1’s ability to bind to its consensus sequence is affected both positively and negatively by specific modifications and their locations.

A G/U mispair positioned at G1 did not interfere with CREB1’s ability to recognize and bind to the CRE-site ($K_d$ G1 G/U 4.2 ±0.7 nM, SE,3). However, when a G/U mispair was positioned at G2 within the CpG islet, CREB1 binding was enhanced 6-fold ($K_d$ G2 G/U 0.8 ±0.1 nM, SE,3) relative to the unmodified control (Figures 4A and B). Unlike the case with a single G/U mispair, CREB1 binding was enhanced only 3-fold ($K_d$ G2 G/U 0.8 ±0.1 nM, SE,3) when C residues in the CpG islet in both strands were replaced with U. The differences in binding between control and a single U, control and two U residues, and between a single U and two U residues were significant at $p \leq 0.001$. The ability of purified CREB1 to recognize and bind to the CRE-site was not affected by the presence of a single °G modification at either the G1 or G2 positions ($K_d$ G1 °G 3.7 ±1.0 nM, SE,3; $K_d$ G2 °G 5.0 ±0.5 nM, SE,3) relative to the unmodified control sequence (Figures 4A and B). However, when both CRE-site G residues on the same strand or G residues on both strands in the CpG islet were replaced with °G residues, saturable binding could not be achieved (Figure 6A). $K_d$ values, calculated from the initial binding slope at low concentrations, were 36 ±3.3 nM (SE, 3) and 36.7 ±2.5 nM (SE, 3) respectively (Figure 6A). An AP site at either the G1 or G2 positions (Figure 4A and B) was detrimental for CREB1 binding ($K_d$ G1 AP 8.1 ±0.5 nM, SE,3; $K_d$ G2 AP 13.9 ±1.5 nM, SE,3). Other modifications to the G1 or G2 residue, or the C residue within the CpG islet were deleterious to total CREB1 binding, except for the presence of a nick in the backbone on the 5’ side of the G1 guanine ($K_d$ G1 5’Nick 5.5 ±1.2 nM, SE,3) (Figure 4A and B). This lesion would have been generated by the 5’ incision repair activity of APEX1 (Timofeyeva et al. 2011). Methylation of both C residues on both strands in the CpG islet, e.g. methylation on both upper and lower strands, resulted in extremely poor CREB1 binding ($K_d$ hypermethylated 99.1 ±15.7 nM, Figure 6A), while hemimethylation resulted in reduced binding affinity ($K_d$ hemimethylated 17.4 ±3.2 nM, SE,3, Figure 4B). Failure of CREB1 to bind to its CRE sequence when both C residues in the CpG context are methylated has been reported previously (Iguchi-Ariga and Schaffner 1989).

3.6.3. DNA modification and spatial position within the CRE-site affects CREB1’s ability to form a dimer.

Since CREB1 is only transcriptionally active when two protein molecules (either CREB1 or CREB1 and one of its binding partners) have bound to each CRE-site (Shaywitz and Greenberg 1999), we investigated how modifications to selected residues in the CRE-site might affect dimer formation, which was easily detected under our conditions (Figures 2 and 3). When 50 nM CREB1 was added to the reaction mix, 62% of the total amount of bound CREB1 was bound as a dimer on the control consensus sequence
(0.28 ±0.02 nM, SE,3) (Figure 5A and B). Several G1 modifications including a nick in the DNA backbone 5’ to the G1 residue (0.26 ±0.02 nM, SE,3) and a U opposite the G1 residue (0.26 ±0.01 nM, SE,3) did not interfere with dimer formation despite changes in overall K_d values (Figure 5A). However, all other modifications to the G1 residue that we tested resulted in reduced dimer formation, including a single °G at G1 (0.23 ±0.01 nM, SE,3) (Figure 5A). When modifications were made to residues in the CpG context, dimer formation was generally suppressed (Figure 5B and 6B) except for a G/U mispair (Figure 5B). A G/U mispair enhanced dimer formation relative to the unmodified CRE-site (0.33 ±0.01 nM, SE,3).

3.7. Discussion

3.7.1. Major Conclusions

Here we have shown that DNA modifications of guanine residues in the CREB1 consensus sequence, the presence of U, or lesions repaired through the BER pathway alter CREB1 binding to its CRE consensus sequence positively or negatively depending on the modification and its location. Selected G or C residues were oxidized (°G), converted to U, T, methylated C (°mC), 5-hydroxymethylated C (°hmC), or a range of intermediates arising during BER. The effects were specific: conversion of a single C residue within the CpG context to U enhanced total binding 6-fold, while a G/T mispair, oxidation of both G residues or a methylated C was highly deleterious, decreasing binding up to 25-fold of the control level, thus expanding the range of CREB1 binding over two orders of magnitude. Furthermore, substitution with all intermediates of the BER pathway was also deleterious. In agreement with others (Iguchi-Ariga and Schaffner 1989), CREB1 binds poorly to its target site when the cytosine residues in the CpG islet of the CRE-site were hyper- or fully methylated, presumably preventing in discriminant and wasteful transcription. We also examined the ability of highly purified recombinant CREB1 to form dimers on the CRE sequence. While others have noted effects of a G/U mispair on CREB1 or activator protein-1 (AP-1) binding from a nuclear extract (Verri et al. 1990), the ability to discriminate monomer and dimer by purified CREB1 allowed us to quantify useful binding (Rogstad et al. 2002), because CREB1 recruits its binding partners only when it has formed a dimer on the CRE.

3.7.2. Response to oxidative DNA damage is pleomorphic and probably extends to other transcription factors.

It is widely accepted that oxidative DNA damage is harmful. Oxidative base lesions have been linked to carcinogenesis and neurodegenerative diseases such as Alzheimer's and Parkinson's disease
However, the response to oxidation is, in fact, graded and depends on the gene in question: in our study CREB1 binding tolerated conversion of either G residue individually to °G without change in recognition, while conversion of both G residues to °G and all BER intermediates were highly deleterious. Sensitivity to guanine oxidation depends on the TF as well, since single °G lesions in the binding sites of the TFs AP-1 and Sp1 are sufficient to inhibit binding affinity (Hailer-Morrison et al. 2003; Mitchell and Ghosh 2007; Ramon et al. 1999). In contrast, °G lesions in select residues of the NF-kB consensus site enhance binding affinity (Hailer-Morrison et al. 2003), while an 8-oxo-A lesion positioned at the first adenine residue of the CREB1 consensus sequence enhances binding affinity (Abraham and Brooks 2011). Since oxidized adenine occurs at much lower frequency than °G, it may or may not play a role in modulating transcription. No data on dimerization are available in these studies.

Conformational changes to DNA imposed by the various modifications examined in this study account for many of the changes to CREB1 binding kinetics. For example, an AP site and its complimentary residue often adopt an extrahelical conformation (Chen et al. 2008; Sicoli et al. 2009), while DNA containing a T:G mispair displays a distorted conformation with the T often flipped out (Imhof and Zahran 2013). Both would interfere with CREB1 binding and dimer formation. The fact that single °G lesions do not appear to overly disturb DNA structure (Sicoli et al. 2009) is consistent with the observation that the overall binding of CREB1 when CRE contains a single °G does not differ with that of control. On the other hand, NMR or electron resonance structural studies of other structures indicated negligible perturbation to DNA when a nick or a gap is present (Sicoli et al. 2009). However, these studies were performed in the absence of protein, suggesting that nicks, gaps, or an AP site in CRE may increase flexibility that is incompatible with strong CREB1 binding.

CREB1 acts as an anchor protein that enables transcription of ~25% of the eukaryotic genome (Impey et al. 2004; Zhang et al. 2005). Thus, one would expect that genes containing CRE sites in their promoters, including FOS, NFkB, and genes whose promoters include consensus binding sites for those TFs would be more sensitive to oxidative damage than others. Transcription factors whose consensus sequences are enriched for G residues would also be prime targets. Sp1 is one such TF whose consensus sequence is composed solely of G and C residues. Consequently, select genes containing one or more Sp1 binding sites as well as one or more CRE in their promoter sequences could be especially responsive to oxidative damage. Genes with both Sp1 and CREB1 binding sites include nitric oxide synthase 1 and 3, dihydrofolate reductase, interleukin 10 and the major enzymes involved in base excision repair including 8-oxoguanine glycosylase, AP endonuclease 1 and DNA polymerase β. At least one of these genes (APEX1) has 8 CREB1 binding sites and 11 AP-1 sites. Since AP-1 consists of FOS and JUN hetero
homodimers and, since FOS has 8 CREB11 sites in its promoter, transcription of such genes would be even more CREB1 dependent and, therefore, would be even more subject to potential regulation by oxidation of G residues or the presence of U. CREB1 itself is autoregulated with 13 self-binding sites in its own promoter. The CREB11 promoter also contains 7 AP-1 sites, 1 Sp1 site, 4 NFkB sites, 2-p300 sites, and an ATF site. The importance of oxidation and potential U in these sites to regulate BER enzymes is further underscored by the fact that the sites in BER promoters are conserved among the human, mouse and rat genomes (Rebhan et al. 1998).

3.7.3. DNA damage repaired by BER is likely to be epigenetic.

Small changes to gene expression in a single prominent TF like CREB11 can have a considerable impact on cell physiology and development. Given the broad importance of CREB1 and the fact that oxidative modification and the presence of U affect other critical TFs as well, we propose that the lesions repaired by the base excision repair pathway serve as modulators for transcription efficiency. It is even possible that AID/APOBEC enzymes might convert C residues to uracil selectively, much as in the case of mutation and recombination of immunoglobulin genes (Smith et al. 2012). While higher levels of oxidative damage might alert the cellular machinery of impending oxygen toxicity, lower levels or the presence of U could serve as a titration mechanism to regulate the degree of activation of selected genes. Consequently, we propose that selective oxidative alterations and the presence of U in DNA act epigenetically and separately from modification of chromatin proteins.

3.8. Acknowledgments

The authors thank Dr. Jennifer Nyborg (Colorado State University) for the gift of purified human CREB1.

3.9. Conflict of Interest

The authors declare that there are no conflicts of interest.
strand resulted in a gap or nick being present at the desired position. These are indicated by | for a nick and by synthesizing two top or bottom strands that when annealed together with a full length complementary sequence column. Probes that had nicks between nucleotides or were missing nucleotide modifications in both top and bottom strands; top and bottom strands are differentiated by T or B in the CRE site in bold (5’- TAC CAT GCC TTG1 ACG2 TCA GAG AGC ATT CGT AAT CAT GGT-3’). Numerals 1 and 2 indicate the guanine residues that were modified. Some probes had modifications in both top and bottom strands; top and bottom strands are differentiated by T or B in the sequence column. Probes that had nicks between nucleotides or were missing nucleotides were constructed by synthesizing two top or bottom strands that when annealed together with a full length complementary strand resulted in a gap or nick being present at the desired position. These are indicated by | for a nick and ■ for a gap. The 5’ oligonucleotide ends at the nick or gap sites did not have phosphate groups present.

### Table 1. Modifications to CRE used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Position of modification</th>
<th>Strand</th>
<th>Sequence</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>N/A</td>
<td>T</td>
<td>TG ACG TCA-</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>G1 →G</td>
<td>G</td>
<td>G1 T</td>
<td>TG ACG TCA-</td>
<td>8-oxo-dG</td>
<td></td>
</tr>
<tr>
<td>G1 Abasic</td>
<td>—</td>
<td>G1 T</td>
<td>THF ACG TCA-</td>
<td>THF</td>
<td></td>
</tr>
<tr>
<td>G1 5’ dRP</td>
<td>—</td>
<td>G1 T</td>
<td>THF ACG TCA-</td>
<td>THF</td>
<td></td>
</tr>
<tr>
<td>G1 3’ dRP</td>
<td>—</td>
<td>G1 T</td>
<td>THF ACG TCA-</td>
<td>THF</td>
<td></td>
</tr>
<tr>
<td>3’ Nick</td>
<td>T</td>
<td>G1 T</td>
<td>TG ACG TCA-</td>
<td>Nick 3’ side of dG</td>
<td></td>
</tr>
<tr>
<td>5’ Nick</td>
<td>G</td>
<td>G1 T</td>
<td>TG ACG TCA-</td>
<td>Nick 5’ side of dG</td>
<td></td>
</tr>
<tr>
<td>G1 Gap</td>
<td>—</td>
<td>G1 T</td>
<td>TG ACG TCA-</td>
<td>dG deleted</td>
<td></td>
</tr>
<tr>
<td>G1&amp;G2 5’G</td>
<td>—</td>
<td>G1 + G2 T</td>
<td>TG ACG TCA-</td>
<td>8-oxo-dG</td>
<td></td>
</tr>
<tr>
<td>G2 5’G</td>
<td>G</td>
<td>G2 (CpG) T</td>
<td>TG ACG TCA-</td>
<td>8-oxo-dG</td>
<td></td>
</tr>
<tr>
<td>G2 Abasic</td>
<td>—</td>
<td>G2 (CpG) T</td>
<td>TG ACG TCA-</td>
<td>THF</td>
<td></td>
</tr>
<tr>
<td>G2 Gap</td>
<td>—</td>
<td>G2 (CpG) T</td>
<td>TG ACG TCA-</td>
<td>dG deleted</td>
<td></td>
</tr>
<tr>
<td>5°C</td>
<td>—</td>
<td>CpG T</td>
<td>TG ACG TCA-</td>
<td>5-hydroxymethyl-dC</td>
<td></td>
</tr>
<tr>
<td>Hypermethylated</td>
<td>—</td>
<td>CpG T, B</td>
<td>5’-TG ACG TCA-</td>
<td>5-methyl-dC</td>
<td></td>
</tr>
<tr>
<td>Hyperoxidised CpG</td>
<td>—</td>
<td>CpG T, B</td>
<td>5’-TG ACG TCA-</td>
<td>8-oxo-dG</td>
<td></td>
</tr>
<tr>
<td>Hypermethylated and oxidized</td>
<td>—</td>
<td>CpG T, B</td>
<td>5’-TG ACG TCA-</td>
<td>5-methyl-dC + 8-oxo-dG</td>
<td></td>
</tr>
<tr>
<td>Hemimethylated and oxidized</td>
<td>—</td>
<td>CpG T, B</td>
<td>5’-TG ACG TCA-</td>
<td>8-oxo-dG</td>
<td></td>
</tr>
<tr>
<td>G1/C uracil</td>
<td>—</td>
<td>G1 B</td>
<td>TG ACG TCA-</td>
<td>dU</td>
<td></td>
</tr>
<tr>
<td>G2/C uracil</td>
<td>—</td>
<td>G2 (CpG) B</td>
<td>TG ACG TCA-</td>
<td>dU</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>—</td>
<td>CpG B</td>
<td>TG ACG TCA-</td>
<td>dT</td>
<td></td>
</tr>
<tr>
<td>Hemimethylated</td>
<td>—</td>
<td>CpG B</td>
<td>TG ACG TCA-</td>
<td>methyl-dC</td>
<td></td>
</tr>
<tr>
<td>CpG 5’nick</td>
<td>—</td>
<td>CpG B</td>
<td>TG ACG TCA-</td>
<td>Nick 5’ side of dC</td>
<td></td>
</tr>
<tr>
<td>CpG 3’ nick</td>
<td>—</td>
<td>CpG B</td>
<td>TG ACG TCA-</td>
<td>Nick 3’ side of dC</td>
<td></td>
</tr>
</tbody>
</table>

The residue/s and positions within the CRE-site that were modified are indicated in the position of modification column, whereas the modification column indicates what actual modifications were made. An unmodified control double stranded oligonucleotide is shown with the CRE site in bold (5’- TAC CAT GCC TTG1 ACG2 TCA GAG AGC ATT CGT AAT CAT GGT-3’). Numerals 1 and 2 indicate the guanine residues that were modified. Some probes had modifications in both top and bottom strands; top and bottom strands are differentiated by T or B in the sequence column. Probes that had nicks between nucleotides or were missing nucleotides were constructed by synthesizing two top or bottom strands that when annealed together with a full length complementary strand resulted in a gap or nick being present at the desired position. These are indicated by | for a nick and ■ for a gap. The 5’ oligonucleotide ends at the nick or gap sites did not have phosphate groups present.
Figure 1. Co-crystal of the TF CREB1 bound as a dimer to the CRE-site of a ds-oligonucleotide. The G1, G2, and C positions of the CpG islet on a single strand of DNA are shown in space fill to indicate proximity to the protein. This structure was obtained from PDB 1DH3.
Figure 2. Purified recombinant CREB1 binds to a 39-mer oligonucleotide containing the 8-mer CRE site first as a monomer and then as a dimer. Purified CREB1 at increasing concentrations was incubated on ice with a 39-mer $^{32}$P-labelled CRE-containing probe (0.5 nM) and resolved on a 5% nondenaturing PAGE. The experiment displayed contains a G/U mispair at the G2 position (CpG). Note the clean separation of monomer and dimer.
Selected DNA modifications alter the overall affinity of CREB1 for the CRE-site both positively and negatively. Graph showing quantitative binding of CREB1 to 39-mer oligonucleotide containing select modifications as a function of CREB1 concentration. ●, control; ▼, G1 6G; ◊, G2 U; ▐, CPG islet with C residues on both upper and lower strands methylated and one G residue as 6G. Each experiment was performed three times with error bars representing ± SE.
Figure 4. The type of DNA modification and its position within the CRE-site has a dramatic effect upon recognition and binding of purified CREB1. (A) Modifications at G1. Modifications at the G1 position are generally detrimental to total CREB1 binding except when "G, a nick on the 3' side of G1, or a G/U mispair is present. (B) Modifications at G2. While a single "G did not affect total binding, a G/U mispair in the G2 position enhanced total CREB1 binding. All other modifications in the G2 position were deleterious to total CREB1 binding. Each experiment was performed independently three times, except for when a nick on the 3' side of G1, or a nick on the 5' side of the C in the CpG islet were present, which were performed twice. In the former, error bars represent ± S.E., in the latter error bars represent ± range and significance was not calculated. P-values are represented as *** ≤ 0.001, ** ≤ 0.01, * ≤ 0.05.
Figure 5. Only a U in the CpG context enhances dimer formation on DNA. (A) Dimer formation after modification at G1. CREB1 ability to form a dimer remained unchanged when a nick in the DNA backbone 5' to the G1 residue, or a G/U mispair were present at the G1 position. Other modifications to the G1 position including the presence of "G resulted in impaired dimer formation by CREB1. (B) Dimer formation after modification of G2. The presence of a G/U mispair at the G2 position enhanced CREB1’s ability to form a dimer on the CRE-site, while all other modifications to the CpG islet inhibited CREB1 dimerization relative to the unmodified control. The amount of DNA.CREB2 as dimer was calculated as part of the kinetic binding studies, using the data from the 50 nM CREB1 concentration. Each experiment was performed independently three times, except for when a nick on the 3' side of G1, or a nick on the 5' side of the C in the CpG islet were present, which were performed twice. In the
figure 6. Some DNA modifications within the CRE-site were especially deleterious to CREB1 binding so that saturation was not achieved under our conditions. (A) K_d values in these cases were estimated from the initial linear binding slope and calculating the concentration of CREB1 at which half concentration would occur. Binding was reduced up to 20-fold with these modifications. (B) Dimerization at 50 nM was severely reduced up to 9-fold for substrates where saturation did not occur. All experiments were performed independently three times except for fully methylated CpG, which was performed twice. In the former, error bars represent ± S.E., in the latter error bars represent ± range and significance was not calculated. P-values are represented as *** ≤ 0.001, ** ≤ 0.01, *≤ 0.05.
Chapter 4. DNA glycosylases and transcription factors compete for damaged transcription factor binding sequences.

In the previous chapter I showed that binding of the TF CREB1 to its consensus sequence could be impacted both negatively and positively by the presence of DNA lesions. We proposed, based on our results, that DNA lesions occurring in the DNA binding sequences of TFs may function in an epigenetic fashion. In fact, we have been recognized as the first group to propose this mechanism (Fleming and Burrows 2017). Since our publication, the Olinski laboratory demonstrated that transcriptionally active DNA contained more °G than repressive heterochromatin and surmised that °G could be epigenetic (Zarakowska et al. 2013), while the Park laboratory determined that °G could occur in vivo at specific sites suggesting that °G could act epigenetically (Park et al. 2016). Likewise, the Burrows laboratory demonstrated that °G occurring in G rich regions of gene promoters in vivo induced transcription. This observation led them to speculate that °G served as an epigenetic regulator (Fleming and Burrows 2017; Fleming et al. 2017). While these groups have generated further evidence in support of our initial proposal, no one has investigated TF binding in the presence of active BER. Since BER is an active and efficient repair pathway, it raised the question as to whether a DNA lesion occurring within a TF’s binding sequence would persist long enough to alter binding efficiencies of the TF, or whether TFs binding the damage could impede repair of the lesion?

The following research paper was published in the journal Journal of Biological Chemistry in 2016 (Moore et al. 2016) and represents the first attempt ever, to our knowledge, to examine the interplay between TFs and BER repair enzymes for TF DNA binding sequences when they contain DNA lesions. This is an extremely important question: if DNA damage can alter the binding kinetics of a TF, misregulation of transcription is a likely consequence. This would be particularly important during times of large scale transcription, such as during embryogenesis when many genes important for development are activated. In this article, we investigated the in vitro kinetics of purified recombinant CREB1 binding to a CRE site when it contained a DNA lesion in the presence of the appropriate DNA glycosylase. We found that DNA glycosylases and CREB1 could compete with each other and the results depended on which had first access to the site. We then investigated the physiological relevance of our in vitro results in vivo using zebrafish embryos.
4.1. Damaged Promoter Sites

Stephen P.G. Moore, Joshua Kruchten, Kimberly J. Toomire, and Phyllis R. Strauss

From the Department of Biology, Northeastern University, Boston MA 02115

Keywords: DNA base excision repair, CREB1 transcription factor, glycosylase/transcription factor competition, uracil DNA glycosylase, 8-oxoguanine glycosylase 1, DNA damage repair.

Background: DNA damage located in transcription factor consensus binding sites and repaired by the base excision repair (BER) pathway modulates ability of transcription factors to recognize their consensus sequences.

Results: BER proteins and transcription factors compete for damaged consensus sequences. Only the G/U mispair results in abnormal development.

Conclusion: DNA repair enzymes can block transcription factor function, while transcription factors can block DNA repair.

Significance: DNA repair enzymes, especially those repairing G/U mispairs located in a consensus sequence, modulate TF function.

4.2. Abstract

Transcriptional regulation is a tightly regulated, vital process. The transcription factor (TF) CREB1 controls ~25% of the mammalian transcriptome by binding the CRE sequence (TGACGTCA). DNA lesions within CRE modulate CREB1 binding negatively and positively. Because appropriate DNA lesions also interact with base excision repair (BER) proteins, we investigated whether CREB1 and repair glycosylases compete with each other. We incubated 39-mer CRE-containing ds oligonucleotides with recombinant CREB1 alone or with UNG2 or OGG1, followed by EMSA. The CpG islet within CRE was modified to contain G/U or 8-oxoG (G)/C mispair. OGG1 and CREB1 reversibly competed for CRE-containing an G/C pair. Also, OGG1 blocked CREB1 from dimerizing by 69%, even when total CREB1 binding was reduced only by 20%-30%. In contrast, bound CREB1 completely prevented access to G/U-containing CRE by UNG2, and, therefore, to BER repair, while UNG2 exposure prevented CREB1 binding. CREB1 dimerization was unaffected by UNG2 when CREB1 bound to CRE, but was greatly reduced by prior UNG2 exposure. To explore physiological
relevance, we microinjected zebrafish embryos with the same oligonucleotides, as a sink for endogenous CREB1. As predicted, microinjection with unmodified or lesion-containing CRE, but not scrambled CRE or scrambled CRE with a G/U mispair, resulted in increased embryo death. However, only the G/U mispair in native CRE resulted in substantial developmental abnormalities, thus confirming the danger of unreppaired G/U mispairs in promoters. In summary, CREB1 and DNA glycosylases compete for damaged CRE in vitro and in vivo, thus blocking DNA repair and resulting in transcriptional misregulation leading to abnormal development.

4.3. Introduction

The study of DNA damage is usually focused on mutation and cellular misregulation that lead to aberrant development or cancer. However, oxidative damage is ongoing and selective DNA oxidation is associated with transcription activation, such as estrogen-dependent and cMYC-dependent transcriptional initiation (Amente et al. 2010; Perillo et al. 2008). Less well known is that DNA damage, specifically substrates and intermediates in the BER pathway, can modulate TF binding itself (Moore et al. 2013). The relative lack of nucleosomes in promoter regions ensures that these regions are especially prone to ongoing oxidative damage or accumulation of uracil/guanine or thymidine/guanine pairing resulting from demethylation immediately after fertilization.

Cyclic AMP response element binding protein (CREB1) is a critical TF that regulates ~25% of the eukaryotic genome (Impey et al. 2004; Zhang et al. 2005). CREB1 binds to the palindromic cAMP-responsive element (CRE) site TGACGTCA or the half CRE site CGTCA, and acts as a required pleiotropic effector (Rudolph et al. 1998). CREB1 is also self-regulating with 13 CRE sites in its own promoter. It is responsible for initial development of the nervous system, memory formation, and neuronal protection (Dworkin et al. 2007; Lonze and Ginty 2002). Misregulation of CREB1 is implicated in a range of congenital and acquired central nervous system disorders, including Alzheimer’s disease and Parkinson’s disease (Sakamoto, Karelina, and Obrietan 2011). Methylation of the central CpG in the consensus sequence prevents binding of CREB1 and consequently prevents transcription of CREB1-dependent genes (Mayr and Montminy 2001). Transcription activation of the central methylated CpG requires demethylation that involves AID/APOBEC or TET enzymes followed by DNA repair via the base excision repair (BER) pathway (Dominguez and Shaknovich 2014; Ramiro and Barreto 2015). Indeed, uracil DNA glycosylase 2 (UNG2) is so important in early zebrafish

---

2 Abbreviations used in this manuscript: AD, Alzheimer’s disease; AP-1, activator protein-1; BER, base excision repair; BSA, bovine serum albumin; cMYC, myelocytomatosis viral oncogene; CREB1, cyclic AMP response element binding protein 1; CRE, CREB1 binding site consensus sequence; ds, double stranded; NFkB, nuclear factor kappa light chain enhancer of activated B cells; OGG1, 8-oxoguanine DNA glycosylase; PD, Parkinson’s disease; ROS, reactive oxygen species; Sp-1, specificity protein-1; TF, transcription factor; UNG1, Uracil DNA glycosylase 1
embryogenesis that knockdown is embryonic lethal, whereas overexpression is sufficient to reduce global DNA methylation (Wu et al. 2014).

Oxidative stress, an imbalance between the production and biological detoxification of reactive oxygen species (ROS), has also been linked with several CNS diseases, including Alzheimer disease and Parkinson disease (Evans, Dizdaroglu, and Cooke 2004). Although a variety of external sources can contribute to ROS levels, endogenous ROS are produced continuously as a byproduct of oxidative phosphorylation (Lindahl 1993). Although ROS damage all cellular components, oxidatively damaged DNA must be repaired quickly or genomic instability results (Evans et al. 2004). Small, non-bulky DNA lesions such as those produced by ROS are recognized and repaired by the BER pathway (reviewed in Ref. 16). Several neurodegenerative diseases display increased levels of oxidative DNA damage, coupled with reduced DNA repair capability (Martin 2008). Furthermore, many neurological disorders, including neurodegeneration, are associated with inherited defects of DNA repair pathways including BER (Caldecott 2008). In a prior publication, we showed that specific DNA lesions within the CRE site that are repaired by BER can both positively and negatively modulate CREB1 recognition and binding to CRE, depending upon the specific site and type of damage (Moore et al. 2013). We therefore proposed that oxidative damage or repair intermediates in BER might modulate TF efficacy, a proposal consistent with that of others (Evans et al. 2004). In this study, we use both in vitro and in vivo methods to investigate whether and how BER enzymes and CREB1 compete for access to CRE site substrates containing DNA lesions typically repaired by BER. We show that CREB1 can block access of DNA glycosylases to DNA lesions and, in turn, DNA glycosylases can interfere with or prevent access of CREB1 to CRE. These results are relevant to studies of developmental biology, neurodegeneration, and aging.

### 4.4. Experimental procedures

**Oligonucleotides:** 39-mer double-stranded (ds) oligonucleotides (see Table 1) containing an unmodified or modified CRE site were sourced either from Midland Certified Reagent Co., Inc. (Midland, TX) or Integrated DNA Technologies (Coralville, IA). Either upper or lower strands were labelled at the 5’ end with \( [\gamma^{32}\mathrm{P}]\mathrm{ATP} \) (Perkin-Elmer, Waltham, MA) by polynucleotide kinase (New England Biolabs, Ipswich, MA) as previously described (Fortier et al. 2009; Strauss et al. 1997) and stored at -20°C until use.

**Source of Proteins:** Purified recombinant human CREB1 protein was the gift of Dr. Jennifer Nyborg (Colorado State University, Fort Collins, CO), and purified recombinant uracil-DNA glycosylase (UNG2) and 8-oxoguanine DNA glycosylase (OGG1) were the gifts of Dr. Nicole Noren Hooten (National Institute of Aging, Baltimore, MD). Purified recombinant mutant UNG2 (His-UNG2-D154N)
was the gift of Dr. Bodil Kavli (Norwegian University of Science and Technology, Trondheim, Norway). AP endonuclease 1 (APEX1) was purified as previously described (Fortier et al. 2009).

**EMSA Assays:** Competition EMSAs were modified from those previously described (Lopez et al. 2007; Moore et al. 2013). Briefly, 8.5-nM purified CREB1 protein was incubated with varying concentrations of purified UNG2 (0 to 64 nM) or OGG1 (0 to 64 nM) protein in a reaction buffer with 5-ng poly(dA-dT), 0.25-µg/µL BSA, and 0.5-nM 32P end-labelled ds oligonucleotide probe, in a final volume of 20 µL. All incubations were carried out at 4°C for 90 min, with variations in the order of addition of protein and oligonucleotide. In the first series CREB1 and a DNA glycosylase were mixed together for 30 min followed by addition of the appropriate oligonucleotide for 60 min. In the second series CREB1 was allowed to bind with the appropriate oligonucleotide for 30 min after which the glycosylase was added for an additional 60 min. In the third series the glycosylase was allowed to bind to the appropriate oligonucleotide for 30 min, after which time CREB1 was added for an additional 60 min. Figure 1 is a representative EMSA showing resolution of CREB monomer and dimer as well as binding of OGG1 in the absence of CREB1. All experiments were performed independently a minimum of three times. In order to demonstrate specificity, 6G containing substrates were incubated with UNG2 and CREB1 and G/U containing substrates were incubated with OGG1 and CREB1. No displacement of CREB1 was observed. Samples without tracking dye were resolved on pre-electrophoresed, native 5% (49:1 acrylamide:bisacrylamide) gels in buffer composed of 40 mM Tris-HCl (pH 8.5), 306 mM Tris-glycine and 0.1% Nonidet P-40. Electrophoresis was performed at room temperature at 90 V for 2.5 h.

**Analysis:** Distribution of isotope was visualized on a PhosphorImager screen with a GE Storm™ PhosphorImager™ (GE Healthcare, Little Chalfont, United Kingdom); data were quantified with Molecular Dynamics ImageQuant (Molecular Dynamics, Sunnyvale, CA) (Fortier et al. 2009). CREB1 binding was determined by the equation:

\[
\frac{[\text{Total bound DNA}]}{[\text{Total unbound DNA}] + [\text{Total DNA bound to CREB1}] } \times \text{CREB1}
\]

**Cleavage Activity in the Presence and Absence of CREB1:** Purified recombinant CREB1 (8.5 nM) was added to UNG2 (85 nM) or OGG1 (85 nM) in reaction buffer with 5-ng poly(dA-dT), and 0.25-µg/µL BSA. The 32P end-labelled ds oligonucleotide probe (Table 1) was added and the reaction incubated on ice for 1 hr in a final volume of 20 µL. DNA was isolated by phenol:chloroform extraction as previously described (Fortier et al. 2009) and resuspended in 5 µL of buffer containing 50-mM HEPES and 0.1-mM EDTA. Purified APEX1 protein (10 nM) was added to 2 µL of resuspended DNA samples in a final reaction volume of 5 µL and incubated for 30 min at room temperature. The cleavage assay was stopped by addition of formamide loading buffer (80% formamide, 2% 0.5 M EDTA, and 10% xylene
cyanol/bromophenol blue). Cleavage products were resolved by denaturing gel electrophoresis as previously described (Fortier et al. 2009).

**Zebrafish Culture and Decoy CRE site Oligonucleotide Microinjections:** Wild-type zebrafish (*Danio rerio*), purchased from Aquatica BioTech (Sun City Center, FL), were maintained and bred using standard protocols in accordance with approved Northeastern University IACUC policies as previously described (Wang et al. 2006). Embryos were microinjected directly into the yolk prior to the 4-cell stage with 2 nL of 39-mer (0.18 nM) ds oligonucleotide substrate (Table 1) in Danieau buffer. Phenol red (0.05%) was used as an injection indicator (Pei et al. 2011). Phenol red (0.05%) in Danieau buffer served as the control. Injected embryos were maintained at 29°C until reaching desired developmental stages and fixed in 4% paraformaldehyde (Nusslein-Volhard and Dahm 2002). Embryos were examined and photographed using a Leica MZ16FA stereomicroscope (Leica, Wetzlar, Germany).

**Statistical analyses:** Data in Figures 2 and 3 were analyzed by mixed model analyses of covariance (ANCOVA) using SPSS 19.0 (IBM, Armonk, NY). Prior to analysis, data were transformed using a natural logarithmic transformation to obviate heteroscedasticity. Homogeneity of variance was verified by Levene’s test. Tukey’s post-hoc analysis was conducted on the factor in the absence of a significant interaction between factor and covariate. When a significant interaction was present between the factor and covariate, simple linear regression was performed.

Data in Figures 4-7 were analyzed by one-way ANOVA using SPSS 19.0. Before analysis was conducted, a Shapiro-Wilk’s test for normality and a Levene’s test for equal variance were conducted. When a significant difference was detected in any analysis, Tukey’s post-hoc analysis was performed to identify significantly different means. Significance for all statistical tests was set *a priori* at 0.05.

**4.5. Results**

*Lesion-specific DNA glycosylases compete with CREB1.* CRE has a CpG islet that, when oxidized or contains a G/U mispair, either negatively or positively alters binding of CREB1, depending on the lesion (Moore et al. 2013). These lesions are repaired by BER. Because glycosylases are the primary entry point to the BER pathway, we asked here whether and how the two major glycosylases, OGG1 and UNG2, might affect CREB1 access to a CRE site that contains 5-G or U in the CpG islet portion of the CRE site, respectively. We also asked whether CREB1 might block access to these same glycosylases.

When the CRE site contained an 5-G lesion within the CpG position, total CREB1 binding to the CRE site was reduced by a maximum of 34% ± 9% (S.E.,3) in the presence of OGG1 (Figure 2A-C). The decrease in CREB1 binding depended on the OGG1 concentration and the order in which OGG1 was exposed to the substrate. The ability of OGG1 to displace previously bound CREB1 was somewhat less efficient than its ability to prevent CREB1 from binding (p = 0.035, n=3) (Figure 2A&B).
Therefore, the presence of this glycosylase and the order of exposure to CRE had a small but substantial effect on total binding.

When the CRE site contained a U residue in the CpG islet, overall binding of CREB1 in the presence of UNG2 was reduced to a far greater extent by a maximum of 60% ± 2% (S.E.,3) (Figure 2D-F). However, the decrease depended strongly on the order in which the proteins were added. Bound CREB1 completely blocked access to UNG2 regardless of the presence of 8-fold excess of UNG2 (Figure 2E). In contrast, total CREB1 binding was extremely sensitive to prior exposure of the CRE site to UNG2 (60% ± 2% inhibition decreases relative to control [S.E.,3]) (Figure 2D) and somewhat sensitive to simultaneous addition of UNG2 and CREB1 (23% ± 6% inhibition relative to control [S.E.,3]) (Figure 2F). In short, unlike the case with °G/C the order of addition made a substantial difference to total binding of CREB1 when the CRE site contained a G/U mispair in the CpG islet of CRE.

Dimerization of CREB1 on its CRE site is strongly impacted by lesion specific DNA glycosylases. CREB1 is transcriptionally active only as a dimer, which forms on the CREB.DNA monomer complex (Shaywitz and Greenberg 1999). The ability of CREB1 to dimerize on a CRE site substrate containing an °G lesion was heavily influenced by the presence of OGG1 regardless of the order of addition (Figure 3A-C). Thus, the presence of OGG1 prevented the addition of a second molecule of CREB1 to the °G-containing CRE site and displaced previously bound CREB1. We presume that the second CREB1 molecule was unable to bind efficiently because OGG1 binds its substrate in a flipped-out configuration that distorts the CRE site and is slow to dissociate (Chen et al. 2002). In fact, we could detect bound OGG1 on our gels, indicating that OGG1 remained associated with the substrate and physically blocked the second molecule of CREB1 from forming a dimer (Figure 1). Consequently, the presence of the repair glycosylase was likely to prevent CREB1 from forming the dimer necessary to recruit TFs required to activate RNA polymerase II.

The presence of UNG2 also negatively affected CREB1’s ability to form a dimer on a G/U mispair-containing CRE site substrate. However, the competition kinetics were very different from those involving OGG1. The reduction depended on the order that the proteins were exposed to the substrate (Figure 3D-F). The effect was both more selective and profound than that on an °G-containing substrate. When UNG2 had first or equal exposure to the substrate relative to CREB1, the ability of CREB1 to form a dimer was markedly reduced up to 77% (Figure 3D&F). In contrast, if the CRE-containing substrate was allowed to bind CREB1 first, CREB1’s ability to form a dimer was not altered relative to control regardless of an 8-fold excess of UNG2 (Figure 3E). In other words, UNG2 was capable of preventing the recruitment of a second molecule of CREB1 to the G/U-containing CRE site, but, if CREB1 had already formed a dimer, then UNG2 could not displace the TF from the U-containing CRE site.
Catalytically inactive UNG2 at any concentration had no effect on the ability of CREB1 to bind to a G/U mispair-containing CRE site, whether measured as the total amount of CREB1 bound or CREB1 dimer formation (Figure 4). Therefore, unlike exposure to OGG1 where we were unable to detect glycosylase activity, the reduction in CREB1 binding and dimerization to a CRE site containing a G/U-mispair was due to excision of U through the catalytic action of UNG2 and not the physical presence of the enzyme blocking access. This result is consistent with the known rapid turnover of UNG2 (Kavli et al. 2002).

CREB1 impedes UNG2 from excising a U residue present within a CRE site. We then asked whether CREB1 binding blocked UNG2 enzymatic activity. Incubation of UNG2 alone with the oligomer containing the G/U mispair at the CRE site resulted in loss of 61% ± 5% (S.E.,3) of the U residues (Figure 5). When CREB1 and UNG2 were co-incubated with this same substrate, UNG2’s ability to remove U was reduced to 30% ± 3% (S.E.,3). Thus, when CREB1 binds to a CRE site containing a G/U-mispair, it blocks UNG2 from removing the U residue, which is the first step in BER. Furthermore, these data indicate that the presence of the U residue stabilized CREB1 binding because if CREB1 were constantly associating with and dissociating from the CRE site, then there would be renewed opportunity for UNG2 with its rapid turnover (Kavli et al. 2002) to remove the U residue. Therefore, these data are consistent with prior results indicating that the presence of U in the CpG site of CRE stabilizes CREB1 binding to CRE (Moore et al. 2013).

°G and U in CREB1 CRE affect early development. To establish the physiological relevance of the in vitro results, we explored whether the presence of DNA lesions repaired by BER in CRE affects normal embryonic development. In order for normal development to proceed, embryos need to activate select genes at the correct time and location. Because ~25% of the vertebrate genome depends on functional CREB1, we hypothesized that exposure of developing zebrafish to a modified CRE site would result in abnormalities, depending on the affinity of CREB1 for the lesion. To that end, we microinjected the same oligonucleotides used in our in vitro studies (Table 1) into 2-4–cell-stage zebrafish embryos and harvested them at 24 hpf. Embryos at this stage contain ample OGG1 and UNG2 (Fortier et al. 2009) and increasing amounts of CREB1 (Pei et al. 2011). Because CREB1 binds more strongly to a U-containing CRE site than to an unmodified CRE site (Moore et al. 2013), we predicted that the developing embryo would be sensitive to microinjection of an oligonucleotide containing a CRE site, but would be particularly sensitive to one containing a G/U mispair within the CpG of the CRE site. As expected, survival of zebrafish embryos to 24 hpf was reduced after microinjection of CRE-site containing oligonucleotides with or without BER-repairable modifications (Figure 6). To explore whether the decreased survival of microinjected embryos was due to the presence of a modified or unmodified CRE site, and not simply oligonucleotide toxicity, we also microinjected embryos with an oligonucleotide in which the CRE site sequence was scrambled. Embryos microinjected with the scrambled sequence did not display any difference in survival from that of control (Figure 6). Only
microinjection of CRE site oligonucleotides containing the G/U mispair resulted in more developmental defects in 24 hpf embryos (Figure 7). These included curvature of anterior-posterior axis, decreased axis length, and abnormal yolk extension formation (Figure 8). When G/C in the scrambled site was replaced with a G/U mispair, there was no increase in developmental failure or defects in comparison with scrambled CRE. These data rule out the possibility that UNG2 was stably recruited to a random G/U mispair. In other words, CREB1 had a higher affinity for CRE containing a G/U mispair not only in the test tube but also in the developing embryo. This demonstrated that CRE containing a G/U mispair is capable of soaking up more endogenous CREB1 than the undamaged CRE so that proper development failed in a significant percentage of embryos.

4.6. Discussion

Here we demonstrate that a TF can alter access to repair of its consensus sequence containing damaged DNA; conversely, the presence of repair enzymes can block access to the TF. That is, the TF CREB1 competed with DNA repair glycosylases for its cognate CRE site when that sequence contained a DNA lesion in the CpG islet and vice versa. However, the type of repair glycosylase in addition to the lesion affected CREB1’s binding ability. When CREB1 was in competition with the repair glycosylase UNG2 for a CRE site that included a G/U mispair in the CpG islet, CREB1 entirely prevented UNG2 recognition and removal of the U lesion, while UNG2 removal of a U residue prevented CREB1 binding. In contrast, the repair glycosylase OGG1 competed with CREB1 for an 8G containing CRE site so that CREB1 binding and dimerization were reduced.

As UNG2 was largely incapable of displacing CREB1 after the TF had bound to a G/U mispair, these data are consistent with our previous results demonstrating that CREB1 binding and dimerization are enhanced when CRE contains a G/U mispair (Moore et al. 2013). However, when the G/U-containing CRE site was exposed to enzymatically active (but not inactive) UNG2 before exposure to CREB1, binding and dimerization were reduced in a concentration-dependent manner. This is unsurprising considering that UNG2 is extremely efficient in processing U lesions when they are mispaired with G, leading to a rapid and robust conversion of U to an AP-site product to which CREB1 binds poorly (Kavli et al. 2002; Moore et al. 2013). Therefore, the TF interfered with entry into the repair pathway; and entry into the repair pathway was diminished by the binding and activation of the TF.

In contrast to results with UNG2, CREB1 binding and particularly dimerization were reduced as OGG1 concentration increased, regardless of the order in which CREB1 and OGG1 were added to the 8G lesion-containing CRE site substrate. Moreover, OGG1 was capable of displacing CREB1 from CRE containing an 8G lesion, despite the fact that 8G within the CpG islet of the CRE site by itself did not affect total CREB1 binding or dimerization (Moore et al. 2013). Thus, OGG1 but not UNG2 was
apparently capable of displacing CREB1 that had already bound to a CRE site. These results are consistent with the fact that OGG1 has a very low turnover number (Hill et al. 2001; Sidorenko et al. 2009; Zharkov et al. 2000). This result suggests that OGG1-mediated reduction in CREB1 binding and dimerization was due to the physical presence of OGG1 blocking access to the CRE site itself rather than the existence of any AP-site product that might have formed.

As noted above, CREB1 is a pleiotropic effector responsible for controlling ~25% of the eukaryotic genome. During embryonic development, CREB1 regulates many genes and is essential for gastrulation (Sundaram et al. 2003). Decoy CRE site oligonucleotides are known to compete for CREB1 protein in vivo (Hara et al. 2003; Lee et al. 2000; Park et al. 1999). Therefore, we microinjected CRE site oligonucleotides with or without a BER-repairable DNA lesion into zebrafish embryos to examine the in vivo effect of the lesion on CREB1 binding. While all CRE-containing oligonucleotides acted as a sink and resulted in 30-40% failure to develop, only microinjection of those containing a G/U mispair resulted in developmental defects. These results are consistent with the critical role that CREB1 plays in early embryonic development and the ability of lesions repaired by BER to act in an epigenetic fashion. Therefore, the presence of a G/U mispair not only results in CREB1 binding its consensus sequence more tightly in vivo, but also, once bound, it blocks access to the repair pathway, resulting in deleterious physiological consequences.

Most research conducted on the effects of DNA lesions, particularly oxidative damage or the presence of uracil, has focused on mutation within transcribed genomic regions (Cooke et al. 2003; Evans et al. 2004). However, several recent studies suggest that DNA lesions within the cognate sequences of activator protein 1 (AP-1), specificity protein 1 (Sp1) and the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB) complex can also alter TF binding (Abraham and Brooks 2011; Ghosh and Mitchell 1999; Ramon et al. 1999), indicating that DNA damage within promoter regions may have serious repercussions for cellular physiology. In addition, binding of the p50 subunit to the NF-kB sequence has been shown to shield 5′G lesions from recognition and repair by OGG1 and Fapy glycosylase (Fpg) suggesting that oxidative DNA damage may be able to persist long enough to affect gene expression (Hailer-Morrison et al. 2003). Furthermore, oxidative DNA damage to genomic promoter sites may be a common occurrence. For example, lysine specific demethylase 1 (LSD1) generates hydrogen peroxide primarily at promoters as a byproduct of demethylation of Histone H3 during the initiation of active transcription (Amente et al. 2010; Li, Braganza, and Sobol 2013). Peroxide in such close proximity to DNA is likely to result in oxidation.

The promoters for CREB1 itself and its modulating proteins including CREB-binding protein (CBP), cAMP response element modulator (CREM), transducer of regulated CREB-binding proteins 1 and 3 (TORC1 & 3), and activating transcription factor (ATF) all contain CRE sites that could be subject to oxidative damage. AP-1 is a homodimer or heterodimer of FBJ murine osteosarcoma viral oncogene
homolog (FOS) and Jun proto-oncogene (JUN), both of which contain multiple CRE sites, as does Sp-1. In addition, cMYC, an important TF involved with metastasis and tumorigenesis, also contains a CRE site in its promoter, so that oxidation could alter MYC-dependent transcription and prevent DNA repair. While NF-κB itself does not contain a CRE site, its promoter contains four FOS sites and one JUN site, which render it sensitive to BER-repaired damage. Thus, DNA damage to promoters, particularly CRE sites, can have profound, far-reaching consequences.

Our studies not only delineate alterations in both binding and displacement of CREB1 but also demonstrate the physiological relevance of unrepaired lesions normally repaired by BER. Small changes to gene expression of a major regulator like CREB1 can have major physiological ramifications. Gene expression is normally tightly regulated, but perturbations in expression are amplified either positively or negatively with profound consequences for cellular development and physiology. These results indicate that DNA repair enzymes can compete effectively with TFs and vice versa, further complicating the cellular response to damage repaired through BER.

4.7. Acknowledgements

We thank Dr. Jennifer Nyborg (Colorado State University, Fort Collins, CO), for purified recombinant CREB1 protein, Dr. Nicole Noren Hooten (National Institute of Aging, Baltimore, MD) for purified recombinant uracil-DNA glycosylase (UNG2) and 8-oxoguanine DNA glycosylase (OGG1), and Dr. Bodil Kavli (Norwegian University of Science and Technology, Trondheim, Norway) for purified recombinant mutant UNG2 (His-UNG2-D154N). We also thank Dr. Tarik Gouhier (Marine Sciences Laboratory, Northeastern University, Boston, MA) for advice with statistical analyses and Mr. Benjamin Snow for technical review of this manuscript.

This work was supported by the G. Harold and Leila Y. Mathers Foundation, Aid for Cancer Research and funds from Northeastern University.

4.8. Conflict of interest

The authors acknowledge that there is no conflict of interest.
Table 1. Modifications made to CRE site used in this study. The Modification column shows the actual residue modification made to the CRE site. The residues and positions modified within the CRE site are denoted in the Location column. Modifications were made to, or opposite, the G2 residue within the CpG islet. The Strand column indicates the modified strand. T = top strand; B = bottom strand. The Sequence column shows the complete CRE site sequence including the modified residue(s). The full length unmodified 39-mer CRE oligonucleotide is shown with the CRE site in bold in the bottom row. Numerals 1 and 2 indicate the guanine residues within the top strand of the CRE site. All modifications in this study were located at the G2 position or its complement. NA, non applicable.

<table>
<thead>
<tr>
<th>Modification</th>
<th>Location</th>
<th>Strand</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRE</td>
<td>NA</td>
<td>T</td>
<td>T: 5’-TGAC G TCA-1’</td>
</tr>
<tr>
<td>8-oxoguanine (°G)</td>
<td>G2 (CpG)</td>
<td>T</td>
<td>B: 5’-TGAC G TCA-1’</td>
</tr>
<tr>
<td>G/U mispair (G/U)</td>
<td>G2 (CpG)</td>
<td>B</td>
<td>T: 5’-TGAC G TCA-1’</td>
</tr>
<tr>
<td>Scrambled CRE</td>
<td>Entire CRE-site</td>
<td>Both</td>
<td>T: 5’-TGAC G TCA-1’</td>
</tr>
<tr>
<td>Scrambled CRE (G/U)</td>
<td>Entire CRE-site U in bottom strand</td>
<td>Both</td>
<td>T: 5’-CA GTA UTG-1’</td>
</tr>
<tr>
<td>Full oligonucleotide sequence</td>
<td></td>
<td></td>
<td>5’-TGAC G TCA-1’, GAG AGC ATT CGT AAT CAT GGT-3’</td>
</tr>
</tbody>
</table>

Figure 1. Monomer and dimer CREB1 binding to the CRE site contained within a ds 39-mer oligonucleotide are cleanly resolved by EMSA and are differentially impacted by UNG and OGG1. In these representative gel images, UNG2 and CREB1, or OGG1 and CREB1 were mixed together followed by addition of the labelled oligonucleotide (oligo) and incubated for 90 min at 4°C.
prior to EMSA analysis. Note detectable OGG1 binding to the °G-containing oligonucleotide. Apparent reduction in unbound substrate at low concentrations of OGG1 is due to direct binding by the repair glycosylase. On the other hand, binding of UNG2 to U-containing CRE site is not sufficiently stable to detect in the absence of CREB1. Typical gels obtained when UNG2 is added to substrate containing a G/U mispair (left) or when OGG1 is added to substrate containing a °G/C pair (right) in the CpG site of CRE.

Figure 2. Both OGG1 and UNG2 impede CREB1 from binding to a CRE site substrate when the CRE site contains the specific DNA lesion recognized by each glycosylase. Interference depended on the lesion, concentration of glycosylase, and order of exposure. A-C, OGG1 in increasing concentrations competes with CREB1 for an °G containing substrate. A, OGG1 added before CREB1; B, CREB1 added before OGG1; C, OGG1 and CREB1 added together. Total bound CREB1 measured in the control experiments was 0.15 ± 0.01 nM (n = 9). D-F, CREB1 and UNG2 each prevent binding of the other. D, UNG2 added before CREB1; E, CREB1 added before UNG2; F, UNG2 and CREB1 added together. Total CREB1 bound in these control experiments was 0.35 ± 0.01 nM (n = 9).
Figure 3. CREB1 dimerization is heavily impacted by the presence of DNA repair glycosylases. As the concentration of glycosylase was increased, CREB1 dimerization decreased, except for when the G/U mispair-containing CRE site had prior exposure to CREB1. A-C, regardless of whether the substrate was exposed to OGG1 before or after CREB1, total CREB1 dimerization was reduced up to 69% ± 2% (n = 3) relative to control. A, OGG1 added before CREB1; B, CREB1 added before OGG1; C, OGG1 and UNG2 added together. CREB1 dimerization was significantly reduced when OGG1 and CREB1 had concurrent access to the substrate compared to when OGG1 had prior access to the substrate (p=0.012, 3). CREB1 dimer in the control experiments was 0.02 nM ± 0.003 nM (n = 9). D-F, Dimerization of CREB1 on a G/U mispair containing CRE site. D, UNG2 added before CREB1; E, CREB1 added before UNG2; F, UNG2 and CREB1 added together. Measured CREB1 dimer in control experiments was 0.14 ± 0.01 nM (n = 9).
Figure 4. Inhibition of CREB1 binding by UNG2 depends upon UNG2 enzymatic activity. Catalytically inactive UNG2 (His-UNG2-D154N) had no effect on total CREB1 binding or dimerization on a CRE site substrate containing a G/U mispair. Wild type or mutant UNG2 was incubated with the substrate prior to addition of CREB1. □, total bound CREB1 compared to control; ■, total CREB1 as a dimer compared to control. i = inactive mutant UNG2, a = active UNG2; numbers in parentheses represent UNG2 concentration. Error bars represent ± S.E., *** p ≤ 0.001 relative to control.
Figure 5. CREB1 interferes with UNG2 enzymatic activity when the CRE site contains a G/U mispair in the CpG islet. In the presence of 85 nM UNG2, 61% ± 5% (n = 3) of U residues within the CRE site were cleaved. However, the presence of 8.5 nM CREB1 together with UNG2 reduced cleavage by 50% ± 3% (n = 3).
Figure 6. Microinjection of oligonucleotide substrates results in reduced survival of 24 hpf zebrafish embryos. Substrates consisted of 39-mer ds oligonucleotides encompassing an unmodified or modified CRE site (Table 1). Control consisted of 0.05% phenol red in 1x Danieau buffer. The presence of a U residue in CRE reduced survival by 42% ± 5% (n = 7), an O-G reduced survival by 35% ± 5% (n = 5), and unmodified CRE reduced survival by 34% ± 4% (n = 13) of control values. There were no differences in survival among those three substrates. The scrambled CRE site substrate failed to reduce survival. Error bars represent ± S.E., *** p ≤ 0.001 relative to control. NS, not significant.
Figure 7. Only the G/U mispair within the CpG islet of the CRE site causes developmental defects in surviving 24 hpf zebrafish embryos. Embryos were microinjected with the indicated 39-mer ds oligonucleotide substrate (Table 1). Viability and developmental progress were determined at 24 hpf. No differences from control were noted except for those microinjected with the CRE site containing a G/U mispair (31% ± 11%, n = 8). Control consisted of phenol red in 1x Danieau buffer. *Error bars* represent ± S.E., **p ≤ 0.01.
Figure 8. Developmental defects occur in zebrafish embryos microinjected with G/U mispair containing CRE site oligonucleotide. A, control zebrafish embryos microinjected with 1x Danieau buffer and 0.05% phenol red dye displayed normal development; the lone exception (lower left in this image) displays curvature of the anterior-posterior axis. B, zebrafish embryos at 24 hpf displayed developmental defects of varying degrees after microinjection with a decoy CRE site oligonucleotide containing a G/U mispair in the CpG islet. Defects included kinked tail ends, curvature of anterior-posterior axis, and abnormal yolk extension. * denotes abnormal development.
Chapter 5. General discussion and future directions of study.

The work described in the previous chapters lays the foundation for future studies to unravel the role of BER during embryogenesis. A key concern is why knockout or knockdown of many components of the BER pathway are embryonic lethal, while differentiated, adult cells remain viable without these same BER participants. This fact alone indicates that something important occurs during embryogenesis that requires the presence of functional BER. We, in the Strauss lab, have long been interested in the effect that loss of the BER enzyme Apex1 has on zebrafish embryogenesis, and the molecular changes that underpin that loss. Here, I have shown that during normal zebrafish embryogenesis the amount of DNA damage occurring to embryos increases as the embryo activates its own genome during entry to the MBT. Damage remains high as the embryo transitions through the MBT followed by a return to pre-MBT levels by 6 hpf. This observation leads me to hypothesize that during zygotic genome activation a process occurs that generates DNA damage requiring BER for normal development to proceed. When Apex1 protein level is reduced by MO, the same trend is evident, only the damage is much more elevated. This implies that the bulk of the DNA damage occurring in the developing embryos is comprised of BER substrates, that is, the lesions are small, non-bulky lesions such as those generated through oxidative DNA damage or spontaneous deamination of cytosine. Furthermore, by quantifying AP sites, I have shown that there is ~2-fold increase in the number present in the Apex1 K/D embryos in comparison to that in normal embryos at the same stage. This indicates that BER is initiated in both normal and K/D embryos, but the process stalls after excision of the lesion due to the lack of Apex1 protein resulting in AP site accumulation. This further supports the assertion that the DNA damage produced during zebrafish embryogenesis is largely comprised of BER substrates.

The traditional view of DNA damage, particularly the small, non-bulky BER repairable lesions such as oxidative damage arising through ROS, is that it is randomly distributed across the genome (Nikitaki et al. 2015), and detrimental to cellular physiology (Fleming and Burrows 2017; Fleming et al. 2017). Recent findings suggest that the situation may not be so simple and that DNA damage may, in some cases, be required for gene activation. For example, oxidative DNA damage was shown to activate estrogen-induced gene expression with DNA damage placed specifically within estrogen-receptive gene promoters. This process involves a mechanism whereby demethylation of histone 3 (H3K9me2) by lysine demethylase 1 (LSD1) produces H2O2 that then reacts with DNA and oxidizes guanine. The resultant DNA lesion recruits OGG1 and topoisomerase IIβ (TOPIIβ) causing the conformational change to DNA required for gene activation (Perillo et al. 2008). Likewise, oxidative damage to the vascular endothelial growth factor (VEGF) promoter enhances binding of hypoxia-inducible factor 1α (HIF1α) with the subsequent assembly of BER components activating gene expression (Gillespie, Pastukh, and Ruchko 2009, 2010; Pastukh et al. 2015). A similar process occurs in the promoters of tumor necrosis factor α (TNF-α) (Pan et al. 2016) and sirtuin 1 (SIRT1) (Antoniai...
et al. 2014). Controlled induction of gene transcription then requires activation of BER through OGG1 and APEX1. The Burrow Lab, investigating the guanine-rich VEGF promoter, showed that OGG1 removes the damaged base leaving an AP site. An AP site changes the thermodynamics of the sequence in which it is embedded, allowing it to adopt a G-quadruplex structure. APEX1 binds the AP site but the reaction kinetics are reduced by the G-quadruplex structure causing sustained binding. APEX1 could then recruit other factors to induce gene transcription (Fleming and Burrows 2017; Fleming et al. 2017). These results show a distinct coupling between DNA damage, acting in a secondary signalling capacity, and DNA repair. Thus, while DNA repair ultimately functions to maintain genome integrity, this process may have an ancillary, but equally important function: when damage occurs in specific genomic regions, DNA repair may contribute to transcriptional activation.

During embryogenesis, a well-coordinated, complex process of genome activation occurs. The timing of and types of genes being activated is crucial as incorrect regulation of this process can lead to developmental arrest. Embryonic genome activation involves many disparate processes, including changes in histone variant expression, histone post translation modifications, changes to chromatin structure, and regulation of transcription factors (Latham and Schultz 2001). Another potential process of genomic activation control involves methylation/demethylation of cytosine in CpG dinucleotides. Due to the process of active DNA demethylation generating potential DNA damage via the BER pathway, I hypothesized that DNA demethylation occurring at CpG islands in the promoters of developmentally important genes would result in elevated DNA damage, and a reduced repair capacity by loss of Apex1 protein would result in even higher levels of DNA damage. Since Creb1 loss is concurrent with that of Apex1 (Snow 2012), I hypothesized that during entry to the MBT active DNA demethylation of the CpG island in the promoters of both zebrafish creb1 genes leads to gene activation. Furthermore, the CpG islands of both creb1 promoters contain half CRE sites (CGTCA), four for creb1a and three for creb1b. Since the half CRE sites feature CpG dinucleotides, it is highly likely that methylation/demethylation of the CpG island includes the half CRE site CpG dinucleotide. When the ability to repair the resultant damage in the CpG island is reduced by loss of Apex1 protein, there would be a corresponding loss of Creb1, and eventually a loss of developmental co-ordination. I have shown that there is a strong inverse correlation between DNA damage in zebrafish embryos and DNA methylation levels. As DNA methylation levels decrease, DNA damage increases and vice versa in both Apex1 K/D and controls, a result suggesting that there is a link between active DNA demethylation and DNA damage during zebrafish embryogenesis. I conducted a qPCR assay for DNA damage across the CpG island of both zebrafish creb1 genes, but did not detect any difference in the number of lesions between the treatments and control. It is possible that most DNA damage in the creb1 promoters occurs at the MBT which we could not sample. Future work should investigate DNA damage in specific sequences before and during the MBT. Additionally, bisulfite DNA conversion and sequencing studies could be undertaken in control and Apex K/D embryos at various embryonic development times.
spanning the MBT to build a comprehensive picture of the changes to the promoters of CpG island containing genes compared to those without CpG islands.

Despite the loss of BER capacity in the Apex1 K/D embryos, DNA damage levels at 8 hpf had returned to the pre-MBT level and were at the same level as the 8 hpf control embryos, even though Apex1 protein levels have dropped to 46% of that in controls (Snow 2012). These data indicate that DNA repair is occurring and while there may still be sufficient Apex1 to suffice, it is not possible to rule out repair through a different pathway. Since 5mC occurs in CpG dinucleotides, unprocessed AP sites could accumulate at those dinucleotides and potentially result in double-strand DNA breaks (DSB). By 6 hpf, monoubiquitination of lysine 120 on histone H2B (H2Bub1) increases by 7-fold in Apex1 K/D embryos (Snow 2012), though I was unable to replicate this result. Double-strand breaks induce modification of H2B to H2Bub1 through ATM-dependent phosphorylation of the RNF20/40 E3 ubiquitin ligases (Moyal et al. 2011). Thus, the increase in H2Bub1 in Apex1 K/D embryos could indicate that a DSB repair pathway had been activated. Activation of DSB repair occurs during active DNA demethylation, though the reasons are currently unclear (Wossidlo et al. 2010); it is possible that DSB repair operates here as a backup to BER during demethylation in case DSBs occur at CpG dinucleotides. Future work should further investigate the possibility that a different repair pathway, particularly DSB repair, is activated in Apex1 K/D embryos.

Further investigation of H2Bub1 during zebrafish embryogenesis is also warranted. Despite the fact that I could not replicate our original findings, when I utilized a method to isolate chromatin and strip-off histones I did find that H2Bub1 levels appeared elevated compared to controls. However, the results were inconsistent among replicates. I also found that usp22, responsible for deubiquitinating H2Bub1 (Lee et al. 2006), transcript increased 7-fold at 6 hpf in Apex1 K/D embryos (p = 0.0003) matching that of H2Bub1 modified H2B (data not shown). Another potential explanation as to why repair has occurred in Apex1 K/D embryos by 8 hpf is that replication is still ongoing. However, upon encountering DNA damage normal replicative polymerases are replaced by Y-family polymerases, which can bypass the DNA damage (Yang 2014). DNA damage bypass by the Y-family polymerases leads to increased nucleotide misincorporation, frameshifts, and blocked replication, all of which result in increased genomic mutation (Liu et al. 2016). If DNA damage is occurring predominately in the promoter regions of developmentally important genes via active DNA demethylation, continued replication could lead to genomic sequence alterations that would permanently alter expression of those genes. Since cell fate determination in zebrafish occurs at gastrulation (~5.3 hpf) (Solnica-Krezel et al. 1996), any mutation to the genome by that time would be passed down to all daughter cells. If any of the affected genes were required during anlage and/or organogenesis, developmental processes would be impaired. This may cause incorrect development of various organs, and could explain the brain and heart abnormalities evident in Apex1 K/D embryos at 24 hpf (Wang et al. 2006).
I have shown that when the CpG dinucleotide of the CRE site, the CREB1 recognition sequence, is hemi-methylated or fully methylated, recombinant CREB1 cannot bind (Moore et al. 2013). The methylation status of the CRE site appears to be dynamic and varies depending on tissue type and developmental times (Zhang et al. 2005). It is likely that transcriptional control by transcription factors is mediated by the methylation status of their binding sites. For example, the binding of many transcription factors, including myelocytomatosis viral oncogene (c-MYC), nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), metal-regulatory transcription factor 1 (MTF-1), and activating enhancer-binding protein 2-alpha (AP-2), are blocked by methylation of CpG dinucleotides within their recognition sequences (Baron 2012). Thus, methylation/demethylation of a CpG dinucleotide in a TF’s recognition sequence could act to control transcriptional activation and repression. When an AP site is present in the CpG dinucleotide of the CRE site, recombinant CREB1 binding is also significantly reduced (Moore et al. 2013). This highlights just how critical the requirement is for ongoing, fully functional BER during embryogenesis. Since BER is required during active DNA demethylation, an absence or reduction in BER capacity will potentially result in substantial changes to developmental processes if TF recognition sites are affected.

I have shown that when uracil is present in the CpG dinucleotide of the CRE site, recombinant CREB1 binding is enhanced significantly (Moore et al. 2013). Furthermore, BER cannot displace CREB1 once it has bound to the uracil containing CRE site, and these results are relevant in vivo (Moore et al. 2016). This raises the intriguing question as to whether uracil might act as a transcriptional activator? Can uracil function to increase transcription by increased CREB1 binding, and could this mechanism occur during embryogenesis? An initial thought was that if DNA demethylation occurs at the level of a single CpG dinucleotide within a TF recognition site during embryogenesis as a means of controlling gene expression, then replacing 5mC with uracil could act to bolster or prolong expression of that gene at a critical time in development. Currently, active DNA demethylation is thought to result in iterative modification of 5mC to T. Thymine and a number of the intermediates generated by modification of 5mC are substrates for Tdg to initiate BER. However, developing zebrafish embryos do not possess tdg transcript (Wu et al. 2014) or protein (Strauss lab, unpublished data); likewise, mouse zygotes do not express Tdg (Xue et al. 2016) even though Tdg is embryonic lethal. It is believed that zebrafish and mouse embryos may use the monofunctional uracil DNA glycosylase (Ung) in active DNA demethylation (Wu et al. 2014; Xue et al. 2016). At the present time, no pathway is known that can replace 5mC with uracil, and Ung is not believed to recognize any of the intermediates of 5mC modification (Fritz and Papavasiliou 2010). Future studies should focus on unravelling the apparent discrepancy between what glycosylases are present during embryogenesis and the BER substrates that
are generated by active DNA demethylation. Once this issue is resolved it will be possible to determine the plausibility that uracil might enhance direct gene expression within the appropriate context.

Furthermore, it is possible that the increase in DNA damage is due to direct oxidative DNA damage, rather than a derivative of DNA demethylation. However, during early embryogenesis in teleosts most energy production is through anaerobic glycolysis, rather than oxidative phosphorylation (Boulekbache 1981). This pathway is often utilized during rapid cell division, such as that occurring before the MBT (Wales, Martin, and Leese 1995). In zebrafish embryos, early total oxygen consumption increases with age, but does not correlate with an increase in mass. Prior to 3 hpf no mitochondrial respiration occurs, but mitochondrial respiration increases from ~3 to 12 hpf along with mitochondrial biogenesis (Stackley et al. 2011). Thus, there is a potential for oxidative DNA damage to occur as developing embryos switch to aerobic metabolism. Another potential source of oxidative DNA damage could occur through changes to histone post-translational modifications (PTM). Many important developmental genes are marked at their promoters by trimethylation of lysine 4 on histone H3 (H3K4me3) prior to the MBT in zebrafish (Lindeman et al. 2011). This PTM is considered permissive for transcription, but in pre-MBT embryos these genes are silenced. Many of these promoters are co-marked with repressive H3K9me3 and/or H3K27me3, which appear to take precedence. When genes are to be transcribed, the repressive PTMs are removed (Lindeman et al. 2011). Demethylation of H3K9me3 is stepwise, H3K9me3 to H3K9me2 to H3K9me1. Demethylation of H3K9me2/1 is mediated by LSD1, which generates H$_2$O$_2$ as a byproduct. Peroxide in turn can react and damage local DNA (Perillo et al. 2008). Perhaps large scale demethylation of H3K9me3 around the time of the MBT leads to an increase in ROS production and oxidative DNA damage in the promoters of developmentally important genes? Future work should examine the levels, rates, and localization of ROS generation during embryogenesis.

In conclusion, I have shown that there is a distinct, dynamic and inverse correlation between the level of total DNA damage and 5mC during zebrafish embryogenesis. Total DNA damage maximally increases when embryos enter the MBT, which is the time when 5mC is maximally reduced. The amount of total DNA damage is further elevated in embryos that have had the BER enzyme Apex1 knocked down, indicating that much of the damage acts as a substrate for BER. This interpretation is supported by the fact that the number of AP sites present in the Apex1 K/D embryos is ~2-fold higher than in controls. I have also shown that when the CpG dinucleotide in CRE, the Creb1 recognition sequence, is fully or hemi-methylated, contains an AP site, or a thymine, binding of recombinant CREB1 is severely impaired. Based on these observations I propose that during zebrafish embryogenesis active DNA demethylation takes place during the MBT at the CpG islands of many developmentally important genes including creb1 itself. This leads to elevated DNA damage, mainly as AP sites and single-strand DNA breaks from Apex1 mediated cleavage of the AP sites. When Apex1 protein is reduced, AP sites accumulate further resulting in greater levels of DNA damage. Since many
AP sites may occur next to each other in the context of CpG dinucleotides or CpG islands, DSBs may also arise with activation of a secondary DNA repair system, likely DSB repair. Since the two zebrafish creb1 gene promoters contain CpG islands and those CpG islands contain CRE sites, demethylation of creb1 promoters is likely to affect CRE sites too. Since Creb1 does not bind methylated CRE sites or CRE sites containing AP sites, loss of Apex1 is likely to affect creb1 expression. With reduced creb1 expression, many other developmentally important Creb1 dependent genes will be under-expressed or never activated, resulting in hugely perturbed developmental programs. This would explain why complete abrogation of Apex1 is lethal at the MBT and why partial Apex1 K/D leads to abnormal development and death. Without the ability to process AP sites generated by active DNA demethylation, genome activation would likely not occur, while genome integrity and stability would be negatively impacted. In addition, this could explain why other BER components are embryonic lethal. Most, if not all BER components that have an embryonic lethal phenotype would be required to repair DNA damage that occurs due to active DNA demethylation.
References

Abraham, Jessy and Philip J. Brooks. 2011. “Divergent Effects of Oxidatively Induced Modification to the C8 of 2′-deoxyadenosine on Transcription Factor Binding: 8,5′(S)-Cyclo-2′-deoxyadenosine Inhibits the Binding of Multiple Sequence Specific Transcription Factors, while 8-Oxo-2′-deoxyadenosine Increases Binding of CREB and NF-kappa B to DNA.” Environmental and Molecular Mutagenesis 52:287–95.


“Global and Gene Specific DNA Methylation Changes during Zebrafish Development.”


Larsen, Elisabeth, Christine Gran, Barbro Elisabet Sæther, Erling Seeberg, and Arne Klungland. 2003. “Proliferation Failure and Gamma Radiation Sensitivity of Fen1 Null Mutant Mice at the Blastocyst Stage Proliferation Failure and Gamma Radiation Sensitivity of Fen1 Null Mutant Mice at the Blastocyst Stage.” 23(15):5346–53.


Li, Jianfeng, Andrea Braganza, and Robert W. Sobol. 2013. “Base Excision Repair Facilitates a
Functional Relationship between Guanine Oxidation and Histone Demethylation.” Antioxidants & Redox Signaling 18(18):2429–43.


Park, Y. G., M. Nesterova, S. Agrawal, and Y. S. Cho-Chung. 1999. “Dual Blockade of Cyclic AMP Response Element- (CRE) and AP-1-Directed Transcription by CRE-Transcription Factor Decoy


Zarakowska, Ewelina, Daniel Gackowski, Marek Foksinski, and Ryszard Olinski. 2013. “Are 8-Oxoguanine (8-oxoGua) and 5-Hydroxymethyluracil (5-hmUra) Oxidatively Damaged DNA Bases or Transcription (Epigenetic) Marks?” Mutation Research 6–11.


