The Genome as a Drug Target: Sequence Specific Minor Groove Binding Ligands

P. R. Turner* and W. A. Denny+

*Hygiene Institut, Abteilung Parasitologie, Im Neuenheimer Feld 324, 69120 Heidelberg, Germany
+Auckland Cancer Society Research Centre, Faculty of Medical and Health Sciences, The University of Auckland, Private Bag 92019, Auckland, New Zealand

Abstract: The ability to target defined sequences on the DNA molecule would be of enormous benefit to the treatment of human disease. Towards this goal much research has been invested in examining the DNA binding and biological mechanisms of action of sequence selective minor groove binding ligands. These compounds act in a variety of ways to inhibit gene expression and DNA replication and also alter nuclear architecture. Concomitant with this, minor groove adducts formed by certain compounds are inefficiently removed by cellular DNA repair systems and are extremely cytotoxic. Additionally compounds targeting A.T rich DNA sequences have found clinical use in the treatment of particular parasitic infections.

INTRODUCTION

DNA is popularly referred to as the blueprint of life. Encoded within its nucleotide sequence lie instructions not only for the manufacture of proteins and RNAs essential for cell survival, but information which ensures these are produced only under tightly regulated conditions. In addition to these instructions DNA also has regions which are crucial to its own replication and its spatial organisation in the cell nucleus. The expression of a specific gene is dependent on a multitude of transcription factors which must recognise and bind specific sequences on the DNA molecule to allow the transcriptional complex to form. DNA replication during cell division is also dependent upon specific sequences, origins of replication, to interact with DNA binding proteins to allow multi-component replication complexes to form. It has also been shown that DNA in the eukaryotic nucleus is spatially organised by its attachment to the proteinaceous nuclear matrix, and that the attachment sites occur at specific DNA sequences. All of these diverse processes have a common theme, that of stretches of nucleotides of varying length acting as a specific recognition site for DNA binding proteins.

Every organism possesses its own specific genetic makeup which identifies it both as a species and often as an individual within a species.

*Address correspondence to this author at the Hygiene Institut, Abteilung Parasitologie, Im Neuenheimer Feld 324, 69120 Heidelberg, Germany

Indeed, many human diseases can now be effectively diagnosed by searching for organism specific DNA in patient samples using DNA amplification technologies [1]. The differences between the DNA sequence of most disease causing organisms and human DNA is significant and many genes in these organisms are unique and play pivotal biochemical roles. Aside from genetic differences the base content itself differs markedly in some organisms, for example Plasmodium falciparum, the causative agent of the most deadly form of malaria, has a genome which is comprised of approximately 82% adenine and thymine [2], compared to 60% for the human genome.

Neoplastic cells can also be distinguished from normal tissue by differences in their DNA, and cancer can be considered a disease of the genes. Unlike other inherited genetic disorders where a defective gene is present in every cell of the body, cancer results from defects in specific genes in possibly only one original body cell [3].

Neoplastic transformation is thus viewed as a multi-step process involving sequential mutation of several key genes [4]. The progression from normal tissue to metastatic colon carcinoma has been shown to involve loss or mutation of at least 4 genes [5]. Mutations in specific genes are associated with particular neoplasms and differ between cancer types, although some key genes are involved in a range of different tumour types. The implication of this is that, each individual cancer has a subset of damaged genes that define it
and set it apart, genetically, from normal tissue. The molecular mechanisms of carcinogenesis are complex but many genes have been identified which contribute to the process in a range of different cancers [6]. As the human genome project proceeds with an ever accelerating pace it is predicted that more of these genes will be identified and detailed molecular mechanisms of carcinogenesis will be divulged.

In terms of human disease, DNA represents an ubiquitous and yet extremely specific drug target. The ability to target specific nucleotide sequences and prevent their normal cellular function would mean that any disease causing organism could be selectively attacked, as could subpopulations of genetically distinct cells containing deleterious mutations such as those found in neoplastic tissue.

TARGETING DNA

Since the discovery of the structure of the DNA molecule by Watson and Crick nearly 50 years ago, an enormous amount of research has been devoted to producing high resolution x-ray crystallographic and NMR based structures of various short DNA molecules which varied in base sequence. These studies determined that in the B-DNA form (which makes up the majority of cellular DNA) the antiparallel helix is right handed, the glycosidic bonds are in the anti conformation and the ribose units are C2′-endo puckered. The net effect of this is the creation of two grooves, the major groove (=12Å wide) and the minor groove (=6Å wide), [7]. The base sequence also imparts variations to this regular structure and specific base sequences confer a huge structural flexibility to the molecule. A relevant example of this is the enhanced ability of A.T base pairs to propeller twist which allows the minor groove to narrow considerably at sequences rich in these base pairs [8]. Other sequences permit a variety of structural changes to occur such as helical bending and kinking. Further studies demonstrated that DNA structure can be drastically altered by the binding of various regulatory proteins. The structural modifications imposed on DNA in such complexes is the combined effect of protein binding and the intrinsic deformability of specific base sequences [9].

The maintenance of coding information in DNA stems from its ability to form a complementary double stranded structure. Complementarity results in a large part from formation of hydrogen bonds between specific opposing bases. Thus, adenine pairs with thymine, and guanine with cytosine by the formation of 2 and 3 hydrogen bonds respectively. The atoms involved in formation of these specific H-bonds are, thus, inaccessible unless the helical structure is disrupted. However, on either side of the planar bases lie additional H-bond donating and accepting groups specific for each base and which protrude into the relatively accessible major and minor grooves of the helix. Thus in the major groove the C6 amino group of adenine or the C4 amino group of cytosine can act as hydrogen bond donating groups while the adenine N7, thymine O4, guanine N7, O6 can act as hydrogen bond accepting groups, and the thymine methyl as an hydrophobic site. Similarly in the minor groove the adenine N3, thymine O2, guanine N3, and the cytosine O2 can act as H-bond acceptors and the C2 amino group of guanine can act as a donating group. Both grooves, therefore, carry base dependent sequences of potential H-bonding atoms that can be used as a target in the design of sequence specific DNA binding compounds.

MAJOR GROOVE ALKYLATING AGENTS

These have a long history in the treatment of human disease. The first clinically used antitumour agent, the nitrogen mustard mechlorethamine [10], exerts its effects by alkylating cellular DNA, preventing its use as a template for cellular proteins, and preventing strand separation for replication. The sequence selectivity of DNA alkylation of mechlorethamine and most other clinically used alkylating agents is relatively poor, with most monoaalkylating in the major groove at the N7 position of guanine in runs of guanines, and cross-linking guanines at 5′-GNC sites (where N is any nucleotide). The antitumour activity of most DNA alkylating compounds resides in their ability to kill rapidly dividing cells (kinetic specificity) rather than any tumour cell selectivity per se. These compounds therefore cause a high degree of collateral damage to dividing cells in normal tissue which is manifest as the side effects of chemotherapy - toxicity to bone marrow and gut epithelia, hair loss and sterility.

DNA INTERCALATORS

These are another class of compounds with major clinical significance, but which also generally show poor sequence selectivity. These poly(hetero)cyclic aromatic compounds bind reversibly but tightly to DNA by slotting the
chromophore in between the stacked base pairs [11]. These compounds work by inhibiting or poisoning the action of topoisomerases, cellular enzymes which relieve topological strain on the DNA molecule during transcription and replication [12]. Again, the sequence selectivity of these compounds is generally very poor, although many contain side chains that interact with limited specificity in one of the grooves. Certain intercalator-linked mustard compounds have also shown interesting changes in base alkylation specificity [13].

SEQUENCE SPECIFIC DNA BINDING COMPOUNDS

The potential of compounds which bind cellular DNA in a sequence specific manner has long been realised and much research has been invested in developing them. Such DNA targeting drugs must meet several criteria to be effective in treating human diseases. They must first be membrane permeable of actively imported, to be able to reach their target within the cell, and must be resistant to intracellular degradation. The sequence specificity must be very stringent and the length of sequence targeted must be long enough to ensure effective gene targeting. The human genome is approximately 3 billion base pairs in size so that a sequence of 16 base pairs should be unique on a statistically random basis. Concomitant with this, the binding affinity must be high enough to permit the drug to inhibit the normal functional role of the sequence which it is targeting. Two broad classes of such sequence-specific compounds can be distinguished i) nucleic acid mimicking compounds and ii) groove binding ligands.

NUCLEIC ACID MIMICKING AGENTS

Compounds which mimic nucleic acids contain naturally occurring nucleotide bases linked to a chemically modified backbone. The nature of the modified backbone varies considerably but in general must be sufficiently different to the sugar phosphate backbones of DNA and RNA to prevent degradation by intracellular nucleases. However, the basic structure must be similar enough to DNA to enable the linked bases to form H-bonding interactions with cellular DNA. Although much work has been done with phosphorothioates, probably the most successful of these compounds are the peptide nucleic acids (PNAs) which can form very stable hybrids with both RNA and DNA [14]. PNAs have a charge neutral peptide-like backbone, (Fig. 1), and the linked bases confer sequence selectivity to these molecules which form several types of hybrid structures. PNAs that are composed of homopyrimidines form a triplex with double stranded DNA where two PNAs bind the complementary homopurine strand of DNA and the homopyrimidine DNA strand is displaced to form a single stranded loop (PNA₂DNA). Other less stable hybrid structures also form depending on base sequence, including duplex invasion complexes where a single PNA forms a complementary duplex with DNA and displaces the non-complementary strand, double duplex invasion complexes where sense and antisense PNAs bind the same DNA on opposite strands, and triplexes where a cytosine rich PNA forms Hoogstein base pairs with the DNA molecule [reviewed in 15]. PNAs have been shown to inhibit DNA polymerisation, RNA polymerisation, reverse transcription and transcription factor binding [15, 16, 17]. Two problems must be overcome for these compounds to be considered for clinical use, sequence specificity and cellular uptake. The former is perhaps a lesser problem as relevant targets possessing homopurine or homopyrimidine tracts can be found in many genes but cellular uptake has presented a very real problem and much research is now focussed on delivery systems for these compounds.

MINOR GROOVE BINDING LIGANDS

There is a plethora of both naturally occurring and synthetic, cell permeable, sequence selective compounds which bind in the minor groove of DNA and research into new compounds is proceeding at a rapid pace. Several features of the minor groove make it amenable to targeting. Sequence information is represented in the minor groove as unpaired base specific hydrogen bond donating and accepting groups, variations in molecular electrostatic potential, variations in Van der Waals contacts and variations in hydrophobicity. Minor groove binding ligands have a number of characteristic features that distinguish them from other types of reversible DNA binding agents such as the intercalators. These include an overall annular shape made up of aromatic rings which matches the curvature of the minor groove of DNA, cationic charges which provide affinity for the tunnel of negative molecular electrostatic potential in the groove, and in addition many ligands possess H-bond donating or accepting atoms. These compounds tend to bind in the minor groove with relatively little distortion of the phosphate backbone, and in fact stabilise the regular B-DNA structure.
The molecular determinants of sequence specificity of minor groove binding ligands vary with ligand structure. The vast majority of these compounds selectively bind to A.T rich DNA sequences and such sites possess several unique characteristics. The width of the minor groove is often considerably narrower than mixed sequence DNA so that the planar aromatic rings of A.T selective ligands can slot into the groove and form stabilising van der-Waals contacts. The electrostatic potential of the minor groove varies with base sequence and is most electronegative at A.T rich sequences [18]. Thus positively charged minor groove binding ligands have a higher affinity for this tunnel of electronegativity on the floor of the groove. Additionally, A.T base pairs have two H-bond accepting groups, the O2 of thymine and the N3 of adenine, which can further stabilise ligands with appropriately situated H-bond donating groups, (Fig. 2). In addition, A.T-rich sequences contain highly-structured networks of coordinated water molecules that add to overall stability, and displacement of these by ligands provides an substantial entropy component to the binding.

Netropsin, (Fig. 1), a naturally occurring polypyrrole-carboxamide binds A.T rich DNA sequences and forms H-bonds via its carboxamide NH groups, with the O2 atom of thymines and the N3 atom of adenines [8]. Netropsin also possesses two terminal cationic diamino groups which provide electrostatic attraction for the negative floor of the groove. The importance of cationic groups in binding to A.T rich sequences is further demonstrated by pentamidine, (Fig. 1), which cannot form stabilising H-bonds with the adenine N3 or thymine O2 atoms, but nevertheless selectively binds A.T rich sequences [19]. Van der Waals interactions between minor groove binding ligands and the minor groove of DNA can also contribute DNA binding affinity and specificity. Analogues of Hoechst 33258, (Fig. 1) which varied in both their electrostatic properties and their benzoimidazole ring structure had differing binding affinities for A.T rich DNA which could be correlated with the degree of van der Waals interactions [20]. These studies with different classes of minor groove binding ligands demonstrate that A.T sequence selectivity stems from a combination of various factors, and the relative importance of each factor varies between ligands.

In contrast to the large number of minor groove binding ligands which bind selectively to A.T rich DNA, compounds which bind to G.C rich DNA or mixed sequence DNA are scarce. Pioneering work by Peter Dervan's research group at Caltech has resulted in compounds which can specifically target all four base pairs in B-DNA and which are conveniently synthesised using solid phase techniques [21]. Structural studies of the complexes formed between distamycin and A.T rich DNA revealed two distamycin molecules can lie side by side in the minor groove in an antiparallel manner, and the normally narrow groove at these sequences can widen considerably [22]. Modulation of the basic distamycin structure at key sites involved in H-bonding with DNA resulted in changes in sequence selectivity, and targeting of G.C and C.G base pairs became possible when the importance of the dimer binding mode was realised, (Fig. 2), [23]. The amide linker groups which join the aromatic pyroles and imidazoles, provide affinity for DNA by forming H-bonds with purine N3 or pyrimidine O2 atoms, while specificity for G.C base pairs is accomplished by a specific H-bond between the imidazole N3 and the exocyclic guanine N2-H. Formation of this H-bond is dependent on an imidazole arranged side by side with a pyrrole which widens the groove and suitably positions the imidazole so the H-bond can form [23]. Distinguishing an A.T base pair from a T.A base pair is also possible with further modification of the pyrrole moiety. Side by side pyroles target both A.T and T.A base pairs in an equivalent manner, but attaching a hydroxyl group to the N3 position of one pyrrole enables recognition of thymines [24]. The O2 atom of thymine can form two H-bonds, one with the amide NH group and the other with the hydroxy-pyrrole OH group. The hydroxyl group also sterically interferes with the adenine C2-H atom, unlike the simple pyrrole moiety. Thus, a hydroxy-pyrrole side by side with a pyrrole selectively targets T.A, while the reverse pair targets an A.T base pair. This sequence specific binding is crucially dependent on side by side dimer formation and as such linking two molecules together to produce a hairpin shaped molecule by using a γ-aminobutyric acid group markedly improves DNA binding [25]. Other modifications have also enhanced the binding of these compounds, notably the addition of flexible β-alanine groups which offset the tighter curvature of the rigid carboxamide linked pyroles and imidazoles [26]. Addition of β-alanine linkers thus permits the targeting of longer DNA sequences.
Fig. (1). Structures of the minor groove binding compounds mentioned in the text.
originated from natural products, such as the extremely cytotoxic cyclopropylidenone CC-1065, (Fig. 1). This alkylates adenines in the minor groove, exclusively at the N3 position, in A.T rich sequences. [27]. Whilst the alkylating group has some modulatory influence on the sequence specificity of the ligand as a whole, in general the minor groove binding portion of the ligand is the major determinant of sequence selectivity. A large number of synthetic analogues, including dimeric compounds that have binding sites of up to six base pairs, have been studied.

Other DNA minor groove alkylators have been deliberately designed, by the linking of alkylating functions to reversible minor groove binding ligands. The polybenzamide mustards studied in our laboratory illustrate this point, (Fig. 3), [28, 29, 30]. Structurally the polybenzamides consist of benzene rings joined by carboxamide linkages which are postulated to form stabilising H-bonds via the carboxamide NH group and the adenine N3 and thymine O2 atoms, (Fig. 2). Cationic dimethyl amine groups attached to the benzene rings provide electrostatic attraction for the minor groove and van der Waals contacts probably also contribute to the sequence specific binding of these compounds. We have used the general polybenzamide framework as a carrier molecule to target nitrogen mustard alkylators to the minor groove at specific sequences. Compound 1 which forms no covalent attachments, binds specifically to adenine rich sequences of DNA [29]. Attachment of a mustard group to the end of this molecule results in alkylation of adenines (presumably at the N3 position) at these same sequences, in contrast to the isolated mustard which alkylates guanines in a relatively non-sequence specific manner [30]. Similar conclusions have been reached with other types of alkylating moieties attached to various types of minor groove targeted molecules and has resulted in the generation of molecules capable of targeting damaging alkylation events to specific sequences [31]. The best known example of this approach is the polypyrrolecarboxamide mustard tallimustine, which is currently in clinical trial [32].

The primary site of action of minor groove binding ligands is cellular DNA, but due to the diverse nature of this class of compound actual mechanisms of action are quite variable. However, some interesting possibilities have become apparent and are discussed below and shown schematically in (Fig. 4).

**TARGETING GENOME CONTENT**

Certain organisms possess genomes with a strong nucleotide bias. The protozoan parasite *Plasmodium falciparum* which is responsible for the most lethal form of human malaria, has a genome comprised of 82% A.T base pairs and has large tracts of DNA comprised almost entirely of these bases. Several A.T selective minor groove binding compounds have been demonstrated to be effective against the blood stages of this
intracellular parasite \textit{in vitro}, including distamycin A, netropsin, DAPI and Hoechst 33258 \cite{33}. These compounds were selectively toxic to the parasite when compared to cultured mammalian cells, in contrast to the G.C selective minor groove binding chromomycin A$_3$. The A.T selective minor groove binding compound pentamidine is also used in treatment of \textit{Pneumocystis carinii} infection in AIDS patients \cite{34}, and these organisms also have an enriched A.T content of about 67\% \cite{35}.

### TARGETING GENE EXPRESSION

Gene expression and its regulation is extremely complex and understanding the molecular mechanisms and biochemical pathways involved in these processes is a fundamental goal in biological research. Gene transcription is dependent on the binding of transcription factors to specific DNA sequences to enable formation of the multicomponent transcriptional complex. Expression of tissue specific genes is probably

#### Table 1: Compound Structure and IC$_{50}$ Values

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>\textsuperscript{a}IC$_{50}$ (\textmu{}M)</th>
<th>Sequence specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>0.39</td>
<td>A.T. Rich</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>0.007</td>
<td>A.T. Rich</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>2.91</td>
<td>5'-A/T A' G/C A/T N</td>
</tr>
</tbody>
</table>

\textsuperscript{a}IC$_{50}$ values are the concentration of drug required to reduce \textit{in vitro} growth of murine P388 cells to 50\% of control values.

\textbf{Fig. (3).} Structure, cytotoxicity and sequence specificity of some polybenzamides.
achieved both by the availability of transcription factors and by alterations in nuclear architecture which modulate transcription factor access to DNA [36].

(A) Transcription Factors

Transcription factor binding sites on DNA are attractive potential targets for minor groove binding ligands, as their blockade results in modulation of gene expression. Another advantage of targeting transcription factor binding sites is the importance of conservation of integrity at these sequences. Mutagenetic studies have demonstrated that changes in the sequence of these cis acting elements usually results in a complete abrogation of transcription factor binding [37]. Thus, changes in sequence that result in the abolition of binding of a sequence targeted ligand will also lead to the inactivation of gene function. This is not the case when targeting sequences in the protein coding region of genes due to the degeneracy of the triplet coding rules. In this case it would be possible for a targeted sequence to mutate at several sites and prevent ligand binding, but without affecting the protein coding sequence.

Modulation of transcription factor binding using sequence specific binding compounds has received intensive investigation in recent years and it has been shown that compounds with different sequence selectivities inhibit the binding of various transcription factors to different extents. An examination of the ability of distinct classes of DNA binding compounds in in vitro assays demonstrated that intercalators either prevented most transcription factors from binding, or were ineffective [38]. The intercalators which were effective at inhibiting transcription factor binding possessed a sugar moiety which protruded into the major groove, whilst ineffective compounds lacked such major groove located groups. The limited sequence specificity or the DNA binding affinity of these intercalators was not correlated with inhibition of transcription factor binding.

Fig. (4). Biological sites of action of minor groove binding ligands. Gene expression, replication, DNA repair and nuclear architecture are inter-related cellular processes and compounds affecting one process may also affect others indirectly.

Studies with chromomycin A₃ showed that ligand induced structural modification of the DNA molecule is also important to inhibition of transcription factor binding. Chromomycin A₃ is a minor groove binder which recognises G.C rich DNA sequences and binds as a side by side dimer and can probably also bind by default to A.T rich sequences when G.C sequences are unavailable. Inhibition of binding of all transcription factors examined was observed with chromomycin
treatment, regardless of their sequence specificity [38]. Chromomycin A₃, (Fig. 1), alters the structure of the DNA molecule by widening the minor groove and compressing the major groove in an analogous manner to the polypyrrole/imidazole compounds. This implies that significant distortion of the major groove by a compound binding in the minor groove will prevent transcription factor binding, even if these factors only interact in the major groove.

In contrast, A.T selective minor groove binding compounds such as distamycin, netropsin and Hoechst 33258 selectively inhibit binding of the TATA binding protein (TBP), which recognises A.T rich DNA sequences via minor groove interactions. The DNA binding of transcription factors which recognised G.C rich sequences or mixed sequence DNA remained unaffected by these compounds, demonstrating the importance of sequence specificity in these interactions. Similar studies with the A.T selective compounds distamycin and tallimustine, (Fig. 1), have shown that these compounds specifically inhibit the binding of the octamer binding factor (OTF-1) and the erythroid specific NFE1, to their A.T rich recognition sequences [39]. Most proteins which bind sequence specifically to A.T rich DNA have extensive contacts within the minor groove, and it is likely that inhibition of binding of these factors to DNA by minor groove binding ligands is mediated by direct steric interference.

The effectiveness of the polypyrrole/imidazole compounds in inhibiting transcription factor binding and subsequent RNA polymerase II mediated transcription, has been demonstrated recently. By using two polyamides which bind sequences adjacent to the binding sites of the transcription factors ETS-1, LEF-1 and TBP, in the HIV-1 genome, specific inhibition of transcription was achieved [40]. LEF-1 (lymphoid enhancer-binding factor-1) and TBP (TATA binding protein) both recognise their target sequences via minor groove interactions, but ETS-1 (E-twenty-six-specific factor), a winged helix-turn-helix protein, binds DNA in the major groove and also makes contacts with the phosphate backbone. Inhibition of ETS-1 binding is probably a result of ligand induced effects with the phosphate backbone contacts of the protein. Despite targeting sequences only 6-7 base pairs in length these molecules were surprisingly non-toxic to peripheral blood mononuclear cells in vitro, and reduced HIV-1 levels to below detectable limits. When used in combination these two compounds were both more effective in inhibiting HIV-1 replication and considerably less cytotoxic, than the currently used anti-viral compound azidothymidine [40].

Inhibition of transcription factor binding by minor groove binding ligands is most effective when the ligand occupies sites where the transcription factor makes contacts in the minor groove. However not all transcription factors make significant contacts in the minor groove and many have crucial interactions in both the major and minor grooves. Compounds, which do not significantly alter the structure of the major groove following binding, but are sequence specific can be chemically modified to enhance their inhibitory potency. Enhanced inhibition of transcription factor binding has been demonstrated to occur with the microgonotropen class of compounds which form bonds in both the minor and major grooves. Microgonotropens are similar in their minor groove targeting portion to distamycin and bind A.T rich sequences, but also possess a protonated polyaniline moiety which extends from the minor groove and binds with the phosphodiester backbone of DNA in the major groove [41]. One such microgonotropen, (Fig. 1), was three orders of magnitude more effective than distamycin, at inhibiting the DNA binding of the growth regulating E2F1 transcription factor, which interacts with both grooves of the DNA molecule, [42].

Neutralisation of phosphate groups on the DNA backbone at specific sequences probably alters the electrostatic properties of the transcription factor binding site to prevent recognition and binding of the cognate factor. GCN4, a member of the leucine zipper family of transcription factors, binds DNA via contacts made solely in the major groove of DNA and this binding is unaffected by a polypyrrole/imidazole located in the minor groove. Attachment of a positively charged Arg-Pro-Arg residue to the end of such a polypyrrole/imidazole transforms it into an efficient inhibitor of GCN4 binding [43]. This general approach of attaching major groove binding moieties to sequence selective minor groove binding molecules provides a potential means of enhancing the potency of other sequence specific minor groove binding compounds.

Another way to enhance the efficiency of transcription factor binding inhibition could be to use alkylating groups attached to minor groove binding compounds. Preliminary results in this respect have demonstrated that the alkylating compound CC-1065 is more efficient at inhibiting
the binding of TBP to the TATA box than any of the other non-covalently binding compounds examined, which all possessed similar sequence specificities [44]. TBP (TFII D) is one of the first factors that binds to the promoter region of genes to activate transcription and the crystal structures [45, 46] of the protein complexed with the TATA box, have shown intimate contacts with the minor groove of the DNA molecule. Extending this concept with the use of bis-alkylating compounds also merits further investigation. DNA interstrand crosslinks formed by bis-alkylating compounds are several orders of magnitude more cytotoxic than monoadducts, probably as a result of the difficulties involved in repairing such lesions. It can be postulated that such crosslinked adducts directed to specific transcription factor binding sites will also be orders of magnitude more effective at inhibiting gene expression than monoadducts or non-covalently bound ligands.

(B) Helicases

Transcription factors are not the only proteins which are involved in gene expression that can be targeted by sequence specific minor groove binding ligands. Helicases form part of the transcriptional complex and are a group of enzymes involved in separating the DNA duplex into single strands to allow not only transcription to proceed, but also DNA repair and DNA replication. Certain compounds have been shown to inhibit helicase activity in vitro, notably the minor groove binding bis-benzimidazoles, and also the intercalating anthracyclines [47, 48]. Sequence specific minor groove binding ligands may therefore present a novel approach to inhibiting expression of defined genes by preventing RNA elongation rather than targeting the earlier step of transcription factor binding and as such may be extremely useful for targeting particular oncogenes which vary in their protein coding sequence rather than their promoter regions.

(C) Topoisomerases

Topoisomerases I and II are enzymes crucially involved in DNA transcription and replication and act by relieving topological stress on the DNA molecule. The topological state is altered by transient cleavage of one or both DNA strands (topo I and topo II, respectively) and the subsequent relaxation of either positively or negatively supercoiled duplex [49]. Several A.T selective minor groove binders inhibit topoisomerase II activity in vitro seemingly by modulating the DNA binding ability of the enzyme [50]. Some topo II targeted intercalating compounds act by binding to the enzyme/ DNA complex and stabilising it, thus preventing its catalytic activity. Several minor groove binding ligands can inhibit the formation of stabilised enzyme-drug-DNA complexes (cleavable complexes) formed by topoisomerase II targeted compounds, providing further evidence that these compounds can inhibit binding of topo II to DNA [51]. Another way of targeting topo II in a sequence specific manner has recently been demonstrated by attaching the topo II targeted compound 4′-demethylepipodophyllotoxin to a minor groove binding polypyrrole carrier. The resulting molecule induced cleavable complex formation with topo II at novel sites on DNA in vitro compared to the parent compounds [52].

In contrast with topoisomerase II, some minor groove binders can form cleavable complexes with topoisomerase I and DNA. A range of bisbenzimidazole (Hoechst) analogues were found to be particularly effective in forming these complexes and could also direct them to new sequences on the DNA molecule [53]. The possibility exists therefore, of stalling transcription via topoisomerase I mediated cleavable complexes in specific genes at defined nucleotide sequences.

MODULATION OF NUCLEAR ARCHITECTURE

The eukaryotic nucleus is spatially organised by a proteinaceous framework termed the nuclear scaffold or matrix, which can be conveniently isolated by salt extraction of the nucleus [54]. The nuclear matrix is the site of many essential nuclear processes such as DNA replication, RNA processing and transcription. Interplay between the organisation of the nuclear matrix and the DNA bound to it can modulate gene activity, and most likely is a key factor in the tissue specific expression of genes [36].

DNA sequences which bind the matrix (scaffold attachment regions) are thought to be located at the base of chromatin loops and many such sequences have been demonstrated to possess A.T rich sites [55]. The importance of A.T rich sequences in chromatin architecture is also implied by the large number of chromatin associated DNA
binding proteins that possess a motif (A.T hook) which interacts in the minor groove at such sequences [56]. Additionally matrix associated proteins specific for neoplastic cells have been detected and one of these (HMG-1(Y)) which is expressed in colorectal tumours can modulate transcription of genes by interacting with specific transcription factors [57]. Several types of minor groove binding ligands, including distamycin and netropsin, have been shown to inhibit the binding of DNA to such scaffold associated proteins [58, 59]. Additionally, most A.T specific ligands induce chromatin decondensation in cultured cells providing further evidence they can modulate nuclear architecture. The structure of assembled nucleosomes in vitro is also affected by A.T specific ligands, which cause a 180° rotation of the DNA on the nucleosome core particle [60].

When taken together, these observations suggest that the biological effects of minor groove binding ligands may, in part, be attributable to the modulation of nuclear architecture through interactions at A.T rich sequences. Further work in this area seems warranted.

DNA REPLICATION

Another possible sequence specific effect of A.T selective ligands may be at the level of DNA replication as many eukaryotic origins of replication possess A.T rich sequences [61]. Many enzymes involved in transcription such as topoisomerases and helicases are also integral to DNA replication. Many scaffold attachment regions are close to or part of origins of replication [62] and minor groove binding ligands which modulate nuclear architecture probably also interfere with DNA replication. Inhibition of replication by A.T specific ligands is general and probably cannot be utilised to produce selective toxicity to particular cell types or pathogens. However, direct inhibition of the polymerisation process by alkylating minor groove binders has also been demonstrated, both in vitro and in intact cells [63]. In an analogous manner to major groove alkylating agents, minor groove crosslinking adducts are also extremely effective at stalling replication [64]. Therefore, specific targeting of replicatory processes by directing alkylator mediated damage to unique genes may eventually be possible.

DNA REPAIR

Both prokaryotic and eukaryotic cells possess systems to repair DNA damage induced by environmental toxins, UV light and oxidative metabolism. Repair of DNA adducts induced by DNA binding compounds modulates their toxicity, and it has been shown that cellular resistance to some antitumour agents is a result of an enhanced DNA repair capacity [65].

In our studies of a series of DNA targeted polybenzamide mustards, we showed that the sequence specificity of base alkylation was dependent on many factors, including charge positioning, mustard reactivity, the backbone structure of the DNA targeting moiety, and the positioning of potential hydrogen bonding groups [30]. Two compounds (2 and 3, (Fig. 3)), which had different sequence selectivities of alkylation showed large differences in cytotoxicity against murine P388 cells in vitro. The more cytotoxic compound 2, had selectivity for runs of consecutive adenines while the less cytotoxic compound 3, had a high alkylation affinity for the sequence A/T A* G/C A/T N. Compound 2 probably forms H-bonds between the inward facing carboxamide NH moieties and the N3 adenine and O2 thymine atoms. In contrast compound 3 probably has outward facing carboxamide NH moieties and is postulated to form an H-bond via the carbonyl oxygen with the guanine C2 amino group. This opposite orientation of carboxamide linkages allows the curvature of compound 3 to more closely match the curve of the minor groove, although the match is still not as good as that of compound 2.

Both the mustard reactivity and the DNA alkylation efficiencies of these compounds are very similar, suggesting that the differences in cytotoxicity are not due to differences in the affinity of binding or alkylation by these compounds. Furthermore, both compounds were also found to crosslink DNA in vitro to a similar extent suggesting that the diversity in cytotoxic effect of this pair of compounds was due either to their sequence selectivity or the nature of the adducts formed. We, therefore, postulated that the enhanced cytotoxic effect of compound 2 is related to the difficulty involved in recognising and repairing its adducts [28].

Further studies with a more extensive series of polybenzamide crosslinking compounds, verified a correlation between in vitro cytotoxicity and the radius of curvature of the compounds. Importantly in this respect one compound with an almost identical sequence specificity of alkylation to compound 2 was significantly more cytotoxic and
possessed a radius of curvature which more closely matched the curvature of the minor groove. Other studies of alkylating minor groove binding ligands have also shown that such adducts can be difficult to repair. Distamycin linked nitrogen mustard compounds are inefficiently repaired by the bacterial nucleotide excision repair pathway [66]. Repair efficiency was found to be inversely related to the number of pyrrole rings of the groove targeting portion of these compounds suggesting that larger adducts which stabilise B-DNA structure are more difficult to repair. Another class of minor groove binding crosslinking compounds, based on the pyrrolobenzodiazepines form adducts in cellular DNA which remain unrepaired 48 hours post-exposure, further demonstrating the difficulty minor groove located adducts present to repair systems [67]. These studies demonstrate the importance of adduct structure in the cytotoxic effects of minor groove binding ligands rather than merely sequence selectivity and provide useful information for future drug design.

CONCLUSIONS

Advances in the understanding of the biological targets of minor groove binding ligands suggests these compounds exert their cytotoxic effects by a variety of mechanisms. Minor groove binders can modulate gene expression by inhibiting transcription factor binding, inhibiting the activity of both topoisomerase I and II and by inhibiting helicase activity. Structure activity relationships have identified important ways of altering ligand structure to optimise these effects. These ligands can also modulate nuclear architecture and probably inhibit both transcription and replication as a result. Some covalently binding minor groove targeted compounds produce adducts which are insufficiently repaired by cellular repair systems and as such may represent an invaluable class of antitumour compounds.

Clinical trials to date with several, mainly A.T selective compounds have been somewhat disappointing [68, 69], but recent improvements in sequence selectivity and understanding of the biological mechanisms of action predicts an exciting time ahead in rational drug design and clinical application.

REFERENCES


