CHAPTER 1

Biological contexts of DNA-mediated charge transport
The vast majority of intracellular deoxyribonucleic acid (DNA), the molecule that houses the information necessary for life as we know it, is stored within the nucleus of eukaryotic cells as chromatin. The chromatin is composed of double stranded DNA wrapped around histone proteins, condensing the DNA in an ordered manner that allows for accessibility of the genetic material when needed. Serving as the primary library of information in the central dogma of life, the genetic information stored within DNA is transcribed into single stranded ribonucleic acid (RNA), which is then translated into a corresponding string of amino acids to ultimately function as a protein.¹ Proteins are the catalytic and structural workhorses of the cell, and all of the information to make these proteins is housed in the DNA. With three DNA bases creating a single codon to which a specific amino acid is ascribed, the four DNA bases can therefore be arranged as 64 unique 3-base combinations. However, with only 21 naturally occurring amino acids, the DNA genetic code allows for redundancy.

DNA exists intracellularly in its physiologically relevant B-form structure. B-form DNA conforms to a 20 Å wide right-handed double helical structure, with the bases stacked centrally along the helical access (grey), with the negatively charged sugar-phosphate backbone circling the exterior (black), as represented in Figure 1.1.² Such B-form DNA is comprised of two antiparallel single strands of DNA that associate through the formation of specific hydrogen bonds among four distinct bases: adenine pairing with thymine via two hydrogen bonds, and guanine pairing with cytosine via three hydrogen bonds.³ The bases pair so that a two-membered ring purine, G or A, always interacts with the corresponding a one-membered ring pyrimidine, C or T, such that the width of this molecule is consistently uniform.²,³ Chemical structures of the individual bases are
FIGURE 1.1 — Structural representation of double stranded B-form DNA. Top: The DNA bases are represented in a space filling model (gray) and the sugar phosphate backbone as a ribbon model (black), highlighting the intimae stacking of the bases. Middle: Looking down the helical axis of B-form DNA, (center) the extensive degree of overlap among the base-paired core is depicted (right, yellow), and the structural similarity of the stacking to graphene (left, yellow), a known charge-conducting substance. Bottom: Stick representations of the four canonical DNA bases, where the purine adenine pairs with the pyrimidine thymine via two hydrogen bonds, and the purine guanine pairs with the pyrimidine cytosine via three hydrogen bonds.
depicted at the bottom of Figure 1.1. Due to the base pairing geometry orienting the two strands not directly opposite of one another, the DNA double helix contains a wider major groove and a narrower minor groove. The major groove is 22 Å wide and allows access to the bases, and is known to act as sequence-specific binding sites for many transcriptions factors. The minor groove is much more narrow at 12 Å, making base access and sequence specific protein binding more difficult from this location.²

Of greatest interest to the research conducted in the Barton laboratory is the ability of DNA to act as a molecular wire. The potential for conductivity through DNA was first suggested in 1962, when structural characterization determined that the interplanar spacing of the aromatic bases in B-form DNA is similar to the spacing between individual sheets of graphite, a known conductive material. This similarity of DNA to stacked graphite sheets could therefore form a conductive path of overlapping π-orbitals extending parallel to the helical axis, as depicted in Figure 1.1.⁴ This property of DNA mediated charge transport (CT) was tested using many platforms and illustrated in ground state electrical experiments, where it was found that well stacked DNA has the same conductivity as charge traveling perpendicular to sheets of graphite.⁵ It was also found that the graphite-like stacked bases of DNA also allow for the conduction of both electrons and electron holes along the helical axes.⁶

**Long range DNA damage**

The properties of DNA with respect to its conductive ability have been probed in solution through the use of tethered oxidants and electrochemically through attaching DNA to electrode surfaces. From the numerous investigations conducted it has been
determined that DNA is able to conduct both electrons and electron holes along the helical axis. DNA must be double stranded for this conductivity to occur, whereby single stranded, poorly stacked, oligonucleotide counterparts are unable to convey CT. The transport of charge in double stranded DNA is also extraordinarily sensitive to the integrity of the DNA base stack. Perturbations such as a single DNA mismatch, an abasic site, or an oxidatively damaged base adduct severely attenuate CT. However, nicks in the DNA sugar-phosphate backbone do not attenuate CT, confirming that the conductive nature is dependent upon the base-stacking and does not involve the backbone. We have also exploited this property in electronic devices to detect base mismatches, base lesions, and to characterize DNA-binding proteins. These investigations have determined that DNA CT occurs with a shallow distance dependence, meaning that charge can be conducted over long molecular distances with low resistance. By using a variety of distally bound photooxidants, we have also measured effective CT through DNA over a distance of 20 nm; much longer distances for CT are expected given the very shallow distance dependence observed. Through in vitro experiments we have also found that oxidative damage to DNA can occur from a distance due to the migration of electron holes through the DNA base stack, as depicted in Figure 1.2.

The Barton laboratory has focused our studies on the properties, usefulness, and biological implications of long-range charge transport (CT) through DNA. With respect to the research conducted herein, the focus has been on the biological context of DNA-mediated oxidation. The eukaryotic genome incurs thousands of oxidative events daily and may arise from such sources as ionizing radiation, exogenous chemicals, and metabolic side products. But how does the DNA respond to these oxidative events?
**FIGURE 1.2** — *In vitro* DNA-mediated charge transport experimental construct. Synthetic oligonucleotide conjugated with a Rh or Ru photooxidant (red) that then intercalated into the double stranded DNA (blue). Photoexcitation ($h\nu$) of the tethered and intercalated photooxidant abstracts an electron from the DNA, creating an electron hole which than equilibrates among the $\pi$-stacked bases. Guanine doublets (yellow) within the DNA sequence are efficient electron hole traps due to their low redox potential, leading to oxidative DNA lesions at these locations. Oxidative lesions have been observed over 200 Å from the DNA bound photooxidant *in vitro*, depicted as the white flare at the far right of the oligonucleotide.$^9$
Research in the Barton lab and other groups has found, based on DNA CT, that once oxidized, the electron hole within the DNA can equilibrate among the bases and localize to sites of low redox potential.\textsuperscript{9,16,17}

The redox potentials of the individual bases are key in determining the location to which electron holes will localize in oxidized DNA. The one-electron oxidation potentials for the bases are as follows: C (1.7 V), T (1.6 V), A (1.42 V), and G (1.29 V).\textsuperscript{16} With the lowest redox potential of the four canonical bases, guanine is the most easily oxidizable. Guanine doublets and triplets are even more readily oxidizable than single guanine residues, and guanine oxidation at the 5’ end of such sites has become a known hallmark of one-electron oxidation of DNA.\textsuperscript{17} As shown in Figure 1.2., \textit{in vitro} photooxidation of a long oligonucleotide shows that the electron hole preferentially localizes to guanine doublet sites over 200 Å away from the DNA tethered photooxidant.\textsuperscript{9} Given the ease of electron hole migration through DNA, we expect holes to localize to DNA sites of lowest reduction potential: particularly guanine doublets and triplets.\textsuperscript{17} Guanine radicals can yield a myriad of mutagenic lesions as a result of reacting with water or dioxygen.\textsuperscript{18}

**Long range DNA CT in the presence of DNA-bound proteins**

One avenue of research within the Barton lab has been to explore how DNA CT may be used \textit{in vivo}. In accordance with the central dogma, DNA is primarily useless if it cannot be transcribed and subsequently translated, indicating that histone proteins or transcriptional proteins are continually within close proximity of DNA.\textsuperscript{1} One can imagine that proteins intimately involved with DNA may be able to couple into the CT
pathway of DNA and potentially utilize this property as a means of cellular signaling. In certain cases, DNA-bound proteins may react with the base radicals to form covalent adducts.\textsuperscript{19} If the DNA-bound proteins are also redox active, they may be able to modulate their activity upon oxidation, and not become covalently attached to the DNA. While most studies conducted thus far have used synthetic oligonucleotides tested \textit{in vitro}, we have also seen been able to observe that long-range oxidative damage can occur in chromatin and in the nucleus of HeLa cells.\textsuperscript{20-22}

One example of a protein that affects the DNA CT properties is the TATA binding protein. The main function of the TATA protein kinking the DNA is to destabilize the bases to allow for other transcriptional machinery to access a transcriptional start site.\textsuperscript{23} However, the kink made by the binding of TATA protein to the DNA is so severe that charge transport is attenuated and can be detected electrochemically.

Another transcription factor formerly studied in the Barton group includes SoxR, which is an \textit{E. coli} stress response protein that contains a [2Fe2S] cluster.\textsuperscript{24} SoxR is a transcriptional regulator for the SoxS gene pathway, and the activation of SoxR only occurs once the protein is oxidized. Since the SoxR protein has similar binding affinities for its response element in the apo, reduced, and oxidized forms, binding of SoxR to DNA is not the source of its oxidation. However, it appears that only the SoxS downstream products are activated in the presence of oxidized SoxR and the DNA mediated oxidation of SoxR leads to a conformational change that elicits the transcription of downstream SoxS.\textsuperscript{25} Experiments \textit{in vitro} have shown that SoxR can be oxidized from
a distance through DNA CT, leading ultimately to the transcriptional activation of SoxS.\textsuperscript{25}

Even more complex DNA-protein interactions have been studied and a model has been proposed whereby DNA CT plays an integral biological role in DNA damage, sensing for the first step of DNA damage repair.\textsuperscript{26} This DNA-mediated genomic repair process is made possible by the prevalence of [4Fe-4S] clusters in base excision repair enzymes, such as MutY, EndoIII, and DinG.\textsuperscript{27,28} More recently, [4Fe-4S] clusters have been found in the full range of DNA-processing enzymes, suggesting a general role for DNA CT within the cell in long range signaling of genomic integrity.\textsuperscript{8,26} In this model, the proteins communicate to one another through electron injection to the DNA. If the proteins can communicate, that means the intervening region between the two proteins has been scanned and is free of damage; in this case, the proteins can dissociate and move on to investigate another location within the genome. If DNA damage is located between the two proteins, the scan cannot be completed due to the attenuation of CT, and both of the proteins remain bound within the local area. The proteins may then process around that general region to find and repair the site of damage.\textsuperscript{8,26}

DNA CT recognition by proteins does not necessarily require an iron-sulfur cofactor; other redox-active moieties within a protein can participate as well. Cysteine residues can be oxidized to form disulfide bonds at physiological redox potentials. As shown both by \textit{in vitro} photooxidation and electrochemical experiments, thiols incorporated into the DNA backbone can be oxidized to disulfides at a distance through long range DNA CT.\textsuperscript{29,30} Cysteine redox chemistry is often harnessed \textit{in vivo} by DNA-
bound proteins as a redox switch in regulation; DNA CT chemistry would offer the ability to carry out such reactions from a distance \textit{in vivo}.

**Transcription factor p53**

Transcription factor p53 was initially thought to be an oncogene, due to its marked upregulation in numerous human cancers. It was however determined that p53 is a transcription factor whose mutation leads to a predisposition to cancer. Therefore, p53 is a tumor suppressor and not itself an oncogene. Human transcription factor p53 transduces a variety of cellular stresses into transcriptional responses. The pivotal role which p53 plays in human cells classifies this protein as a tumor suppressor. Intracellularly, p53 has a short half life due to its negative regulator murine double minute 2 (MDM2), which is an E3 ubiquitin ligase that sequesters p53 and targets it for proteolytic degradation through multiubiquitination. When some cellular stress signal is sensed, such as oxidative stress, hypoxia, or oncogene activation, p53 is activated and escapes MDM2 control. This increases intracellular p53 levels leading to the regulation of p53 target genes or other protein-protein interactions. Overall, many of the pathways in which p53 is involved revolve around decisions of cellular fate, including responses like apoptosis, senescence, cell cycle arrest, or DNA repair (Figure 1.3).\textsuperscript{31-36}

The importance of p53 integrity for proper biological function is highlighted by the fact that mutations in this protein are observed in over half of all human cancers. The most common type of mutations observed in human cancers involving mutant p53 are point mutations, resulting in a single amino acid substitution within the protein.\textsuperscript{37} Such mutations may cause improper protein folding, disruption of integral protein-protein interactions, or alteration of protein-DNA contacts.\textsuperscript{38} Of the known p53 point mutations
FIGURE 1.3 — The p53 response to cellular stress. The p53-MDM2 feedback loop is the primary means of intracellular p53 regulation. Activating signals (top) inhibit the p53-MDM2 interaction, leading to increased intracellular p53 concentrations and the subsequent activation and repression of various transcriptional targets. Under physiologically normal levels of cellular stress, p53 tends to promote repair processes. However, in the case of severe cellular stress, in which repair attempts may be futile, cellular senescence and apoptosis are preferentially promoted.⁴⁰
FIGURE 1.4 — Frequency of point mutations within p53 observed in human cancers. The top chart represents the percent of point mutations as observed in human cancers (n=24,210) per individual codon. Of the cancer relevant mutations observed, over 80% of these occur within the conserved DNA-binding domain. The structural domains of p53 are depicted below the plot. p53 contains an N-terminal trans activation domain (TAD), followed by a proline rich domain (PRD), the highly conserved and structured DNA binding domain (DBD), followed by the tetramerization domain (TD) and the C-terminal domain (CTD).57
observed in human cancer, the majority of these mutations occur within the DNA-binding domain, as seen in Figure 1.4. This finding strongly suggests that the proper function of the DNA binding domain is of the utmost importance for proper p53 function. Human p53 is the most highly researched human transcription factor due to its association with cancer.

**Structure of p53**

At 393 amino acid residues long, human p53 contains many highly conserved residues within the DNA binding domain. Human p53 is comprised of a loosely structured amino-terminal transactivation domain (TAD—residues 1-63) containing two transactivation subdomains (TADI—residues 1-42, TADII—residues 43-63), to which many different post translational modifications can be appended. The TAD is followed by a proline-rich domain (PRD—residues 64-92), which is a common feature in many transcription factors. Following the PRD is a flexible and unstructured region that leads into the highly structured and evolutionarily conserved core of the protein, the DNA-binding domain (DBD—residues 102-292). The DBD is followed by the tetramerization domain (TD—residues 307-355), which allows for the protein to assemble as a tetramer when binding response element DNA. The TD contains a flexible linker region (residues 307-315) as well as a nuclear localization signal domain (residues 316-325). The p53 protein is then terminated with an unstructured basic C-terminal domain (CTD—residues 356-393). This general landscape of the p53 domains is depicted in Figure 1.4. Corresponding point mutation frequencies at each codon of p53 as observed in human
cancers is also listed, demonstrating that the majority of cancer relevant mutations occur within the DNA binding domain of the protein.\textsuperscript{37}

Transcription Factor p53 binds to its response element as a tetramer. The DNA sequence to which p53 was found to recognize and bind is comprised of two copies of the 10 base pair half site motif 5′-RRRCWWGYYY-3′, separated by 0-13 base pairs, with R representing a purine, Y representing a pyrimidine, and W being either an adenine or thymine.\textsuperscript{41} Each monomer of p53 also contains one Zn\textsuperscript{2+} that appears necessary for structural integrity, allowing for response element binding. p53 makes several direct contacts with bases within the major groove of the response element, as well as direct backbone contacts and several water-mediated contacts. As depicted in Figure 1.4, the p53 DNA binding domain assembles to the response element as a tetramer.\textsuperscript{42,43} Within each p53 monomer, three cysteine residues (C176, C238, and C242) and one histidine (H179) coordinate a zinc ion that is believed to be structurally necessary for DNA binding.\textsuperscript{38,42-44} Located close to the Zn\textsuperscript{2+}, but not participating in metal binding, is C182. Closer to the DNA-p53 interface are the remaining conserved residues of interest: C124, C135, C141, C275, and C277. Nestled into the major groove, C277 is capable of forming a hydrogen bond within the purine region of the p53 response element quarter site.\textsuperscript{42,43} C275 is located 7.0 Å away from C277, from sulfur atom to sulfur atom. Residues C124, C135, and C141 are found clustered deeper inside the core of the DNA binding domain, with C275 7.0 Å angstroms away from C135. Chen and coworkers have reported these residues as reduced in their structural characterizations of the p53 DNA binding site; however, disulfide formation is plausible based on the proximity of these residues with respect to one another.\textsuperscript{42,43}
FIGURE 1.5 — Crystallographic representation of the p53 DNA-binding domain. Structural representation of p53 binding as a tetramer (blue and purple) to a full response element. The spherical representation of DNA and surface representation of p53 (top) display the tight interaction between the p53 monomers and the response element DNA. Looking down the helical axis of the stick figure DNA (bottom) one can see the symmetry of the DNA-bound p53 tetramer and how deeply it binds into the DNA major groove. Images based on PMID: 3KMD crystal structure.42
Transcriptional activity of p53

Most commonly, p53 serves as a transcription factor in the promotion of RNA polymerase II transcribed genes. Some of the most noteworthy and important genes under direct p53 regulation include genes playing roles in cell cycle arrest (\textit{p21, 14-3-3}), apoptosis (\textit{pig, Bax, puma, noxa}), senescence (\textit{pai-1}), and autophagy (\textit{dram}).\textsuperscript{45} Interestingly, p53 also promotes its negative regulator MDM2 and itself, p53. The majority of p53 response elements cluster within noncoding regions of the genes they regulate; it has been found that they can be located nearly anywhere within the target gene locus.\textsuperscript{46} Response elements for p53 are most commonly found in the upstream promoter regions from the target gene transcription start site, within about 300 base pairs. A general trend appears and response elements tend to decrease in transactivation potential as they increase in distance from the transcription start site.\textsuperscript{46} In several cases, p53 response elements have also been found in early intronic sequences of the target genes, as well as within exons.\textsuperscript{46}

Once p53 binds its designated response elements, histone modifications within the region are necessary to relax the chromatin and enable general transcription machinery accessibility. In response to DNA damage, p53 is involved in the recruitment of the histone variant H2A.Z, an event which is required for full activation of p21.\textsuperscript{47} The relationship of p53 with its most well-studied HATs, p300 and CBP, is fairly complex.\textsuperscript{48} Once the local chromatin has been modified and remodeled, components of the preinititation complex may then be recruited or somehow altered to allow for the initiation of transcription.\textsuperscript{49} TFIID is recruited to the promoter’s TATA region to nucleate the formation of the PIC, followed by TFIIB, and finally by the assembly of the
other transcription initiation factors (TFIIF, TFIIE, TFIH) complexed with unphosphorylated RNAPII. However, for many genes, such as p21, it is clear that the levels of p53 bound are not the sole determinant of the ensuing transcriptional response.

Repression of certain genes by p53 has been proposed to occur in several different ways. One suggestion is that the binding of p53 to certain response elements recruits corepressors to the site and results in overall downregulation. Another idea is that p53 can secondarily inhibit expression of certain genes by promoting the activation of certain repressor proteins. p53 may also bind to its response element and occupy the site so that other transcriptional activators cannot gain access. Lastly, it is thought that p53 may also repress genes that do not contain a p53 response element through protein-protein interactions that inhibit the promotion of those genes.

Other roles of p53

There are many post-translational modifications that affect p53 and how it functions within the cell. It has been observed that upon DNA damage, p53 is phosphorylated on its NTD, and such damage-inducible phosphorylation then enhances p300/CBP-mediated acetylation and methylation of lysine and arginine residues of the CTD. Arginine methylation has also been observed in the tetramerization domain. Interestingly, unlike the MDM2 ubiquitination leading to p53 degradation, mono-ubiquitination of k320 appears to be for transcriptional regulation.

Although p53 primarily serves as a transcription factor in response to cellular stress, many other roles of p53 have been investigated. Such roles have been found to
include transcriptional repression, translational regulation, recognition of DNA double strand breaks, as well as playing a role in homologous recombination and enabling a transcription-independent apoptotic response. The C-terminal domain of p53 has also been suggested in specifically recognizing and binding to unique and biologically relevant DNA structures such as single-stranded DNA over hangs, hemicatenated DNA, minicircular DNA, and supercoiled DNA.

**Focus of this thesis**

Much research has focused on the transcriptional role that p53 play through its recognition and binding of response element sites. However, with each new study, the network of roles played by p53 just becomes more and more complex. One facet of this research that has gone uninvestigated by other laboratories is determining if and how p53 can directly sense DNA damage, seeing that it is known that p53 is activated in this case. What also is not greatly known is how p53 selects binding to one response element over another. Also, many researchers focus on p53 recruitment, activation, and modification, but there is little understanding of what the deciding factor is and how p53 then returns to a signal-off state. Specifically we ask how p53 senses oxidative genomic stress and whether p53 senses it directly. The following work described in this thesis are investigations on the direct sensing of genomic oxidative stress by p53, and how it may accordingly respond. While p53 is generally known to sense oxidative stress as one of its inputs, its function as a redox-active DNA-binding protein remains to be fully elucidated.
REFERENCES


