Design and synthesis of DNA-binding agents using dynamic combinatorial chemistry

Author:
Sherman Durai, Chandramathi

Publication Date:
2012

DOI:

License:

DESIGN AND SYNTHESIS OF

DNA-BINDING AGENTS USING

DYNAMIC COMBINATORIAL CHEMISTRY

This thesis is submitted in fulfilment
of the requirements for the degree of

Doctor of Philosophy

by

Chandramathi R Sherman Durai

School of Chemistry
The University of New South Wales
Sydney, Australia

July, 2012

PREFACE

This thesis is a summary of the work carried out in the School of Chemistry at The University of New South Wales, under the supervision of Professor Margaret M Harding between March 2008 and December 2011.

Except where reference is made in the text this thesis contains no material previously published or extracted in whole or in part from a thesis presented by me for a another degree or diploma. No other person's work has been used without due acknowledgement in the mail text of the thesis.

Sections of this work have been published elsewhere:

Targeting Nucleic Acids Using Dynamic Combinatorial Chemistry,
Sherman Durai, C. R.; Harding, M. M.
Aust. J. Chem. **2011**, *64*, 671-680.

This work has been presented at the following meetings:

Targeting DNA Using Dynamic Combinatorial Chemistry,
Harding, M. M.; Sherman Durai, C. R.; Abeysinghe, P. M.; Garner, J.
2nd Conference on Supramolecular Chemistry in New Zealand and Australia (SCiNZaA-2), Sydney, October, **2011**.

Targeting DNA with Dynamic Combinatorial Chemistry,
Sherman Durai, C. R.; Garner, J.; Harding, M. M.
RACI 31st Annual One day Organic Chemistry Symposium, Wollongong, December, **2010**.

Targeting DNA with Dynamic Combinatorial Chemistry,
Sherman Durai, C. R.; Bodkin, J. A.; Jeanette Stock; Harding, M. M.
RACI 23rd Conference on Organic Chemistry "ORGANIC08", Tasmania, December, **2008**.

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Signature:

Date:

COPYRIGHT STATEMENT

AUTHENTICITY STATEMENT

ACKNOWLEDGEMENTS

First and foremost, I thank God Almighty for his creation of this wonderful planet with amazing things in science to discover every second. It's been one of the most marvelous and challenging journeys in this exploration for the past three and half years. God Almighty has helped me immensely in every step in this path of exploration and has given me the strength and direction to move forward during obstacles. I thank him for all his support and guidance in this journey.

I thank my Supervisor Prof. Margaret M Harding for giving me the opportunity to work on Dynamic Combinatorial Chemistry at UNSW. It's been an honor to be her PhD student. I appreciate all her contributions of time, ideas and funding to make my PhD productive and exciting. The enthusiasm and passion she has for research was motivational and constructive for me, even during tough times in my research. I thank her for the valuable guidance and help throughout the course of the project. Also, I thank my co-supervisor Dr Pall Thordarson for his valuable comments on molecular modeling studies and on my thesis writing.

The award of UNSW TFS (Tuition Fee Remission Scholarship) and Faculty Research Stipend provided during the entire course of the work is greatly acknowledged.

I am thankful to the past and present members of the Harding research group. I owe my thanks to James for his help with molecular modeling and answering my queries during the DCL analysis and Mano for her valuable comments and

ABSTRACT

This thesis reports the application of dynamic combinatorial chemistry (DCC) to identify new DNA-binding compounds, and provide insight into the factors that are important in DNA recognition. The discovery of new binding motifs and DNA-binding compounds are important to understanding the rules of DNA recognition, and in the long term has the potential to assist on the design of new compounds with therapeutic and biotechnological applications.

DCC was used to generate heterocycles functionalised with amidines and carbohydrates in the presence of different oligonucleotides, in order to identify functionality that would increase the affinity of the heterocycles for DNA. Water soluble heterocycle building blocks of quinolines, imidazoles and naphthalamide were synthesised as thiol and disulfide derivatives. The thiol substitution on the quinoline ring was varied in these building blocks, and in the case of 4-thioquinoline, included an electron withdrawing trifluoromethyl group. The carbohydrates included glucose and aryl derivatives of the deoxysugar fucose, while the amidines included alkyl, benzyl and aryl groups. Flexible bisamine and bisthiol derivatives with the potential to form DNA bisintercalators were also studied.

DCC experiments were conducted using thiol disulfide chemistry in aqueous methanol using either GSSG/GSH at neutral pH or disulfide exchange at basic pH, conditions that have not been used previously for studies with nucleic acids. DCC experiments were conducted to assess the effect of overall charge, substitution of the quinoline ring, the importance of deoxysugars *versus* glucose, aryl and imidazole rings on DNA-binding. Analysis of both the DNA-

bound and unbound solutions provided important insights into the features that are important for DNA recognition and allowed the effect of subtle structural features on DNA-binding to be identified. Molecular visualisation of the selected DNA-bound and unbound compounds were used to rationalise the results and propose minor groove and intercalation binding motifs.

Flexible amino quinoline **Q4**-**Y** and guanidine disulfides **Q4**-**A1** and **Q4**-**A2** interacted with DNA. In contrast, neither of the aromatic guanidine disulfides **Q4**-**A2** and **Q4**-**A3** interacted with DNA, suggesting that the aryl groups may interfere with positioning of the amidine near the phosphate backbone. In the case of 7-trifluoromethyl-4-thioquinoline **Q2**, the thioglucose derivate **Q2**-**S1** was amplified with DNA, and the relative binding affinity **Q2**-**Cys**>**Q2**-**A1**>**Q2**-**S1** was determined. This result is consistent with proposed models of intercalation of the structurally related compound, chloroquine, with DNA. In contrast, the rigid arylfucose with 2-thio quinoline **Q1**-**S2**, was amplified in preference to the corresponding benzylic disulfide or glucose derivative. The fucose sugar was shown to be important for DNA-binding, consistent with DNA minor groove binding. The flexible bisthiol derivatives failed to produce any DNA-binding compounds, and experiments with naphthalimides were unsuccessful due to precipitation during the course of the assay.

Biostable mimics of the two lead compounds **Q1**-**S2** and **Q2**-**S1** were studied. The thioether analogue of **Q1**-**S2** interacted more strongly with DNA compared to the amide, consistent with minor groove binding. Both 1,4- and 1,5-triazole analogues of **Q2**-**S1** bound to DNA, with the similar binding profile of the 1,-4-triazole to the parent disulfide supporting intercalation as the binding mode.

TABLE OF CONTENTS

ABBREVIATIONS

DABCO	1,4-Diazabicyclo[2.2.2]octane
DCC	Dynamic Combinatorial Chemistry
DCL	Dynamic Combinatorial Library
DMAP	4-(Dimethylamino)pyridine
DMF	N,N-dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DTT^{OXD}/DTT^{RED}	Oxidised and reduced dithiothreitol
EDC	N-(3-Dimethylaminopropyl)-N-ethylcarbodiimide
EDTA	Ethylenediaminetetraacetate
ESI	Eletrospray ionization
GSSG/GSH	Oxidised and reduced glutathione
HOBt	1-Hydroxybenzotriazole
HPLC	High pressure liquid chromatography
HRMS	High resolution mass spectroscopy
LC	Liquid chromatography
MS	Mass spectroscopy
m.p.	Melting point
NMR	Nuclear magnetic resonance
PDA	Photodiode array
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidino phosphoniumhexafluorophosphate
RNA	Ribonucleic acid
RP	Reverse phase
Rt	Retention time
TEG	1-Dimethoxytrityloxy-3-O-(N-biotinyl-3-aminopropyl)triethyleneglycolglyceryl-2-O-(2-cyanoethyl)-(N,N-iisopropyl)phosphoramidite
TFA	Trifluoroacetic acid
TFO	Triplex-forming oligonucleotide
THF	Tetrahydrofuran
Tris	Tris(hydroxymethyl)aminomethane
UV-Vis	Ultraviolet-Visible

Chapter 1

Introduction

1.1 Structure of DNA

Nucleic acids DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are polymeric biological molecules essential for life. Nucleic acids are found in all living things, where they function in encoding, transmitting and expressing genetic information. The structure formed by double-stranded molecules of nucleic acids such as DNA are double helix.[1-2] The structure of the DNA double helix was described by Watson and Crick in 1953.[3] The critical feature of DNA is its linear array of monomers called nucleotides, which are made of an *N*-glycoside of a 2-deoxyribophosphate ester and one of four nitrogenous bases adenine (A), guanine (G), cytosine (C) and thymine (T).[1,3-6] The nucleotides encode all the genetic information necessary for cellular functioning and development.

The physical structure of DNA contains two antiparallel polynucleotide chains that form a double helical structure that involves two identical strands linking together by hydrogen bonding and π-π stacking interactions of the DNA base-pairs (Figure 1.1a).[3-4,7-9] The bases are stacked near the center of the helix, which provide considerable stability to the double helix. The driving force for the assembly of the structure could be due to the double helical structure enables the hydrophobic portion of the molecules present in the nitrogenous bases to avoid the aqueous solvent by stacking in the centre of the helix. The sugar and the phosphate groups are on the outside of the helix that forms a backbone for the DNA. The diameter of the double helix ranges from 2.0-2.4 nm and there are ten base pairs in each complete turn of the helix (3.4 nm); each base pair is thus twisted 360° relative to the preceding base pair in the molecule.[1-2]
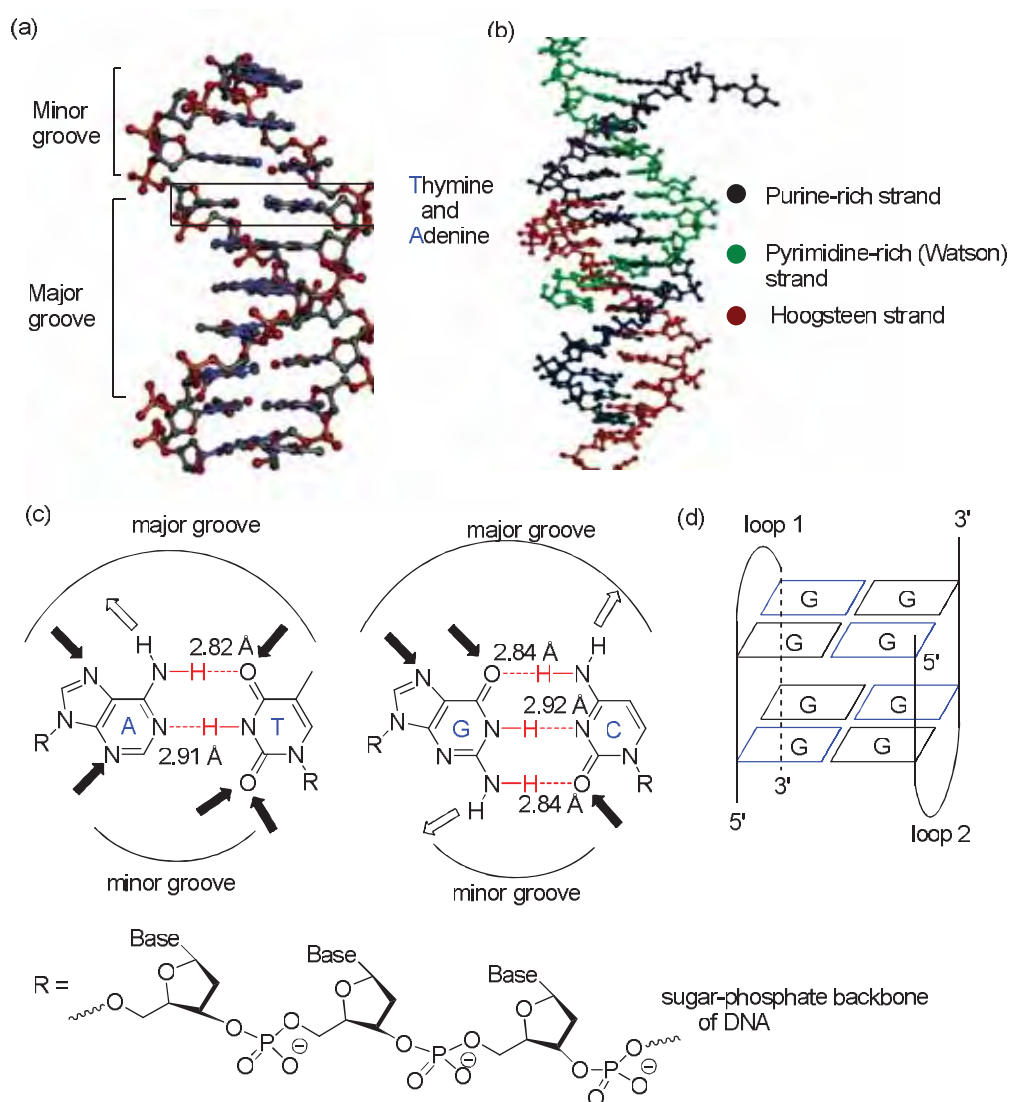
Figure 1.1 (a) Crystal structure of the duplex DNA d(ATATATATAT)₂ [PDB ID =3EY0][10] highlighting the major and the minor grooves and the bases thymine (T) and adenine (A), (b) the crystal structure of the triplex DNA [PDB ID=1D3R][11] highlighting the puring-rich strand (black), pyrimidine-rich (green) and Hoogsteen strand (red), (c) structures of Watson-Crick base pairs adenine (A); thymine (T) and guanine (G); cytosine (C) showing hydrogen bond donor (plane arrow) and hydrogen bond acceptor (bold arrow) in the major and minor grooves[1-2,12] and (d) G-quadruplex DNA.[13]

The negatively charged sugar-phosphate groups in the double helix are not equally arranged, and as a result, there are unequal sizes in the grooves, called the major and minor groove, the major groove is wider than the minor groove (Figure 1.1a).[14] The structure of DNA physically protects many of the nucleophilic and electrophilic sites on the nucleic bases from chemical modification, as these sites are protected from interactions with reagents by base stacking and hydrogen-bonding. The DNA helix is chiral, as the polynucleotide chains form a right-handed double helix due to the D-ribose sugar ring. The homo-chirality in the monomeric sugar building blocks thus leads to homo-chirality in the right-handed DNA secondary structure.[2,14-16]

Diffraction studies on heterogeneous DNA fibers have identified two distinct conformations for the DNA double helix. A-DNA is the favored structure at low humidity and high salt conditions, whereas at high humidity and low salt condition, the dominant conformation is B-DNA (Figure 1.1a).[5,16-19] However, the exact shape and dimensions of the double helix are highly dependent on the base sequence and environment and irregular shapes such as bulges and hairpins can be formed. Nucleic acid strands are able to generate intrastrand hairpin loops, which occur in single stranded DNA or RNA having about six bases in the loop.[1-2] The hairpin loops are formed by rapid unimolecular process and are thermodynamically less stable than the duplex DNA. DNA bulges are unpaired stretches of nucleotides, which are located within one strand of a nucleic acid duplex. Bulges are formed by hydrogen-bonded bases including canonical (Watson-Crick) and non-classical base pairs. However, the sizes of the bulges can vary from a single unpaired residue to several nucleotides, which form flexible extrusion from double helices, the interior loops and bulges are more stable than hairpin loops.[1,20]

3

Hence, while DNA is commonly represented in the B-DNA conformation with a regular repeating structure and groove sizes (Figure 1.1a), bulk DNA contains a highly variable structure. In addition, triple-stranded (or triplex) and four-stranded (or quadruplex) DNA can be formed with certain sequences under specific conditions. Triplex DNA is a structure of DNA in which a third oligonucleotide strand interacts with the major groove of DNA *via* hydrogen bonding to form a triple helix (Figure 1.1b). Quadruplex DNA are rich in guanine (G) and are capable of forming four stranded structure and are also called as G-quadruplexes or G4-DNA (Figure 1.1d).[2,13]

1.2 Forces stabilizing the DNA double helix

The hydrogen-bonded pairs that stabilize DNA are of specific type termed Watson-Crick base-pairs, which are specific hydrogen bonds either adenine (A) and thymine (T), or guanine (G) and cytosine (C) (Figure 1.1c). Hydrogen bonds are mainly electrostatic in character and are weakly directional. There are three hydrogen bonds in G:C base pairs and are separated by 2.84-2.92 Å, whereas the two hydrogen bonds in the A:T base pairs are separated by 2.82 and 2.91 Å.[1-2,20]

Stacking interactions are also highly important in stabilizing the double helical structure of DNA and in molecular recognition. Hunter and Sanders developed a model to explain the nature of π-π interactions on porphyrin systems. The π-π interactions occur in aromatic ring systems when the attractive interactions between the π-electrons and the σ-framework, called as π-σ favorable attractions, which leads to a coplanar geometry of DNA bases.[21] The geometry

observed during the attractive interactions between two porphyrin rings are,[21-22]

- The π-electron system of two neighbouring porphyrins are parallel with an interplanar separation of 3.4-3.6 Å;
- In the π-stacked porphyrins, the nitrogen-nitrogen axes are parallel;
- One porphyrin ring is offset relative to the other ring by 3-4 Å with respect to the nitrogen-nitrogen axes.

Similar to the stabilization energy provided by hydrogen bonding interactions, stacking interactions also provides energies of stabilization of the DNA helix. Stacking interactions are mainly contributed by electrostatic and van der Waals interactions, however, electrostatic interactions conclude the geometry of the interaction with the π-electron density above and below the planes of aromatic molecules.[21-22] Electrostatic interactions between the positively charged σ-framework sandwiched (face-to-face) between the two regions of negatively charged π-electron density are unfavorable due to the π-π repulsive forces. However, edge-to-face i.e. T-shaped perpendicular stacking geometries between the π-electrons and the σ-framework are attractive.[23]

Van der Waals interactions are weakly attractive interactions, which constrain the base pairs to remain in van der Waals contact. When, the electron distribution around the nuclei results in weak dipoles, as a result the atoms induce the dipoles in the neighboring atom and these dipoles cause a weak attractive interactions called van der Waals interactions. In double DNA helical structures, the interactions between the DNA nucleotides are controlled by van der Waals interactions and the characteristics associated with the geometrical

constraints due to sugar-phosphate backbone and interbase interactions due to stacking and base-pairing.[2]

Hydrophobic and van der Waals interactions are involved in the stacking interactions. Hydrophobic interactions are important for the DNA double helix and also for folding of proteins. These interactions illustrate the relations between water and the molecules which are not compatible with water. In solution, the hydrophobic interaction between nucleic acid bases significantly contributes to the stabilization of the DNA helix.[1-2,20] The hydrophobic interaction is mostly an entropic effect originating from the interference of hydrogen bonds between the molecules of water by the non-polar solute. Hydrocarbons are incapable of forming hydrogen bonds with water, introduction of such a non-hydrogen bonding surface into water causes disruption of the hydrogen bonding network between water molecules.

1.3 Molecular Recognition of DNA

The molecular recognition of DNA is fundamental to many biochemical processes related to transcription, regulation and gene expression.[24-26] The study of small molecules that interact with high affinity (high binding constant) and base sequence selectivity (G:C or A:T rich sequences) to duplex, triplex and quadruplex DNA, as well as hairpins, bulges and RNA loops has attracted significant interest due to the involvement of many of these structures in disease.[13,27-32] These molecules can interact with DNA using either covalent or non-covalent interactions.[1,33-35] Of most relevance to work in this thesis are the different types of non-covalent interactions, which can be classified as (i) electrostatic interactions (ii) groove binding and (iii)

intercalation. Many compounds interact with DNA using a combination of these interactions shown schematically in Figure 1.2, and illustrated with selected examples **1**-**4**. The chiral nature of the helix is important in DNA molecular recognition and as a result the chirality and shape of small molecules molecules are also important in molecular recognition.

1.3.1 Electrostatic Interactions

DNA is stabilized by electrostatic interactions that occur when positively charged ions like Na^+, K^+, Mg^{2+} and cations of organic amines interact with the sugar-phosphate DNA backbone. These interactions neutralize the phosphate charges and result in the release of condensed counterions. The ions retain their inner sphere water of hydration and move rapidly along the sugar-phosphate backbone of DNA. In general, multiple charged cations interact with DNA more strongly than monovalent cations.[1,35]

1.3.2 Groove Binding Interactions

As shown in Figure 1.1a, DNA contains two grooves, referred to as the major groove and the minor groove. Small molecules generally interact with the minor groove of DNA, while proteins typically interact with the major groove of DNA. Groove binding interactions involve direct interactions of the bound molecule with the edges of the base pairs in either the major or the minor groove of the nucleic acids. Major groove recognition by proteins generally involves cylindrical binding motif based on α-helices, and the size and the shape of the protein motif fits snugly into the major groove by stabilizing hydrogen-bonding interactions. In addition, the methyl group at the C5-

(a)

1

2

3

4

(b)

(c)
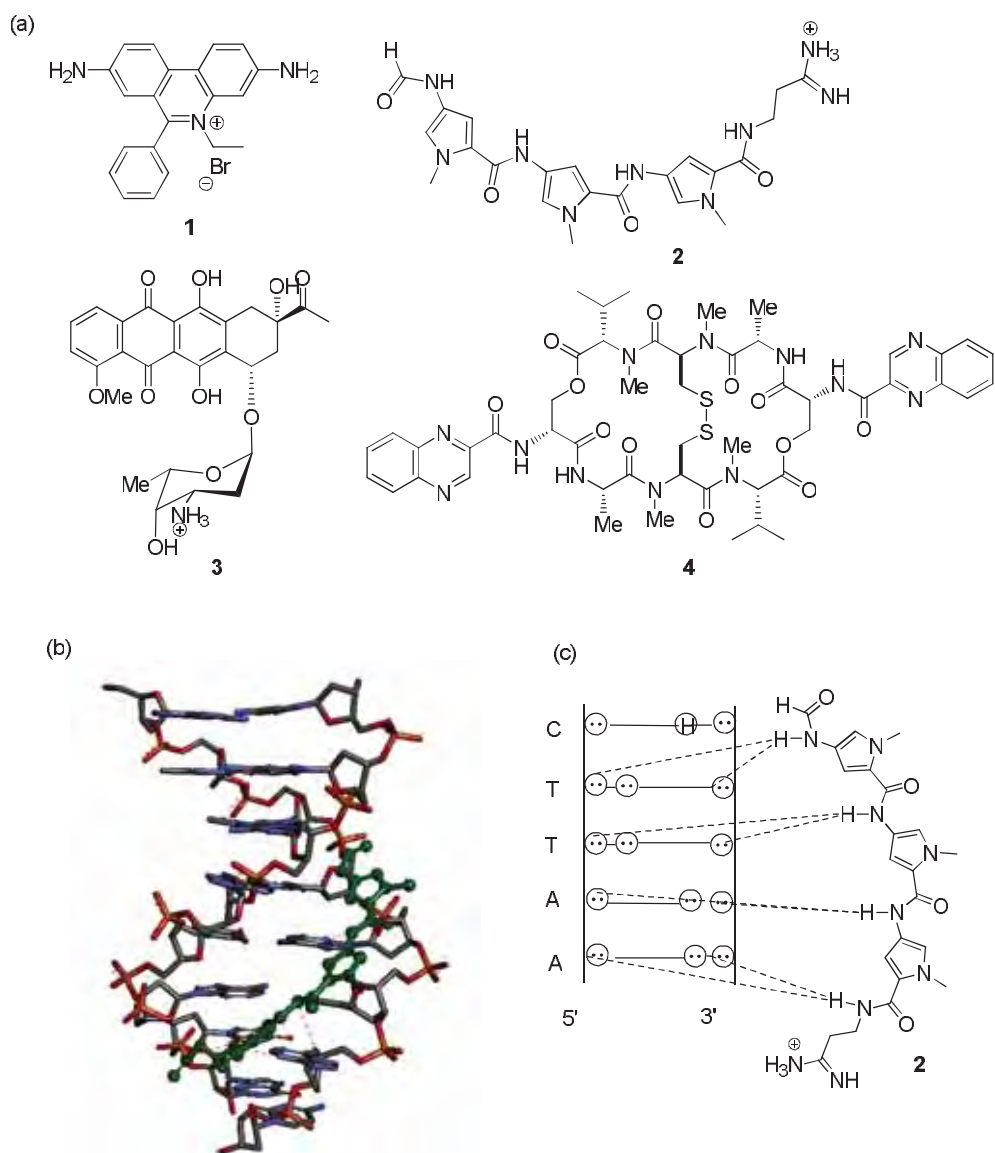
Figure 1.2 Examples of non-covalent DAN interactions shown by (a) intercalation: ethidium bromide **1**, minor groove binding: distamycin **2**, intercalation-groove binding: daunomycin **3** and bisintercalation:triostin A **4**, (b) crystal structure of distamycin **2** bound to d(GTATATAC)2 [PDB ID = 378D][36] distamycin is indicated in green and hydrogen bonds are indicated in pink and (c) schematic representation of the DNA-binding mode of distamycin **2** highlighting the hydrogen bonds in dashed lines and the circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines.[37]

position of thymine (T) can also participate in van der Waals interactions (Figure 1.1c).[38-39] Most synthetic molecules reported to date interact with DNA *via* minor groove. Distamycin **2**, (Figure 1.2b) is an example of a well-characterized DNA minor groove binder. The long and crescent shaped distamycin **2** antibiotic binds to the minor groove of A:T rich DNA sequences as a dimer, stabilized by hydrogen bonding between the amide groups and the A:T base pairs. Figure 1.2c shows the distamycin **2** binding with A:T rich minor groove sequences that form complexes with DNA in 1:1 stoichiometry.[37,40] The molecule makes close van der Waals contacts with the walls of the minor groove and all solvent is displaced from the groove at the binding site, the binding is also supported by the amidines functional groups *via* electrostatic interactions.[37,41] Structurally related minor groove binders share similar features, and contain small aromatic rings (e.g. pyrrole, furan or phenyl groups)[40,42-45] connected by amides or functional groups with partial rotational freedom, allowing the appropriate twist to complement the shape of the minor groove with the displacement of water molecule from the groove of DNA.[1,34,46] The features of typical minor groove binders include,[12,47]

- Sufficient flexibility to allow a conformation in which the overall shape fits the curvature of the minor groove;
- Positively charged functional groups to enhance electrostatic interactions;
- Appropriate hydrogen-bonding groups for sequence recognition of N3 (A) and O2 (T) and NH (G) and O (C) in the in minor groove (Figure 1.1c).

1.3.3 Intercalation

The most common way that small aromatic molecule interacts with DNA *via* intercalation. Intercalation occurs when cationic aromatic or planar molecules

containing two or more fused rings (example, compounds **1**, **3** and **4**, Figure 1.2a) insert between the DNA base pairs and are stabilized by aromatic stacking interactions with the DNA base pairs.[1,48] During intercalation, several changes in the shape and flexibility of DNA occurs, since two adjacent base-pairs must physically separate to accommodate the intercalated molecule. Hence the sugar-phosphate backbone is distorted, which results in a lengthening of the DNA double helix. Figure 1.2a shows an intensively studied cationic intercalator ethidium bromide **1**. Upon intercalation, ethidium unwinds DNA by about 26°, which leads to structural changes such as lengthening of the DNA strand or twisting of the base pairs, leading to functional changes, including inhibition of transcription, replication and DNA repair processes.[35,49]

DNA intercalators are an important class of chemotherapeutic agents. Daunomycin **3** (Figure 1.2a), which is a widely used anthracycline antitumor antibiotic binds to DNA *via* intercalation.[1,50-51] The cationic amino-sugar attached to the ring facilitates the well-defined fit of the molecule into the right handed minor groove and displaces water molecule and ions from it.[52-53] In addition the molecule is stabilized by hydrogen-bonding interactions involving hydroxyl and carbonyl groups at C9 of the daunomycin chromophore. The conformation of the amino-sugar attached to the aromatic ring of daunomycin and the conformation of the DNA are also significantly changed relative to the B-DNA helical structure and these changes facilitate the snug fit of the antibiotic into the right handed minor groove. The binding affinity of daunomycin **3** for DNA increases with increasing G:C sequence.[1,19,54] Most intercalators display either no binding preference or G:C base-pair preference. However, there are some examples of intercalators as a preference for A:T

base-pair. The G:C base-pair preference of intercalators are due to the larger instrinsic dipole moment of the G:C base- pair relative to A:T base-pairs.[1,20]

DNA bis-intercalators are also an important class of naturally occurring anti-tumor antibiotics, due to their increased binding affinity for DNA and slow off rates.[1,55-59] For example, triostin A **4**, a quinoxaline antitumor antibiotic binds to DNA by bisintercalation of the quinoxaline rings into DNA (Figure 1.2a), with the cyclic depsipeptide backbone located in the minor groove.[55,60-62] The two aromatic rings are oriented in an optimum configuration for bisintercalation and are separated by 10-11 Å and can accommodate two base-pairs between the rings for binding. The relatively rigid geometry of the cyclic peptide plays an important role in the biological activity of the antibiotic, by preorganising the aromatic rings for bisintercalation and makes specific contacts with the minor groove of A:T rich DNA sequence.[63-64]

1.4 Synthetic DNA-Binding Molecules

Medicinal and synthetic chemists have incorporated the features present in natural products to design new DNA-binding molecules by linking heterocycles with chains of varying length and flexibility in order to elucidate the rules that govern nucleic acid sequence recognition.[13,65-66] The design of these small molecule DNA-binding compounds has been guided by structure-activity studies, combined with molecular modeling or examination of X-Ray or NMR data of drug-oligonucleotide complexes. Incremental changes to molecules have been made by modifying rigidity and/or positioning of functional groups, and the compounds have been synthesized and their properties have been

examined.[59,67] Based on these results, iterative changes have been made to structures in order to refine the design and alter DNA-binding characteristics.

Using these traditional approaches, significant progress has been made by modifying rigidity and/or positioning of functional groups and also repetitive changes have been made to structures in order to refine the design and alter DNA-binding characteristics with predictable binding affinity and sequence selectivity that bind to the minor groove of DNA (discussed in detail in Chapter 4, section 4.1).[68] However, this traditional approach relies on total synthesis of complicated molecules and is unable to take into account the highly sequence dependent conformational flexibility of DNA, long-range effects, the involvement of bound water molecules, and the interplay between hydrophobicity and electrostatic potentials along the helix.

Several recent examples have highlighted the difficulty in designing molecules with predictable DNA-binding affinity from X-ray crystal structures, and the identification of new classes of minor groove binders that suggest that new modes of DNA-binding are yet to be discovered. For example, Figure 1.3 shows the structures, DNA-binding constants and biological activities of two synthetic anthracycline bisintercalators WP631 and WP762, which differ by the linker L.[59,67] WP631 was designed based on the X-ray crystal structure of two daunomycin monomers bound to a DNA oligonucleotide, which suggested that crosslinking the daunomycin monomers with a *para*-disubstituted phenyl linker would generate a bisintercalator with increased affinity for DNA.[59]

While strong DNA-binding and biological activity of WP631 were observed, the *meta*-linked derivative WP762 showed picomolar DNA-binding affinity and higher biological activity. The enhanced DNA-binding of WP762 compared with WP631 was attributed to the positioning of the daunosamine sugar near one side of the minor groove, which added more favorable van der Waals interactions with the DNA backbone. These interactions also resulted in a decrease in the width of the minor groove. This example illustrates the challenges in predicting the changes in binding affinity from structural changes to small molecules that bind to DNA. The flexibility of the double-helix does not allow precise positioning of functional groups to enhance van
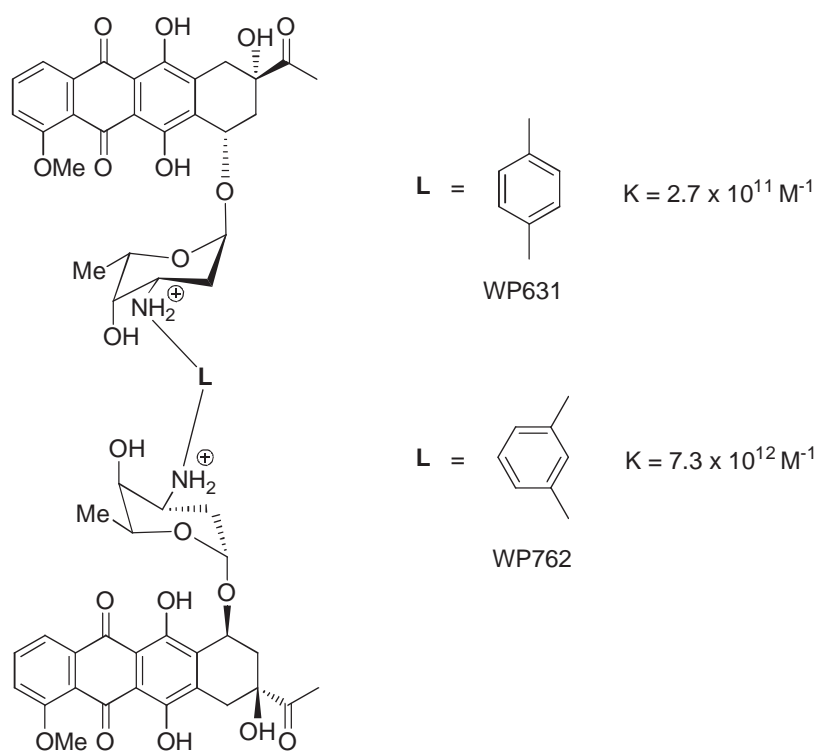
Figure 1.3 Synthetic anthracycline bisintercalators showing the effect of the linker L on DNA-binding affinity and biological activity.[59,67]

der Waals, hydrophobic and hydrogen bonding interactions, as the groove sizes and conformation of DNA can change on binding to maximize binding interactions.

A number of recent studies of heterocyclic amidines have provided unusual structures that do not fit the features of classical minor groove binders in terms of shape and molecular recognition of DNA, which suggest that minor groove binding molecules with new structural features are yet to be discovered.[12,69-70] For example, the linear the bisamidines RT29, DB950 are strong minor groove binders (Figure 1.4b).[12,47] The benzimidazole diamidine RT29 has a highly twisted diphenyl ether linkage that has too much curvature to fit the minor groove of DNA. In contrast to the distamycin **2** (Figure 1.3b), Hoechst 33258 and DB75 (Figure 1.4a) which are crescent shaped and match the curvature of the minor groove,[12,71-74]
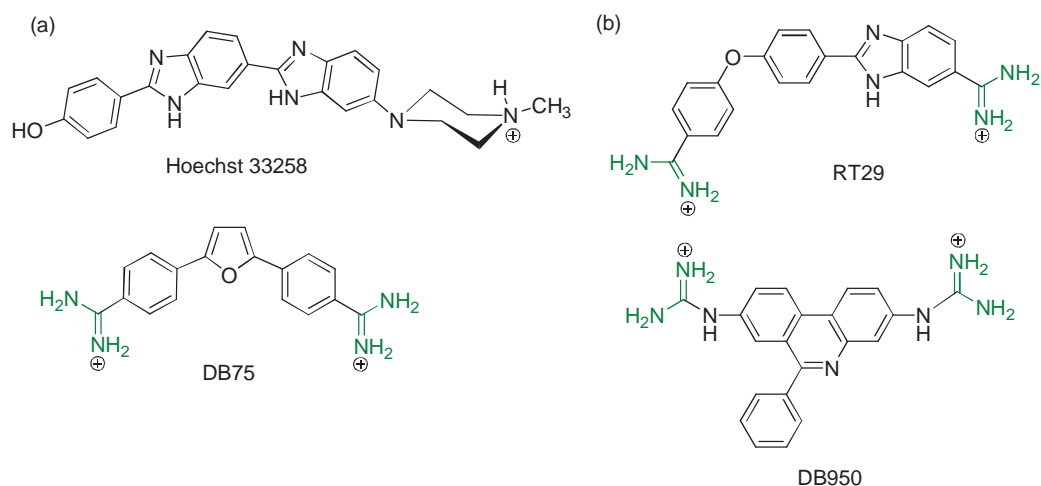
Figure 1.4 Examples of (a) "classic" DNA minor groove binders that match the curvature of the groove[47] and (b) bent and planar DNA minor groove binders.[47,75]

The compound displays strong and selective binding to A:T sequences, by undergoing significant conformational changes to enhance minor groove interactions. In addition, bound water molecules are essential for strong DNA-binding. Similarly, the replacement of the two amino groups in ethidium bromide **1** (Figure 1.2a) with diguanidine functional groups converted ethidium from a DNA intercalator to a strong AT selective DNA minor groove binder.[75] The DNA interaction of DB950 was ~ 50 times greater than that of the ethidium **1** and was comparable with the known minor groove binder DB75. The authors concluded that conversion of a well-known intercalating agent into a minor groove binder by an amine guanidine substitution may provide an interesting strategy to optimize the targeting of DNA sequences using small molecules.[75]

1.5 Dynamic Combinatorial Chemistry

Dynamic Combinatorial Chemistry (DCC) is a powerful method for the identification of novel ligands for molecular recognition. In DCC, building blocks that incorporate functional groups that undergo reversible reactions are equilibrated to generate a dynamic combinatorial library (DCL), which comprises all possible combinations of the building blocks that interchange constantly (Figure 1.5).[76-78] At equilibrium, all library members are interconverted to give a distribution of a mixture of library members, which are under thermodynamic control. The amounts of each library member are directly related to their thermodynamic stability.[78-81] Addition of a guest molecule that can selectively bind to the receptor molecule present in the library. After quenching the equilibrium, the successful receptor molecule is

identified and isolated. The scope and applications of DCC are the subject of a number of excellent reviews.[76,78-79,82-87]

A schematic of the DCC process is shown in Figure 1.5. The DCC process has been explained in three steps:[88-91]

- Selection of the building blocks depends on the functional groups such as aldehydes and amines, thiols, etc.(Table 1.1) that undergo reversible reactions;
- Under the established library condition, the selected building blocks are allowed to equilibrate to give the dynamic combinatorial library (DCL) of all possible combinations;
- The receptor molecule selects the best binder, which depends on the binding strength of the receptor molecule present in the DCL.
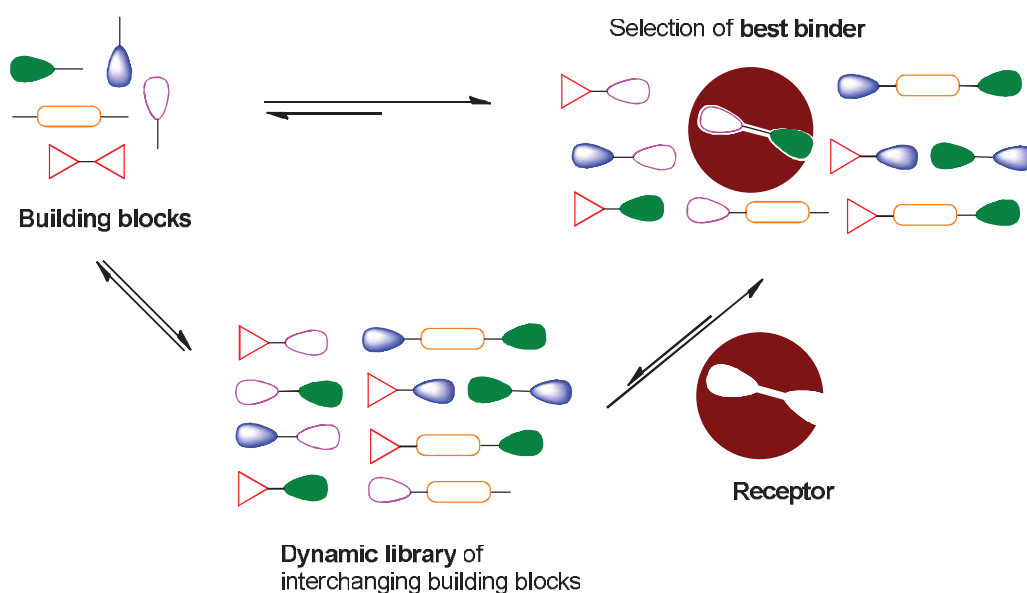
Figure 1.5 Schematic of DCC.

1.5.1 Reversible Reactions in DCC

Reversibility is the feature of DCC, which mediates exchange of the building blocks between the different library members. Hence, the building blocks which have reversible functional group undergo DCC to generate different combinations of DCL mixtures.[76,89] The reversible reaction needs to meet a number of requirements[78] such as:

- Be reversible on a defined time scale for a given experiment;

- Be compatible with the experimental conditions, such as functional groups present in the building blocks, pH, solvent and template;

- Be performed under mild (temperature, pressure, concentration) conditions;

- Ensure the solubility of all library members at equilibrium; ideally all library members should be isoenergetic in order to prevent the production of reaction mixtures.

Table 1.1 summarizes potential reversible reactions[76,88,91] that could be used in DCC. Thiol-disulfide,[92-96] imines,[97-98] Diels-Alder,[99-100] hydrazone chemistry,[101-102] aminal[103] and thioester formation,[104-105] imines and metal coordination[106-107] and metathesis[108] (in the case of metathesis reactions, the reactivity is highly dependent on the substrate R1 and R2 present in the alkene system) DCLs have been reported in the literature. Eliseev and Lehn reported the combined study on hydrozone chemistry with metal-ligand coordination involving a labile Co(II) center that could be oxidized to form a stable and kinetically inert Co(III) species.[107] Nitschke *et al.* extensively explored systems that combine both imine and metal-ligand interactions[109] and also reported a combination of three distinct dynamic linkages having three separate reversible chemistries such as disulfides, imine and metal

coordination were shown to be capable of simultaneous dyamic exchange within a single system.[106] Though Table 1.1 reports all possible literature DCC examples, only a few DCC reactions such as thiol/disulfide[92-96] and metal-ligand coordination[110-111] reactions have been studied in the presence of biomolecules.

Table 1.1 Dynamic processes for potential use in DCC systems.[76,88,91]

1.5.2 Applications of DCC

DCC is a powerful tool for synthesizing and identifying thermodynamic minima in complex mixtures and has been applied in the development of receptors for guest molecules.[78,106] Recently, high levels of enantioselectivity have been identified using deracemization, a process of transforming a racemate into enantiopure products in the presence of two receptors for (-) cytidine or (-)-2-thiocytidine by hydrazone-based dipeptides.[112] DCC offers a number of advantages in the generation and investigation of compounds. The first advantage is that the interactions can be used to extend the diversity of the library, the rapid generation and investigation of compounds. The second advantage is in the process of self screening i.e. analyzing in the presence of a target that can be effectively performed with pre-equilibrated libraries, which involves generating the dynamic libraries under reversible conditions.

In recent years, DCC has been actively pursued to identify and develop new synthetic receptors,[81,113-114] for biologically active molecules[88,95,115] and catalysts[116-117] metallo-organic self-assembly for the formation of covalent (carbon-heteroatom) and dative (heteroatom-metal) bonds simultaneously,[118] for the discovery of donor-acceptor [2]-catenanes[119-120] and [2]-rotaxanes,[121] for the generation of disulfide and hydrazone exchange simultaneously,[122] water-soluble disulfide-linked cages,[123] macrocyclic disulfides,[124] polymer-supported cationic templates using anionic hosts.[125-127] In recent years, DCC has been utilized for analytical purpose. The concentrations of the DCL members depend on the physical and chemical environment (pH, solvent, concentration, etc.) of the particular system. Very recently, Severin *et al.* have succeeded in using DCLs as colorimetric sensors for peptides[128-129] and nucleotides.[130] In contrast to idependent sensor units, a DCL sensor is comprised of compounds

that are connected by reversible exchange functional groups. In addition, the various sensors of an array have to be analysed separately, whereas a single UV-Vis measurement is sufficient for the DCL sensors. More recently, Lehn *et al*.[131] have reported the synthesis of dynamic polymers with oligosaccharides, called as glycodynamers, which were found to be strongly fluorescent and are applied in the field of biosensing.

Figure 1.6 shows an example of disulfide DCL and illustrates the power of this method to identify unique molecules that would be difficult to synthesize using traditional host-guest methods. A dynamic library of 45 different macrocyclic disulfide species of unique mass were formed,[78,81,132] when three different aromatic dithiol building blocks were allowed to oxidize reversibly. From HPLC and mass spectrometry techniques, only two major species **5** and **6** were identified in the library, when two different guest molecules, 2-methyliso-quinolinium iodide **7a** and *N*-methylated morphine **7b** respectively, were added into the DCL. The amplification was effectively quantitative and the templated macro-cyclization was under kinetically and thermodynamically control. The macro-cycle **5** exhibited no catalytic activity, whereas, the larger macro-cycle **6** induced ~ 10-fold increase in the rate of the reaction.[78,81]
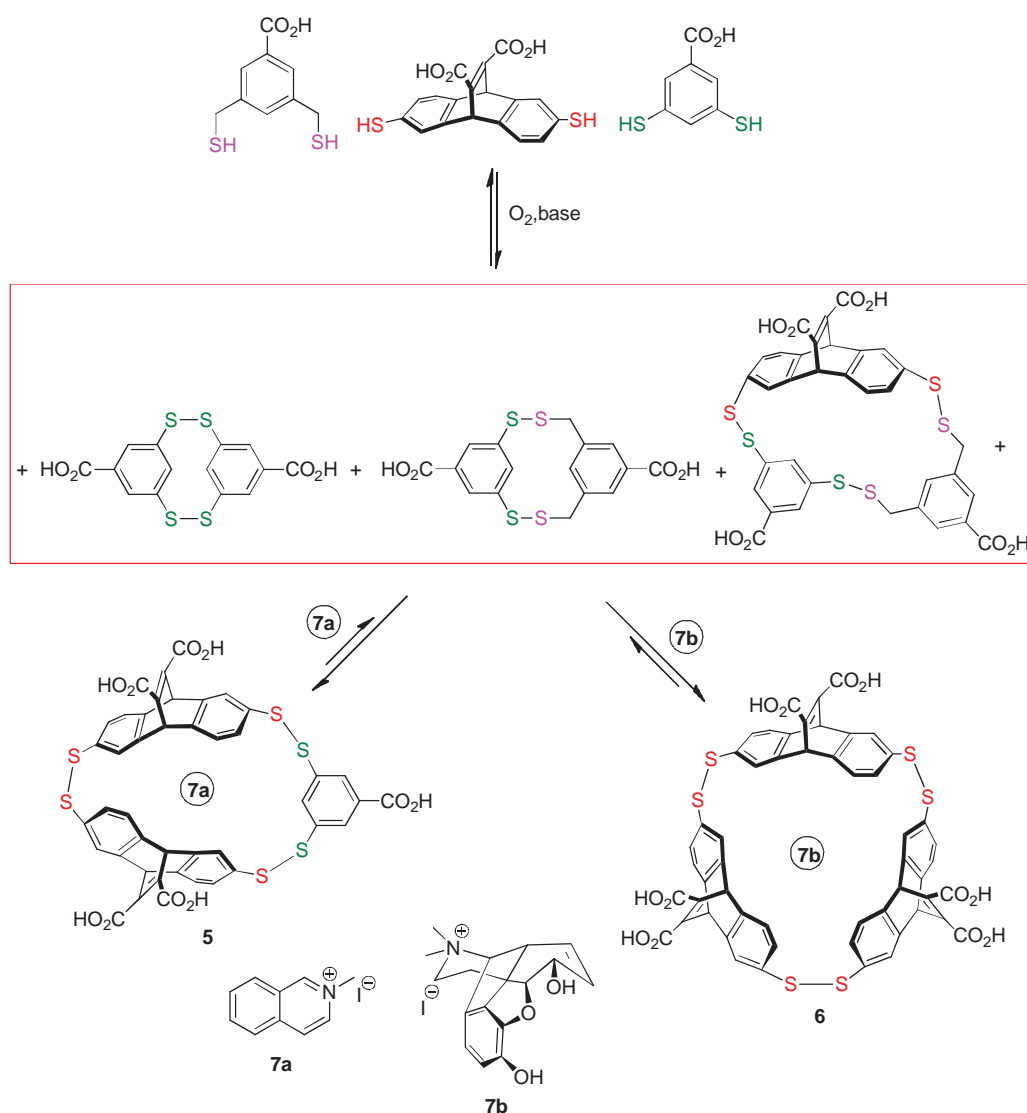
Figure 1.6 Example of a thiol-disulfide DCL used to identify **5** and **6** as macrocyclic hosts for 2-methylisoquinolinium iodide **7a** and *N*-methylmorpholine **7b** respectively.[132]

1.6 Targeting Nucleic Acids Using DCC

To date, there have been limited applications of DCC for the identification of new molecules that interact with nucleic acids. While a large number of DCC studies have been reported, the great majority of these reports have been conducted in organic solvents, under reaction conditions that cannot be used

in the presence of DNA or water. Table 1.2 summarizes the reported applications of DCC with nucleic acid targets and is discussed in detail in the following section 1.6.2. Studies have been performed to identify small molecules with affinity for duplex DNA, quadruplex DNA and RNA. In addition, several examples of DCC applied to identify modified oligonucleotides that are stabilized by appended groups have been reported.

As shown in Table 1.2, almost all examples have utilized the reversible thiol-disulfide oxidation reaction, which can generally be conducted in aqueous media, in the physiological pH range, initiated and modulated by oxidized and reduced glutathione (GSSG/GSH).

Table 1.2   Summary of DCC reactions and conditions used to study recognition of nucleic acid targets.

Nucleic acid	Sequence		DCC Reactions and Conditions			Ref
	Target	Control	Start. Materials	Buffer	Initiation	
	Poly d(AT)	Poly d(AT) in the absence of Zn^{2+}	Salicyl-aldimines Zn^{2+}	Tris.HCl, KCl, 1% DMSO (pH 7.5)	Incubate 25 °C	110,133
	[A]d(CTTTTATTTTG).(GAAAATAAAAC)	random duplex; sequence not given	R-SH	Tris.HCl, KCl (pH 7.4)	GSSG/GSH	93
Duplex DNA	d(CGCGAAATTTCGCG).(GCGCTTTAAAGCGC) & d(CGTACGGCCGTACG).(GCATGCCGGCATGC)	-------	R-SH R$_1$-SS-R$_2$	Tris.HCl, NaCl, MgCl$_2$ (pH 7.4)	GSSG/GSH	94
	[B]d(CCATGATATC).(GGTACTATAG)	d(TCTAGACGTC).(AGATCTGCAG)	R-S-resin	Phosphate (pH 7.4)	R-SH	134
	[A]d(GTTAGG)$_5$	[A]d(AGTTAG)$_5$	R-SH	Tris.HCl, KCl (pH 7.4)	GSSG/GSH	92
	[A]d(GTTAGG)$_5$	random duplex; sequence not given	R-SH	Tris.HCl, KCl (pH 7.4)	GSSG/GSH	93
Quadruplex DNA	[A]d (CGGGCGGGCGGCGCGAGGGAGGGGG) & [A]d(TGAGGGTGGGTAGGGTGGGTAA)	[A]d(GGGCATAGTGCGTGGGCGTTAGC). (CCGTATCACGCACCCG CAATCG)	R-SH	Tris.HCl, KCl (pH 7.4)	GSSG/GSH	95
	[A]d(GTTAGG)$_5$	[A]d(AGTTAG)$_5$	R$_1$-SS-R$_2$	Tris.HCl, KCl (pH 7.4)	Cystamine hydrochloride	96

23

RNA	UAGUCUUUCGAGACUA	TAGTCTTTCGAGACTA	Salicylamides, Cu^{2+}	H15-Mg[C] (pH 7.5)	Equilibrium dialysis	111
	[D]UUUUUAGGGAAGAAGAUCUGGCCUUCCCA CAAGGGAAGGCCAGGGAAU	RNA in the presence of R-S-resin only	R-S-resin (Cys S-[t]Bu protected)	Phosphate (pH 7.4)	R-SH	115
	[D]CCG(CUG)$_{10}$CGG	[E]GGG(CUG)$_{109}$GGG	R-S-resin (Cys S-[t]Bu protected)	Phosphate (pH 7.2) $MgCl_2$	R-SH	135
	[F](ACGCGU),(ACGCGU)	[F]TTTCGU	RCHO, Oligo-NH_2	Phosphate (pH 8.0)	Equilibrium dialysis	136
Oligo Conjugates	CACUGGGAUC	[G]GAUCCCAGUG	RCHO, oligo-NH_2	Phosphate (pH 6.0), NaCl, KCl, $MgCl_2$	Equilibrium dialysis	137
	CCAGAUUUGAGCCUGGGAGCUCUCUGG	GGGAGGACGAAGCGGACGC-AGAAGACACGCCCGA	RCHO, RNH_2	Phosphate (pH 6.0), NaCl, KCl, $MgCl_2$	Incubate	138

[A] 5'-Biotinylated oligonucleotide
[B] 5'-TAMRA (tetramethyl-6-carboxyrhodamine) labeled with a six-carbon spacer separating DNA and fluorophore
[C] Na-HEPES (25 mM); HEPES (25 mM); KCl (135 mM); MgCl$_2$ (15 mM) at pH 7.5
[D] Labelled at 5'-end with Cy-3 water-soluble cyanine dye
[E] Labelled with Fluorescein Amidite dye
[F] Oligonucleotide strands contain 2'-deoxy-2'-aminouridines at 3' ends
[G] Oligonucleotidestrands contain 2'-deoxy-2'-aminouridines at U3 & U9

1.6.1 DCC Reaction Conditions and Analysis

Two methods have been commonly used to detect and analyze the amplified products in DCCs with nucleic acids (Figure 1.7). In the most common method (Figure 1.7a), the biotinylated oligonucleotide target molecule is immobilized onto streptavidin functionalized magnetic beads, by the strong biotin-streptavidin interaction, allowing ready separation of the DNA from the DCL members.[92-93] Upon denaturing the DNA and removing the beads, the compounds bound to the DNA are identified. In the second method, termed resin-bound DCC (RB-DCC) (Figure 1.7b), a library of TentaGel S resin-bound monomers
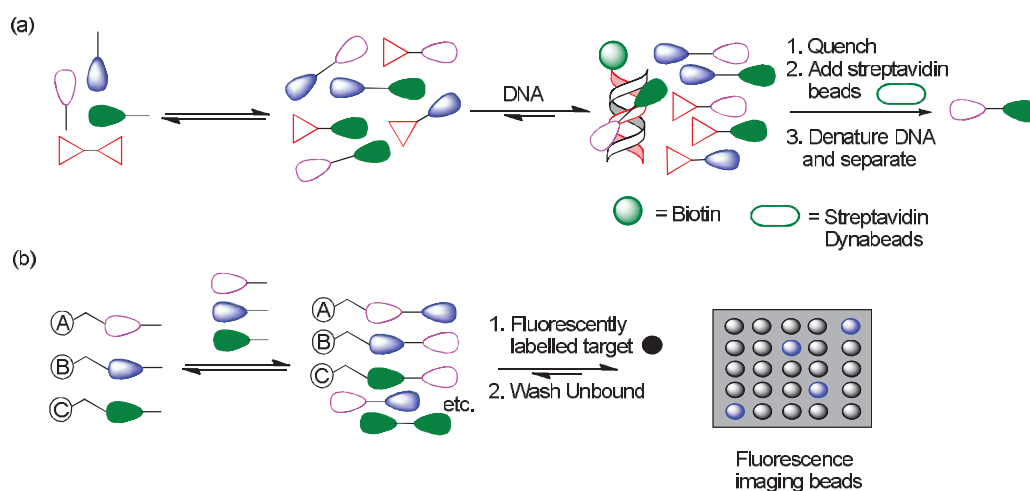
Figure 1.7 Schematic of formation of a DCL library and detection of amplified products in the presence of a nucleic acid target using (a) biotinylated probes and streptavidin beads;[92-93] and (b) RB-DCC.[134]

are combined with an identical library of monomers in solution to generate a resin-bound DCL.[134] When the fluorescently tagged target oligonucleotide was added, the reaction was allowed to equilibrate to undergo exchange reaction.

After equilibration, the resin was drained under vacuum and washed with buffer, then the resin was immediately positioned on a microscope slide and subjected to exposure under a fluorescent filter provides ready identification of the selected library members. An important feature of RB-DCC is that the solution phase and resin-bound components are in competition for target binding, which needs to be taken into account in the analysis of the beads, as diminished fluorescence may be a result of a strong solution binder being washed out of solution. Equilibrium dialysis has also been used in DCC studies involving metal ions with both DNA and RNA.[111]

1.6.2 Studies with Nucleic Acids

(a) Duplex DNA

Miller and colleagues were the first group to report the selection and amplification of novel DNA-binding compounds using DCC.[110,133] The DCL of 36 metal complexes were generated from a library of salicylaldimine ligands and zinc (II), and was performed in 10 mM Tris.HCl, 100 mM KCl, 1% DMSO at pH 7.5. The DCL was screened against poly d(AT) immobilised on a cellulose resin. While this study was important in establishing the identification of novel DNA-binding compounds, the formation of unstable imines in water, partial hydrolysis of the metal complexes and participation of the buffer in the reactions complicated the analysis of the results.

Balasubramanian and colleagues designed several thiol functionalised polyamides **8**, **9** and **10** based on distamycin **2** (Figure 1.8) with the goal of identifying the optimum number of heteroaromatic units for binding to an AT

rich duplex DNA sequence identified in the promoter region of oncogene c-kit, a possible site for intervention of transcriptional regulation.[93] Distamycin like polyamides, which contains *N*-methylpyrrole and *N*-methylimidazole functional groups can bind to the minor groove of double-stranded DNA with an affinity similar to naturally occurring DNA-binding proteins in a sequence–specific manner.[65,139-140] In the presence of an 11-mer duplex, the heterodisulfide formed from oxidation of **9** and **10** and the homodisulfide of **10** were amplified.
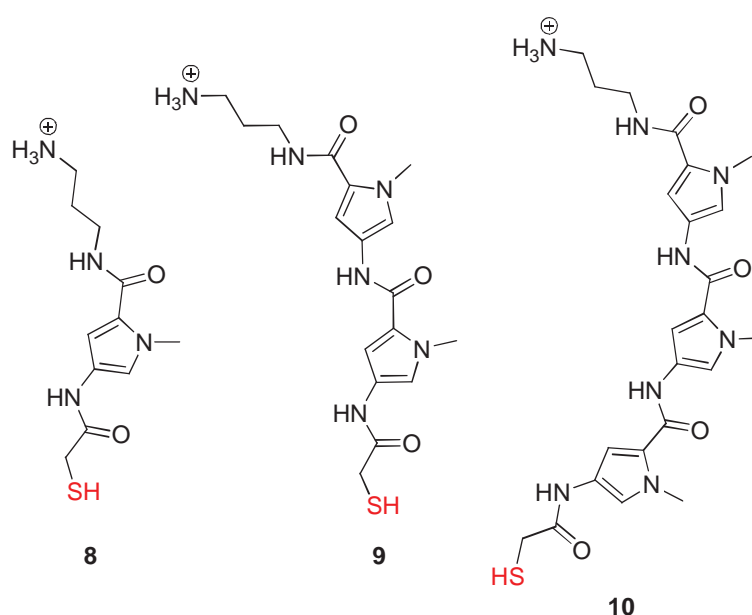
Figure 1.8 Structures of thiol-functionalised polyamides designed to mimic distamycin **2**.[93]

Thiol derivatives of Hoechst33258, a known DNA minor groove binder with high A:T selectivity were studied using an equilibrium shift assay.[94] Thiol, carboxylated and alkylated derivatives of Hoechst33258 as well as the glutathione derivatives were studied and the DCL mixtures were analyzed in the presence of DNA sequences containing an A_3T_3 binding motif. The ligands connecting Bis-Hoechst33258 units showed selective binding towards the DNA

sequence with two A_3T_3 binding sites as well as three-way junction DNA. Hoechst 33258, a rigid molecule (Figure 1.4a) binds more strongly with A:T rich sequence in the minor groove than G:C sequence. The DNA-binding of Hoechst 33258 requires a minimum of four consecutive A:T base pairs. The molecule with a crescent shaped fits the curvature of the minor groove by van der Waals interactions between the benzene rings and the backbone of CH and CH_2 groups of deoxyribose in DNA. Hydrophobic interaction of the flanking phenol and *N*-methylpiperazine rings with the walls of the minor groove makes steric interactions at the intercalative site; hence, the molecule prefers minor groove binding.[68]

McNaughton and Miller illustrated the concept of RB-DCC by selecting DNA-bisintercalators based on the naturally occurring quinoxaline depsipeptide triostin A **4** (Figure 1.2d).[134] A number of studies of synthetic variants have shown that the composition of the depsipeptide linker has a profound effect on the sequence selectivity of the bisintercalators.[55] A family of nine Cys-containing quinoline tripeptides (Figure 1.9a) was oxidized to form a library of 45 unique disulfides dimers. The library was screened for binding against a DNA sequence reported to be preferably bound by triostin A (5'-TCTAGACGTC-3') and a sequence reported to be preferentially bound by a synthetic analogue(5'-CCATGATATC-3'). Both oligonucleotide sequences were labeled at the 5'-end with the fluorophore tetramethyl-6-carboxyrhodamin (TAMRA). The serine (Ser) bisquinoline disulfide **11** (Figure 1.9b) was identified as a lead compound, and further independent DNA-binding experiments confirmed the high binding affinity of **11** for DNA. While this report demonstrated the concept of RB-DCC with a relatively small library of components, the authors
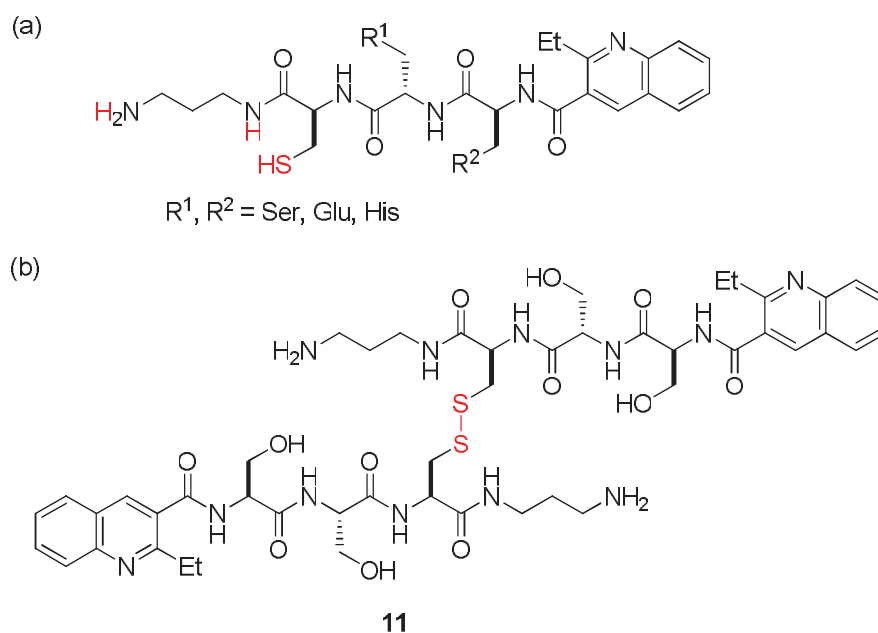
Figure 1.9 (a) Peptide mimics of Triostin A **4** used for generation of a DCL of solution and resin-bound disulfides; and (b) highest affinity ligand **11** for d(CT$_4$AT$_4$G).(GA$_4$TA$_4$C) identified using RB-DCC.[134]

highlighted the potential of RB-DCC for significantly larger libraries by the use of microarrays and beads as mixtures using an encoding system.[134] This approach has been used to generate a library of around 11,000 members for detection of RNA-selective small molecules[115] (see later section on RNA).

(b) Quadruplex DNA

The identification of small molecules that bind selectively and with strong affinity to quadruplex DNA has attracted significant interest in recent years.[29-30,141-142] G-quadruplex formation has been linked to telomer formation and has implications in cancer biology and ageing and the stabilization of quadruplexes has the potential to control gene expression.[29-30,32]

The potential of combining a number of distinct recognition units to recognise a human telomeric quadruplex formed from the deoxyoligonucleotide-(GTTACG)$_5$ was investigated using the thiols shown in Figure 1.10.[92] The hydrophobic acridone unit **12** was designed to intercalate in the terminal G tetrad of the quadruplex, and the thiol derivative of the tetrapeptide FRHR **13** has been reported to have quadruplex recognition properties. In the presence of GSH, a library of nine species were generated, including adducts of **12** and

12

13
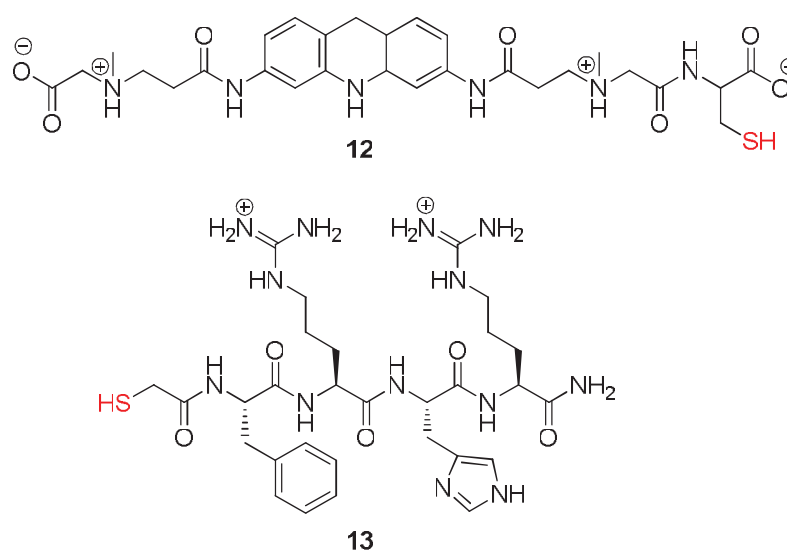
Figure 1.10 Acridone **12** and peptide thiol building block **13** used to generate a DCC in the presence of GSH and quadruplex DNA.[92]

13 with GSH. The disulfide formed from the oxidation of **12** with **13**, as well as the hexapeptide dimer of **13**, showed greater binding affinity with G-quadruplex DNA than other species present in the library. The amplification of the hexapeptide is particularly an interesting result, as there are no other reports of short peptides that bind quadruplex DNA with high affinity. Based on these results, it was proposed that the binding of **12** and **13** could be mediated by acridone π-π interactions with the top tetrad of the quadruplex

and quadruplex-loop/groove interactions with the appended peptide. Also, the molecular modeling studies suggested that the amide-linked acridone-peptide conjugates with the quadruplex.

The DNA-binding properties of DCLs generated from thiol functionalised polyamides based on distamycin **2** (see Figure 1.8) was also contrasted with the quadruplex DNA.[93] These experiments showed that the homodimers formed from **9** and **10** have a much higher greater affinity for duplex DNA than for quadruplex DNA.

DCC has provided important insight into the requirements for specificity in G-quadruplex recognition by oxazole-peptide macrocycles.[95] Based on previous results that indicated both the number and length of simple alkylamine side chains appended to the macrocycle are strong determinants of quadruplex affinity,[143] the library building blocks included derivatives of *p*-benzylic thiols and neutral carbohydrate derivatives with different potential for H-bonding and electrostatic interactions (Figure 1.11). The DCL was screened against two intramolecular quadruplex forming sequences (c-Kit21, c-Myc22) and a 22-mer duplex DNA for comparison. In the presence of c-Kit21 and c-Myc22, and the charged building blocks (Figure 1.11b), the two guanidinium disulfides formed from **14** and **15**, **14** and **16** were amplified. Although all the side chains are positively charged at physiological pH, the DNA-binding of **14** and **15**, **14** and **16** were attributed not only by electrostatic interaction between the charged molecule and the polyanionic DNA target, demonstrating that the geometry and/or the hydrogen bonding potential of the side chains play a role, and it is not just overall charge that determines affinity. Of interest was the

fact that under the same conditions in the presence of DNA there was no evidence for any binding. In the case of the carbohydrate building blocks (Figure 1.11c), there was also no perturbation of the DCL composition in the
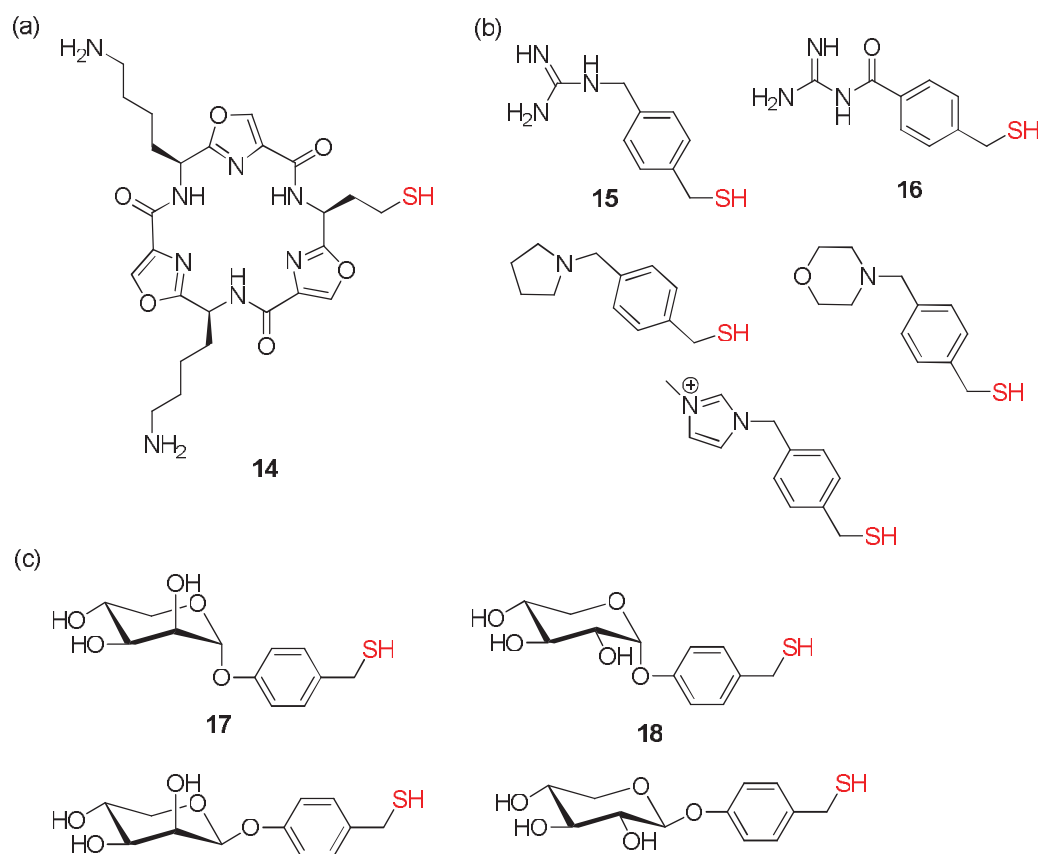
Figure 1.11 Structures of (a) oxazole-based peptide macrocycle **14**, (b) cationic benzylic thiols and (c) neutral carboydrate benzylic thiols used to generate a DCL for recognition of quadruplex DNA targets.[95]

presence of DNA. However, the disulfide formed from **14** and **17** was most strongly amplified in the presence of c-Kit21. The disulfides **14** and **17** as well as the disulfide formed from **14** and **18** bound to c-Myc22 with similar K_d values. The DCL studies with oxazole-based peptide macrocycle showed that

the subtle chemical and stereochemical variations in the positively charged and also in the carbohydrate side chains altered the affinity of the ligand for a particular quadruplex target, hence leads to differential recognition of G-quadruplex DNA.

In contrast to the examples above, Nielsen and Ulven employed disulfide scrambling to generate a DCL by the use of a central scaffold designed to carry and equilibrate with several side chains (Figure 1.12).[96] Three aromatic
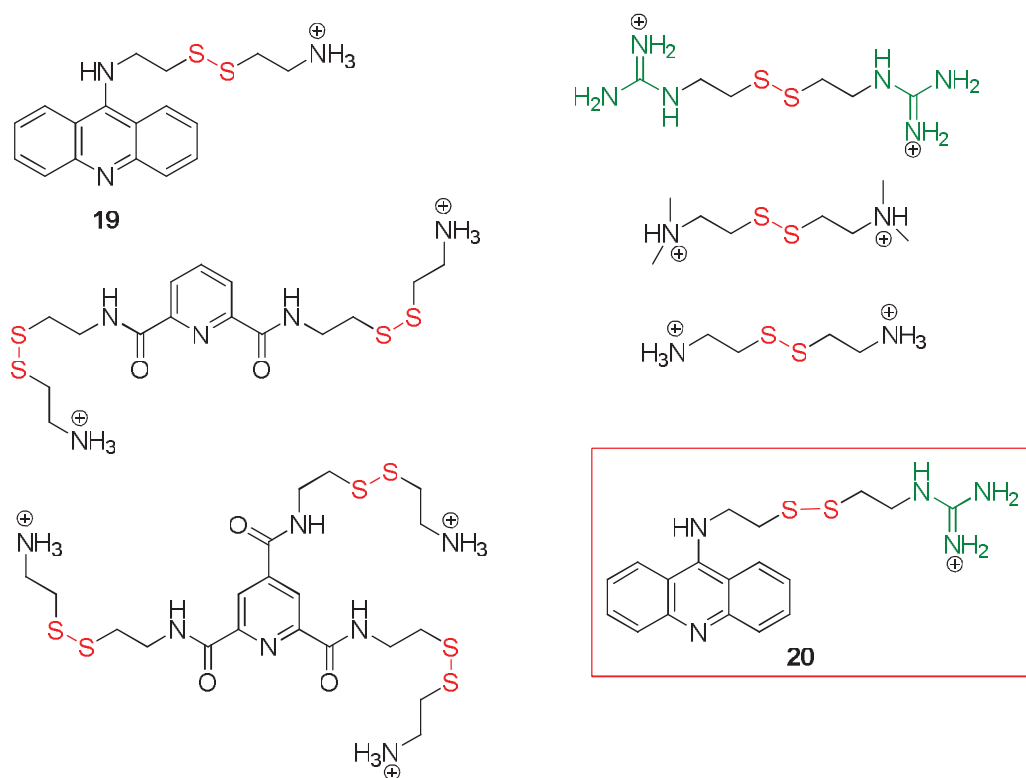
Figure 1.12 Structures of aromatic scaffolds, and sidechains used to generate a DCL in the presence of the scrambling initiator, cystamine hydrochloride, for selective extraction of G-quadruplex DNA; acridine **20** identified as strongest binder to d(AGT$_2$AG)$_5$.[96]

scaffolds, each carrying positively charged side-chains, were designed based on the structures of the majority of G-quadruplex ligands, which generally contain a central aromatic scaffold with one or several positively charged side chains and cysteamine hydrochloride was used to initiate the reaction. From the equilibrium distribution of 18 species, the acridine **20** was extracted with high selectivity with a lower amount of the acridine **19** also extracted.

(c) RNA

The rules that govern the design and synthesis of small molecules that bind with high sequence specificity and affinity to RNA are much less advanced than for DNA. The majority of RNA-binding molecules are natural products based on derivatives of amino glycosides,[28] but some synthetic peptide threaders have been successfully designed and synthesized to interact with RNA *via* a threading mechanism.[144]

The first example of DCC applied to RNA involved the use of metal ion coordination as the reversible chemistry to generate the DCL.[111] Metal ion coordination with ligands provides significant diversity in a DCL as a range of different stoichiometries and geometries of the metal complexes can result depending on the coordination preference of the metal. However, the lability of metal-ligand coordination required to generate the DCL presents challenges for isolation of the stable, selected metal complex. Specifically, the metal-ligand bond is labile; the target complex is often unstable and dissociates during the isolation process. Karan *et al* used derivatives of salicylamide ligands for the construction of a RNA binding library in the presence of metal ions (Figure 1.13). The ligands included a variable position for incorporation of potential

RNA-binding moieties.[111] It was assumed that square-planar mono and bis(salicylamide) complexes would form in the presence of Cu^{2+}, with other coordination modes possible, giving at equilibrium, a minimum of 27 metal complexes (Figure 1.13). The library of metal complexes was equilibrated in the presence of an RNA hairpin derived from the GTP-binding P7 helix from the *Pneumocystis carinii* Group I intron, as well as the homologous DNA

Figure 1.13 DCL formed from salicylamide ligands and copper (II) and the high affinity copper complex **22** detected in the presence of RNA.[111]

sequence for comparison. Using equilibrium dialysis and a 3500 molecular weight cut-off membrane, in the absence of Cu^{2+} none of the ligands bound to either RNA or DNA. However, in the presence of Cu^{2+}, rather surprisingly, a neutral peptide ligand **21** was selected in preference to the positively charged amidine peptide ligands. UV titrations confirmed the strong affinity of the

selected metal complex **22** for the RNA hairpin, which exhibited greater than 300 fold selectivity over the DNA sequence. It should be noted however, that while binding to DNA was detected in the DCC experiment, there was no evidence for binding by UV, possibly due to non-specific binding or because the binding mode of the metal complex with DNA could not be detected by UV. While this communication demonstrated the potential of DCC to identify RNA binding molecules, the use of paramagnetic metal complexes, and the inability to fully characterize the overall shape and stoichiometry of the bound metal complex **22**, the literature example (Figure 1.13) highlights the difficulties of using metal ion coordination and isolation of the metal-complex in DCC with nucleic acids.

In a series of papers, Miller and colleagues have demonstrated the very high potential of RB-DCC to identify novel RNA binding molecules (Figure 1.14).[31,115,135] The 22 nucleotide hairpin sequence in the HIV-1 frameshift stem-loop RNA was targeted by screening against a DCL with a theoretical size of 11,325 members.[115] The library was constructed from 150-resin attached Cys-containing building blocks and an identical set of solution-phase building blocks (Figure 1.14a). Dimers formed from three of the building blocks were identified as possible RNA-binding compounds, and following synthesis of the pure dimers and measurement of affinity constants, the sequence-selective, high affinity dimer **23** formed from a single Cys-building block was identified (Figure 1.14b). In an independent study, the same library was screened for target compounds able to inhibit muscular dystrophy type 1 (MBNL1) binding to (CUG) repeat RNA.[135] Four lead compounds were identified and inhibited

the interaction of GGG(CUG)$_{109}$GGG RNA with MBNL1 *in vitro* with K_i values in the low micromolar range.

Figure 1.14 Design of the 11,325-member RB-DCC using three different bead sizes to encode the Cys position for studies with RNA;[135] and (b) The lead compound **23** detected using this assay for HIV-1 frameshift stem loop RNA,[115] and hydrocarbon isosteres **24** and **25**.[31]

In a very recent study, the olefin and hydrocarbon isosteres of **23**, compounds **24** and **25** respectively (Figure 1.14b), were synthesized and tested for activity, as a first step to improving the biostability of the disulfide **23**.[31] The olefin

analogue **24** had similar activity to the parent disulfide, and thus provides the basis for production of compounds that may be suitable for cellular assays of frameshifting.

(d) Stabilised Oligonucleotide Conjugates

Rayner and colleagues have applied imine libraries, formed from an oligonucleotide functionalised with an amino group with a set of aldehydes, to the identification of covalently appended residues that stabilize oligonucleotides (Figure 1.15).[136] The library design involved incorporation of
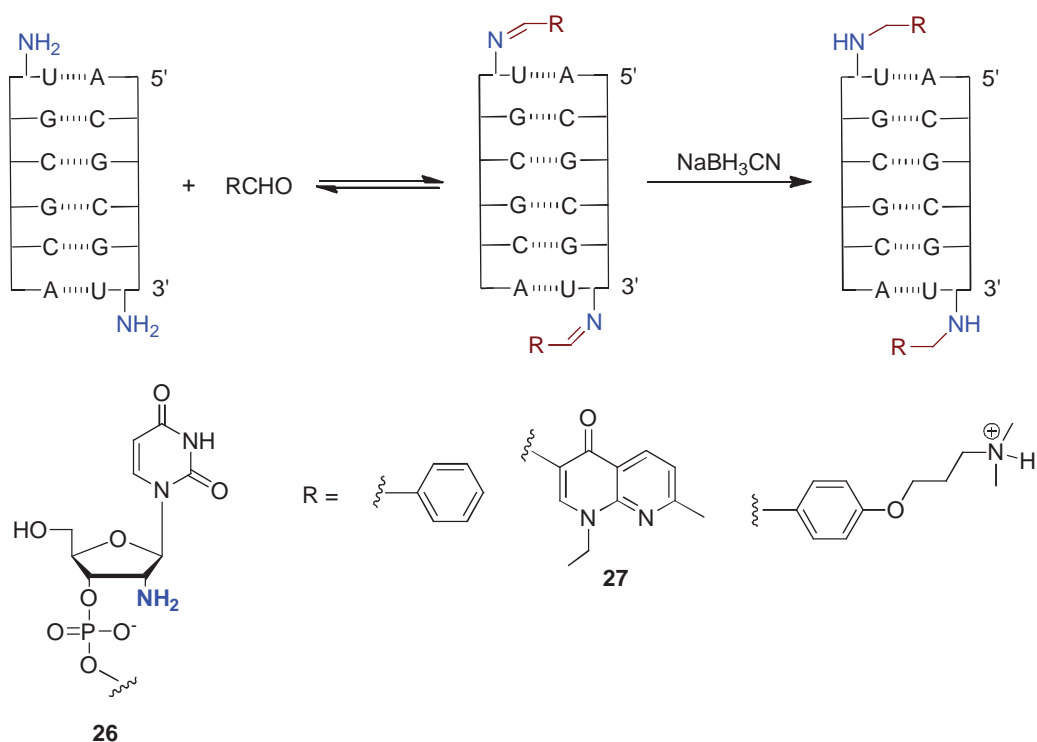
Figure 1.15 Selection of 3'-appended residues that stabilise a DNA duplex using DCC.[136]

the reactive amine group as the 2'-amino-2'-deoxynucleotide **26**. The higher nucleophilicity of the amino group at this position (pK_a 6.2) ensured selective

reaction at this position only and no reaction at the amino groups in the nucleic bases. The dynamic mixture of conjugated duplexes was generated in aqueous solution, the reaction was quenched with sodium cyanoborohydride to give the stable amines, and the products were analyzed by HPLC. A significant enrichment of the nalidixic conjugated product **27** was obtained compared to the other possible products. These initial results with the DNA model system were applied to a tertiary structured RNA complex formed by a loop-loop interaction between an RNA hairpin aptamer and its target the TAR RNA hairpin element of HIV-1.[136] A 14-nucleotide version of the aptamer and a 27-nucleotide from TAR (miniTAR) were used for the study. The nalidixic derivative **27** was amplified by 20% in the presence of minTAR. A later study extended this result to the incorporation of the appended stabilised ligand within the oligonucleotide ligand,[137] and allowed entry into DCC combined with Systematic Evolution of Ligands by Exponential enrichment (SELEX) for the *in vitro* selection of modified aptamers. Using this approach, conjugated RNA aptamers that bind tightly to the transactivation-responsive (TAR) element of HIV-1 were identified.[137-138]

Very recently, DCC has been applied to the selection of triplex-forming oligonucleotides (TFOs).[145] TFOs designed to bind to a DNA target and stabilise triple-helix formation were selected from a DCL of amines and polyamines.

(e) DNA Analysis

Very recently, DCC has been applied to single-nucleotide polymorphism (SNP) in DNA analysis.[146] While there are a number of methods available for SNP analysis, DCC was an attractive method for non-enzymatic genotyping of

genomic DNA. The imine chemistry and approach is similar to that used for the identification of stabilized oligonucleotide conjugates shown in Figure 1.15, but relied on the novel use of the amine group in a PNA backbone (Figure 1.16b), and selection of the appropriate aldehyde functionalized nucleic base (Figure 1.16a), followed by reduction and analysis.
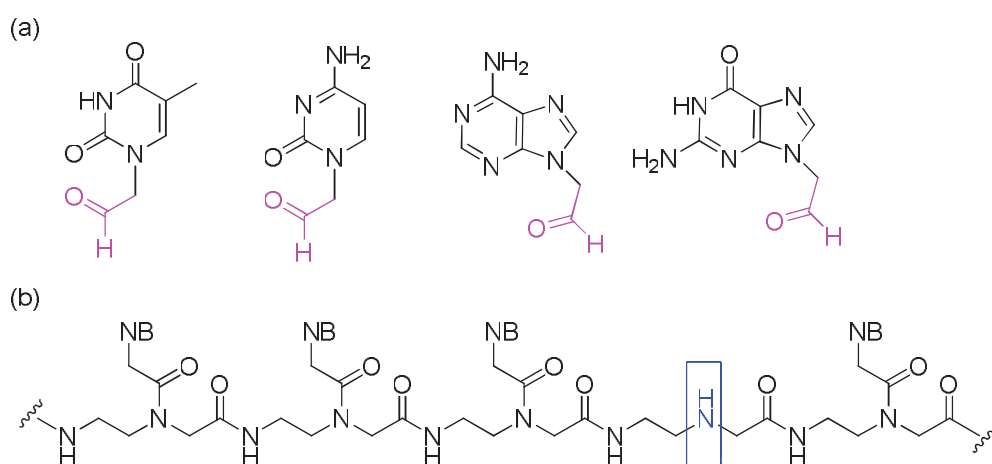
Figure 1.16 Aldehyde modified nucleobases; and (b) The general structure of a modified blank PNA strand (vacant site with free amine indicated in box) used to generate a DNA/iminium PNA DCL, which is reduced to a DNA/PNA duplex.[146]

1.7 Aims of Research

The broad aim of this research is to investigate the potential of DCC to identify new DNA-binding compounds. A schematic of the overall DCC approach is shown in Figure 1.17. The discovery of new binding motifs and DNA-binding compounds are important in understanding the rules of DNA recognition, and in the long term has the potential to assist in the design of new compounds with therapeutic and biotechnological applications.

To date, there have been very limited studies in which DCC has been applied to duplex DNA. These studies have focused on identifying dimers and simple derivatives of the well-characterised DNA minor groove binders such as distamycin **2** and Hoechst33258, and used only a small number of building blocks to generate the DCLs. As a result, the DNA products that were
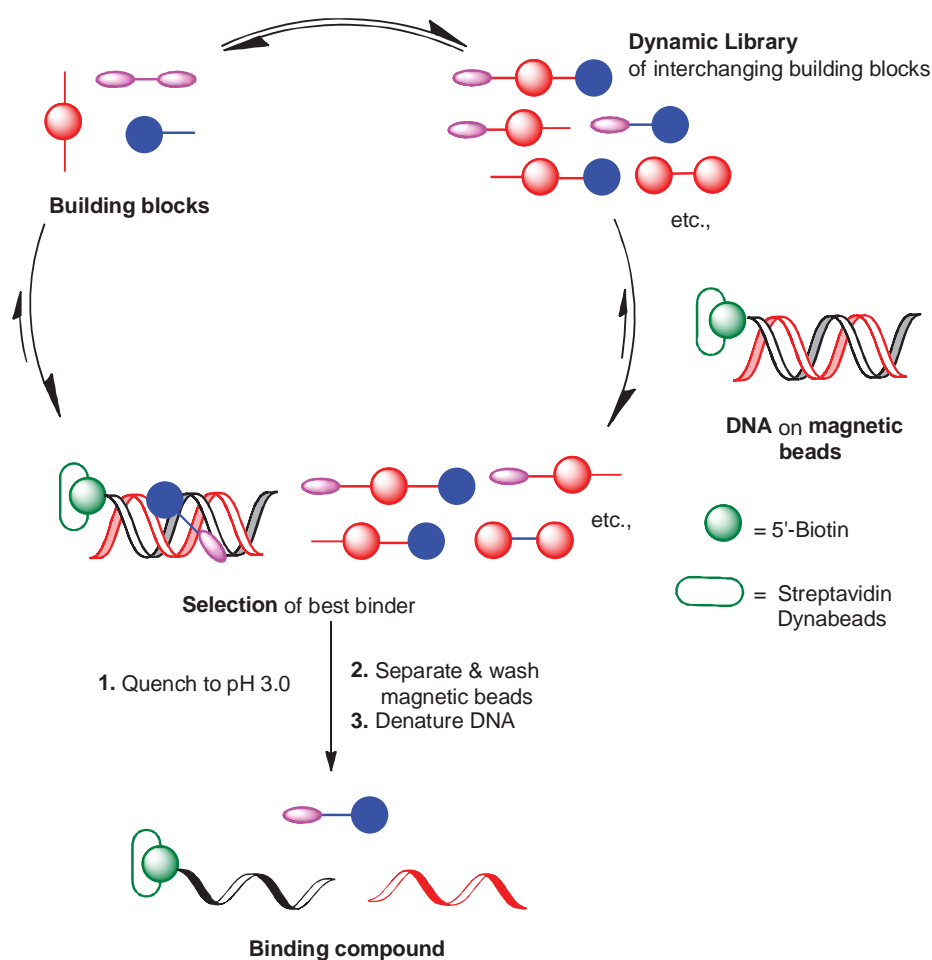
Figure 1.17 Schematic of the DCC approach to identify DNA-binding compounds.

identified using DCC would have been easily predicted to interact with DNA based on current understanding of DNA minor groove binders. A further

limitation of the studies reported to date is that all experiments were conducted in the presence of GSSG/GSH and as a result, in most cases the amplified products were disulfides of GSH. None of the studies addressed the issue of the limited stability of the identified disulfides in biological media where they are susceptible to redox reactions with biological thiols, and hence applications of these compounds in medicine or biotechnology are not possible.

The major goal of this research was to extend these initial studies to more diverse libraries that include building blocks that have not previously been investigated with duplex DNA. In particular, building blocks were designed that would form compounds whose DNA-binding characteristics could not be easily predicted, and included intercalators and intercalator-groove binding conjugates. These studies were used to provide information on the scope and limitations of DCC to identify novel duplex DNA-binding compounds through systematic variation of the reaction conditions, DNA sequences, and number and types of building blocks. The specific goals of this work were to:

- Design and synthesize building blocks that would allow formation of DCLs in water and incorporated heterocyclic and aromatic intercalators based on quinoline, naphthalene and imidazole, as well as disulfides of amidines, thio sugars, aliphatic and aromatic compounds. The solubility, compound rigidity, hydrophobicity, spacing of functional groups and overall charge was systematically varied in order to assess the potential for DCC to identify the relative importance of these features on DNA-binding. The design and synthesis of these building blocks is presented in Chapter 2.

- Establish reaction conditions for conducting DCC in the presence of duplex DNA in the absence of GSSG/GSH, in order to avoid the formation of GSSG/GSH adducts. Such conditions would allow more diverse libraries to be generated by the use of a larger number of building blocks, as the formation of GSH adducts complicate the analysis and interpretation of results. These GSH adducts have been favoured in the studies reported to date, due to the high concentrations of GSSG/GSH that have been required for the experiments, which can mask the detection of the other DNA-binding compounds under equilibrium conditions. Chapter 3 presents the DCL methods, reaction conditions and generation of DCLs under conditions that are suitable for studies with DNA, including two methods that allow generation of DCLs under conditions that are compatible with DNA, and do not require GSSG/GSH.

- Establish the scope and limitations of DCC to identify new DNA binding compounds that would not be readily predicted using current knowledge of DNA-binding compounds. Chapter 4 summarises the current understanding of the rules for DNA molecular recognition, followed by the results of the DCC experiments performed using the methods reported in Chapter 3. The DCC experiments were conducted with three different oligonucleotide sequences as models for DNA, and the results were analyzed using LC-MS. The mode of binding with DNA of the compounds amplified in the DCC experiments was analysed using molecular visualisation, in order to rationalise the results, propose models for how the molecules interacted with DNA, and comment on whether DCC allowed conclusions to be made on the importance of rigidity, solubility, charge, etc., on DNA-binding for the compounds that were studied.

- Since disulfides are unstable in biological media, design and synthesize an appropriate analogue of the disulfide moiety with enhanced biostability; biostable analogues of promising lead DNA-binding disulfides identified using DCC are essential for drug development or applications in biotechnology. Chapter 5 outlines studies on two disulfide mimics of selected quinoline derivatives identified using DCC. The disulfide mimics incorporated a triazole ring and a thioether group in place of the disulfide bond, based on the successful use of these mimics in peptide chemistry. The DNA-binding properties of the analogues were compared with the parent disulfides in order to make conclusions regarding the suitability of these mimics to deliver biostable DNA-binding compounds.

Chapter 2

Design and Synthesis of

Building Blocks

2.1 Design of Target Building Blocks

The design of water soluble building blocks for the application of DCC to identify molecules that bind to DNA presents a number of challenges. The building blocks need to contain a functional group that undergoes reversible chemistry that is compatible with the experimental conditions. In addition, the experimental conditions, including the conditions required to initiate and quench the reversible reaction, cannot denature the DNA secondary structure, and must permit isolation of the selected product. The building blocks as well as the library members need to be soluble in aqueous solution in the approximate pH range 5-8, which can be difficult to achieve with aromatic chromophores that are typically present in intercalators. The reversible chemistry must also be compatible with the functional groups present in DNA, notably the exocyclic amino groups, the phosphodiester backbone and the C-4 and C-6 positions of the pyrimidines (Figure 1.1).

From the reversible covalent reactions that are suitable for DCC (Table 1.1, Chapter 1), most of the reactions are not suitable for studies with DNA, as they are unable to be conducted in water under conditions where DNA is stable, or they are incompatible with the functional groups present in the DNA bases. Disulfide formation is the exception, and all DCC studies to date with duplex and quadruplex DNA have used thiol-disulfide DCLs (Table 1.2).

In this work, the target building blocks required to synthesize DNA-binding molecules using DCC were designed with the following features:

- Hetereocycles and aromatic compounds that are present in established DNA-binding compounds as either intercalators or minor groove binders;

- Functional groups and binding motifs that are present in natural and synthetic compounds that interact with nucleic acids using non-covalent interactions e.g. hydrogen bonding, electrostatic and hydrophobic interactions;
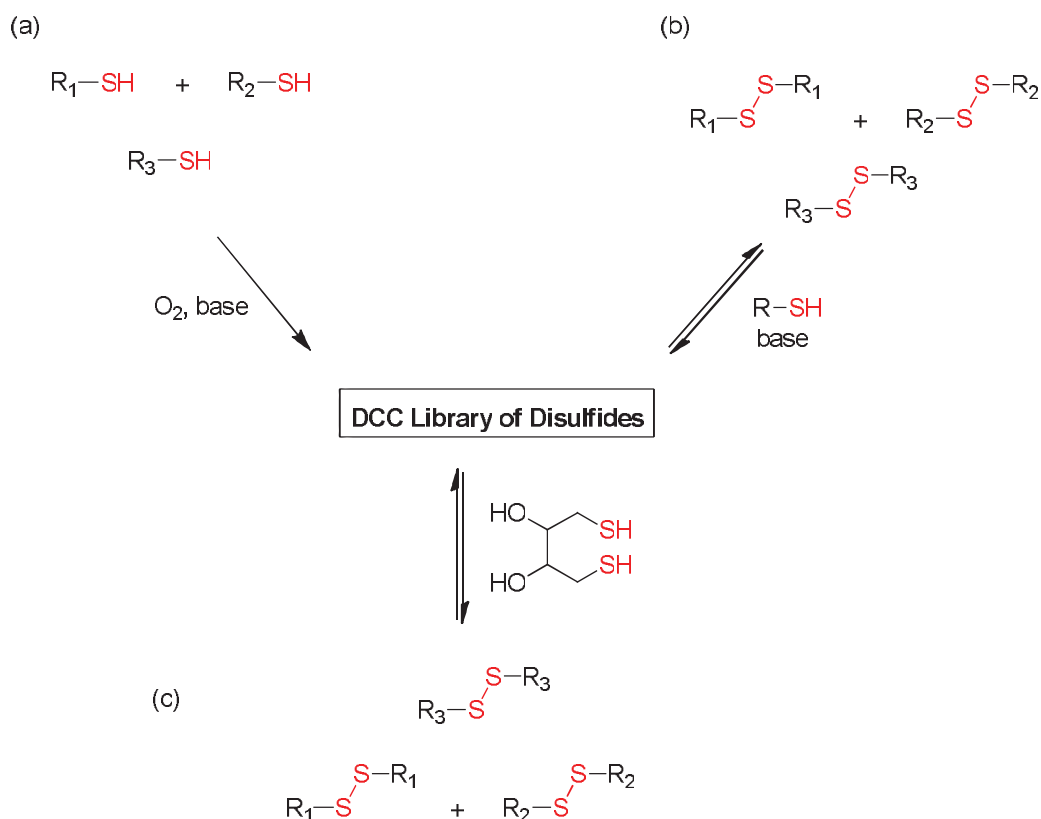
- Thiol and disulfide functional groups in order to facilitate the reversible chemistry needed for DCC in aqueous solutions;

- Good solubility in aqueous solutions at pH 7-9 or solubility in aqueous methanol solutions i.e. conditions that are compatible with DNA.

Figure 2.1 summarizes the target thiol, disulfide and bisthiol building blocks used to generate the DCLs with DNA. While the building blocks were designed as thiols, DCLs can be prepared starting from the corresponding symmetrical disulfides or from a mixture of thiols and disulfides (Figure 2.2).[78] Hence, for synthetic ease, most of these building blocks were prepared as the disulfides. The unsymmetrical disulfides **Q4**-**Y** and **N1**-**Y** were derivatives of cysteamine hydrochloride, which improved aqueous solubility of the building blocks, and allowed DCL formation using a scrambling initiator approach.[96]

Quinolines were chosen as they are present in the naturally occurring DNA cyclic depsipeptide bisintercalators such as sandramycin **28** (Figure 2.3).[55-57,147] Naphthalimides **N1** and **N1**-**Y** were based on the synthetic bisintercalator
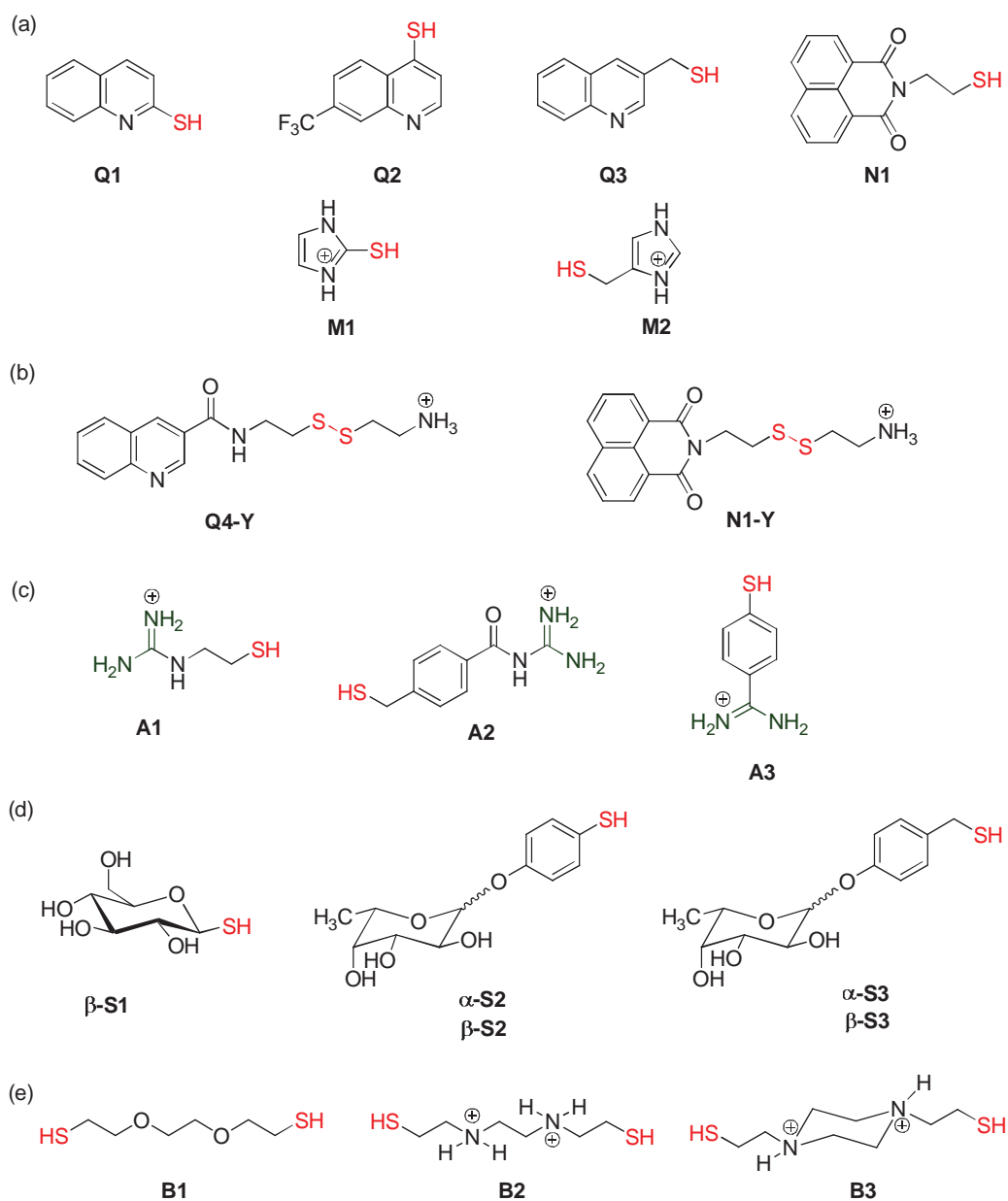
Figure 2.1 Target building blocks (a) aromatic thiols, (b) aromatic disulfides of cysteamine, (c) amidines, (d) carbohydrates and (e) alkyl bisthiols.

(a)

R_1-SH + R_2-SH

R_3-SH

O_2, base

(b)

R_1-S $S-R_1$ + R_2-S $S-R_2$

R_3-S $S-R_3$

$R-SH$
base

DCC Library of Disulfides

HO SH
HO SH

(c)

R_3-S $S-R_3$

R_1-S $S-R_1$ + R_2-S $S-R_2$

Figure 2.2 General methods for the preparation of disulfide DCLs.[78]

elinafide **29**, which is composed of two naphthalimide chromophores and has undergone clinical trials against solid tumors.[148-150] Although quinolines **Q1**, **Q2** and **Q3** are neutral compounds, the protonated forms of the corresponding ligands were predicted to have improved aqueous solubility at acidic pH, and thus allow DCC experiments to be conducted in aqueous mixtures. Imidazoles **M1** and **M2** were included as charged heterocycles that are structurally related to pyrrolic units present in minor groove binders such as distamycin **2** (see Figure 1.2).

(a)

28

(b)

29

30
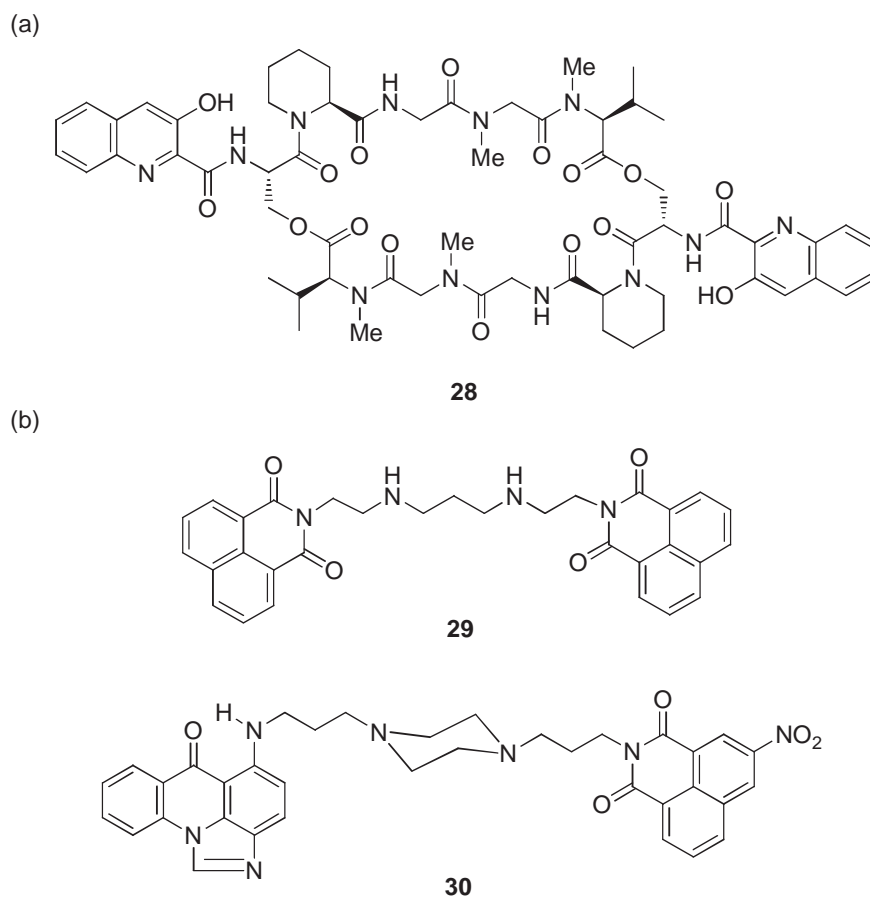
Figure 2.3 Structures of DNA-bisinteracalators (a) the natural product sandramycin **28** and (b) synthetic compounds elinafide **29** and WMC-79 **30**.

As discussed in Section 1.4, amidines are important components of many DNA-binding compounds including proteins where the amino acid arginine often is involved in important hydrogen bonding and electrostatic interactions with DNA.[151-152] Amidine functional groups also play an important role in the DNA recognition of many DNA minor groove binders.[12,44,70] Amidines **A1**, **A2** and **A3** varied the rigidity of the building blocks and the distance between the thiol and amidine functional groups, in order to assess the relative importance of these features on DNA-binding.

Thiosugars **S1**, **S2** and **S3** included the commercially available β-D-thioglucose **S1**, which is supplied > 95% as the β-anomer. The deoxy sugars **S2** and **S3** are derivatives of fucose, and were studied based on the importance of hydrophobic deoxy sugars as DNA minor groove binders.[153-154] The thiosugars **S2** and **S3** contained the rigid phenyl group and the more flexible benzylic thiol group respectively, in order to assess the importance of these structural changes on binding. In all cases, the sugars were present as a mixture of α- and β-anomers; no attempt was made to separate the anomers, as the presence of both stereoisomers is highly desirable in DCC, as they give rise to additional diversity in the DCLs generated from these building blocks.

Bisthiols **B1**, **B2** and **B3** were designed as potential linkers that would allow the formation of bisintercalators with the quinoline and napthalimide building blocks shown in Figure 2.1. Bisthiols **B2** and **B3** incorporated amino functional groups that would generate highly water soluble compounds at pH 7.0 and also potentially aid in DNA recognition through H-bonding and/or electrostatic interactions. Similar linkers have been reported in synthetic bisintercalators such as WMC-79 **30**.[155-156] Given the high entropic cost that is required for the fully flexible linkers **B1** and **B2** to adopt a conformation that is suitable for interaction with the DNA minor groove, the conformationally restricted bisthiol **B3** was also studied.
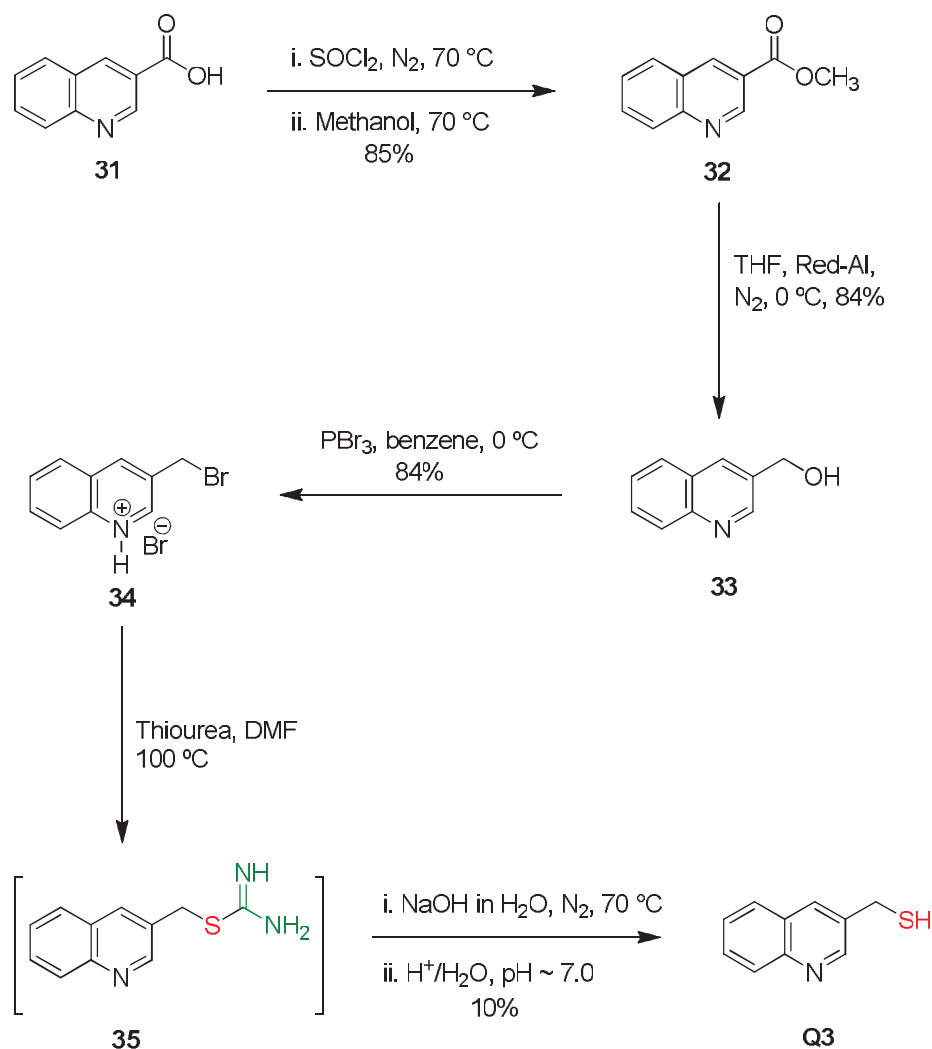
2.2 Synthesis

2.2.1 Aromatic Thiols and Disulfides

Quinolines **Q1** and **Q2** are commercially available. Two routes were investigated to synthesize quinoline **Q3** from quinoline alcohol **33** which was

synthesized in two steps from 3-quinoline carboxylic acid **31** (Scheme 2.1).

Scheme 2.1

The carboxylic acid **31** was heated at reflux with thionyl chloride to give the unstable acid chloride, which was immediately heated at reflux with methanol to give the ester **32** in 85% yield. The next step required selective reduction of the ester **32** in the presence of the quinoline ring which is susceptible to reduction with typical reducing agents such as sodium borohydride.[157-160]

Treatment of the methyl ester **32** with the mild reducing agent, sodium bis(2-methoxyethoxy) aluminium hydride (Red-Al, 1 equiv.), at 0 °C [159] gave the alcohol **33** in good yield (84%), and there was no evidence of any degradation of the quinoline ring. The alcohol **33** was treated with phosphorus tribromide at 0 °C which resulted in precipitation of the hydrobromide **34**.[161] Conversion of the bromide **34** to the quinoline thiol **Q3** was carried out *via* the intermediate **35** using thiourea and standard conditions.[162] The hydrobromide **34** was heated at reflux with thiourea, followed by base hydrolysis. Using this method, the required quinoline thiol **Q3** was obtained in very low yield (10%), a result that was attributed to the high aqueous solubility of **Q3**. While shown as the neutral compound in Scheme 2.1, **Q3** also exists as the zwitterionic compound, which did not allow isolation and purification of the compound using column chromatography. Purification was not investigated due to the poor yield and the fact that the hydrobromide **34** has been reported to be unstable and polymerises readily as the free base compound.[158,163] Hence, an alternative route was investigated.

Given the low isolated yield of the thiol **Q3**, an alternate route to **Q3** was investigated that avoided the bromoquinoline **34** as the intermediate. Thus, 3-hydroxymethylquinoline **33** was converted in high yield to the mesylate under standard conditions, followed by reaction with potassium thioacetate to afford the thioacetate **36** in 82% yield (Scheme 2.2). The thioacetate **36** was hydrolyzed under acidic conditions to give the charged and highly water soluble hydrochloride salt **37**, which was converted to the neutral disulfide **Q3**-**Q3** to allow purification. The hydrochloride salt **37** was subjected to air oxidation, followed by pH adjustment to ~ 7-8 using sodium hydroxide. The

solution was extracted into chloroform and the organic soluble disulfide **Q3-Q3** was isolated in good yield and high purity.
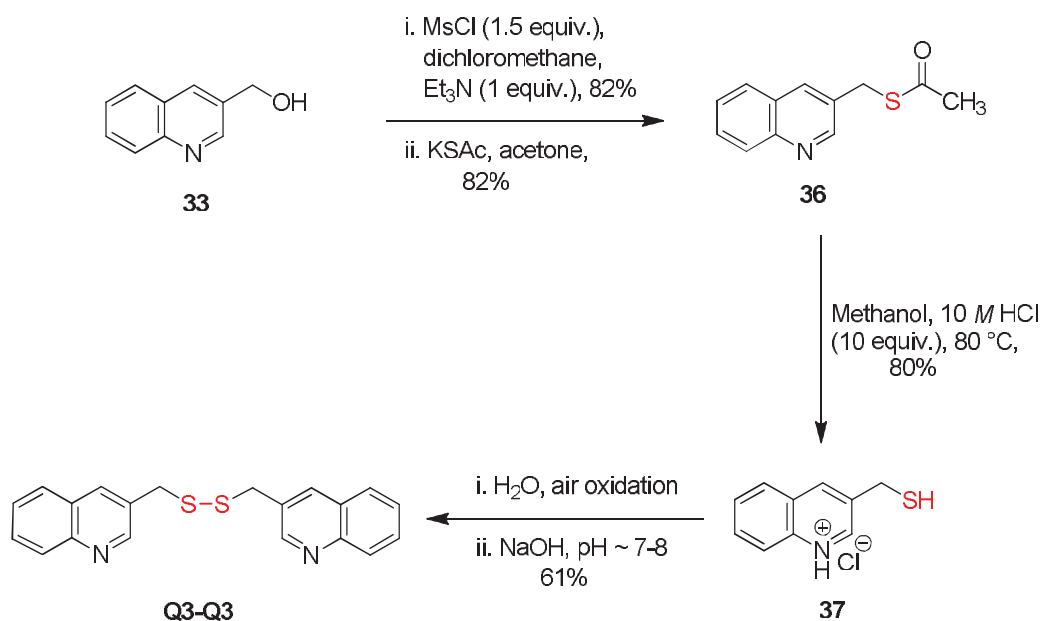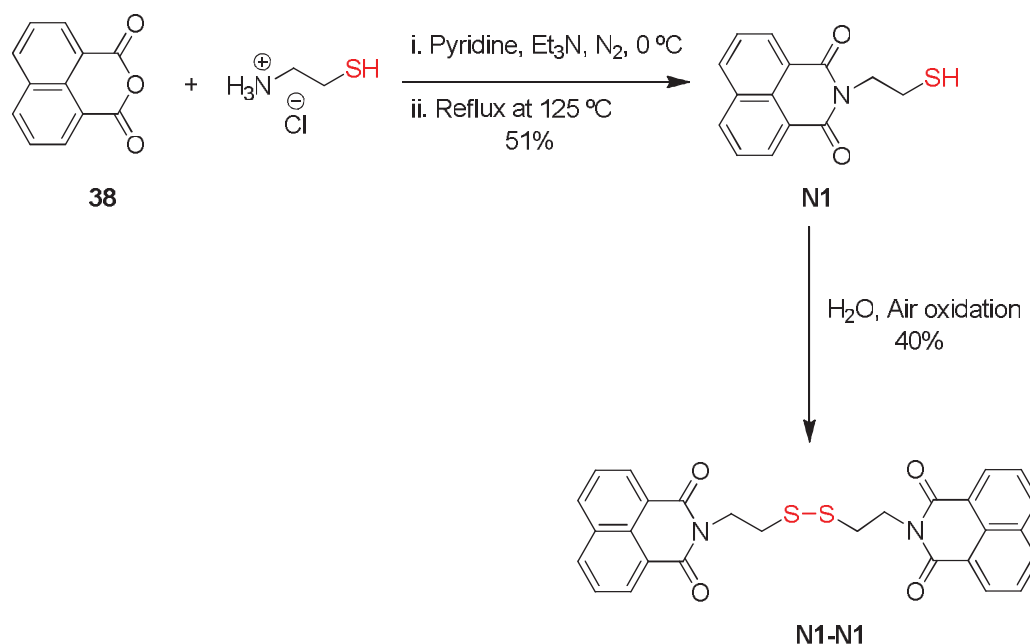

Scheme 2.2

Naphthalimide **N1** was synthesized from naphthalic anhydride **38** according to the literature procedure.[164] Treatment of naphthalic anhydride **38** with cysteamine hydrochloride afforded a mixture of the desired thiol **N1** as well as some of the oxidised product **N1-N1** (Scheme 2.3). In order to allow purification, the crude mixture of products were subjected to air oxidation to give the disulfide **N1-N1** as the exclusive product.

Imidazole **M2** was synthesized from 4-(hydroxymethyl)imidazole **39** *via* the corresponding alkyl chloride **40** using the literature procedure.[165-168] 4-(Hydroxymethyl)imidazole **39** in benzene was heated at reflux with thionyl chloride to give the alkyl chloride **40** in quantitative yield.[165,167-168] The alkylchloride **40** was treated with potassium thioacetate using the adapted

literature conditions[169] to afford the thioacetate **41**. The thioacetate **41** was highly soluble in water; hence, only 29% was isolated from the reaction

Scheme 2.3

mixture (Scheme 2.4). The thioacetate **41** was hydrolyzed under acidic conditions to give the charged and highly water soluble hydrochloride salt of **M2** in quantitative yield. While the synthesis of the imidazole **M2** has been reported,[166] the route used in this work contains fewer steps and gave the product in improved overall yield.

The synthesis of the cysteamine disufides **Q4-Y** and **N1-Y** required the preparation of the Boc protected amine **42**, which was synthesized according to the literature procedure (Scheme 2.5).[170] Treatment of cysteamine dihydrochloride with di-*tert*-butyl-dicarbonate (1 equiv.) gave a mixture of the mono **42** and bis-Boc **43** protected products, as reported in the literature.[170]
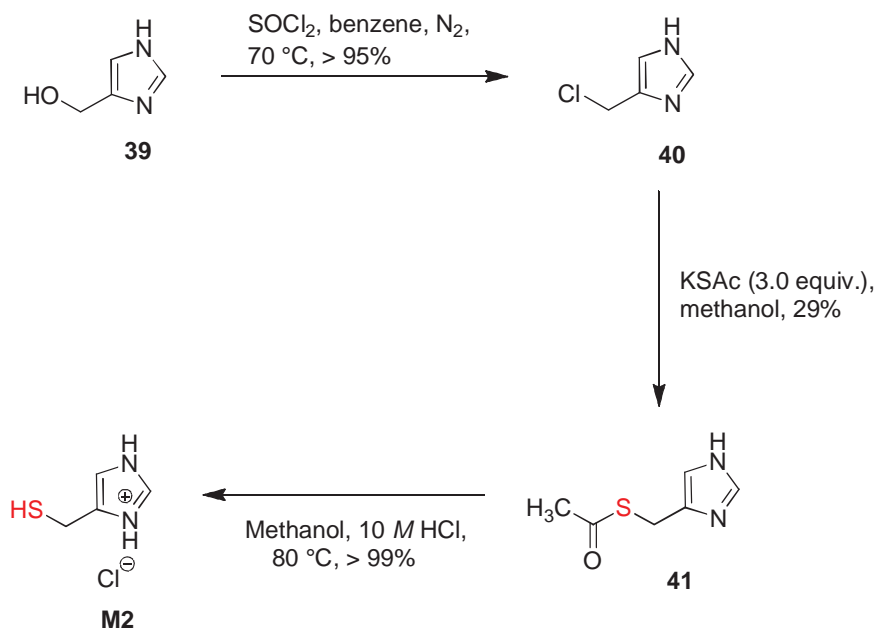
Scheme 2.4

These products were readily separated by simple extraction procedures and the required mono-Boc protected compound **42** was isolated in 36% yield.
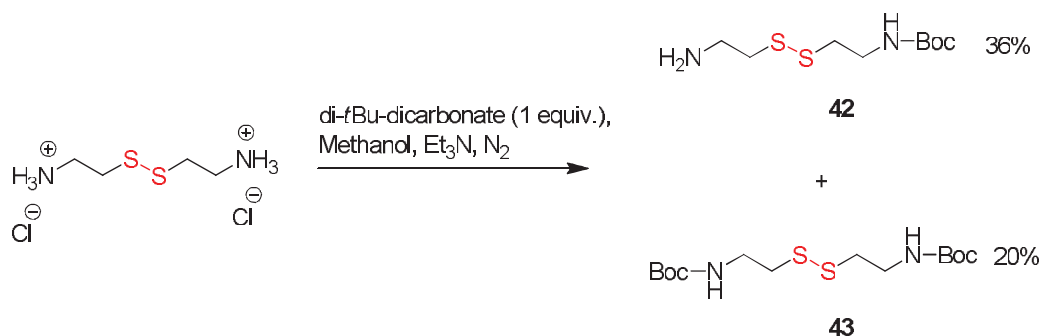
Scheme 2.5

3-Quinoline carboxylic acid **31** was converted to the Boc-protected quinoline in 70% yield using standard coupling chemistry (Scheme 2.6).[96] Removal of the Boc-protecting group by treatment with dichloromethane and TFA afforded quinoline **Q4-Y**, which was purified by semi-preparative RP-HPLC.
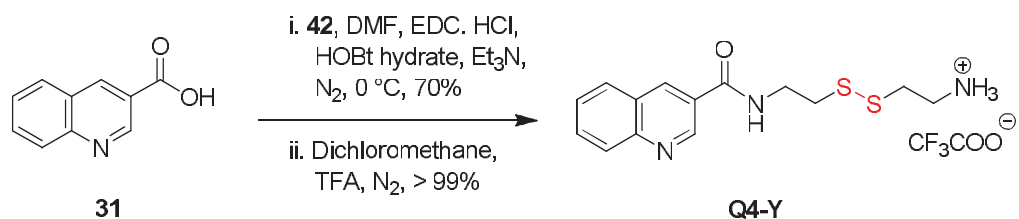
Scheme 2.6

Naphthalene disulfide **N1-Y** was synthesized using similar conditions but starting from the naphthalic anhydride **38** (Scheme 2.7). Deprotection using dichloromethane and TFA afforded the naphthalimide **N1-Y** in 71% yield, which was also purified by semi-preparative RP-HPLC.

Scheme 2.7

2.2.2 Alkyl and Aromatic Amidines

Amidines **A1**, **A2** and **A3** were all synthesized as the corresponding disulfides **A1**-**A1**, **A2**-**A2** and **A3**-**A3** respectively. The amidine **A1**-**A1** was synthesized from cysteamine dihydr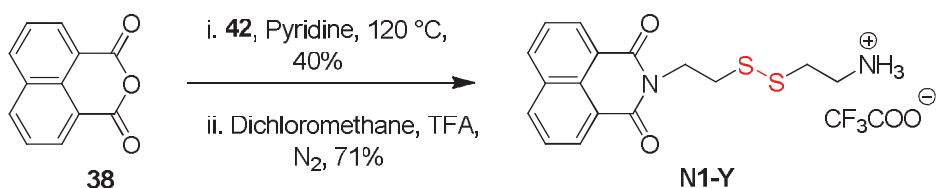ochloride and *N,N'*-di-Boc-*N"*-trifluoromethane-sulfonylguanidine in two steps using the literature procedure (Scheme 2.8).[96] The first step generated the tetra-Boc-guanidine in 85% yield, and removal of the Boc groups using dichloromethane and TFA afforded the bis-amidine **A1**-**A1** in quantitative yield.
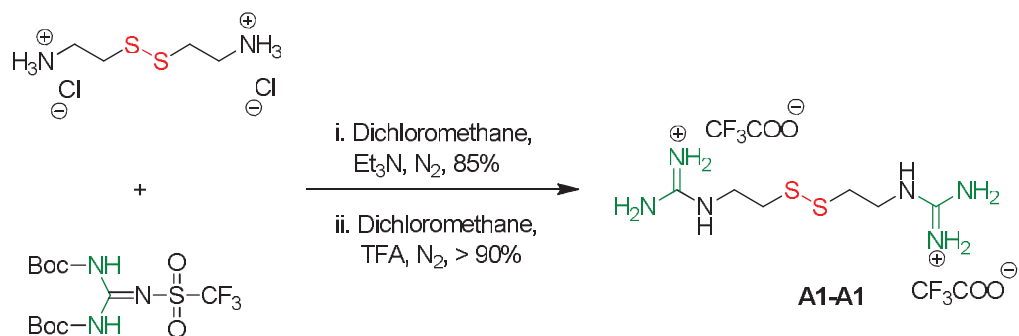
Scheme 2.8

The bisamidine **A2**-**A2** was synthesized as shown in Scheme 2.9. The benzylic bromide **44** was treated with triphenylmethanethiol in the presence of *N,N*-diisopropylethylamine (DIPEA) to give the trityl substituted methyl ester **45** in 75% yield. Base hydrolysis of the methyl ester **45** afforded the corresponding benzoic acid **46**. The next step required conversion of the acid to the guanidinium derivative **A2**. A number of different coupling reagents and conditions have been reported in the literature[95,171-172] for the synthesis of guanidinium derivatives. However, initial attempts to convert the acid to the corresponding guanidinium derivative by treatment of the acid **46** with Boc-guanidine using standard coupling agents (*N,N*-dicyclohexylcarbodiimide and HOBt-hydrate) gave very low yields. The desired Boc-protected compound was successfully synthesized in 41% yield using *N*-methyl-morpholine and the coupling agent PyBOP. Deprotection of the trityl and Boc groups were achieved in one step using TFA and triethylsilane. The highly water soluble crude product **A2** was purified by semi-preparative RP-HPLC, which afforded the amidine as a mixture of the thiol **A2** and the disulfide **A2**-**A2**. The mixture was subjected to air oxidation to give the desired disulfide **A2**-**A2** in 25% yield.
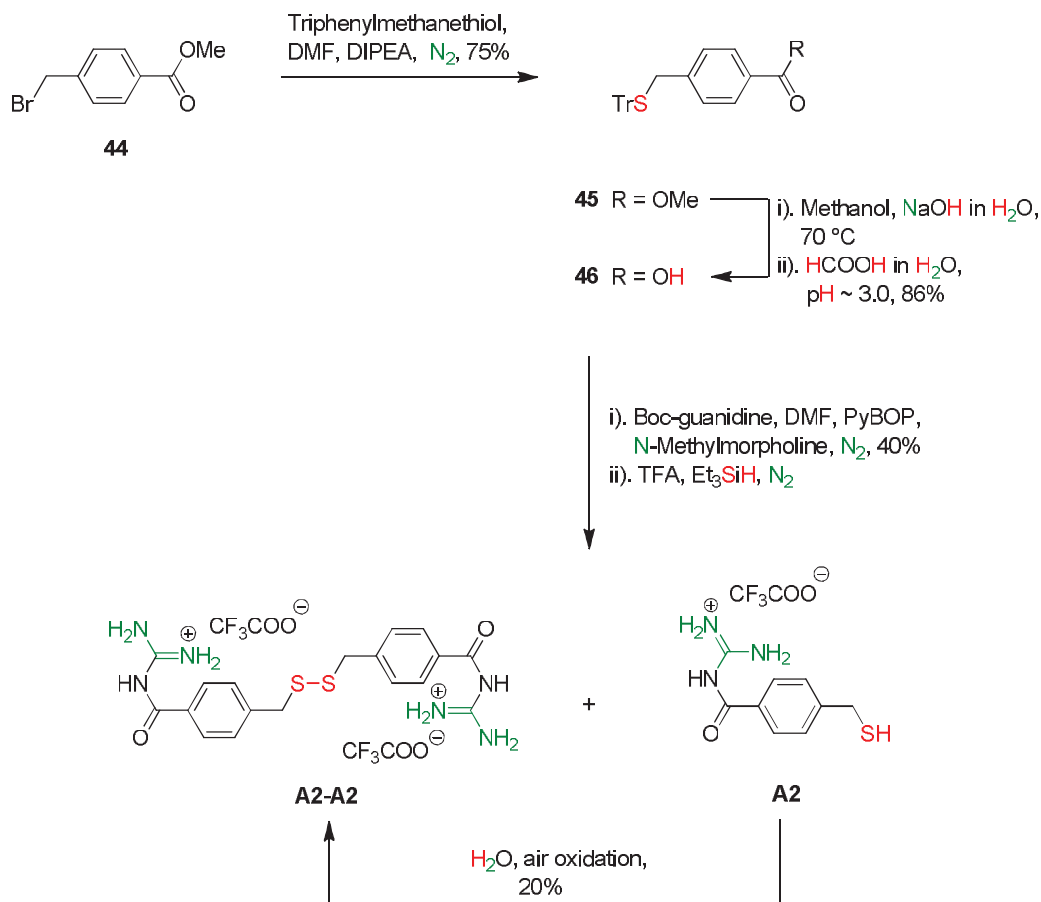
Scheme 2.9

Bisamidine **A3**-**A3** was synthesized using the literature procedures for the synthesis of related compounds[173-177] (Scheme 2.10). Treatment of cyanophenol **47** and thiocarbamoylchloride **48** using the strong hindered base DABCO, followed by acid neutralization gave the thiocarbamate **49** in 92% yield. The neat thiocarbamate **49** was heated at reflux at 210 °C under argon, which afforded the intramolecular rearranged product **50** in quantitative yield. While the two compounds **49** and **50** have identical molecular weight, the compounds were readily identified from the ^{1}H NMR data of the compounds which showed distinct chemical shifts for the NMe_2 and aromatic signals.
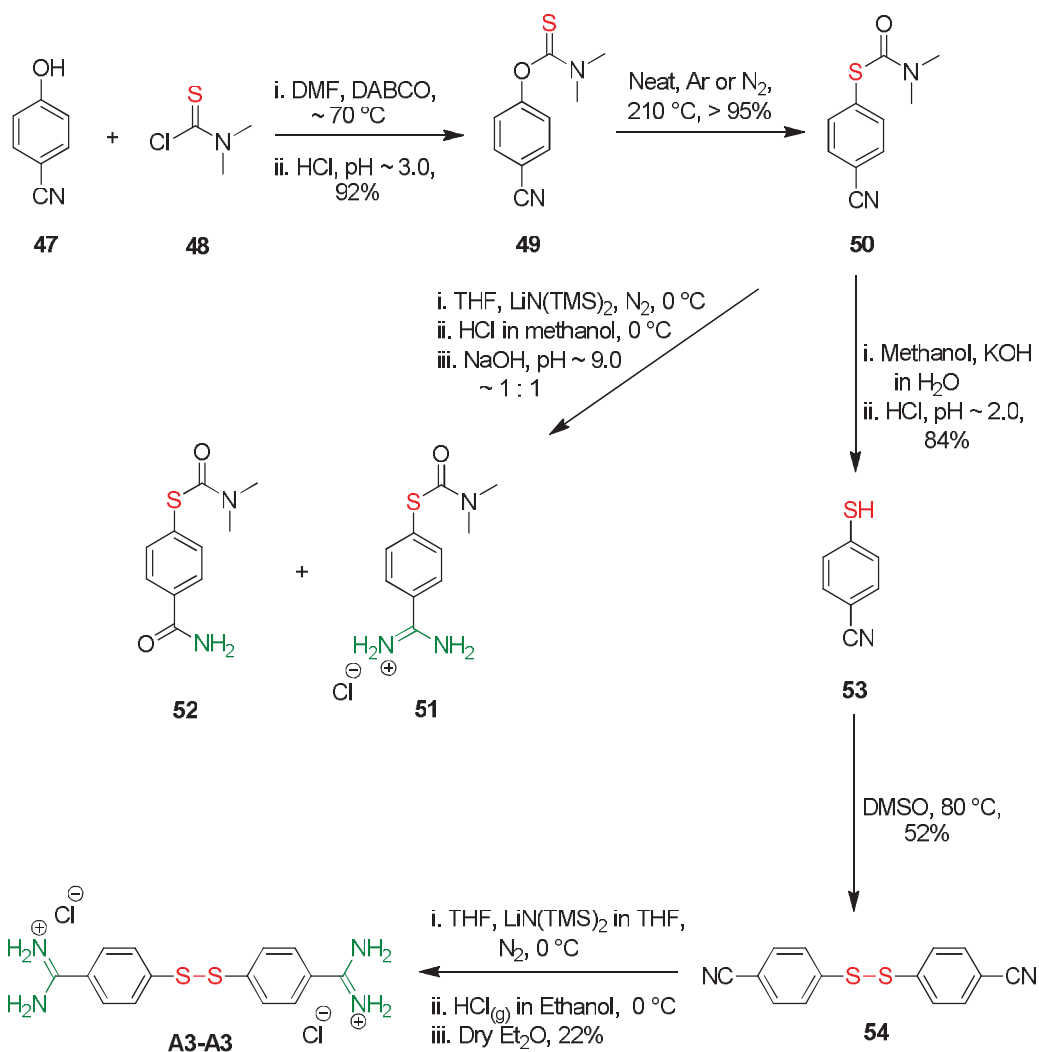
Scheme 2.10

The next step required conversion of the *para*-nitrile to the required amidine functional group. Thiocarbamate **50** was treated with lithium bis (trimethylsilyl)amide according to the literature procedure used for the synthesis of related compounds.[70] However, under these conditions analysis of the crude product by LC and MS techniques showed the presence of two major compounds in a 1:1 ratio. These products were tentatively assigned as the desired product **51** as well as the amide **52**, on the basis of ^{1}H and ^{13}C NMR

spectra and mass spectra which were consistent with these structures. In particular, the presence of a signal at δ 171.53 in the ^{13}C NMR spectrum was strong evidence for the presence of the carbonyl group in the amide of **52**. Attempts to optimize the reaction conditions and avoid formation of the amide by rigorous exclusion of water were unsuccessful, and separation of the amidine **51** and amide **52** was difficult and further complicated by the partial hydrolysis of the thiocarbamoyl protecting group in both **51** and **52**.

Hence, an alternative route to introduce the amidine functional group was investigated with the thiol group protected as the disulfide **54** (Scheme 2.10). Using the literature method[173-174] thiocarbamate **50** was hydrolyzed to the thiophenol **53**, which was oxidised with DMSO to give the disulfide **54** in 52% yield. Conversion of the nitrile **54** to the amidine **A3-A3** was performed using the method of Rogana *et al.*[178] The disulfide **54** was treated with lithium bis (trimethylsilyl)amide followed by treated with saturated hyrochloric acid (gas) in ethanol to give the required amidine **A3-A3**, which was purified by semi-preparative RP-HPLC to afford the desired amidine **A3-A3** in 22% yield. While the synthesis of amidine **A3** has been reported,[179-182] the route used in this work is significantly shorter and avoids any byproduct formation.

2.2.3 Alkyl Bisthiols

The diamino bisthiol **B2** was synthesized using the literature procedure shown in Scheme 2.11.[183] Treatment of commercially available *N,N*-bis(2-hydroxyethyl)ethylene diamine **55** with 4-(dimethylamino) pyridine and *p*-toluenesulfonyl chloride gave the crude tosyl substituted derivative, which was purified by column chromatography to give the tetra-tosyl product **56** in 30% yield. Thioacetate **57** was synthesized in 75% yield by treatment of **56** with

potassium thioacetate in DMF. Treatment of the thioacetate **57** with phenol in 33% hydrobromic acid in acetic acid, followed by heated at reflux with ethanol afforded the bisthiol **B2** in quantitative yield.
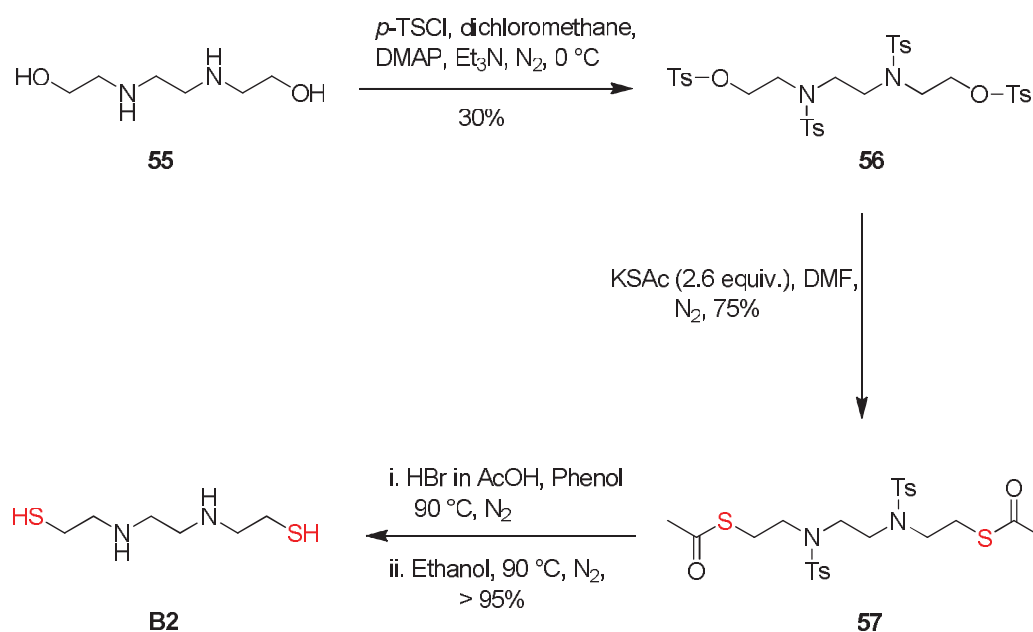
Scheme 2.11

2.3 Summary

The building blocks synthesized in this work contained thiol/disulfide functional groups and were based on the literature DCC studies with nucleic acids. The quinoline and naphthalene building blocks are the first examples of intercalator derivatives that have been used for studies with duplex DNA. The solubility of these aromatic compounds was identified as an important factor that needs to be considered for their use in DCC studies; both **Q3** and **N1** were insoluble in water and aqueous methanol. Hence, **N1-Y** was synthesized as the charged cysteamine derivative in order to increase the aqueous solubility. While this approach improved solubility in polar solvents such as methanol,

the large aromatic chromophore resulted in limited solubility in aqueous mixtures. In contrast, the cysteamine derivative **Q4**-**Y** which contains the heterocyclic 3-substituted quinoline exhibited good solubility in aqueous solutions; this derivative was synthesized as the amide derivative to address the insoluble nature of **Q3**.

The thiol/disulfide building blocks of imidazoles, alkyl and aromatic amidines, thiosugars and aliphatic bisthiols were designed to allow the formation of intercalator-groove binding conjugates and bis-intercalators in the DCLs. These building blocks incorporated features to vary properties including solubility, linker rigidity, hydrophobicity, spacing of functional groups and overall charge in the disulfides formed from these compounds in DCL experiments. These properties were systematically varied in order to assess the relative importance of these features on DNA-binding.

2.4 Experimental

2.4.1 Materials and Methods

Quinolines **Q1**, **Q2,** imidazole **M1**, thiosugar **S1**, bisthiol **B1**, Red-Al and hydrobromic acid in acetic acid were purchased from Aldrich Pty Ltd. Aromatic thiosugars **S2**, **S3** and bisthiol **B3** were synthesised by P. M. Abeysinghe in our research group.

Commercially available reagents and solvents (HPLC quality) were used without further purification, unless otherwise stated. Dichloromethane, THF and methanol were dried using an Innovative Technology Inc., Pure Solv 400-

4.4 mD solvent purification system. All references to water refer to the use of Milli-Q water generated from a Millipore, Milli-Q Bicel A 10 system.

Reactions were monitored by thin layer chromatography (TLC) on Merk silica gel 60 F_{254} precoated sheets (0.2 mm). Visualization of TLC plates were carried out under short and long wave UV light followed by staining with Vanillin dip [vanillin (1 g), methanol (85 mL), acetic acid (10 mL), sulfuric acid (5 mL, 18 *M*)] or Goofy's dip [phosphomolybdic acid (3 g), ceric sulfate (0.5 g), sulfuric acid (5 mL) and water (95 mL)]. Flash chromatography was carried out using Merk silica gel grade 230-400 mesh (40-63 microns) and eluting solvents are quoted as volume/volume mixtures.

Melting points were determined using a Gallenkamp heating stage apparatus (model MPD350 BM 2.5). UV-Vis spectra were measured with a CARY (100 Scan) UV-Visible spectrometer from 500-190 nm using methanol as reference solvent. The pH was measured on a Beckman Instruments Φ210 pH meter. ^{1}H and ^{13}C NMR spectra were performed on Bruker DPX 300 (300 MHz), 400, 500 and 600 MHz spectrometers and referenced to solvent peaks (given in parenthesis). ^{1}H NMR data is reported in chemical shifts (δ in ppm), integration, multiplicity, coupling constant, assignment, in that order. Abbreviations used in multiplicity are: s, singlet; bs, broad singlet; bd, broad doublet; d, doublet; dd, doublet of doublets; t, triplet; dt, doublets of triplet; m, multiplet; Ar, aromatic. ^{13}C NMR data is reported in chemical shifts (δ in ppm) only.

Electrospray ionization (ESI) mass spectroscopy was performed on a TSQ Quantum LC-MS/MS spectrometer and Waters2690 separations module with

a Waters micromassZQ spectrometer. Values of m/z are quoted with intensities expressed as percentages of the base peak in parenthesis. High resolution mass spectroscopy (HRMS) was performed in a Thermo Scientific (San Jose, CA) LTQ-FT Fourier Transform mass spectrometer. The samples were analyzed in positive ion nanospray mode at a resolution setting of 200,000. A minimum of five digital values for the calculated as well as found HRMS values for the ions are given.

2.4.2 HPLC Conditions

The mobile phase consisted of eluents A (water with 0.1% formic acid) and B (acetonitrile) for all runs. Reverse-phase (RP) HPLC was performed on a Shimadzu separations module with a SPD-M20A photodiode array detector (190 nm to 350 nm), FRC-10A fraction collector and LC-20AD pump with 5-channel degasser and SIL-20AHT auto sampler. Analytical RP-HPLC was performed with a Thermo Scientific 5 μm Hypersil GOLD RP C18 column (2.1 × 150 mm column, 5 μm particle size, and flow rate 0.2 mL min^{-1}). Semi-preparative RP-HPLC employed a RESTEK Pinnacle DB C18 column (10 × 150 mm, 5 μm particle size, flow rate 3.5 mL min^{-1}). The solvent from the HPLC fractions were removed using a CHRIST freeze dryer model ALPHA 1-4 LDplus/2-4 LDplus, John Morris Scientific, Australia.

2.5 Synthesis

Methyl-3-quinolinecarboxylate 32
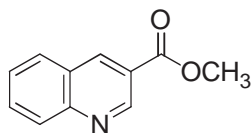
3-Quinoline carboxylic acid **31** (2.50 g, 0.014 mol) was heated at reflux with thionyl chloride (150 mL) for 3 h under nitrogen to give a pale yellow solution. The excess thionyl chloride was removed to yield the corresponding acid chloride as a pure white solid in quantitative yield, which was used immediately in the next step. Methanol (100 mL) was added to the acid chloride and the solution was heated at reflux for 5 h and the solvent was removed *in vacuo* to give a pale yellow solid. Sodium hydrogen carbonate solution (50 mL, 5%) was added and the solid was extracted into chloroform (3 × 50 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate concentrated *in vacuo* to give quinolinecarboxylate **32** as a pale yellow solid (2.30 g, 85%), m.p. 77-80 °C (lit.[184] m.p. 76 °C, lit.[185] m.p. 78-79 °C). δ_H(300 MHz, CDCl$_3$) 9.45 (1H, d, J = 2.3 Hz, H2), 8.86 (1H, d, J = 1.7 Hz, H4), 8.17 (1H, d, J = 9.3 Hz, H5), 7.94 (1H, d, J = 9.0 Hz, H8), 7.84 (1H, dt, J = 8.0 and 9.3 Hz, H7), 7.63 (1H, t, J = 8.5 Hz, H6), 4.02 (3H, s, OCH$_3$); δ_C(300 MHz, CDCl$_3$) 165.86, 150.02, 149.84, 138.80, 131.90, 129.50, 129.15, 127.49, 126.84, 122.99, 52.53.

3-Hydroxymethylquinoline 33
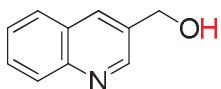
Methyl-3-quinolinecarboxylate **32** (1.0 g, 0.005 mol) in dry THF (90 mL) was stirred with Red-Al (1.5 mL, 1 equiv.) under nitrogen for 5 h at 0 °C. The excess Red-Al was decomposed with hydrochloric acid (2.0 mL, 3 *M*) and the solvent was removed *in vacuo* to give a yellow solid. Sodium carbonate (50 mL, 10%) was added and the solid was extracted into dichloromethane (4 × 50 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate concentrated *in vacuo*. The crude product was purified by column chromatography (0-10% methanol/chloroform) to give the title compound **33** as a pale yellow solid (715 mg, 84%), m.p. 83-86 °C (lit.[157] m.p. 83.5-84 °C, lit.[161] m.p. 65-67 °C; δ_H(300 MHz, CDCl$_3$) 8.91 (1H, d, J = 2.3 Hz, H2), 8.16-8.10 (2H, m, H4 and H5), 7.83 (1H, d, J = 9.0 Hz, H8), 7.71 (1H, dt, J = 9.3 and 7.7 Hz, H7), 7.56 (1H, dt, J = 9.0 and 7.7 Hz, H6), 4.93 (2H, s, CH$_2$), 2.05 (1H, bs, OH); δ_C(300 MHz, CDCl$_3$) 149.90, 146.90, 134.25, 133.97, 129.40, 128.45, 127.90, 127.76, 126.88, 62.21.

3-Thiomethylquinoline Q3
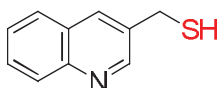
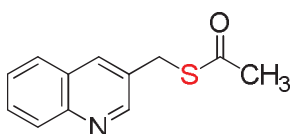
3-Hydroxymethylquinoline **33** (120.0 mg, 0.75 mmol) in benzene (2.0 mL) was treated with phosphorous tribromide (1.0 mL) in benzene (1.0 mL) at 0 °C. The solution was stirred for 3 h, filtered and dried to give the hydrobromide salt **34** as a yellow solid (140 mg, 84%), which was used immediately in the next step. A solution of thiourea (50.0 mg, 0.63 mmol) in DMF (0.27 mL) was added to 3-

bromomethylquinoline hydrobromide **34** (140 mg, 0.63 mmol) in DMF (0.84 mL, 0.75 mol) and the reaction mixture was heated at reflux for 3 h. The reaction mixture was concentrated *in vacuo* and the residue was mixed with sodium hydroxide solution (5.0 mL, 2 *M*) and was heated at reflux for 2.3 h under nitrogen. The mixture was cooled to room temperature, filtered and the filtrate was cooled in ice, acidified with concentrated hydrochloric acid and then neutralized with sodium bicarbonate solution (12.0 mL, 10%) and was extracted into diethyl ether (3 × 50 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate concentrated *in vacuo* to give the title compound **Q3** as a yellow solid (12 mg, 10%). δ_H(300 MHz, D_2O) 8.90 (1H, d, *J* = 2.3 Hz, H2), 8.43 (1H, d, *J* = 2.3 Hz, H4), 8.12-8.05 (2H, m, H5 and H8), 7.89 (1H, t, *J* = 8.6 Hz, H7), 7.74 (1H, t, *J* = 8.4 Hz, H6), 3.39 (2H, s, CH_2). The compound had properties consistent with the hydrochloride salt **37** prepared using the alternate route (Scheme 2.2).

S-Quinolin-3-ylmethyl ethanethiolate 36

3-Hydroxymethylquinoline **33** (150 mg, 0.94 mmol) in dichloromethane (2.0 mL) was stirred with triethylamine (45 μL, 1 equiv.) at room temperature for 10 min. To the stirred solution, methane sulfonylchloride (0.36 mL, 1.5 equiv.) was added and the stirring was continued for 4.3 h. The reaction mixture was extracted into dichloromethane (3 × 30 mL) and washed with cold water (3 × 15 mL) to remove unreacted sulfonyl chloride. The combined organic layers were dried over anhydrous sodium sulfate, filtered and the filtrate

concentrated *in vacuo* to give the methanesulfonate as a dark liquid in quantitative yield, which was used immediately in the next step. The methanesulfonate (240 mg, 1.01 mmol) was added to a stirred solution of potassium thioacetate (1.08 g, 0.009 mol) in acetone (20 mL) and the reaction mixture was stirred at room temperature for 4 h. The solvent was removed *in vacuo* to give a dark residue and the residue was extracted into dichloromethane (3 × 30 mL) and washed with cold water (3 × 15 mL) to remove unreacted potassium thioacetate. The combined organic layers were dried over anhydrous sodium sulfate, filtered and the filtrate concentrated *in vacuo* to give the required thioacetate **36** as a dark brown residue (180 mg, 82%). δ_H(300 MHz, CDCl$_3$) 8.83 (1H, d, J = 2.3 Hz, H2), 8.08 (2H, m, H4 and H5), 7.78 (1H, d, J = 8.7 Hz, H8), 7.69 (1H, dt, J = 8.0 and 9.3 Hz, H7), 7.54 (1H, t, J = 8.3 Hz, H6), 4.28 (2H, s, CH$_2$), 2.37 (3H, s, CH$_3$); δ_C(300 MHz, CDCl$_3$) 190.47, 175.56, 147.22, 132.56, 132.28, 129.20, 128.35, 128.08, 125.07, 30.63, 30.53, 29.84, 20.77; MS (+ESI) m/z 217.96 (M$^+$, 67%); HRMS (ES$^+$) m/z calcd. for C$_{12}$H$_{11}$NOS$^+$ 218.06396, found 218.06357; λ_{max}(methanol)/nm 237 (ε/dm^3 mol^{-1} cm^{-1} 35367) and 203 (34498).

1,2-Bis(quinolin-3-ylmethyl)disulfane Q3-Q3
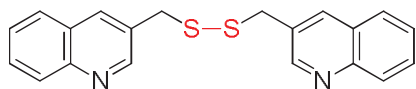

Hydrochloric acid (1.40 mL, 10 *M*) was added to a solution of quinoline-3-methylthioacetate **36** (180.0 mg, 0.83 mmol) in methanol (20 mL) and was heated at reflux for 3.3 h. The solvent was removed *in vacuo* to give the hydrochloride salt of the crude thiol **37** as a black residue (140 mg, 80%). δ_H(300 MHz, CDCl$_3$) 9.25 (1H, d, J = 1.7 Hz, H2), 8.84-8.78 (2H, m, H4 and

H5), 8.10-8.00 (2H, m, H7 and H8), 7.88 (1H, t, J = 8.5 Hz, H6), 4.11 (2H, d, J = 9.0 Hz, CH_2), 3.48 (1H, s, NH), 2.27 (1H, t, J = 9.2 Hz, SH).

The hydrochloride salt **37** (70.0 mg, 0.33 mmol) in water (10 mL) was oxidized through gentle bubbling of air *via* pipette in an open vial for 3.5 days. The pH of the solution was adjusted to ~ 7-8 using sodium hydroxide and the solution was extracted into chloroform (3 × 30 mL). The combined organic layers were washed with water (3 × 10 mL) and dried over anhydrous sodium sulfate, filtered and the filtrate concentrated *in vacuo* to give the disulfide **Q3-Q3** as a dark brown solid (70 mg, 61%), m.p. 143-145 °C. δ_H(300 MHz, $CDCl_3$) 8.78 (1H, d, J = 2.7 Hz, H2), 8.09 (1H, d, J = 9.7 Hz, H4), 7.84-7.69 (3H, 2m, H5, H7 and H8), 7.57 (1H, dt, J = 9.0 and 7.7 Hz, H6), 3.73 (2H, s, CH_2); δ_C(300 MHz, $CDCl_3$) 149.30, 139.02, 131.47, 128.46, 128.03, 126.92, 39.75, 29.81; MS (+ESI) m/z 349.15 (M$^+$, 100%); HRMS (ES$^+$) m/z calcd. for $C_{20}H_{17}N_2S_2$ 349.08331, found 349.08307; λ_{max}(methanol)/nm 233 (ε/dm^3 mol^{-1} cm^{-1} 108398) and 200 (378323).

[*N*,*N'*-Disulfanediylbis(ethane-2,1-diyl)]di-1,8-naphthalimide N1-N1
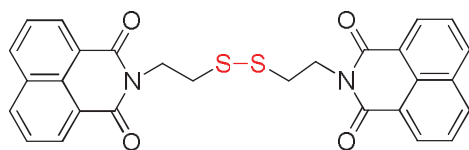

Cysteamine hydrochloride (1.16 g, 0.010 mol) was treated with triethylamine (1.5 mL) and pyridine (20 mL) at room temperature under nitrogen for 30-40 min. 1,8-Naphthalic anhydride **38** (0.50 g, 0.002 mol) was added and the resultant solution was heated at reflux (125 °C) for 25 h. The reaction mixture was cooled and filtered to remove unreacted cysteamine. The filtrate was

concentrated *in vacuo* to yield the crude product as a pale brown solid. Water (30 mL) was added to the solid followed by filtration and washed with water (3 × 10 mL), 1:1 ethanol/water and ethanol. The solid was dried under vacuum to give a mixture of the thiol **N1** and disulfide **N1-N1** as a white solid (330 mg, 51%).

The white solid (40.0 mg, 0.15 mmol) was dissolved in water (10 mL) and oxidized through gentle bubbling of air *via* pipette in an open vial for 4 days. The solvent was concentrated *in vacuo* and the residue was dried in vacuum to give **N1-N1** as a pale white solid (30 mg, 40%), m.p. 204-207 °C. δ_H(300 MHz, DMSO-d_6) 8.46 (4H, dt, J = 8.0 and 7.3 Hz, H2, H7 and H4, H5), 7.86 (2H, t, J = 8.7 Hz, H3 and H6), 4.38 (2H, t, J = 8.0 Hz, N-CH$_2$), 3.08 (2H, t, J = 8.3 Hz, S-CH$_2$); δ_C(300 MHz, DMSO-d_6) 163.40, 163.28, 134.43, 134.16, 131.32, 130.83, 130.65, 127.23, 127.18, 48.78, 36.90, 35.18; MS (+ESI) m/z 534.82 ([M+Na], 77%), 280.13 ([M+2Na], 100%), 256.34 ([M+H/2], 30%). HRMS (ES$^+$) m/z calcd. for $C_{28}H_{20}N_2NaO_4S_2$ 535.07622 found 535.07888; λ_{max}(methanol)/nm 334 (ε/dm^3 mol^{-1} cm^{-1} 11597) and 214 (62465).

4-(Chloromethyl)imidazole 40
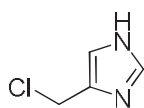
To a suspension of 4-(hydroxymethyl)imidazole **39** (153.0 mg, 1.56 mmol) in benzene (0.2 mL), a solution of thionyl chloride (0.15 mL, 2.10 mmol) in benzene (0.3 mL) was added dropwise. The resultant mixture was heated at reflux for 7 h to give a pale brown solid. The solid was filtered, washed with benzene (3 × 2 mL) followed by diethylether (3 × 5 mL) to give the title compound **40** in quantitative yield, m.p. 137-140 °C (lit.[167], m.p. 138-142 °C,

lit.[165], m.p. 138-141 °C; δ_H(300 MHz, methanol-d_4); 8.99-8.94 (1H, m, H2), 7.65 (1H, brs, H5), 4.82 (2H, s, CH$_2$); δ_C(300 MHz, methanol-d_4) 136.48, 131.79, 119.28, 34.50; MS (+ESI) m/z 117.20 (M+H, 100%).

4(5)-(Mercaptomethyl)imidazole M2
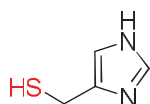
Potassium thioacetate (381.0 mg, 3.33 mmol) in methanol (2.0 mL) was added to a stirred solution of 4-(chloromethyl)imidazole **40** (129.0 mg, 1.11 mmol) in methanol (1.0 mL). The reaction mixture was stirred at room temperature for 1 day and the solvent was removed *in vacuo* to give a bright yellow residue. The residue was washed with cold water (3 × 5 mL) to remove unreacted potassium thioacetate and was extracted into dichloromethane (3 × 25 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate concentrated *in vacuo* to give the required thioacetate **41** as a yellow solid (50 mg, 29%). δ_H(300 MHz, CDCl$_3$) 9.08 (1H, brs, NH), 8.12-8.06 (1H, m, H2), 7.38-7.33 (1H, m, H5), 4.05 (2H, s, CH$_2$), 2.35 (3H, s, CH$_3$); δ_C(300 MHz, CDCl$_3$) 190.47, 137.22, 134.56, 120.26, 30.56, 26.60; MS (+ESI) m/z 157.27 (M+H , 100%), 313.00 (2M+H, 95%); HRMS (ES$^+$) m/z calcd. for $C_6H_9N_2OS^+$ 157.04356 found 157.04278; λ_{max}(methanol)/nm 216 (ε/dm^3 mol^{-1} cm^{-1} 5677).

Hydrochloric acid (0.45 mL, 1.5 mmol, 10 *M*) was added to a solution of thioacetate **41** (23.5 mg, 0.15 mmol) in methanol (2.0 mL) and the solution was heated at reflux for 3.3 h. The solvent was removed *in vacuo* to give the hydrochloride salt of the title compound **M2** as a pale yellow residue (17 mg, > 95%), which had spectral data in accordance with the literature.[166] δ_H(300

MHz, D$_2$O) 8.69 (1H, dd, J = 5.0 Hz, H2), 7.50-7.45 (1H, m, H5), 4.11 (2H, s, CH$_2$); δ$_C$(300 MHz, D$_2$O) 133.94, 129.44, 128.63, 118.00, 117.61, 30.56, 30.27; MS (+ESI) m/z 227.27 [(2M-2)+H, 100%], 115.20 (M+H , 62%).

N-tert-(Butyloxycarbonyl)cysteamine 42
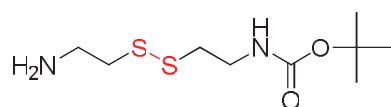
Cystamine dihydrochloride (2.0 g, 0.009 mol) in methanol (60 mL) was treated with di-*tert*-butyldicarbonate (1.94 g, 0.009 mol) and triethylamine (3.7 mL, 0.027 mol) at room temperature for 1 h. The solvent was removed *in vacuo* to give a pale brown solid, the solid was washed with sodium phosphate monobasic solution (2 × 10 mL, 1 *M*, pH 4.16) and was extracted into diethyl ether (3 × 30 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate concentrated *in vacuo* to give the di-*tert*-Boc-cystamine **43** as a pale brown solid (170 mg, 22%), m.p. 105-108 °C (lit.[170], m.p. 106-107 °C). The aqueous layer was basified to pH 9.21 using sodium hydroxide solution (3 *M*) and was extracted with ethyl acetate (3 × 30 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate concentrated *in vacuo* to give the title compound **42** as a pale yellow liquid (820 mg, 36%). δ$_H$(300 MHz, CDCl$_3$) 4.94 (1H, bs, NH), 3.44 (2H, q, J = 6.7 Hz, CH$_2$N), 3.00 (2H, t, J = 7.0 Hz, CH$_2$S), 2.77 (4H, q, J = 6.7 Hz, CH$_2$N and CH$_2$S), 1.44 (9H, s, *t*-Boc); δ$_C$(300 MHz, CDCl$_3$) 155.98, 79.73, 42.63, 40.70, 39.442, 38.56, 28.53; MS (+ESI) m/z 252.99 (M+H, 100%), 196.63 (MH$^+$-*t*-Butyl, 58%), 274.93 (M+Na, 5%); HRMS (ES$^+$) m/z calcd. for C$_9$H$_{21}$N$_2$O$_2$S$_2$ 253.10444 found 253.10378; λ$_{max}$(methanol)/nm 203 (ε/dm^3 mol^{-1} cm^{-1} 8402).

[2-(2-Aminoethyl)disulfanylethyl]quinoline-3-carboxamide.TFA Q4-Y
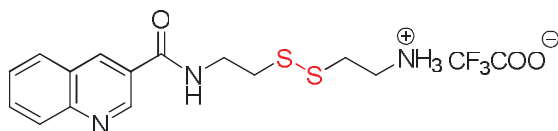
Quinoline-3-carboxylic acid **31** (258.0 mg, 1.49 mmol) was dissolved in DMF (8.0 mL) and was stirred with HOBt-hydrate (402.0 mg, 2.98 mmol), EDC hydrochloride (571.0 mg, 2.98 mmol) and triethylamine (0.41 mL, 2.98 mmol) at room temperature under nitrogen for 5 min and the reaction mixture was cooled to 0 °C. To the cold solution, *N-tert*-(butyloxycarbonyl)cystamine **42** (377.0 mg, 1.49 mmol) in DMF (2.0 mL) was added slowly and the reaction mixture was stirred at room temperature under nitrogen for 1.5 days. The solvent was removed *in vacuo* to give a dark brown solid. The solid was dissolved in dichloromethane (20 mL) and washed with water (2 × 15 mL), hydrochloric acid (2 × 20 mL, 1 *M*), saturated sodium bicarbonate (2 × 20 mL), brine (2 × 20 mL) and was extracted into dichloromethane (2 × 30 mL). The combined organic layers were dried over anhydrous magnesium sulfate and the solvent removed *in vacuo* to give the crude product as dark reddish brown solids. Purification by column chromatography (25-75% ethyl acetate/*n*-hexane) gave the *Boc*-protected compound as a pale yellow solid (425 mg, 70%), m.p.109-112 °C. δ_H(300 MHz, CDCl$_3$) 9.35 (1H, d, *J* = 2.0 Hz, H2), 8.66 (1H, d, *J* = 2.3 Hz, H4), 8.15 (1H, d, *J* = 9.3 Hz, H5), 7.92 (1H, d, *J* = 9.3 Hz, H8), 7.81 (1H, dt, *J* = 7.7 and 9.7 Hz, H7), 7.61 (1H, t, *J* = 7.8 Hz, H6), 7.39 (1H, bs, CONH), 4.96 (1H, bs, NH-Boc), 3.87 (2H, q, *J* = 6.7 Hz, NHCH$_2$), 3.50 (2H, q, *J* = 7.3 Hz, CH$_2$NH), 3.04 (2H, t, *J* = 6.7 Hz, CH$_2$S), 2.84 (2H, t, *J* = 8.0 Hz, SCH$_2$), 1.37 (9H, s, *t*-Boc); δ_C(300 MHz, CDCl$_3$) 165.54, 155.93, 147.94, 147.74, 136.76, 134.09, 131.60, 130.48, 129.16, 128.84, 128.38, 128.10, 127.63, 127.38, 127.03, 126.95, 79.65, 39.41, 39.15, 38.36, 37.83, 28.26; MS

(+ESI) m/z 429.97 (M+Na, 100%), 351.90 (MH$^+$ -t-Butyl, 70%), 407.99 (M$^+$, 60%); HRMS (ES$^+$) m/z calcd. for $C_{19}H_{25}N_3NaO_3S_2$ 430.12350 found 430.12270; λ_{max}(methanol)/nm 234 (ε/dm^3 mol^{-1} cm^{-1} 53651) and 200 (142780).

The *Boc*-protected compound (152.0 mg, 0.373 mmol) was dissolved in dry dichloromethane (5.0 mL) and was treated with TFA (1.0 mL) at room temperature under nitrogen for 2.3 h and concentrated to give the TFA salt of the title compound. The compound was purified by semi preparative RP-HPLC (5 to 85% acetonitrile:water over 25 min, Rt = 10.95 min) to afford the desired compound **Q4-Y** as a pale yellow solid (27 mg, 66%). δ_H(300 MHz, D$_2$O) 8.83 (1H, d, J = 2.7 Hz, H2), 8.45 (1H, s, CONH), 8.40 (1H, d, J = 2.3 Hz, H4), 7.89 - 7.77 (3H, m, H5, H7 and H8), 7.61 (1H, t, J = 8.3 Hz, H6), 3.77 (2H, t, J = 7.0 Hz, CH$_2$N), 3.39 (2H, t, J = 7.3 Hz, NHCH$_2$), 3.03 (4H, t, J = 7.3 Hz, CH$_2$S and SCH$_2$); δ_C(300 MHz, D$_2$O) 171.00, 167.87, 147.46, 147.32, 136.69, 132.16, 129.10, 127.90, 127.22, 126.34, 125.92, 38.67, 37.78, 36.19, 33.94; MS (+ESI) m/z 308.00 (M$^+$, 100%); HRMS (ES$^+$) m/z calcd. for $C_{14}H_{18}N_3OS_2^+$ 308.08858 found 308.08859.

[*N*-(2-Aminoethyl)disulfanylethyl]-1,8-naphthalimide.TFA N1-Y
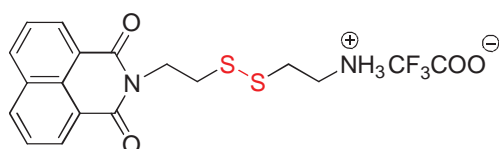
1,8-Naphthalic anhydride **38** (100 mg, 0.50 mmol) was added to a solution of *N*-*tert*-(butyloxycarbonyl)cystamine **42** (500 mg, 1.98 mmol) in pyridine (10 mL) and the resultant solution was heated at reflux (125 °C) under nitrogen

for 1.5 days. The reaction mixture was cooled, filtered and the filtrate was concentrated *in vacuo* to give a pale brown solid. Water (10 mL) was added to the solid which was filtered and washed with water (3 × 10 mL), 1:1 ethanol/water and ethanol. The solid was dried under vacuum to give the crude product as a pale brown solid. Purification by column chromatography (25-75% ethyl acetate/*n*-hexane) gave the *Boc*-protected compound as a cream solid (90 mg, 40%), m.p.150-154 °C. δ_H(300 MHz, CDCl$_3$) 8.62 (2H, d, J = 8.0 Hz, H2 and H7), 8.23 (2H, d, J = 9.0 Hz, H4 and H5), 7.77 (2H, t, J = 8.7 Hz, H3 and H6), 5.29 (1H, bs, CH$_2$NH), 4.55-4.50 (2H, m, NCH$_2$), 3.48-3.42 (2H, m, CH$_2$S), 3.06-3.01 (2H, m, CH$_2$NH), 2.90 (2H, t, J = 7.0 Hz, SCH$_2$), 1.43 (9H, s, *t*-Boc); δ_C(300 MHz, CDCl$_3$) 164.29, 164.20, 155.96, 134.33, 134.15, 131.72, 131.57, 131.46, 128.30, 127.11, 127.05, 122.52, 79.73, 39.68, 39.30, 38.53, 35.54, 28.50; MS (+ESI) *m/z* 454.99 (M+Na, 100%), 374.98 (M- *t*-Butyl, 40%), 432.99 (M$^+$, 20%); HRMS (ES$^+$) *m/z* calcd. for C$_{21}$H$_{24}$N$_2$NaO$_4$S$_2$ 455.10752 found 455.10663; λ_{max}(methanol)/nm 333 (ε/dm^3 mol^{-1} cm^{-1} 25883), 230 (144196) and 200 (497524).

The *Boc*-compound (66.0 mg, 0.15mmol) was dissolved in dry dichloromethane (5.0 mL) and was treated with TFA (0.50 mL) at room temperature under nitrogen for 3 h and concentrated *in vacuo* to give TFA salt of the title compound. The compound was purified by semi preparative RP-HPLC (45 to 85% acetonitrile:water over 25 min, Rt = 8.12 min) to give the title compound **N1-Y** as a white solid (36 mg, 71%). δ_H(300 MHz, D$_2$O) 8.11 (4H, dd, J = 8.0 and 9.0 Hz, H2 and H4, H5 and H7), 7.58 (2H, t, J = 8.7 Hz, H3 and H6), 4.24 (2H, t, J = 8.0 Hz, NCH$_2$), 3.38 (2H, t, J = 7.0 Hz, CH$_2$N), 3.04 (2H, t, J = 7.3 Hz, SCH$_2$), 2.98 (2H, t, J = 8.0 Hz, CH$_2$S); δ_C(600 MHz, D$_2$O) 170.45, 164.94, 134.92, 131.14, 130.48, 126.61, 126.46, 120.06, 38.91, 37.42, 33.83, 33.38;

MS (+ESI) m/z 256.04 (M$^+$ - cysteamine, 100%), 333.08 (M$^+$, 35%); HRMS (ES$^+$) m/z calcd. for $C_{16}H_{17}N_2O_2S_2^+$ 333.07260 found 333.07247.

1,1'-[2,2'-Disulfanediylbis(ethane-2,1-diyl)]diguanidine.2TFA A1-A1
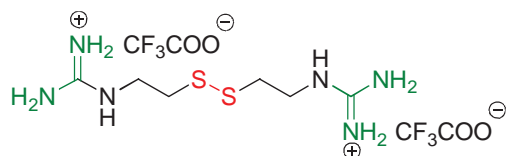
Cysteamine dihydrochloride (51.0 mg, 0.226 mmol) and N,N'-di-*Boc*-N''-trifluoro-methanesulfonylguanidine (177.0 mg, 0.452 mmol) were suspended in dry dichloromethane (6.0 mL) under nitrogen. Triethylamine (0.20 mL, 1.356 mmol) was added and the reaction mixture was stirred at room temperature under nitrogen for 14 h, while stirring which resulted in a clear solution. The reaction mixture was diluted with dichloromethane (30 mL) and the organic layer was washed with sodium bisulfate (20 mL, 2 M), saturated sodium bicarbonate (20 mL) and brine (20 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate was concentrated *in vacuo*. Purification by column chromatography (5-25% ethyl acetate/n-hexane) gave the tetra-*Boc*-protected product as a white solid (123 mg, 85%), which had spectral data in accordance with the literature.[96] δ_H(300 MHz, CDCl$_3$) 11.39 (2H, bs, CONH and NHCO), 8.62 (2H, t, J = 5.7 Hz, 2 × NH), 3.75 (4H, q, J = 7.0 Hz, NHCH$_2$ and CH$_2$NH), 2.87 (4H, t, J = 7.0 Hz, CH$_2$S and SCH$_2$), 1.49 (36H, s, t-Boc); δ_C(300 MHz, CDCl$_3$) 163.62, 156.33, 153.26, 83.38, 79.49, 39.32, 37.20, 28.42, 28.21; MS (+ESI) m/z 637.25 (M$^+$, 100%), 437.08 [M^{3+} - 2(t-Boc), 52%], 337.00 [M^{4+} - 3(t-Boc), 25%], 236.96 [M^{5+} - 4(t-Boc), 18%], 659.18 (M+Na, 18%), 537.14 (MH$^+$ - t-Boc, 7%).

The tetra-*Boc*-bisguanidine (84.0 mg, 0.132 mmol) in dry dichloromethane (2.5 mL) and was treated with TFA (400 μL) at room temperature under nitrogen for 2.3 h and concentrated to give the TFA salt of the title compound **A1**-**A1** as a white solid in quantitative yield, which had spectral data in accordance with the literature.[96] δ_H(300 MHz, D$_2$O) 3.54 (4H, t, J = 7.0 Hz, NHCH$_2$ and CH$_2$NH), 2.91 (4H, t, J = 7.3 Hz, CH$_2$S and SCH$_2$); δ_C(300 MHz, D$_2$O) 163.84, 41.20, 37.20; MS (+ESI) m/z 236.96 (M$^+$, 100%).

4-Tritylsulfanylmethylbenzoic acid methyl ester 45
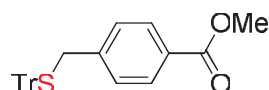
N,N-Diisopropylethylamine (0.61 mL, 3.50 mmol) and triphenylmethanethiol (484.0 mg, 1.75 mmol) were added to a solution of methyl-4-(bromomethyl)-benzoate **44** (400.0 mg, 1.75 mmol) in DMF and the reaction mixture was stirred at room temperature under nitrogen overnight. The reaction mixture was diluted with diethyl ether (20 mL) and was extracted into diethyl ether (3 × 20 mL) and washed with brine (3 × 10 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate was concentrated *in vacuo*. Purification by column chromatography (5-25% ethyl acetate/*n*-hexane) gave the title compound **45** as a white solid (560 mg, 75%), which had spectral data in accordance with the literature.[95] δ_H(300 MHz, CDCl$_3$) 7.90 (2H, d, J = 7.3 Hz, 2 × CH aromatic), 7.49-7.45 (6H, m, 6 × CH aromatic), 7.35-7.15 (11H, m, 11 × CH aromatic), 3.89 (3H, s, CH$_3$), 3.35 (2H, s, CH$_2$); δ_C(300 MHz, CDCl$_3$) 166.98, 144.60, 142.70, 130.30, 129.84, 129.70, 129.32, 129.20, 128.88, 128.12, 128.06, 126.94, 67.77, 52.18, 36.86.

4-Tritylsulfanylmethylbenzoic acid 46
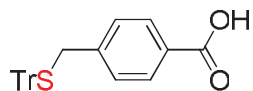
4-Tritylsulfanylmethylbenzoic acid methyl ester **45** (250.0 mg, 0.59 mmol) in methanol (10 mL) was heated at reflux with sodium hydroxide (1.10 mL, 2 *M*) overnight. The solvent was removed *in vacuo* to give a pale yellow solid and the solid was dissolved in ethyl acetate (20 mL), washed with formic acid (30 mL, 5%) and was extracted with ethyl acetate (2 × 20 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate was concentrated *in vacuo*. Purification by column chromatography (5-25% ethyl acetate/*n*-hexane) gave the title compound **46** as a white solid (209 mg, 86%), which had spectral data in accordance with the literature.[95] δ_H(300 MHz, DMSO-d_6) 12.85 (1H, bs, COOH), 7.82 (2H, d, J = 9.3 Hz, 2 × CH aromatic), 7.44-7.18 (17H, m, 17 × CH aromatic), 3.32 (2H, s, CH$_2$); δ_C(300 MHz, DMSO-d_6) 167.13, 144.15, 141.52, 129.46, 129.14, 129.07, 128.17, 126.90, 67.19, 35.93; MS (+ESI) m/z 432.82 (M+Na, 100%).

[4,4'-Disulfanebis(4-methylbenzamido)]diguanidine.2TFA A2-A2
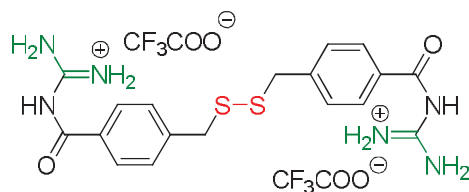
4-Tritylsulfanylmethylbenzoic acid **46** (381.0 mg, 0.93 mmol) was dissolved in DMF (3.0 mL) and the solution was stirred with PyBOP (531.0 mg, 1.021 mmol) and *N*-methylmorpholine (600 µL) at room temperature under nitrogen for 45 min. *Boc*-guanidine (295.0 mg, 1.86 mmol) was added to the stirred solution and the resultant solution was stirred at room temperature overnight.

Water (40 mL) was added and the mixture was extracted with ethyl acetate (3 × 50 mL), washed with water (3 × 25 mL) and brine (3 × 25 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate was concentrated *in vacuo* to give the crude product as a pale brown residue. Purification by column chromatography (5-35% ethyl acetate/*n*-hexane) gave the *Boc*-protected product as a pure white solid (210 mg, 41%), m.p.184-188 °C. δ_H(300 MHz, CDCl$_3$) 7.97 (2H, d, *J* = 8.3 Hz, 2 × CH aromatic), 7.43-7.14 (17H, m, 17 × CH aromatic), 3.30 (2H, s, CH$_2$), 1.38 (9H, s, *t*-Boc); δ_C(300 MHz, CDCl$_3$) 144.63, 129.70, 129.33, 128.12, 128.04, 127.88, 126.92, 67.72, 36.86, 28.03; MS (+ESI) *m/z* 573.98 (M+Na, 100%), 551.98 (M+H, 40%); HRMS (ES$^+$) *m/z* calcd. for $C_{33}H_{34}N_3O_3S$ 552.23209 found 552.23178; λ_{max}(methanol)/nm 258 (ε/dm^3 mol^{-1} cm^{-1} 35527) and 203 (121024).

The *Boc*-compound (127.0 mg, 0.23 mmol) in TFA (2.5 mL) was stirred with triethylsilane (400 μL) at room temperature under nitrogen overnight. The reaction mixture was concentrated *in vacuo* to give the TFA salt of the title compound. The compound was purified by semi-preparative RP-HPLC (10 to 85% acetontrile:water over 20 min, Rt = 10.51 min) to give a mixture of thiol **A2** and disulfide **A2**-**A2** as a white solid. The crude product (12 mg, 0.03 mmol) was dissolved in water (10 mL) and oxidized through gentle bubbling of air *via* pipette in an open vial for 2 days. The solvent was removed and the residue was dried *in vacuo* to give the disulfide **A2**-**A2** as a white solid (12 mg, 25%). δ_H(600 MHz, D$_2$O) 7.85 (2H, d, *J* = 2.5 Hz, 2 × CH aromatic), 7.42 (2H, d, *J* = 2.3 Hz, 2 × CH aromatic), 3.77 (2H, s, CH$_2$); δ_C(600 MHz, D$_2$O) 170.41, 169.45, 155.24, 144.12, 129.75, 129.39, 129.08, 127.83, 40.89, 29.41; MS

(+ESI) m/z 209.22 [(M+2H)$^{2+}$, 80%]; HRMS (ES$^+$) m/z calcd. for $C_{18}H_{22}N_6O_2S_2^{2+}$ 209.12347 found 209.12178.

O-4-Cyanophenyl-*N*,*N*-Dimethylthiocarbamate 49
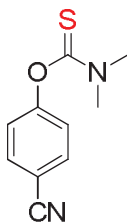
4-Cyanophenol **47** (2.0 gm, 0.0168 mol) and DABCO (4.70 gm, 0.042 mol) in DMF (25 mL) were stirred at room temperature for 15 min. To the well stirred solution, *N*,*N*-dimethylthiocarbamoylchloride **48** (2.60 gm, 0.021 mol) was added and the resultant solution was heated at reflux under nitrogen for 5 h. The reaction mixture was poured onto crushed ice and acidified to pH ~ 3.0 with hydrochloric acid (6 *M*). The resultant precipitate was filtered and dried to give the title compound **49** as a pale yellow solid (3.2 gm, 92%), which had spectral data in accordance with the literature.[173-174] δ_H(300 MHz, CDCl$_3$) 7.69 (2H, d, J = 9.7 Hz, aromatic), 7.19 (2H, d, J = 9.7 Hz, aromatic), 3.45 (3H, s, *Me-N-Me*), 3.36 (3H, s, *Me-N-Me*); δ_C(300 MHz, CDCl$_3$) 186.55, 157.16, 133.63, 133.51, 124.30, 118.48, 109.96, 43.51, 39.07.

S-4-Cyanophenyl-*N,N*-Dimethylthiocarbamate 50
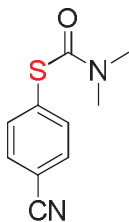
O-4-Cyanophenyl-*N,N*-dimethylthiocarbamate **49** (2.50 g, 0.012 mol), in a flask fitted with a reflux condenser, was immersed in a preheated metal bath (Wood's metal) under argon. The flask was maintained at 210 °C and was stirred very slowly for 6 h. The reaction was cooled to give the pure rearranged title compound **50** as a pale brown solid (2.40 g, > 95%), which had spectral data in accordance with the literature.[173-174] δ_H(300 MHz, CDCl$_3$) 7.68-7.59 (4H, m, aromatic), 3.07 (6H, d, J = 8.7 Hz, NMe$_2$); δ_C(300 MHz, CDCl$_3$) 165.00, 135.74, 135.67, 135.63, 132.90, 132.26, 126.67, 118.49, 112.66, 37.11.

S-4-Phenyl-*N,N*-dimethylthiocarbamate-4-guanidine.HCl 51
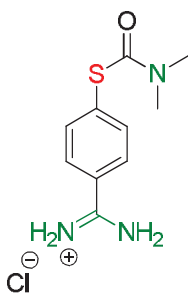
Lithium bis(trimethylsilyl)amide in THF (5.0 mL, 1 *M*) was added to a solution of *S*-4-cyanophenyl-*N,N*-dimethylthiocarbamate **50** (354.0 mg, 1.72 mmol) in THF (8.0 mL) under nitrogen at 0 °C and the resultant solution was stirred at room temperature for 19 h. The reaction mixture was cooled to 0 °C and was acidified with hydrochloric acid in methanol (10.0 mL, 1.25 *M*) and the stirring was continued at room temperature for 4 h. The solvent was removed *in vacuo*

to give a yellow residue and the residue was triturated with dry ether (10 mL) and filtered. The resultant solid was dissolved in cold water (50 mL) and basified with sodium hydroxide (1.0 M) to pH ~ 9.0 and the precipitated yellow solid was filtered, washed with cold water (5.0 mL) and dried. The yellow solid was suspended in absolute ethanol (8.0 mL) and the suspension was treated with a solution of ethanol saturated with hydrochloric acid, and filtered. The filtrate was concentrated to give a pale yellow residue and the residue was stirred with anhydrous ether (5.0 mL) for 30 min, filtered, washed with dry ether (2 × 5 mL) and dried. The compound was purified by semi-preparative RP-HPLC (10 to 35% acetonitrile:water over 15 min, Rt = 10.56 min) to give the title compound **51** as a pale yellow solid. δ_H(300 MHz, methanol-d_4) 8.56 (1H, brs, NH), 7.80 (2H, d, J = 9.3 Hz, aromatic), 7.71 (2H, d, J = 9.7 Hz, aromatic), 3.09 (6H, bd, NMe$_2$); δ_C(300 MHz, methanol-d_4) 170.46, 168.11, 167.10, 137.42, 136.99, 130.24, 129.18, 37.26; MS (+ESI) m/z 224.30 (M$^+$ 100%); HRMS (ES$^+$) m/z calcd. for $C_{10}H_{14}N_3OS^+$ 224.08576 found 224.08575.

4-Cyanophenyl Disulfide 54
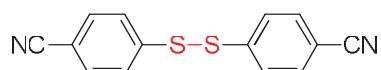
S-4-Cyanophenyl-N,N-Dimethylthiocarbamate **50** (2.31 g, 0.011 mol) in methanol (7.0 mL) was treated with potassium hydroxide (2.0 mL, 0.04 mol) at room temperature for 5 h. The reaction mixture was poured into crushed ice and the pH was adjusted to ~ 2.0 and was stirred well until the oil solidified. The resultant precipitate was filtered and washed with cold water to give 4-cyanobenzenethiol **53** as a pale brown solid (1.27 g, 84%), which had spectral data in accordance with the literature.[173-174] δ_H(300 MHz, CDCl$_3$) 7.50 (2H, d, J = 9.7 Hz, aromatic), 7.31 (2H, d, J = 9.3 Hz, aromatic), 3.67 (1H, s, SH); δ_C(300

MHz, CDCl$_3$) 139.34, 132.68, 128.83, 118.76, 108.94; MS (+ESI) m/z 136.26 (M+H, 90%).

A stirred solution of 4-cyanobenzenethiol **53** (581 mg, 4.30 mmol) in DMSO (4.0 mL) was heated at reflux under nitrogen for 5 h. The reaction mixture was cooled to room temperature and was poured into vigorously stirred crushed ice. The resultant precipitate was filtered, washed with cold water (2 × 20 mL) and dried to give the title disulfide **54** as a pale brown solid (600 mg, 52%), which had spectral data in accordance with the literature.[173] δ$_H$(300 MHz, CDCl$_3$) δ 7.62-7.54 (4H, m, aromatic); δ$_C$(300 MHz, CDCl$_3$) 142.01, 132.71, 126.41, 118.09, 110.81.

4,4'-Disulfanediyldibenzimidamide.2HCl A3-A3
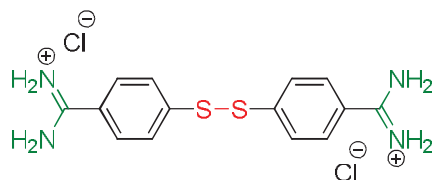
Lithium bis(trimethylsilyl)amide in THF (2.5 mL, 1 *M*) was added to a solution of 4-cyanophenyl disulfide **54** (73.0 mg, 0.27 mmol) in dry THF (2.0 mL) under nitrogen at 0 °C and the resultant solution was stirred at room temperature for 2 days. The reaction mixture was cooled to 0 °C and was acidified with a solution of ethanol (3.0 mL) saturated with hydrochloric acid(g) under nitrogen and the mixture was stirred at room temperature for 4 days. To the reaction mixture, anhydrous diethyl ether (15 mL) was added and the solution was cooled to 0 °C for 40 min, filtered and washed with dry ether to give a pale yellow solid. The crude product was purified by semi preparative RP-HPLC (0.5 to 30% acetontrile:water over 13 min, Rt = 5.97 min) to give the title

compound **A3-A3** as a white solid (22 mg, 22%). δ_H(500 MHz, D$_2$O) 7.91 (2H, dd, J = 8.5 and 6.7 Hz, aromatic), 7.84 (2H, dd, J = 8.5 and 6.7 Hz, aromatic); δ_C(400 MHz, D$_2$O) 166.38, 158.96, 129.57, 128.55, 124.35; MS (+ESI) m/z 152 (M^{2+}, 100%).

2,2'-[Ethane-1,2-diylbis(4-methylbenzenesulfonylazanediyl)]bis(ethane-2,1-diyl)bis(4-methyl-benzenesulfonate 56
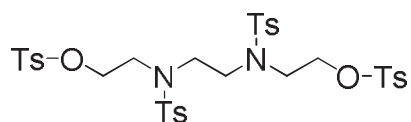
N,N'-Bis (2-hydroxyethyl)ethylene diamine **55** (1.0 g, 0.0067 mol) in dichloromethane (30.0 mL) was stirred with 4-(dimethylamino)pyridine (0.842 g, 0.0067 mol) and triethylamine (8.5 mL, 0.062 mol) at 0 °C under nitrogen for 15 min. To the stirred solution, p-toluenesulfonyl chloride (5.66 g, 0.030 mol) was added and the stirring was continued at 0 °C under nitrogen for 24 h. Additional portions of p-toluenesulfonyl chloride (2.43 g, 0.013 mol) were added and the stirring was continued at 0 °C under nitrogen for 6 h, followed by 24 h at room temperature. The reaction mixture was concentrated *in vacuo* to give the crude product as a white solid which was washed with hydrochloric acid (3 × 20 mL, 0.1 *M*) and extracted into chloroform (3 × 30 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate was concentrated *in vacuo* to give the crude product as a white solid. Purification by column chromatography (20-70% ethyl acetate/n-hexane) gave the title compound **56** as a white solid (1.56 gm, 30%), which had spectral data in accordance with the literature.[183] δ_H(300 MHz, CDCl$_3$) 7.74 (8H, dd, J = 9.0 and 9.3 Hz, aromatic tosyl protons), 7.34 (8H, d, J = 9.0 Hz, aromatic tosyl protons), 4.14 (4H, t, J = 6.0 Hz, -N(Ts)CH$_2$CH$_2$OTs),

3.36 (4H, t, J = 6.0 Hz, -N(Ts)CH$_2$CH$_2$OTs), 3.30 (4H, s, CH$_2$N(Ts)-CH$_2$CH$_2$OTs), 2.44 (12H, s, tosyl-CH$_3$); δ_C(300 MHz, CDCl$_3$) 145.32, 144.15, 135.12, 132.51, 130.17, 128.19, 127.55, 69.14, 50.03, 49.62, 21.82, 21.71; MS (+ESI) m/z 787.09 (M+Na, 47%).

S,S'-2,2'-[Ethane-1,2-diylbis(4-methylbenzenesulfonylazanediyl)]-bis(ethane-2,1-diyl)diethane-thioate 57
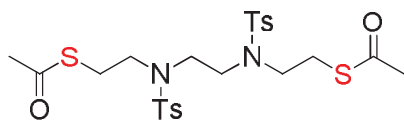
The tetra-tosyl protected compound **56** (1.50 g, 0.002 mol) in DMF (25.0 mL) was stirred with potassium thioacetate (0.56 g, 0.005 mol) at room temperature for 24 h. The reaction mixture was concentrated to give the crude product as a pale brown solid. The solid was washed with water (3 × 20 mL) and extracted into chloroform (3 × 50 mL). The combined organic layers were dried over anhydrous magnesium sulfate, filtered and the filtrate was concentrated *in vacuo* to give the title compound **57** as pale brown solid (840 mg, 75%), which had spectral data in accordance with the literature.[183] δ_H(300 MHz, CDCl$_3$) 7.74 (4H, d, J = 9.0 Hz, aromatic tosyl protons), 7.33 (4H, d, J = 9.0 Hz, aromatic tosyl protons), 3.38 (4H, s, -CH$_2$N(Ts)CH$_2$CH$_2$S), 3.26 (4H, t, J = 8.3 Hz, -N(Ts)CH$_2$CH$_2$S), 3.04 (4H, t, J = 8.3 Hz, -N(Ts)CH$_2$CH$_2$S), 2.44 (6H, s, tosyl-CH$_3$), 2.33 (6H, s, -COCH$_3$); δ_C(300 MHz, CDCl$_3$) 195.29, 143.88, 135.70, 130.03, 127.48, 49.50, 49.02, 30.76, 28.35, 21.68; MS (+ESI) m/z 595.05 (M+Na, 62%), 573.11 (M$^+$, 20%).

N,N'-Bis(2-mercaptoethyl)ethylenediamine B2
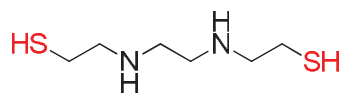
A solution of compound **57** (500 mg, 0.87 mmol) was dissolved in a solution of hydrobromic acid in acetic acid (14.0 mL, 33%) and was heated at reflux with phenol (34.0 mg, 0.36 mmol) under nitrogen for 24 h. The reaction mixture was filtered and the resultant pale brown solid was washed with ethanol (2 × 20 mL). The solid was heated at reflux with ethanol for 1 h, the solution was filtered and the solid was washed with diethyl ether (2 × 20 mL) to give the bisthiol **B2** as a pale brown solid in quantitative yield, which had spectral data in accordance with the literature procedure.[183] δ_H(300 MHz, D_2O) 3.58-3.49 (8H, m, -CH$_2$NHCH$_2$), 3.08 (4H, t, J = 7.3 Hz, -NHCH$_2$CH$_2$SH); δ_C(300 MHz, D_2O) 46.23, 42.97, 31.63.

Chapter 3

Formation
and
Analysis of DCLs

3.1 DCL Methods

Thiol disulfide reversible reactions have been the most widely used reversible reactions to generate DCLs in aqueous conditions.[78,186-187] The interconversion of thiols and disulfides is one of the few reversible reactions that can be carried out with biomolecules, due to the following properties;[92-93]

- The chemistry required to form disulfide DCLs is compatible with water;
- The reactions are relatively fast;
- Initiation and quenching of the DCLs are pH controlled.

Several methods have been reported[78] to prepare disulfide DCLs, starting from thiols, disulfides or a mixture of both thiols and disulfides (see Figure 2.2). As discussed in Section 1.5.1, almost all examples of DCC with nucleic acids have used reversible oxidation of thiol building blocks in the presence of GSSG/GSH at pH 7.4.

In this work, two general methods were used to prepare DCLs by using either redox chemistry (method A) or disulfide exchange (method B) (Figure 3.1). In each method, two different reagents and reaction conditions were used to initiate the reactions at either neutral pH (method A) or under basic conditions (method B). These four sets of reaction conditions were studied prior to any studies with DNA, in order to establish the reactivity of the building blocks, and to optimize the conditions for formation of the DCLs. Particular attention was paid to solubility and ensuring that all building blocks and the disulfides formed in the DCLs remained soluble for a period of time suitable to undertake experiments with DNA.

Method (A):

(a)

$$R_1-S-S-R_1$$

$$+$$

$$R_2-S-S-R_2$$

$$+$$

etc.

GSH

GSSG

$$R_1-S-S-R_2 \ + \ R_2-S-S-R_3$$

$$R_1-S-S-G \ + \ \text{etc.}$$

H^+/H_2O

(b)

$$R_1-S-S-R_1$$

$$+$$

$$R_2-S-S-R_2$$

$$+$$

etc.

DTT^{Red}

DTT^{Oxd}

$$R_1-S-S-R_2 \ + \ R_2-S-S-R_3$$

$$+ \ \text{etc.}$$

H^+/H_2O
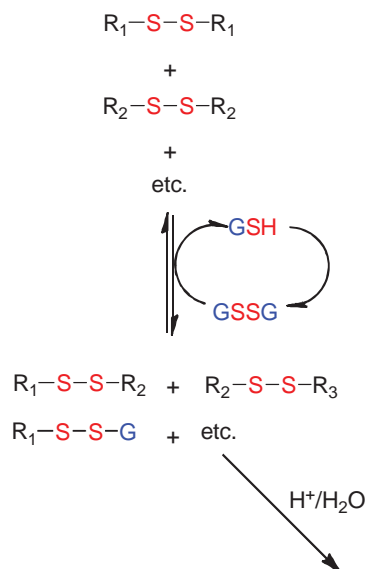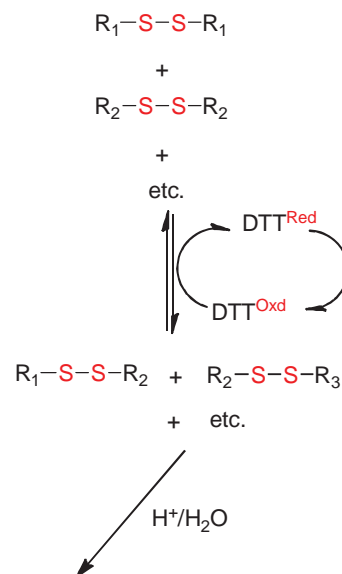
Library of Disulfides

Method (B):

(c)

H^+/H_2O

$$R_1-S-S-R_2 \ + \ R_2-S-S-R_3$$

$$R_1-S-S-R \ + \ \text{etc.}$$

R-SH
Base

etc.

$$+$$

$$R_2-S-S-R_2$$

$$+$$

$$R_1-S-S-R_1$$

(d)

H^+/H_2O

$$R_1-S-S-R_2 \ + \ R_1-S-S-Y$$

$$R_2-S-S-R_3 \ + \ \text{etc.}$$

Y-SH
Base

etc.

$$+$$

$$R_2-S-S-Y$$

$$+$$

$$R_1-S-S-R_1$$
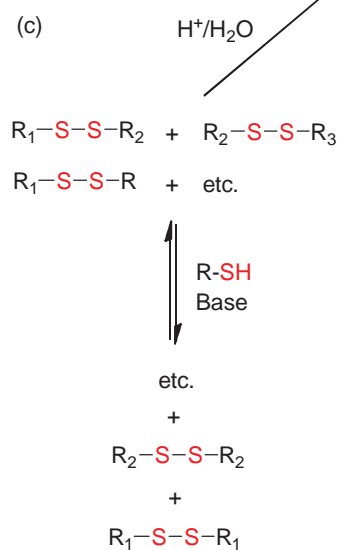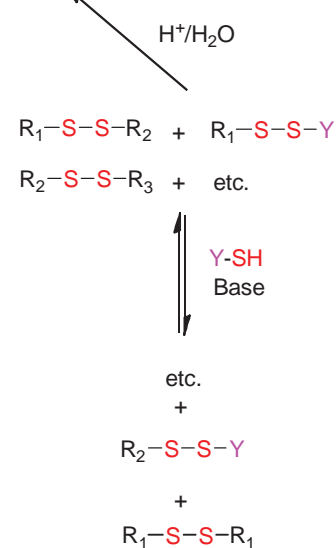
Figure 3.1 Generation of disulfide DCLs using Method A: redox reactions at neutral pH (~ 7.4) employing either (a) GSSG/GSH or (b) DTT^{RED}/DTT^{OXD} and Method B: exchange reactions at basic pH (~ 8.5) employing either (c) RSH or (d) cysteamine as thiol initiators.

In method A, reactions were performed using similar methods to those reported[92-95] with nucleic acids (pH 7.4, GSSG/GSH) (Figure 3.1a) or with dithiothreitol (Figure 3.1b). GSH is the most abundant cellular thiol, which plays a vital role in maintaining the redox equilibrium.[188-189] Both oxidized (GSSG) and reduced (GSH) glutathione (Figure 3.2a) are required to form DCLs, with the ratio of GSH over GSSG essential in regulating the redox potential.[190-191] The advantages of utilizing GSSG/GSH are to facilitate the redox reaction at neutral pH, as well as generating additional diversity to the constituents of the library, as the thiol and disulfide reagents also participate in the DCC reactions.

(a)
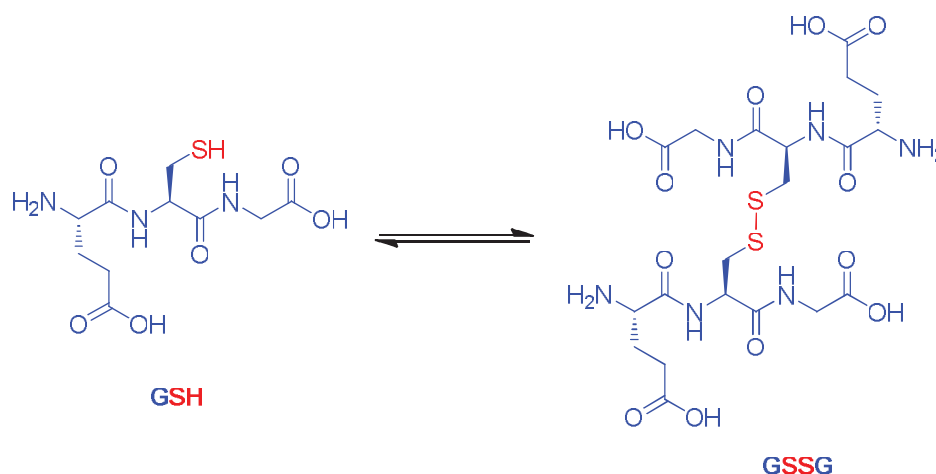
(b)
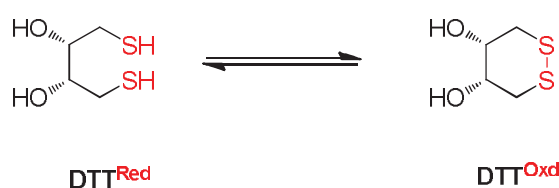
Figure 3.2 Structures of oxidized and reduced forms of (a) glutathione and (b) dithiothreitol.

The reversible chemistry is quenched by acidification of the solution. DCLs were also generated with dithiothreitol (DTT, Cleland's reagent[192]), as an

alternate reducing agent that does not participate in the DCC reactions. Dithiothreitol has been reported in only one DCL study with DNA, but no details were provided.[96] Dithiothreitol has been widely used as a reducing agent in protein chemistry;[193] it rapidly reduces disulfide groups by generating the very stable six membered cyclic oxidized species (Figure 3.2b).[190,194]

In method B, disulfide exchange conditions were employed to generate DCLs using either a thiol building block (R-SH) as the initiator with symmetrical disulfides (Figure 3.1c) or cysteamine (YSH) as the initiator and disulfides including derivatives of cysteamine (Figure 3.1d). These are attractive methods for studies in aqueous solution as the formation of the charged thiolate anion at basic pH to initiate the reaction also improves aqueous solubility. These reaction conditions have not been reported in any studies with nucleic acids.
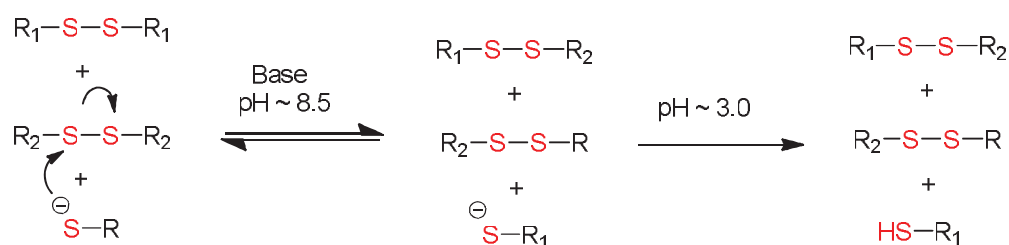
Figure 3.3 Mechanism of disulfide exchange.

The mechanism of disulfide exchange reaction is shown in Figure 3.3. Under basic conditions (typically pH 8-9), the thiolate anion undergoes a nucleophilic substitution reaction with the symmetrical disulfide generating a new thiolate anion, which continues the exchange reaction.[190] As a result of this chemistry, different combinations of disulfides are generated at equilibrium under reversible conditions. The exchange reactions are quenched by acidifying the solution to pH ~ 3.0, allowing isolation of the disulfides in the DCL. These

experiments were carried out using symmetrical disulfides and a thiol building block.

The unsymmetrical disulfides **Q4**-**Y** and **N1**-**Y** were designed to allow the use of cysteamine as a thiol initiator to initiate the exchange reaction. This approach has been reported by Ulven *et al.*[96] The unsymmetrical disulfides were also used as starting materials in experiments conducted under conditions shown in method A (Figure 3.1).

3.2 Preparation and Analysis of DCLs

The thiol or disulfide building blocks were typically equilibrated for 48-72 h after initiation of library formation and quenched with aqueous TFA, prior to analysis using LC-MS. In some experiments, precipitation occurred over time, and hence the DCLs were generated in 10-20% methanol/water, which gave clear solutions. In the case of **Q3** and **N1**, despite modification of the solvent mixtures, at μmol concentrations, precipitation occurred during the course of the reaction. Quinoline **Q3** was also insoluble in 10-20% methanol/water at acidic pH. Hence, no further experiments were carried out with the building blocks **Q3** and **N1**. However, the corresponding cysteamine disulfide derivatives of **Q3** and **N1**, **Q4**-**Y** and **N1**-**Y** respectively, which are positively charged, were soluble in aqueous methanol solutions and the experiments were able to be conducted with these building blocks.

Table 3.1 summarizes the typical reaction conditions used to generate thiol-disulfide DCLs conducted in this work. The reaction conditions and the concentrations of the building blocks were varied to provide traces that

Table 3.1 Summary of reaction conditions used to prepare disulfide DCLs, calculated and observed number of components in DCLs.[a]

Starting Material	(nmol)	Initiation	DCL Members Calcd.[b]	Found
Q4-Y	100			
A1-A1	500	pH ~ 8.5	16	10
A2-A2	90			
S1	200			
Q4-Y	100	1. GSSG/GSH		
A1-A1	500	(1.25/5.0 mM)	23	17
A2-A2	90			
S1	200	2. Adjust pH ~ 7.4		
Q4-Y	100	1. DTT^{Red} (5.0 mM)		
A1-A1	500		16	10
A2-A2	90	2. Adjust pH ~ 7.4		
S1	200			

[a]The pH of the initial solution was adjusted by addition of ammonium hydroxide and the DCLs quenched by addition of TFA to pH ~ 3.0.

[b]The number of species calculated as a liner combination of all species that can form the possible disulfide species with the reactive thiol.

allowed optimal resolution of the peaks by LC-MS and provided sufficient concentrations of products to allow detection and identification by mass spectrometry. The DCC libraries were analyzed by LC-MS using UV detection (total scan PDA) followed by mass (m/z) measurements, which allowed detection of both UV active and UV inactive DCL members. The use of dilute solutions (< 500 μmol) concentrations of the building blocks, where all components remained soluble, resulted in the formation of weak signals that were at the limit of the machine's sensitivity. Hence, most of the experiments were performed at mmol concentrations. However, at these higher concentrations, reduced solubility in water or aqueous methanol (10-20%) solutions resulted in some cloudiness and precipitation over time.
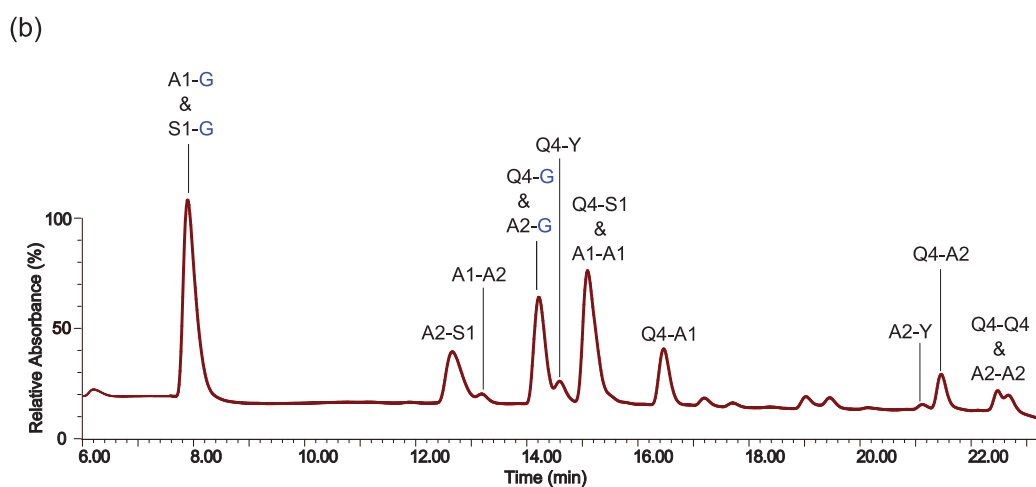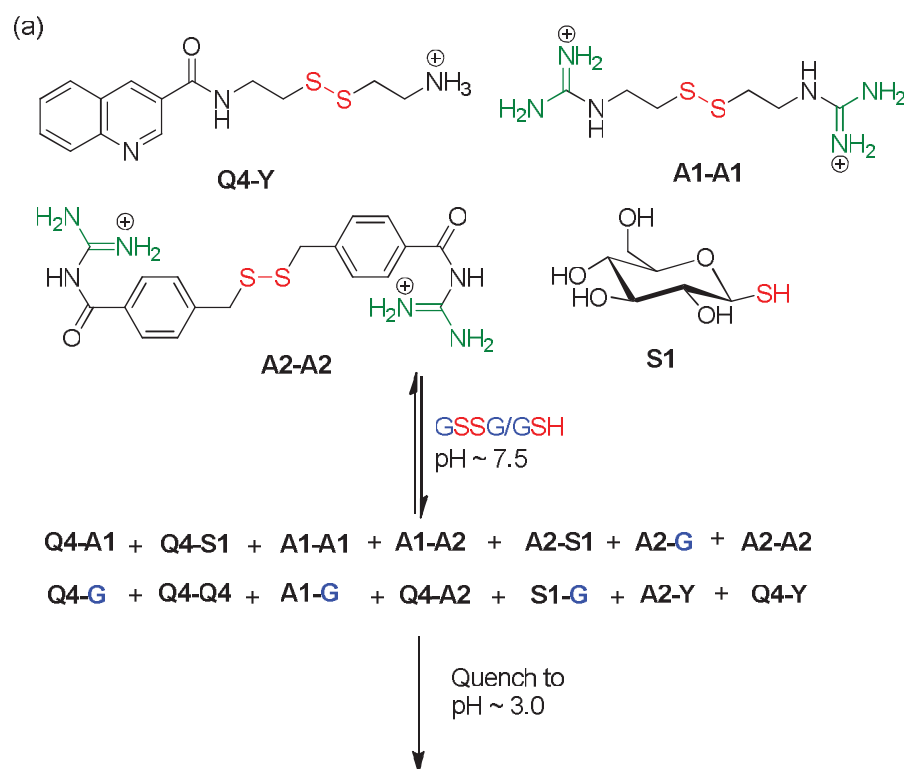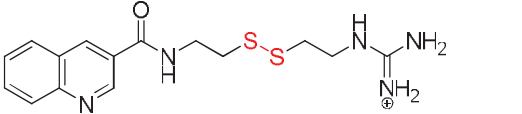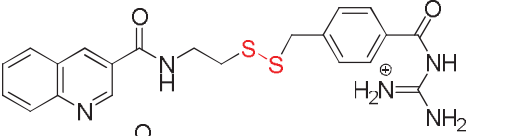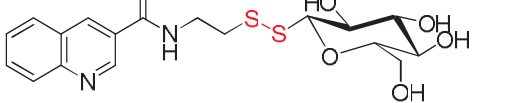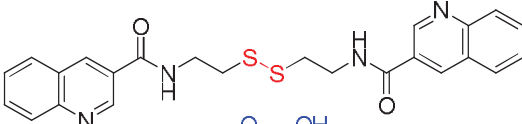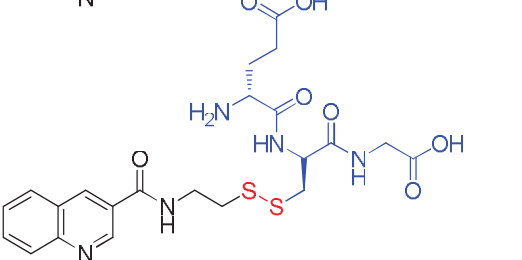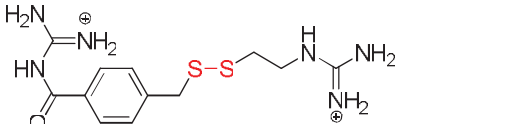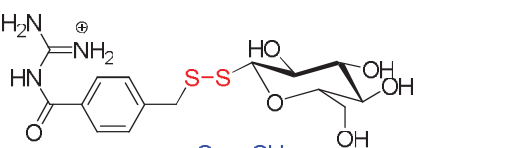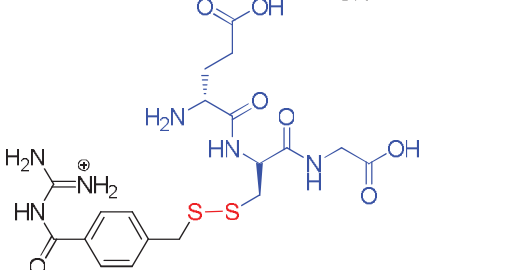
Figure 3.4 DCL experiment generated from (a) thiol **S1** and disulfides **Q4**-**Y**, **A1**-**A1** and **A2**-**A2** using GSSG/GSH, pH ~ 7.4 and (b) section of the LC trace of the DCL with UV detection (210-400 nm); UV active DCL species are labeled including adducts of GSH (blue).

Figure 3.4 shows the results of a typical DCL experiment. Quinoline **Q4**-**Y**, amidines **A1**-**A1**, **A2**-**A2** and thiosugar **S1** were equilibrated for 48 h in the presence of GSSG/GSH (Figure 3.4a) and the reaction was quenched and analysed by LC-MS (Figure 3.4b). The major products identified by mass spectrometry were the disulfides shown in Figure 3.5. However, there were a significant number of minor peaks also detected by UV (Figure 3.4b), which were not present at high enough concentration to allow accurate characterisation by LC-MS. In this experiment, the relative absorbance and the intensity of the peaks in GSSG/GSH adducts were stronger than other disulfide species (see Figure 3.4b). In order to regulate the redox potential, more concentrated solution of both oxidized (GSSG) and reduced (GSH) were used to activate the DCL. The relative amount of each disulfide generated in the DCL experiment reflects the concentrations of the building blocks used, the relative stability and solubility of the products formed under equilibrium conditions.

From this experiment, quinoline **Q4**-**Y** generated all possible disulfides with the building blocks (Figure 3.5). In the case of **A1**-**A1**, which is UV-inactive, the presence of **A1**-**A1** at the same retention time as **Q4**-**S1** was confirmed by analysis of the total ion current trace. Four disulfides of the benzylic amidines **A2**-**A2**, (**Q4**-**A2**, **A1**-**A2** and **A2**-**S1**) were detected along with the glutathione adduct (**A2**-**G**).

3.3 Comparison of DCL Methods

In order to compare the different methods for generating disulfide DCLs, the formation of the DCL shown in Figure 3.4 was repeated using the same

Figure 3.5 Structures of (a) quinoline adducts and (b) amidine adducts identified in the DCL shown in Figure 3.4 with calculated and detected masses.
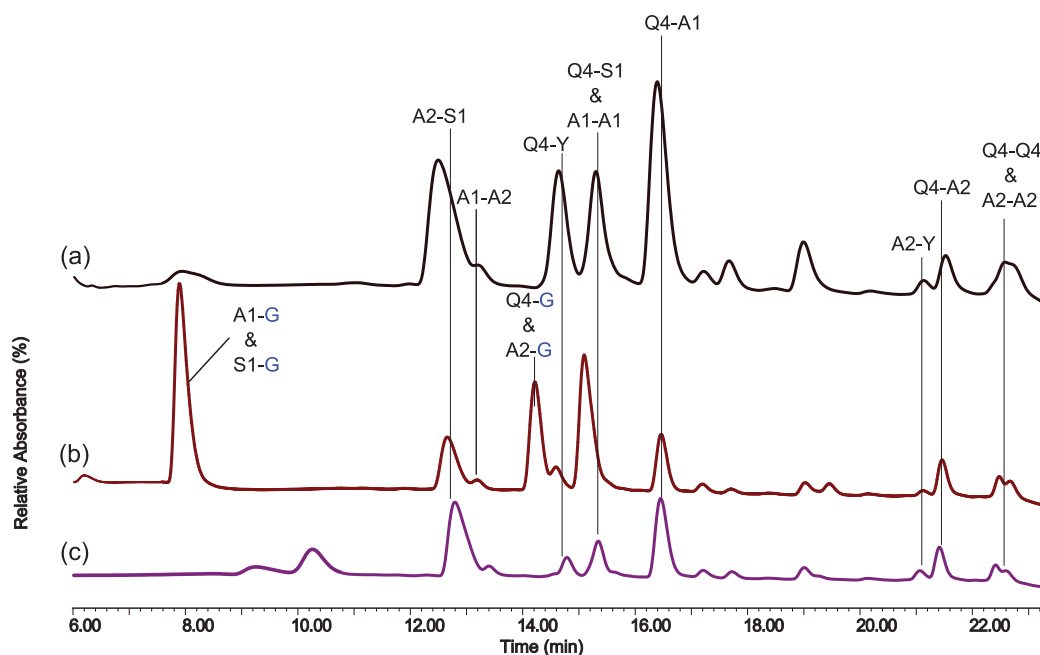
Figure 3.6 LC trace with UV detection (210-400 nm) of DCL generated from thiol **S1** and disulfides **Q4-Y**, **A1-A1** and **A2-A2** using (a) exchange (pH ~ 8.5) (b) GSSG/GSH, pH ~ 7.4 and (c) DTT, pH ~ 7.4.

concentrations of building blocks under both exchange and redox conditions using either GSSG/GSH or DTT and the results are shown in Figure 3.6. There are clear differences in the number and diversity of the species generated in each library. The maximum numbers of disulfides that may be generated using the different reaction conditions were calculated (Table 3.1), to include the participation of GSSG/GSH in the reaction. Several charged and water soluble GSH adducts were identified and hence using GSSG/GSH in the reaction added additional diversity to the constituents of the library. Analysis of the DCLs showed, not surprisingly, that adducts of DTT were not formed. Formation of the stable six membered cyclic ring (DTT^{Oxd}) is strongly preferred to the linear bisthiol (DTT^{Red}) (Figure 3.2b) and other linear adducts of DTT.

3.4 Summary

Four methods for the generation of DCL libraries in the presence of duplex DNA were investigated based on redox chemistry and disulfide exchange experiments. While the redox methods have been reported in the literature, the use of disulfide exchange with nucleic acids has not been studied, as this requires basic conditions, which are not typically compatible with DNA.

The poor aqueous solubility of building blocks **Q3** and **N1** at μmol concentrations, under all conditions investigated, resulted in precipitation during the equilibration time to form the DCLs. These properties indicated that these building blocks are not suitable for DCC studies with DNA. However, the corresponding cysteamine disulfide derivatives of **Q3** and **N1**, **Q4**-**Y** and **N1**-**Y** respectively, were soluble in aqueous methanol and DCLs using these derivates were able to be generated under conditions compatible with duplex DNA.

The formation of DCLs under the same DCC reaction conditions using the same building blocks were compared using the four methods. Based on these results, it was proposed to generate DCLs in the presence of duplex DNA under two sets of conditions using either (i) GSSG/GSH at neutral pH, or (ii) disulfide exchange at basic pH. The first method using GSSG/GSH was selected in preference to DTT, as these conditions have been reported in the literature with nucleic acids.[92-95] In addition, the formation of charged GSH adducts in the reaction provided a mechanism to form water soluble derivatives with the building blocks that had limited aqueous solubility. However, the formation of GSH adducts in these experiments can complicate the interpretation of the results, and can make identification of amplified

products difficult especially when high concentrations of the redox reagents are used.[93] The second method, which has not been used previously with nucleic acids, used disulfide exchange and carefully controlled basic pH to initiate the formation of the DCLs. This method was used as this is a relatively easy experiment to conduct, and allows ready interpretation of the results compared to experiments conducted under redox conditions. In addition, the initial formation of the charged thiolate anion improves aqueous solubility of the compounds, which may assist in ensuring all compounds remain in solution.

3.5 Experimental

3.5.1 Materials and Methods

Commercially available reagents and solvents (HPLC quality) were used without further purification, unless otherwise stated. All references to water refer to the use of Milli-Q water generated from a Millipore, Milli-Q Bicel A 10 system.

3.5.2 LC-MS Conditions

The mobile phase consisted of eluents A (water with 0.1% formic acid) and B (acetonitrile) for all runs. Analytical RP-HPLC was performed with a Phenomenex Synergy 4 μm Hydro RP C18 column (4.6 x 100 mm column, 4 μm particle size, and flow rate of 1.0 mL min^{-1}). The DCL experiment and analysis was performed by LC-MS using a Waters2690 separations module with a Waters996 photodiode array detector (210-400 nm) and was attached to Waters micromassZQ spectrometer. MassLynx software is used with

Micromass mass spectrometry (MS) detectors, using the IEEE-488 interface between the Micromass computer and the Waters 2695 Separations Module, which include two UV detector channels (Waters 2487 and/or 486 Tunable Absorbance Detectors) and one RI detector channel (Waters 2410 or 410 Differential Refractometer). The instrument conditions were optimised for sensitivity on both solvent (background) and compound using MassLynx™ V4.0 software and Micromass® MassLynx™ software (version 3.5 or higher). MS^n product ion scans were performed using collision induced dissociation (30 eV), the data was initially acquired in full scan mode and the ions generated were measured over the mass range 100-800 Da. Data was analyzed using the MassLynx™ V4.0 and the software that uses IEEE-488 Communications.

3.5.3 Formation of DCLs

Stock solutions were prepared at the following concentrations: **Q4**-**Y** (1.0 mM); **A1**-**A1** (5.0 mM), **A2**-**A2** (1.0 mM); **S1** (2.0 mM); GSSG/GSH (1.25/5.0 mM) or DTT^{Red} (5.0 mM). Depending on the solubility of the DCL components, the stock solutions were prepared in water or aqueous methanol (10-20%) in water. Stock solutions were stored in the freezer until required, and allowed to reach room temperature before use.

The DCC reactions were initiated at physiological pH ~ 7.4 (GSSG/GSH or DTT) or by addition of ammonium hydroxide (1.0%) to give pH ~ 8.5. The solutions were stirred for 72 h with a gentle stream of air bubbled in the solution *via* pipette, by which time equilibrium was reached. Typical DCC experiments were conducted with the following conditions (see Figure 3.1):

Method A(a): **Q4-Y** (1.0 mM, 100 μL, 100 nmol); **A1-A1** (5.0 mM, 100 μL, 500 nmol) and **A2-A2** (1.0 mM, 90 μL, 90 nmol); thiosugar **S1** (2.0 mM, 100 μL, 200 nmol); GSH (5.0 mM, 75 μL, 375 nmol) and GSSG (1.25 mM, 75 μL, 93.75 nmol) and ammonium hydroxide (10.0 μL, 1.0%) were mixed to give a total volume of 550.0 μL at pH ~ 7.4.

Method A(b): **Q4-Y** (1.0 mM, 100 μL, 100 nmol); **A1-A1** (5.0 mM, 100 μL, 500 nmol) and **A2-A2** (1.0 mM, 90 μL, 90 nmol); thiosugar **S1** (2.0 mM, 100 μL, 200 nmol); DTT^{Red} (5.0 mM, 100 μL, 500 nmol) and ammonium hydroxide (10.0 μL, 1.0%) were mixed to give a total volume of 500.0 μL at pH ~ 7.4.

Method B(c): **Q4-Y** (1.0 mM, 100 μL, 100 nmol); **A1-A1** (5.0 mM, 100 μL, 500 nmol) and **A2-A2** (1.0 mM, 90 μL, 90 nmol); thiosugar **S1** (2.0 mM, 100 μL, 200 nmol) and ammonium hydroxide (20.0 μL, 1.0%) were mixed to give a total volume of 410.0 μL at pH ~ 8.5.

An aliquot of the equilibrated DCL solution (40 μL) was transferred into an eppendorf tube and was quenched by aqueous TFA (15 μL, 0.1%) to pH (~ 3.0) and was analysed by LC-MS. For each analysis 40 μL was injected per run. The following elution gradients were employed: 95% A for 2 min, then gradient was raised to reach 30% B at 15 min, then increased to 100% B at 20 min and held at 100% B for 2 min, then ramped to reach 95% A at 25 min, then held at 95% A until 35 min.

Chapter 4

DNA-Binding Studies

4.1 Rules for DNA Molecular Recognition

Medicinal and synthetic chemists have designed new ligands in order to elucidate the rules that govern DNA recognition.[65-66,195] While significant progress has been made by modifying the linker rigidity and/or positioning of various functional groups in the design of synthetic molecules that bind to the minor groove of DNA,[68] there remains poor understanding of the precise positioning of functionality that brings about well-defined contacts leading to molecular recognition.[12,41,196-199]

In this section a brief review of the rules for DNA recognition are summarized as well as examples that illustrate the challenges in designing new DNA-binding agents. Of particular relevance to the results in this chapter are the DNA-binding characteristics of carbohydrates, substituted aromatics and heterocycles containing two fused aromatic rings.

(a) Hydrogen bonding and shape characteristics of the grooves

The most important feature involved in recognition of the base sequence of DNA and the grooves of DNA is hydrogen bonding. The Watson-Crick base pairs (Figure 4.1) are differentiated on the minor groove by the positions of the hydrogen bond donor and acceptor groups which are distinct in G:C *versus* A:T base pairs. The exocyclic amine of guanine (G) represents an unsymmetrical hydrogen bond donor group on the minor groove edge of the G:C base pair, and the amino group protrudes into the minor groove providing a steric clash compared to the A:T base pair.[37,200] The A:T base pairs on the
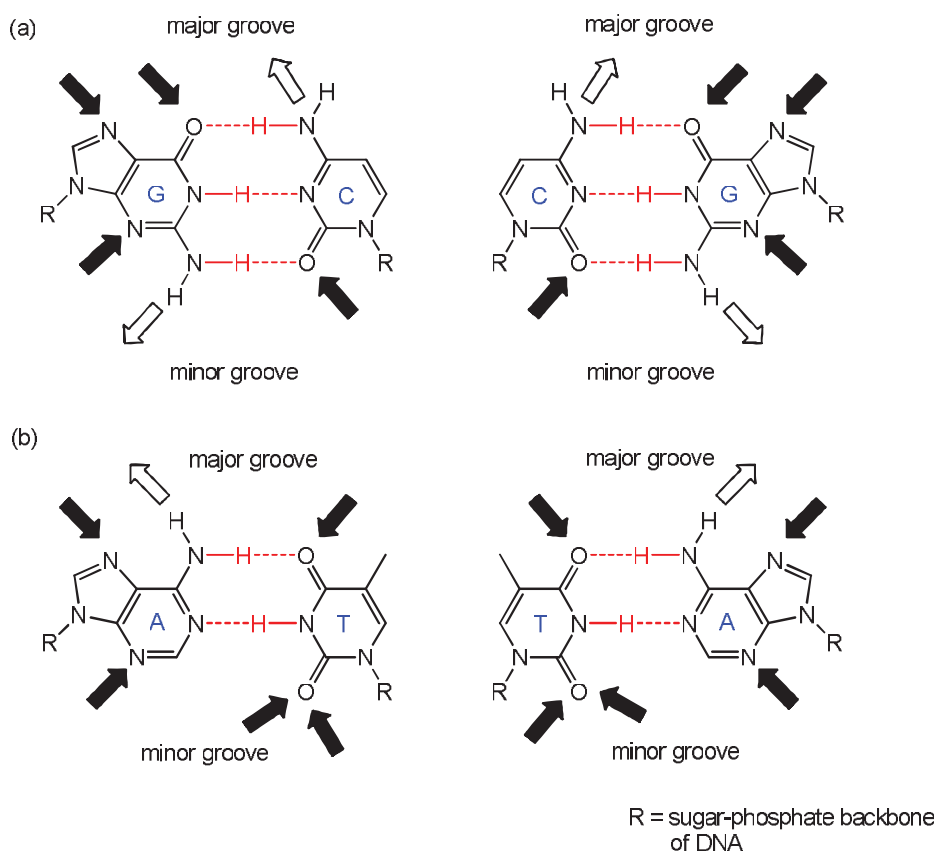
Figure 4.1 Hydrogen bond donor (plane arrow) and hydrogen bond acceptor (bold arrow) sites in the major and minor grooves[1-2,12] in (a) G:C and C:G and (b) A:T and T:A Watson-Crick base pairs.

minor groove are more symmetrical, having a hydrogen bond acceptor on both orientations of the A:T base pairs (Figure 4.1b).

While the hydrogen bond donor and hydrogen bond acceptor sites are well-defined, designing a molecule to match these characteristics is not straight forward, as the exact dimension of the major and minor grooves depends on the base sequence. It is still not possible to accurately predict the shape and properties of the major and the minor groove of a given DNA base pair

sequence and significant conformational changes occur with different base pair sequences upon binding, which affect the mode of binding.[41,196-197] Hence, one of the major difficulties in designing a molecule to match the shape of the minor/major groove to bind to specific DNA sequences is that DNA is not a rigid receptor, but is a conformationally flexible target.

(b) Minor groove binders

Polyamide based heterocyclic small-molecules have been extensively studied by Dervan and others and have provided a set of rules for recognition of A:T rich DNA sequences *via* hydrogen bonding.[37,40,68,198-200] Distamycin **2** interacts with d(GAATT). (CTTAA) sequences using hydrogen bonding and the molecule binds to the minor groove of A:T rich DNA sequences that form complexes with DNA in 1:1 and 2:1 stoichiometries. The key feature of distamycin **2** (Figure 1.2, Chapter 1) that allow binding to this specific sequences are the small heterocyclic pyrrole rings which are linked by amide bonds. The partially restricted amide bonds allow distamycin **2** to adopt a conformation in which the overall shape complements the convex surface of the minor groove; there is an isohelical interaction by hydrogen bonding. In addition hydrophobic and van der Waals interactions are provided by the *N*-methyl group and the positively charged amidine functional groups are involved in electrostatic interactions.[37,41] Based on these studies, lexitropsins have been designed that have the same overall curved shape, but in which the heterocyclic ring has been changed thus allowing different sequences to be recognized.[201-202]

Hoechst 33258 (Figure 1.4a) is also a well-characterized DNA minor groove binder. In contrast to distamycin **2**, the major feature of Hoechst 33258 that is

important in DNA recognition is the crescent shape that fits the curvature of the minor groove.[12,71-74] Hoechst 33258 preferentially binds to A:T rich sequences rather than G:C sequences due to the narrow minor groove of A:T rich DNA, which does not contain the bulky amino group present in the G:C rich DNA sequence (see Figure 4.1). The DNA-binding interaction of the molecule is stabilized by hydrophobic and van der Waals interactions of the terminal phenol and electrostatic interaction provided by the *N*-methylpiperazine rings.[68]

Another important class of DNA minor groove binders are deoxy sugars, which are generally substituted on DNA-intercalators or DNA-threaders. When present as oligosaccharides, these sugars are unusually hydrophobic and twist to follow the unwinding path of the minor groove. The hydrophobicity of the sugar enhances DNA-binding characteristics and can influence the specificity.[153-154,195] An important example of a sequence-specific DNA intercalator-groove binder is daunomycin **3** (Figure 1.2b, Chapter1). The cationic amino-sugar attached to the ring facilitates the well-defined fit of the molecule into the right handed minor groove *via* electrostatic and hydrogen bonding interactions and displace water molecules and ions from it.[1,50-53] In addition, the molecule is also stabilized by hydrophobic and hydrogen bonding interactions during intercalation. The conformation of the daunosamine sugar is also important, as it changes significantly on binding relative to the B-DNA and these changes facilitate the snug fit of the antibiotic into the right handed minor groove.[1,19-20,54]

(c) Major groove binders

In contrast to minor groove binders, the design of small molecules that recognize the major groove DNA is more difficult. The major groove is significantly wider than the minor groove (Figure 1.1a, Chapter1) hence, it is more challenging to design large molecules that match the hydrogen bond characteristics in the major groove.[154,203] The major class of natural compounds that bind to major groove are proteins in which the tertiary structure of the protein positions amino acid side chains to make specific contacts with the base pairs in the major groove.

An important class of synthetic molecules that have been successfully designed to match the shape of the major groove are transition metal complexes based on the pioneering work of Barton and colleagues.[204-205] For example, the octahedral $[Rh(phi)]^{3+}$ metal complex interacts with the major groove. The metal complex is a structurally rigid molecule with a well-defined shape and symmetry, and intercalation of one of the aromatic ligands from the major groove allows the complex to fit the shape of the major groove. The chirality of the metal complex is also important in recognizing the chiral helix and the overall positive charge interacts favorably with the negatively charged sugar-phosphate backbone.[206-207] By changing the one or more of the ligands in the metallo-intercalator, to incorporate hydrogen bonding functionalities the sequence preference of the complex can be changed.

(d) Non-classic minor groove binders

Recent studies on a number of linear amidines have shown unexpected DNA-binding profiles, and in particular, identified new minor groove binders which

do not follow the rules for the polyamide based minor groove binders discussed in the previous section. Linear diamidines, such as CGP 40125A (Figure 4.2) interact with DNA *via* the minor groove, despite the molecule lacking the crescent/curvature shape feature of the polyamides, and surprisingly CGP 40125A shows strong binding to A:T rich DNA sequences. Crystallographic studies have shown that an –NH- group on the central linker forms bifurcated hydrogen bonding with the O2 of two thymine (T), which is important to A:T binding.[12,69] However, the amidine groups at the other two ends of the CGP 40125A molecule forms hydrogen bond bridge with two water molecules to the N-atoms (N2 and N3) of guanine (G) in the DNA-binding site. The presence of bound water molecules has been observed in related diamidine DNA-binding complexes,[12,47] and showed that the inclusion of water molecules are important in molecular recognition and designing DNA-binding molecules.

CGP 40215A
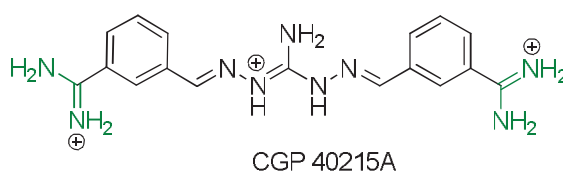
Figure 4.2 Example of linear diamidine DNA minor groove binder.[12,69]

A second example of the unexpected DNA-binding properties of diamidines are illustrated by DB950 (Figure 1.4b), discussed briefly in section 1.4. Ethidium bromide **1** (Figure 1.2a) a classic, strong DNA intercalator was converted to an A:T specific DNA minor groove binder by the replacement of the two amino groups with diguanidine functional groups. The conversion of a well-known intercalating agent (ethidium bromide **1)** into a minor groove binder (DB950)

by substituted guanidine functional group completely altered the DNA sequence selectivity.[75]

(e) Intercalators

The general principles for the design of DNA intercalators are well-understood. Planar, aromatic molecules that contain a minimum of three fused rings almost always act as strong DNA intercalators. Heterocyclic systems are most common, as the hetero atom provides a site for protonation and changes the electronic distribution of the aromatic system, an important factor in maximising stacking interactions. In general, the stacking interactions that occur with G:C base pairs are more favourable than A:T base pairs and hence, most intercalators interact with G:C rich sequences. The DNA-binding mode of planar, aromatic molecules that contain only two fused rings is varied. Many heterocycles (e.g. quinolines, quinoxalines and indoles) are intercalators, but the substitution on the ring is important and can change the binding mode to groove binding or lead to both intercalation and groove binding. In the case of Hoechst 33258 and RT29 (Figure 1.4), the rigid and bulky diaryl rings result in steric interactions that prevent intercalation; hence, the molecule prefers minor groove binding.[12,68,71-74]

In summary, DNA-binders such as ethidium bromide **1**, distamycin **2** and daunomycin **3** (Figure 1.2a) have led to a set of rules for the design of DNA minor groove binders and intercalator-groove binder hybrid molecules. These studies have led to set of rules for "classic" DNA binders and are,

- The molecule contains sufficiently conformational flexibility to allow it to adopt an overall shape fits the curvature of the minor groove;

- Positively charged functional groups to enhance electrostatic interactions;
- Appropriate hydrogen-bonding groups for sequence recognition of N3 (A) and O2 (T) and NH (G) and O (C) in the in minor groove.

The discovery of several linear molecules that bind to minor groove and a derivative of ethidium bromide **1** that is a minor groove binder <u>not</u> an intercalator have highlighted that the discovery of new classes of DNA binders may be possible. In this context, DCC is an ideal technique to discover novel compounds that would not be designed using classic rules. The presence of DNA in solution means that all possible conformations are available for the DCL components to interact with, and water molecules can also be involved in molecular recognition.

4.2 Design of Oligonucleotides

In this work, the potential for DCC to identify new DNA-binding molecules was investigated by using short oligonucleotides as models for DNA. While short oligonucleotides lack the long range effects present in cellular DNA, they provide useful molecular level information about DNA recognition. In particular, many NMR studies and X-ray crystallographic studies of small molecule oligonucleotide interactions have provided to understand clearly the specific interactions that lead to binding and sequence selectivity.[47,69,208-209]

DNA-binding experiments were carried out with three oligonucleotide sequences **O1**, **O2** and **O3** (Table 4.1). All three sequences were purchased

with a 5'-biotin group that was linked to the oligonucleotide with the linker biotinTEG phosphoramidite (L). The sequences included the 10 base pair self-complementary G:C and A:T rich sequences, d(GC)$_5$ **O2** and d(AT)$_5$ **O3**, respectively. The third sequence was the 20 base pair sequence hpGC **O1**, which forms a hairpin sequence in solution. The hairpin sequence contained a single G:C intercalation binding site flanked by A:T base pairs and has been used previously to study the binding parameters, structure and dynamics of a series of DNA and elinafide complexes.[210]

Table 4.1 Oligonucleotides used in DCC Experiments[a]

Name		Sequences	Molecular Weight
O1	**hpGC**	biotin-L-TATGCATATTTTTATGCATA	6665.6
O2	**d(GC)$_5$**	biotin-L-(GCGCGCGCGC)$_2$	3599.6 (ss)
O3	**d(AT)$_5$**	biotin-L-d(ATATATATAT)$_2$	3594.7 (ss)

[a] linker L is   biotinTEG phosphoramidite (L).

Oligonucleotides **O2** and **O3** were chosen as the majority of intercalators showed a preference for binding to G:C rich sequences[55,63-64] and many DNA minor groove binders show a preference for A:T rich sequences.[46,68,211] The hairpin sequence **O1** included a unique G:C intercalation binding site and was designed to allow identification of molecules that interact with DNA *via* intercalator-groove binding. However, it is noted that many DNA-binding molecules exhibit subtle sequence selectivity effects that are not well understood, and hence more than one binding mode is possible with the oligonucleotides.

4.3 DNA-Binding Assay

The protocol for the DNA-binding experiments was based on that reported by Subramanian *et al*,[92,95] who used GSSG/GSH to generate the equilibrating

libraries. However, due to the solubility issues at the concentrations studied, modified conditions were also investigated.

Figure 4.3 shows an overview of the DNA-binding assay. Briefly the biotinylated oligonucleotide was annealed by heating at 100 °C and then slowly cooled to room temperature to ensure formation of duplex DNA. The DCL solution was allowed to equilibrate in the presence of the DNA, thus allowing the duplex to select the optimal binding compounds from the solution. The biotinylated oligonucleotides were immobilized onto streptavidin functionalized magnetic beads by the strong biotin-streptavidin interaction which allowed separation of the DNA from the unbound compounds. The DNA was denatured by heating to 95 °C and the single-strand sequences were removed using a magnet.

4.4 Normalization and Analysis of Spectra

In order to compare the results obtained with each oligonucleotide, it was necessary to normalise the spectra with respect to one another. Experiments were conducted by addition of an internal standard to the initial DCL solution, in order to allow direct comparison of the relative amounts of a specific disulfide in the control library, as well as the relative amount of compound that bound to each oligonucleotide, and the amount of compound that was left (unbound) in solution.

2,2'-Bipyridyl was initially chosen as an internal standard, as this heterocycle does not interact with DNA. However, in the LC-MS analysis of some

i. Mix equi. vol. of TE buffer
   and **anneal** at 100 °C,
ii. **cool** to room temperature (RT)

5'--

Annealed ds DNA

○ = Biotin

◯ = Dynabeads

5'--- TATGCATA ^{T} T
      ATACGTAT $_T$ T

5'--- GCGCGCGCGC
      GCGCGCGCGC

5'--- ATATATATAT
      ATATATATAT

i. Add **DCL**
ii. **Equilibrate**

Selection of best binder

etc.,

i. **Quench** to pH ~ 3.0
ii. Add Streptavidin ◯
    coated **Dynabead**s
iii. Incubate at RT

etc.,

i. Separate the beads
   with **Magnet**
ii. Wash with 0.1 % TFA

etc.,

Unbound

i. **Denature** DNA at 95 °C
ii. Separate with magnet

Binding
compound

Figure 4.3 Schematic of DNA-binding assay.

111

experiments, the bipyridyl peak was broadened suggesting some non-specific interactions with DNA. In addition, the peak co-eluted with some of the products formed in the DCL. Hence, benzoic acid was used as a water soluble internal standard in DCL studies.[212] Experiments were also conducted with this compound as the standard, but the internal standard overlapped with the DCL peaks and also peak broadening was obtained similar to those with 2,2'-bipyridyl. Hence the assay experiments were modified, and a known amount of a standard solution of 3,5-dihydroxybenzoic acid[95] was added to the final solutions in the same volume, thus allowing the spectra to be normalised with respect to one another.

DCLs have been typically analyzed at a particular wavelength near an isosbestic point[92,95-96,213] (a specific wavelength at which the chemical species have the same molar absorptivity (ε) or more generally, are linearly related. For example, Huc and Lehn *et al*[213] used UV-Vis absorption spectra to allow an equally sensitive detection of the different imine products at a given wavelength (230 nm). Bagaut *et al*[95] reported the peaks as percentage proportion changes based on the peaks areas normalised at 220 nm with respect to an internal standard, 3,5-dihydroxybenzoic acid, whereas, Balasubramanian *et al*[92] calculated the change in equilibrium mixture composition, by measurement of the peak area at 220 nm and the differences in extinction coefficients. Nielson and Ulven[96] analyzed library components at three different wavelengths, as in this study, there were distinct wavelengths at which the relative absorbance of one component was very strong and with all other components showing only weak absorption.

In this work, the quinoline-containing disulfides in the DCL did not share a common wavelength with similar relative absorbance to the other components, and hence the analysis using the methods summarised above that have been reported in the literature could not be used. Figure 4.4 shows the UV spectra of the quinolines **Q1**, **Q2** and **Q4-Y,** and the deoxy sugars **S2** and **S3-S3** starting material peaks and 3,5-dihydroxybenzoic acid. From these spectra, there was no single wavelength that could be used to detect the products using UV. Given the overlap of all absorbance curves, it was also not possible
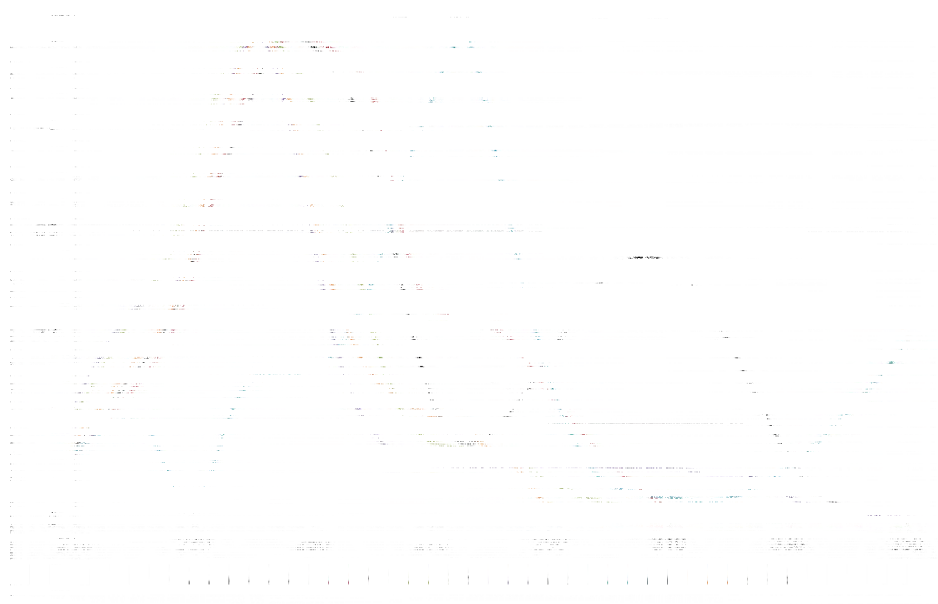
Figure 4.4 UV spectra of quinoline starting materials **Q1**, **Q2** and **Q4-Y** and thiosugars **S2** and **S3-S3**.

to perform the analysis at different wavelengths, the method employed by Nielson and Ulven.[96] Instead, a linear combination of absorbances at two wavelengths were selected near isosbestic points where the non-quinoline components have high relative absorbance and the others low and *vice versa* at the second wavelength, considering the standard deviation of the relative

113

absorbances of the components are minimum. The sum of peak areas of the two wavelengths at 234 and 286 nm were selected for all DCL components peak detection. All the DCL samples were analysed by LC-MS in the UV-Vis region (210-350 nm) followed by mass (m/z) measurements. The UV inactive compounds were identified by m/z analysis of the total ion current spectra.

4.5 Optimisation of DCC Reaction Conditions

The DNA-binding assay conditions were optimised by first conducting the experiment with compounds that do not undergo DCC. These experiments were conducted in order to optimise the equilibration time, concentration of the building blocks, denaturation and isolation of DNA, to give good quality and reproducible spectra. In addition, based on the results reported in Chapter 3, the use of 10-20% aqueous methanol solutions were required to ensure all building blocks remained in solution, and formation of the DCLs at neutral and basic pH were studied. It was therefore important to ensure that these conditions did not denature or degrade the double-stranded DNA during the course of the binding assay.

The DNA assay was tested with daunomycin **3**, an established intercalator-groove binder (See Figure 1.2a, Chapter 1), that binds strongly to DNA with K_a ~ 4.8×10^6 M^{-1}.[67,214] 8-Aminoquinoline was also studied as a weak DNA-binder,[215-216] that was similar to the building blocks **Q1**, **Q2** and **Q4-Y**, 2,2-bipyridyl was included as a compound that does not bind to DNA. The binding specificity of daunomycin **3** is well characterised and it has a preference for binding to CGTACG sequences.[1,53] Experiments were conducted with **O1**, which has a single daunomycin binding site, to ensure that formation of a 1:1

complex could be detected in the DNA-bound fraction of the assay, with the unbound compounds detected in the supernatant fraction. The DNA assay experiments were conducted under the same conditions required for DCC, using disulfide exchange, and redox chemistry with GSSG/GSH (see Figure 3.1) in water and in 10-20% methanol. Under all conditions, similar results were obtained, suggesting that the duplex structure of DNA is not significantly affected at pH 8.5 or with 10-20% methanol present.

As the intercalation of daunomycin **3** into DNA is a reversible process, careful optimisation of the concentrations of **O1** and of daunomycin **3**, molar ratio of **O1** with respect to **3** as well as the ratio of **O1** to dynabeads were required. Experiments were conducted with 100-200 μmol solutions of **O1**, and different ratios of daunomycin (1-3 equiv.). Initial experiments with a ratio of 217 μmol concentration of daunomycin **3** *versus* 130 μmol concentration (2:1) of **O1** gave only 30-50% of daunomycin **3** bound to the DNA. Hence, the experiment was repeated by varying the concentrations of both daunomycin and DNA until almost all daunomycin **3** bound with **O1**. Approx 70% of the daunomycin **3** was detected bound to **O1** in 1:1 ratio using 196 μmol concentrations of daunomycin **3** and 176 μmol concentrations of **O1**.

4.6 Results of Oligonucleotide Binding Experiments

DCC experiments were carried out using disulfide exchange under mildly basic conditions as well as with GSSG/GSH at physiological pH. The DCL experiments with building blocks **Q1**, **Q2**, **Q4-Y**, **S2** and **S3-S3** were carried out in 10-20% methanol/water, in order to avoid mild precipitation. In the case of **Q1** and **Q2**, basic pH was required along with methanol (10-20%) in

water to ensure solubility of all components in the DCL and hence the GSSG/GSH method at physiological pH was not conducted.

In a typical experiment, the building blocks (75-1000 nmol, Table 4.2) were equilibrated for 48-72 h prior to the addition of biotinylated oligonucleotides **O1**, **O2** and **O3**. After 48 h, the reaction was quenched, and the duplex was denatured and separated, providing solutions of the DNA-bound material as well as the unbound compounds. Both samples were adjusted to a defined volume, and an internal standard added to allow LC-MS spectra to be normalised and compared.

4.6.1 Charged 3-Substituted Quinoline Derivatives

While it is well-established that charged compounds interact strongly with the DNA backbone (Section 1.3), unless the precise DNA-binding mode is known, the exact positioning of charge in a molecule is difficult to predict. Of the quinoline building blocks (Figure 2.1), the only charged derivative was the 3-substituted quinoline amide **Q4-Y**, which was designed as a disulfide derivate of cysteamine. The relative affinity of this amine versus guanidine derivatives for DNA was assessed using DCC.

Initial experiments were conducted with the 3-substituted charged quinoline **Q4-Y**, in order to identify quinoline derivatives with a high affinity for DNA. The preference for **Q4-Y** to form a disulfide with the flexible alkyl guanidine **A1-A1** was also compared with the semi-rigid guanidine **A2-A2**. The neutral sugar **S1** was included as a reference compound. **Q4-Y** was equilibrated with

Table 4.2 Summary of reaction conditions for DCC studies with DNA.[a]

Starting Material	nmol	Method		DCL Members	
		Exchange	Redox	Calcd.[b]	Found
Q4-Y	105	i. cysteamine (YSH)			
A1-A1	250	ii. Adjust pH ~ 8.5			
A2-A2	250			20	10
A3-A3	250				
Cysteamine	1000				
Q4-Y	105		i. GSSG/GSH		
A1-A1	250		ii. Adjust pH ~ 7.4		
A2-A2	250			26	14
A3-A3	250				
GSSG/GSH	94/375				
Q1	75				
Q2	75				
Q4-Y	84	Adjust pH ~ 8.5		36	23
S1	150				
S2 (α, β)	157				
S3-S3 (α, β)	100				
Q1	75				
Q2	75	Adjust pH ~ 8.5		22	10
S2 (α, β)	157				
S3-S3 (α, β)	100				
Q1	75				
M1	250				
M2	187.5	Adjust pH ~ 8.5		27	14
A1-A1	125				
A2-A2	125				
A3-A3	125				
Q2	75				
M1	250	Adjust pH ~ 8.5		14	6
M2	187.5				
A1-A1	125				
Q1	75				
Q2	100				
B1	100	Adjust pH ~ 8.5		19	5
B2	100				
B3	100				

[a]The pH of the initial solution was adjusted by addition of ammonium hydroxide and the DCLs quenched by addition of TFA to pH ~ 3.0

[b]The number of species calculated as a liner combination of all species that can form the possible disulfide species with the reactive thiol.

A1-**A1**, **A2**-**A2** and **S1** (Figure 4.5a) for 48 hours, the reaction quenched and the DNA-bound adducts analysed (*APPENDIX*, Figure 1). Not surprisingly, **Q4**-**A1** the flexible guanidinium disulfide was the predominant compound selected by all three oligonucleotides along with the bisguanidine **A1**-**A2** (Figure 4.5b). None of the benzylic quinoline disulfide i.e. **Q4**-**A2** was detected as bound to DNA, or the quinoline sugar derivative **Q4**-**S1** (Figure 4.5c).
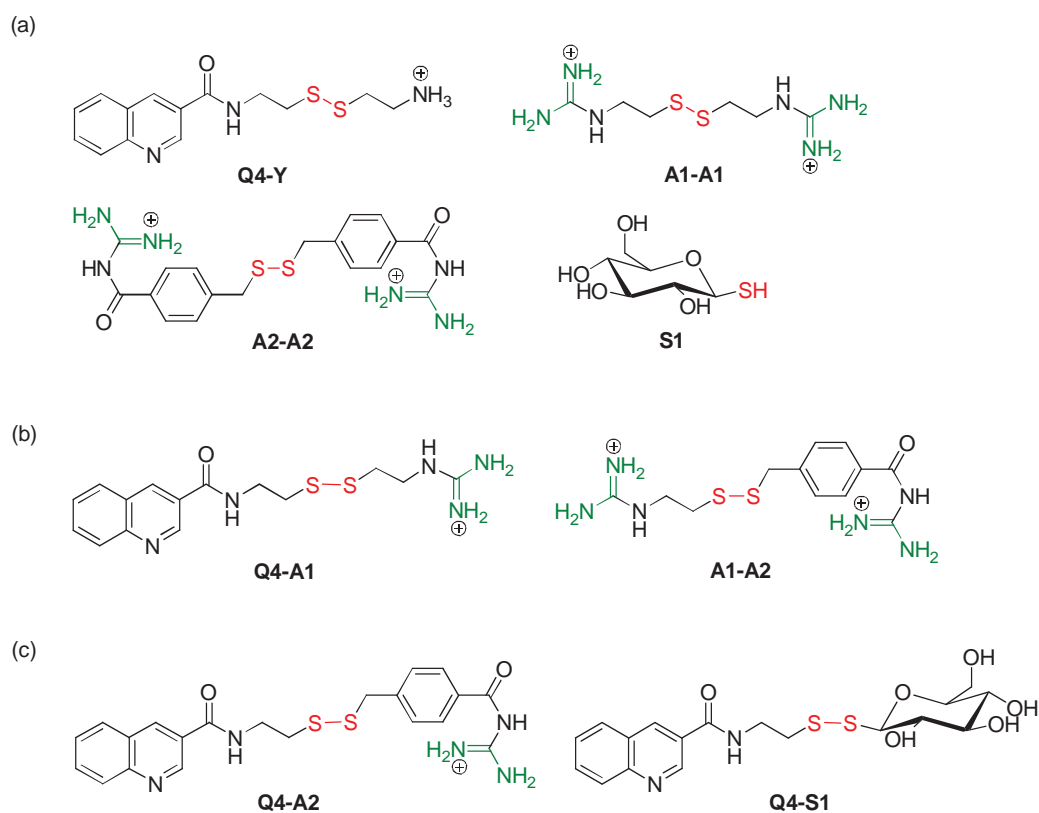
Figure 4.5 (a) thiol **S1** and disulfides **Q4**-**Y**, **A1**-**A1** and **A2**-**A2** used in DCC experiment to generate 3-substituted charged quinoline derivatives, (b) structures of disulfides **Q4**-**A1** and **A1**-**A2** selected by the oligonucleotides **O1**, **O2** and **O3** and (c) structures of disulfides **Q4**-**A2** and **Q4**-**S1** not selected by any of the oligo sequences **O1**/**O2**/**O3**.

In order to establish the structural requirements for the amidine in the quinoline side chain, **Q4**-**Y** was reacted with the three amidines **A1**-**A1**, **A2**-**A2** and **A3**-**A3**, which differ in the degree of flexibility in the side chains and the
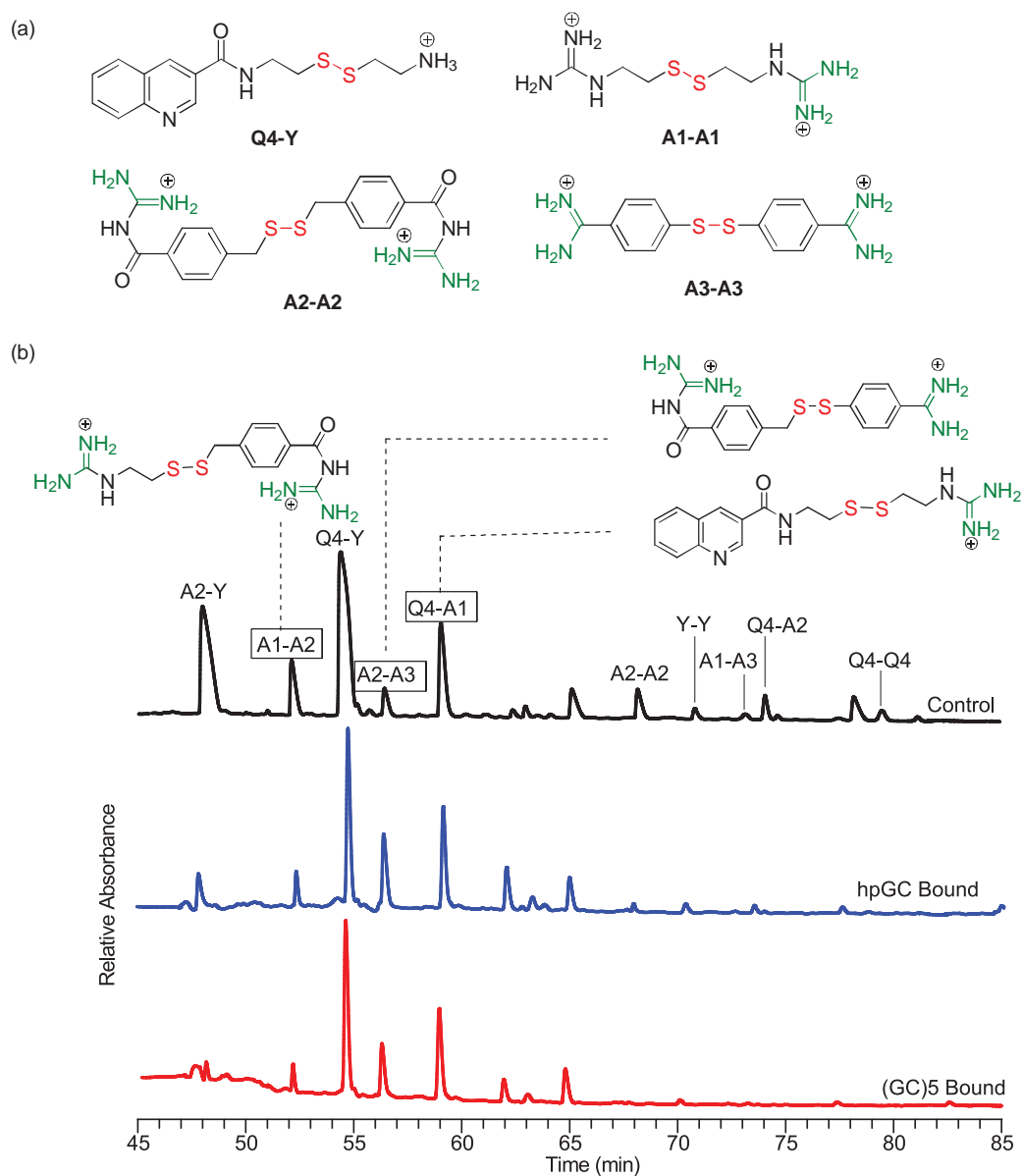
Figure 4.6 DCL generated from (a) **Q4**-**Y**, **A1**-**A1**, **A2**-**A2** and **A3**-**A3** using the scrambling method (YSH, pH ~ 8.5), (b) LC trace with UV detection of (210-350 nm) for the control DCL and DNA-bound spectra of hpGC **O1** and D(GC)$_5$ **O2** oligo sequences and highlighting the structures of the DNA-bound compounds.

presence of the aryl group (Figure 4.6a). These experiments required the use of scrambling initiator, cysteamine (Y-SH), as all the DCL building blocks were disulfides (see Figure 2.1, Chapter 2). Figure 4.6 shows the results of the DCC experiment. Of the 3 amidine building blocks (Figure 4.6a), **Q4-A1** the disulfide formed with the alkyl amidine **A1-A1** was selected by DNA. While the benzylic amidine **Q4-A2** was present in the control library, it was not selected by any of the DNA sequences **O1**, **O2** and **O3**. The starting material **Q4-Y** was also selected by all the three DNA sequences, but it is more likely that this result is directly related to the high concentration of the scrambler (9.5 equiv.) required to generate the initial library. The DCL studies using the scrambler with < 5.0 equivalent the disulfide exchange was minimal and/or sometimes the reaction was not activated by the scrambler. Hence, in order to activate the DCL, higher concentration i.e. 9.5 equivalent of the scrambler were added. The bisguanidine **A1-A2** as well as **A2-A3** were also selected by all three sequences.

The experiment was repeated using GSSG/GSH at physiological pH, which increased the diversity of the DCL and the results are summarised in Figure 4.7. In the presence of DNA, both **Q4-A1** and **A1-A2** were selected as observed in the previous experiment (see Figure 4.6). However, **A2-A3** was not selected in this experiment. A significant amount of the disulfide starting material **A2-A2** was also selected by all the oligonucleotides. This result is consistent with a lower concentration of **A2-A2** forming disulfides in the equilibrating mixture, due to the presence of the additional disulfide GSSG, which was present at 1.5 times higher concentration than the amidines **A1-A1**, **A2-A2** and **A3-A3** in the control DCL. The major compound that bound to the $(GC)_5$ sequence **O2** was
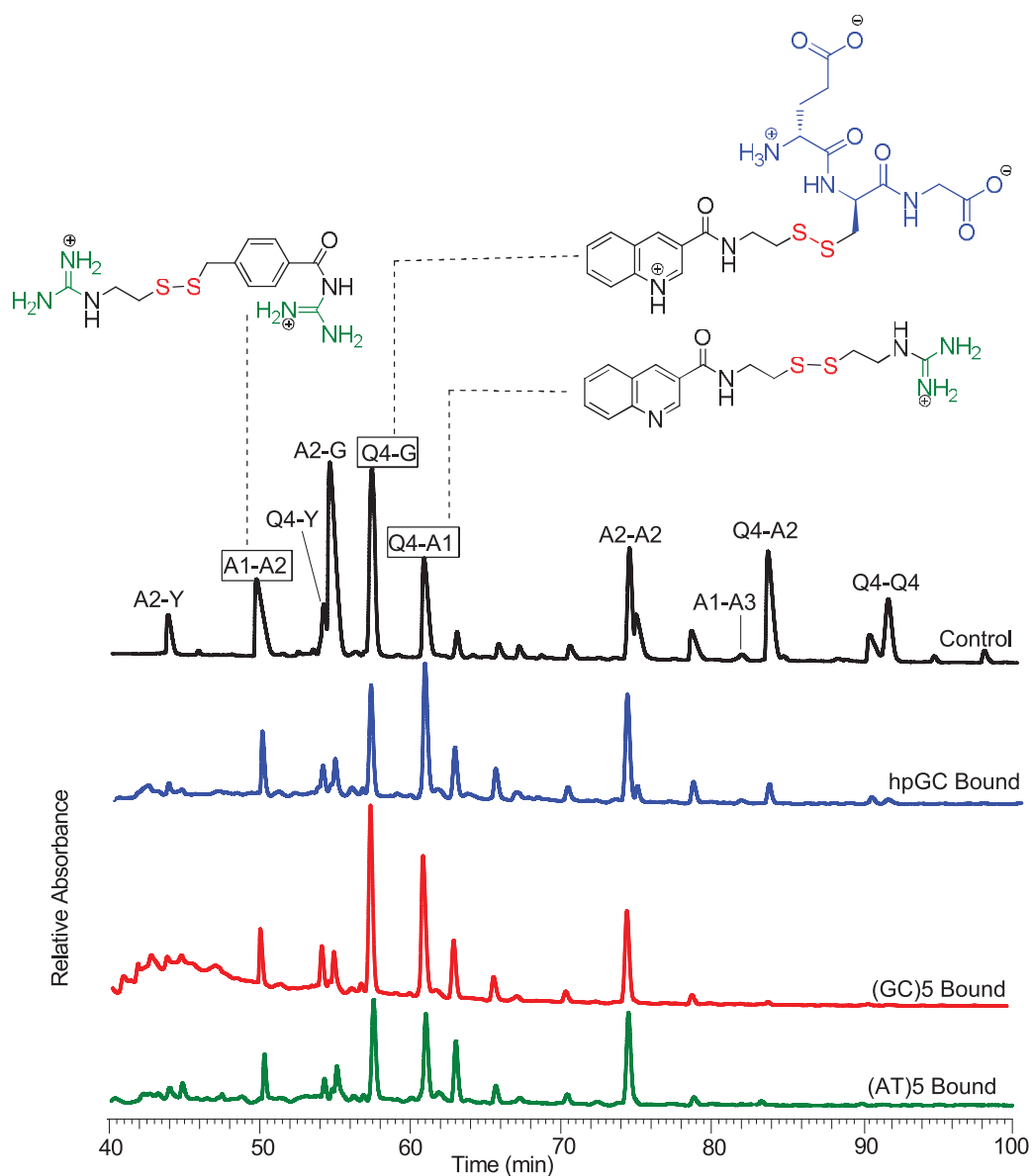
Figure 4.7 LC trace with UV detection of (210-350 nm) of a DCL generated from **Q4**-**Y**, **A1**-**A1**, **A2**-**A2** and **A3**-**A3** using **GSSG/GSH**, pH ~ 7.4 method for the control and the DNA-bound spectra of all the **O1**, **O2** and **O3** oligo nucleotides, highlighting the DNA-bound compounds including the glutathione adduct **Q4-G** (blue).

the neutral, (but doubly zwitterionic) glutathione adduct **Q4**-**G** (Figure 4.7), which was also selected by the other oligonucleotides **O1** and **O3**. Overall, the

experiment confirmed that both flexible amino quinoline **Q4-Y** and guanidine disulfides of **A1-A1** and **A2-A2** interact with DNA. However, neither of the aromatic guanidine disulfides with quinoline i.e. **Q4-A2** and **Q4-A3** interacted with DNA, suggesting that the aryl groups may interfere with positioning of the amidine near the phosphate backbone. Structurally related aryl guanidine functional groups are present in DNA minor groove binders with bound water molecules often assisting in DNA recognition and binding.[12]

As quinoline is a weak DNA binder, and does not act as an intercalator, the DCC experiment was repeated with the charged naphthalamide **N1-Y**. Intercalation is favoured more strongly by the larger aromatic planar group. Hence, it was expected that there may be different profiles of DNA-binding, if the quinoline and naphthalamide interacted with DNA in a different way.

Under the same conditions described above, surprisingly, none of the naphthalamide disulfides were detected in the DNA-bound fractions. However, the poor solubility of the naphthalamide **N1-Y** (even with the positively charged amino side chain) resulted in some cloudiness in the assay studies. Warming the solution, or using more dilute samples did not improve the reaction and it was not possible to make conclusions from these experiments.

4.6.2 Effect of Substitution on the Quinoline Ring

The effect of substitution of the quinoline ring on DNA-binding was assessed by a DCC experiment with 2-, 3-, and 4-substituted quinolines and carbohydrate building blocks (Figure 4.8a). **Q1** and **Q2** differ only in the

position of the thiol group on the ring. In addition, **Q2** contains an electron-withdrawing trifluoromethyl group, which has been reported to enhance DNA-intercalation.[217-218] **Q4-Y** as a disulfide of an alkyl thiol was also included for comparison, and as a charged compound that was expected to interact with DNA. Given that 2-deoxy sugars are often present in DNA-minor groove binders,[153-154] the quinolines were reacted with two different deoxy sugars **S2** and **S3-S3**. Thioglucose **S1** was included for comparison to see whether there was a preference for the deoxy sugars **S2** or **S3-S3** over the fully oxygenated glucose **S1**.

Figure 4.8 shows the results of the DCL experiment generated from the quinoline thiols **Q1**, **Q2** and **Q4-Y** and three thiosugars **S1**, **S2** and **S3-S3** (Figure 4.8a). In the control DCL, 22 disulfides were identified. The thio-sugars **S1**, **S2** and **S3-S3** exist as a mixture of α- and β- anomers. The two peaks of identical masses suggested that both anomers of **Q1**-α-**S1** and **Q1**-β-**S1** were in the control library, only a single peak was detected for the disulfides of the deoxy sugars **S2** and **S3-S3**. In the DNA-binding studies, it was not possible to determine whether the anomers co-eluted, or whether only one anomer was selected. Hence, the glycosides are represented as α- and β- mixtures throughout the chapter. Relative to the control, the disulfides **Q1-S2** and **Q2-S1** were amplified in the presence of all the **O1**, **O2** and **O3** oligonucelotides and the results shown in Figure 4.8c. In addition, the starting material **Q4-Y**, **Q4-S1** and the disaccharide **S3-S3** were selected by DNA (Figure 4.8b). Of particular interest was the fact that only **Q1-S2**, the disulfide of the 2-substituted quinoline **Q1** with the rigid 2-deoxysugar **S2**, was selected by
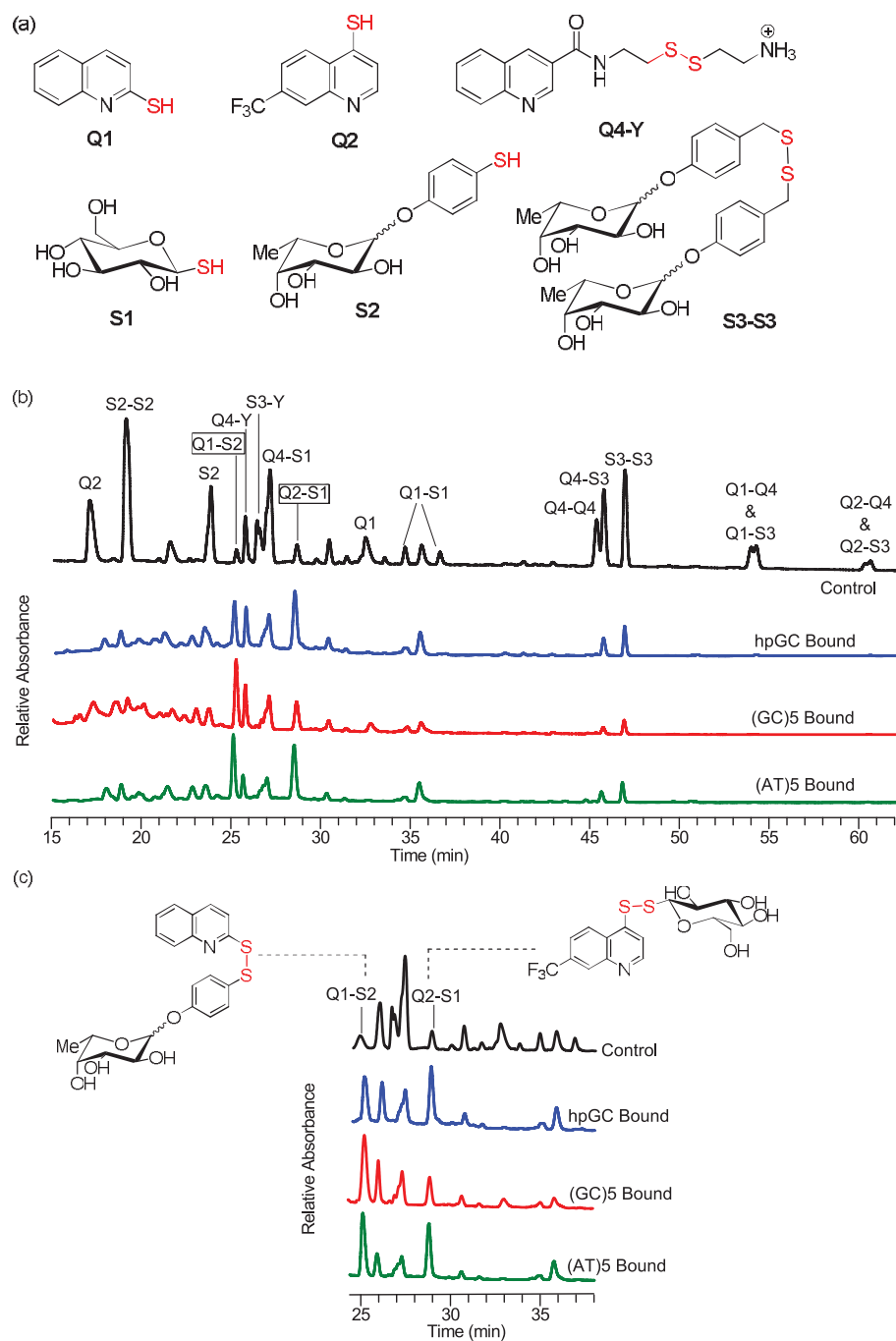
Figure 4.8 DCL generated from (a) quinolines **Q1**, **Q2** and **Q4**-**Y** and thiosugars **S1**, **S2** and **S3**-**S3**, (b) LC trace with UV detection of (234 and 286 nm) for the control DCL and DNA-bound spectra of **O1**, **O2** and **O3** oligo sequence and (c) expansion of DNA-bound region highlighting the structures of the DNA amplified compounds.

DNA. While **Q1**-**S1** and **Q1**-**S3** were both present in the control library, they were not present in either the DNA-bound or unbound solutions. This result is consistent with the building blocks equilibrating to form more stable products in the presence of DNA.

In contrast to the results with the 2-substituted quinoline, only **Q2**-**S1**, the disulfide of the 4-substituted quinoline **Q2** with thioglucose **S1** was observed to bind to DNA. Not surprisingly, there was no evidence for any of the quinoline dimers, such as **Q1**-**Q1**, **Q2**-**Q2**, **Q4**-**Q4**, **Q1**-**Q2**, **Q1**-**Q4** or **Q2**-**Q4** being selected by DNA, but the disaccharide disulfide **S3**-**S3** was selected. Comparing the three **O1**, **O2** and **O3** oligonucleotides, the same compounds were selected, but there appear to be minor sequence preferences. Figure 4.9 shows percentage composition of the library (Figure 4.9a) and percentage change composition between the sample and control library (Figure 4.9b). The analysis clearly showed that **Q2**-**S1** was detected at higher levels in the presence of hpGC **O1** and d(AT)$_5$ **O3** compared with d(GC)$_5$ **O2**. **Q1**-**S2** was the preferred compound selected by both d(GC)$_5$ **O2** and d(AT)$_5$ **O3** with a lower amount present with hpGC **O1** (Figure 4.9b). These experiments show clearly that the substitution of the quinoline ring is important in DNA-binding, with different carbohydrate disulfides formed from **Q1**, **Q2** and **Q4**-**Y** bound to DNA. In the case of 2-substituted quinoline **Q1**, only the rigid deoxy sugar **S2**, **Q1**-**S2**, was selected and none of the benzylic deoxy sugar derivative, **Q1**-**S3**, or the glucose disulfide, **Q1**-**S1**, bound to DNA. The slightly different sequence preferences shown by **Q1**-**S2** are consistent with DNA minor groove recognition involving hydrogen bonding and/or shape recognition. DNA minor groove binders are generally characterised by small aromatic rings directly linked to

(a)

(b)

Figure 4.9 Example of the DCC experiment analysis (a) percentage composition of library and (b) percentage change between the sample and the control library.

one another or with partial rigidity *via* amide functional groups.[74,208,219] The greater flexibility of **Q1-S3** compared with **Q1-S2**, would result in a higher entropic penalty to bind to the DNA minor groove, and may explain why only **Q1-S2** was selected by DNA.

In contrast to the results with the 2-substituted quinoline **Q1**, the only carbohydrate disulfide of the 4-substituted quinoline selected was the glucose derivative, **Q2-S1**, which was selected by all three **O1**, **O2** and **O3** oligonucleotides. None of the deoxysugar disulfides **Q2-S2** or **Q2-S3** were bound to DNA. Glucose is not typically present in DNA-binders, as the glucose sugar is hydrated and prefers to be located in bulk water and it is unlikely that the glucose sugar in **Q2-S1** would be involved in hydrogen bonding with the DNA base-pairs. A more likely explanation for the selection of **Q2-S1** may be the increased water solubility of the glucose derivative.

The identification of **Q1-S2** and **Q2-S1** as new DNA-binding compounds is broadly consistent with a number of other studies on 2- and 4-substituted quinolines that have also highlighted the importance of ring substitution, steric effects and stereoelectronic effects on DNA-binding. A study of 25 phenyl substituted quinoline-8-carboxamides **58**, in which the steric, electronic and lipophilic properties of the phenyl ring were varied (Figure 4.10a), showed that strength of DNA intercalation was strongly affected by the nature of substituents present in the phenyl ring at different positions.[220] The importance of the CF_3 group was emphasized by DNA-binding studies on trifluoromethyl-quinoline derivatives (Figure 4.10b).[217,221-223] The DNA-binding affinity was increased when the molecule was substituted with two CF3

groups **60**, whereas the compound with only one CF3 group **59** showed moderate or no DNA-binding activity.
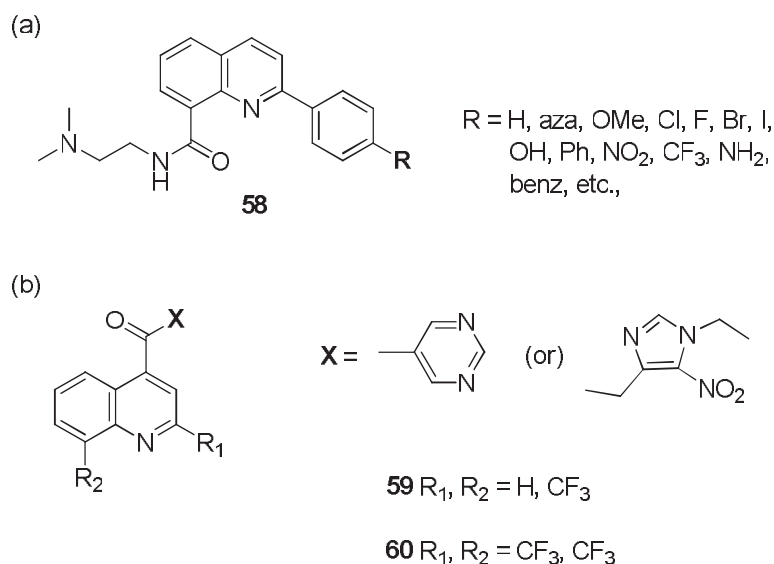
Figure 4.10 (a) Phenyl substituted quinoline-8-carboxamide **58** in which the steric, electronic and lipophilic properties of the phenyl ring were varied[220] and (b) trifluoromethylquinoline derivatives **59**, **60** and the importance of CF$_3$ groups in DNA-binding studies.[217]

Quinoline **Q2** is structurally related to the anti-malarial agent chloroquine. A large number of derivatives of chloroquine have been studied and the DNA-binding properties of some of the derivatives have been reported (Figure 4.11).[218,224-225] These studies showed that the electron withdrawing groups at the 7-position increases the DNA intercalation with X = H<Cl<<CF$_3$ (Figure 4.11a). The cationic amino substituted side chain at the 4-position (Figure 4.11b) has been proposed to be involved in electrostatic interactions with phosphodiester backbone and reduce the overall charge of the DNA.[218]

The structural similarity between **Q2**-**S1** and chloroquine strongly suggests that both molecules would interact with DNA by intercalation, which is

enhanced due to the presence of the CF_3 group at the 7-position of the quinoline ring. Intercalation of **Q2-S1** in a similar manner to chloroquine would position the sugar **S1** into the DNA minor groove and offers an explanation as to why **Q2-S2** and **Q2-S3** are not selected by DNA; the aryl rings in these disulfides introduce significant steric bulk which would likely impede intercalation.
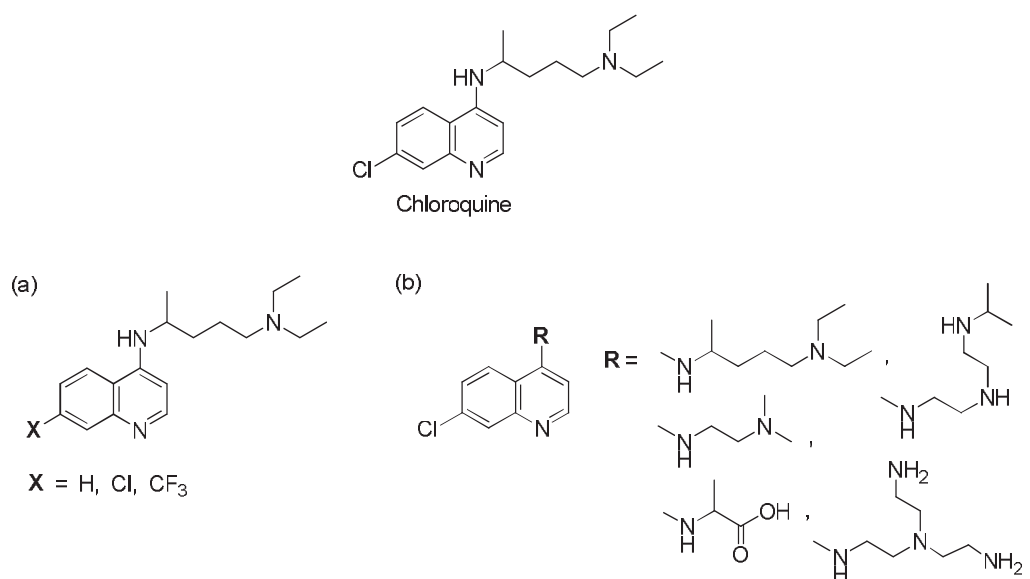
Figure 4.11 Structure of chloroquine and the derivatives (a) by varying electron-withdrawing groups and (b) amino substituted derivatives.[218]

With the exception of the starting material **Q4-Y** and **S3-Y**, all the disulfides formed in the DCL were neutral compounds. The DCL experiment was repeated with the same building blocks and addition of guanidine **A1-A1** (*APPENDIX*, Figure 2), in order to assess whether the charged amidine derivatives such as **Q1-A1** and **Q2-A1** interacted with DNA in preference to the sugar derivatives **Q1-S2** and **Q2-S1**. This experiment confirmed the importance of the sugars to DNA-binding as the same result was obtained; **Q1-**

S2 and **Q2-S1** were selected by DNA and there was no evidence for the corresponding amidine derivatives of **Q1-A1** and **Q2-A1** being selected by DNA.

4.6.3 Quinoline-Carbohydrate Derivatives

The amplification of the glucose conjugate **Q2-S1** in preference to the corresponding deoxysugar conjugates **Q2-S2** and **Q2-S3** was not expected (Figure 4.8b). The experiment was repeated (Figure 4.12a) in the absence of thio glucose **S1** to determine whether the deoxysugar conjugates of **Q2**, **Q2-S2** and **Q2-S3**, showed any affinity for DNA, as well as allowing a direct comparison between the disulfides of the substituted quinolines **Q1** and **Q2**.

Figure 4.12b shows the DCC results from equilibration of quinolines **Q1**, **Q2** and the aromatic deoxy thiosugars **S2** and **S3-S3**. The predominant disulfide selected in this experiment was **Q1-S2** (Figure 4.12b) with a lower amount of the disaccharide **S3-S3** also selected. A similar profile was observed with d(AT)$_5$ **O3**, d(GC)$_5$ **O2** and the hairpin **O1**. In the absence of thioglucose **S1**, there were no significant amounts of any disulfides of **Q2-S2** and **Q2-S3** in the DNA-bound solution, confirming that the small glucose sugar **S1** is essential for **Q2-S1** DNA-binding. As observed in the previous experiment, the disulfides of **Q1-S3** and **Q2-S3** were not selected by either of the DNA sequences.

An unidentified compound (labelled as *, Figure 4.12b) was amplified by all the DNA sequences. This compound was determined to originate from an impurity
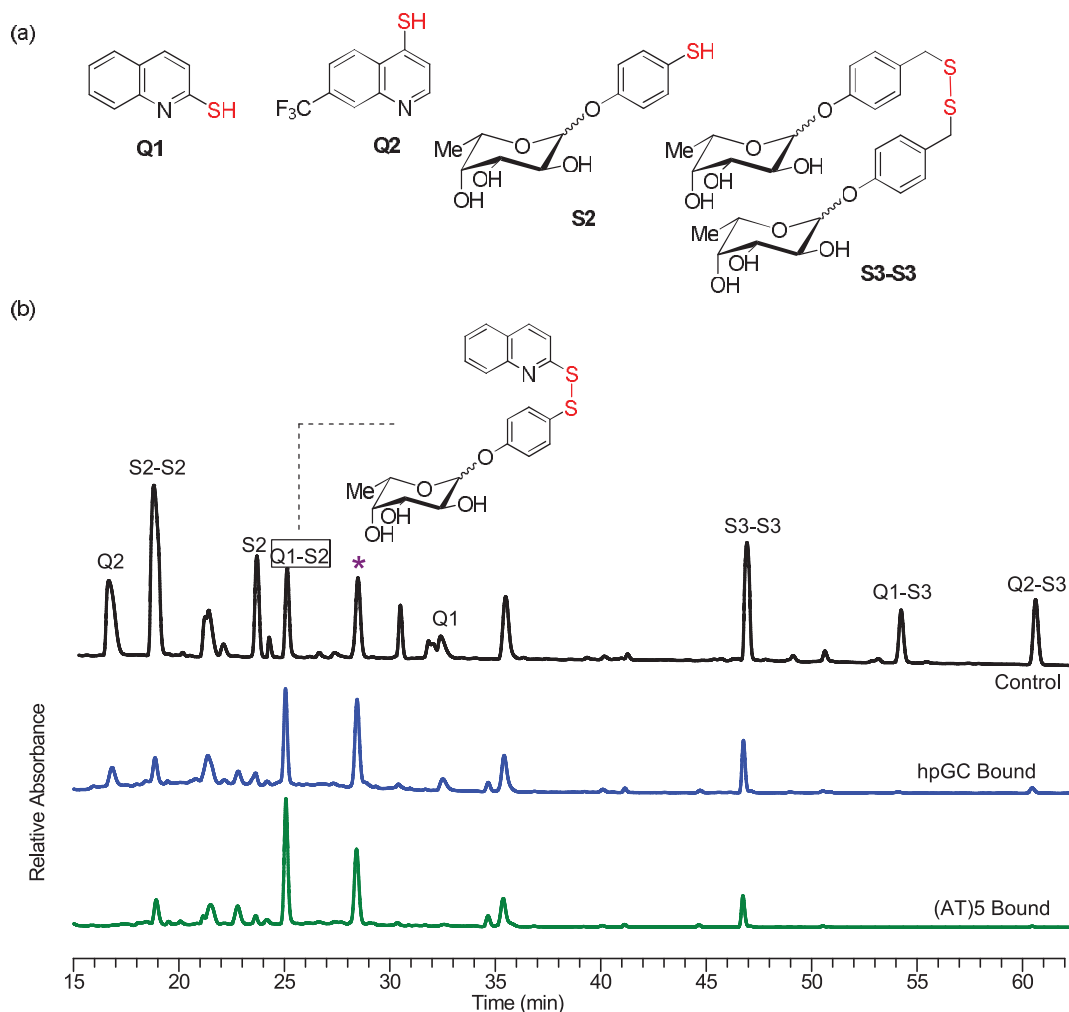
Figure 4.12 Result of the DCC experiment generated from (a) quinolines **Q1** and **Q2**, and thiosugars **S2** and **S3**-**S3** and (b) LC trace with UV detection of (234 and 286 nm) of the control DCL and DNA-bound spectra of hpGC **O1** and $(AT)_5$ **O3** oligo sequences, highlighting the structure of the DNA selected **Q1**-**S2** and unidentified (*) compound.

in the starting material **Q1**. This was confirmed by air oxidation of **Q1** separately at basic pH in 20% methanol/water. After 24 h air oxidation, the LC-MS analysis of **Q1** showed the presence of an additional peak along with **Q1**. A small amount of the unknown compound was isolated by HPLC and was

analysed by NMR, which suggested the presence of a *para*-disubstituted aromatic ring. Unfortunately, a molecular ion for the compound could not be detected and on the basis of the NMR spectrum it could not be identified.

On the basis of the results obtained from the DCL experiments (Figures 4.8 and 4.12) it was proposed that the role of glucose in **Q2-S1** is to improve aqueous solubility. In order to confirm this hypothesis, the DCL experiment was repeated with three water soluble groups **A1-A1** (charged), cysteine (**Cys**, zwitterionic) and **S1** (neutral). In this experiment, **Q2-Cys** and **Q2-A1** were selected by DNA in preference to **Q2-S1** (*APPENDIX*, Figure 3). Overall, the order of preference for DNA-binding was **Q2-Cys**>**Q2-A1**>>**Q2-S1**. The studies support the conclusion that charged water soluble groups in **Q2-S1** are preferred at the 4-position. This result is also in agreement with the structure-activity results on chloroquine (Figure 4.11) that identified charged groups at the 4-position enhancing DNA-binding affinity *via* electrostatic interactions.

The DNA-binding properties of **Q1-S2** were further examined, in order to evaluate the role of the *para*-substituted phenyl ring in DNA-binding as well as the importance of the deoxy sugar in DNA recognition. A DCL experiment was carried out with **Q1**, **S2** and **A3-A3**, which can give rise to the two structurally related disulfides **Q1-S2** and **Q1-A3** (Figure 4.13). The results showed that only **Q1-S2** was selected by DNA and none of the amidine disulfide **Q1-A3** bound to DNA (*APPENDIX*, Figure 4). Hence, the deoxy sugar in **Q1-S2** appears to enhance the DNA-binding interaction.
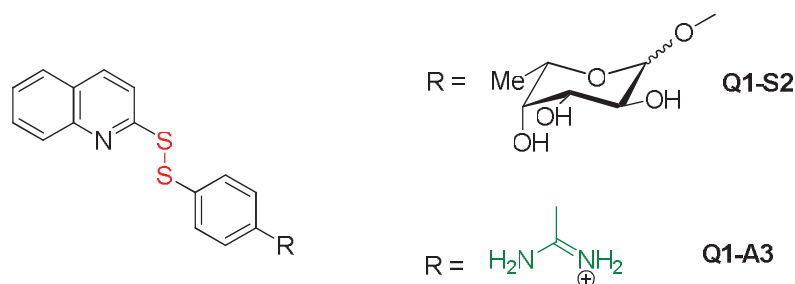
Figure 4.13 Structurally related disulfides of **Q1-S2** and **Q1-A3**.

4.6.4 Quinoline and Imidazole Derivatives

As discussed in Chapter 1, small molecules with charged heterocycles bind to DNA with high affinity and selectivity.[13,29,31] DCC experiments were conducted with quinolines **Q1** and **Q2** and the small aromatic charged and highly water soluble imidazoles **M1** and **M2** and their DNA-binding studies were compared with the charged amidines.

Figure 4.14 shows the results of a DCL experiment generated from quinoline **Q1**, amidines **A1-A1**, **A2-A2, A3-A3** and imidazoles **M1** and **M2** (Figure 4.14a). Relative to the control, **A1-A2** was the predominant disulfide selected in this experiment (Figure 4.14b) as observed in the previous DCL experiment (see Figure 4.5b). The imidazoles disulfides **A1-M1** and **M1-M2** were bound with the hpGC **O1** oligo in almost the same amount. Along with the benzylic amidine starting material **A2-A2**, a small amount of the flexible imidazole **A1-M2** was also selected by the **O1** oligo sequence.

Similar to the previous experiment (Figure 4.12b), the impurity present in the **Q1** starting material (labelled as *, Figure 4.14b) was also bound with the hpGC **O1** sequences. As expected, none of the disulfides of the rigid amidine

A3, A1-A3 and **A2-A3** were selected by the DNA. There was no evidence for quinoline-imidazoles **Q1-M1** and **Q1-M2** interacted with DNA and the selected compounds are all doubly charged and selected with DNA by electrostatic interactions.
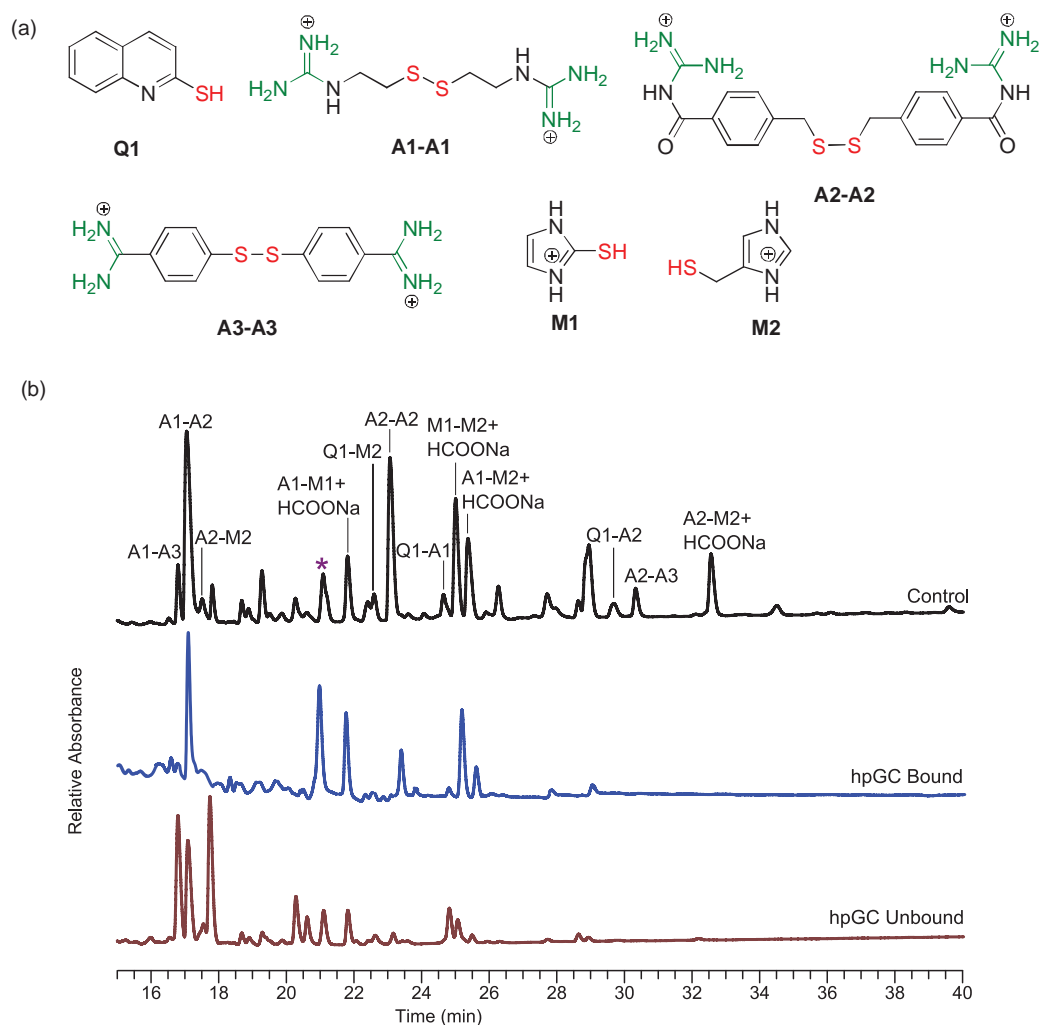
Figure 4.14 Result of the DCC experiment generated from (a) quinoline **Q1**, amidines **A1-A1**, **A2-A2** and **A3-A3**, imidazoles **M1** and **M2** and (b) LC trace with UV detection of (210-350 nm) of the control DCL, DNA-bound and unbound spectra of hpGC **O1** oligonucleotide.

In a second experiment, the imidazoles **M1** and **M2** were equilibrated with the quinoline **Q2** and amidine **A1-A1** (Figure 4.15a). The predominant disulfide selected in this experiment was **M1-M2** (Figure 4.15b), which was also selected in the previous DCL experiment (Figure 4.14b). None of the quinoline **Q2** disulfides with the imidazoles **M1**, **M2** and **A1-A1** i.e. **Q2-M1**, **Q2-M2**, **Q2-A1** were selected by the hpGC **O1** oligo sequence.
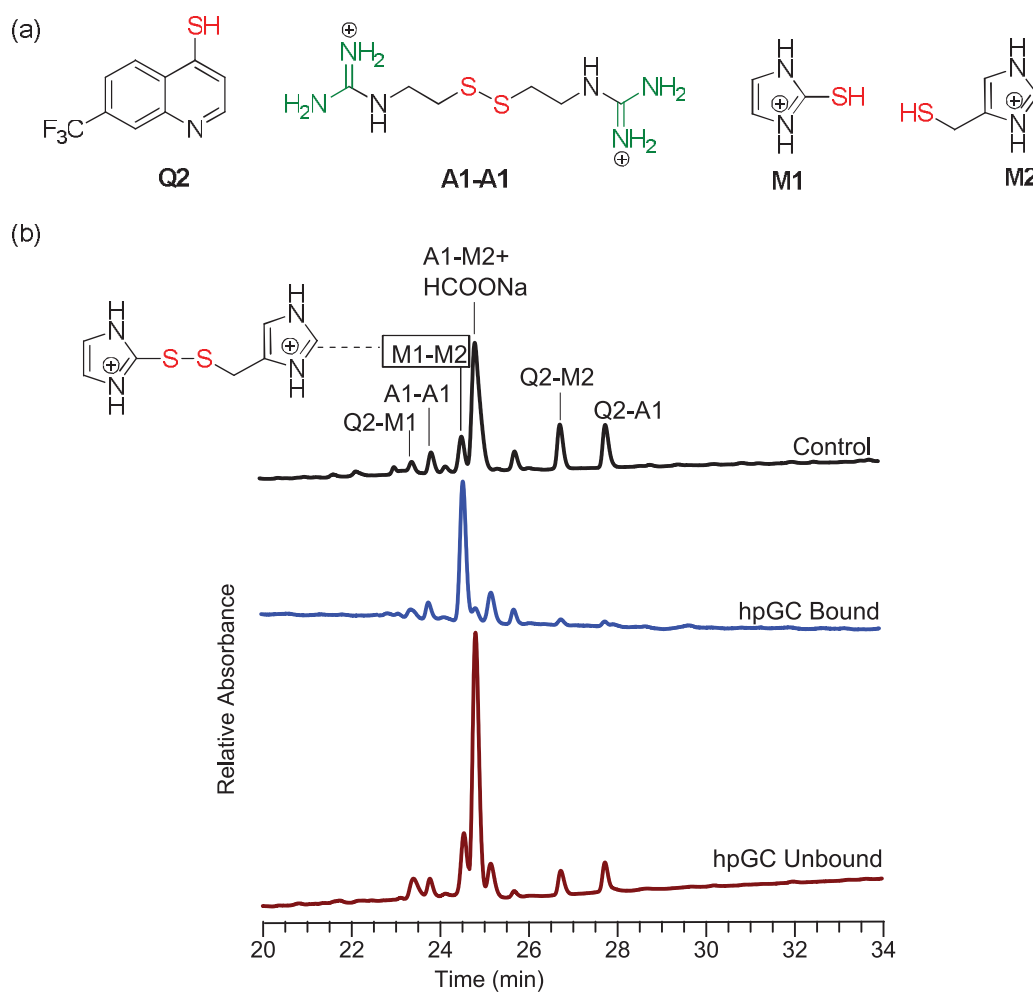
Figure 4.15 Result of the DCC experiment generated from (a) **Q2**, **A1-A1**, **M1** and **M2** and (b) LC trace with UV detection of (210-350 nm) of the control DCL, DNA-bound and unbound spectra of hpGC **O1** oligonucleotide, highlighting the DNA amplified compound **M1-M2**.

135

4.6.5 Quinolines and Bisthiols

The bisthiols **B1**, **B2** and **B3** were designed to allow the formation of potential bisintercalators with **Q1** and **Q2**. The flexible bisthiols **B1** and **B2** varied by the presence of the hydrophilic atoms, oxygen and nitrogen, which may participate in hydrogen bonding with the DNA bases or electrostatic interactions in the case of the protonated amine. The semi-rigid bisthiol **B3** has a piperazine functional group that is protonated at acidic pH, which potentially aids DNA recognition *via* electrostatic interactions. The bisthiols **B1**, **B2** and **B3** were equilibrated with the quinolines **Q1** and **Q2** under a range of different concentrations. Under all the conditions investigated, the predominant species formed in the DCL were the quinoline starting material and/or the quinoline disulfides **Q1**-**Q1**, **Q2**-**Q2** or **Q1**-**Q2** (*APPENDIX,* Figure 5).

The reaction conditions were altered in order to favour oxidative reactions of the aliphatic bisthiols **B1**, **B2** and **B3** with the quinolines **Q1** and **Q2**. However, while low amounts of the hetero-disulfides were present in the DCL, there was strong evidence for polymerisation of the alkyl thiols and the polymerised products **B1**-**B1**-**B1**-**B1** and **B2**-**B2**-**B2**-**B2** were detected in the control DCL. In the presence of DNA, no DNA-binding compounds were detected. While the proposed bisintercalators were not detected in the control DCL experiment, an advantage of DCC is that a very minor (< 5%) compound in the DCL can be amplified in the presence of the target molecule. However, the absence of any DNA-binding molecules suggests that either the reactivity of the building blocks does not generate the desired disulfides that may

interact with DNA, or none of the disulfides in the DCL are DNA bisintercalators.

Figure 4.16 shows selected examples of the quinoline DNA-bisinterclators reported in the literature.[226-227] DNA-binding studies on the rigidity of linkers, strength and specificity of DNA interactions of quinolines with flexible alkyl and piperazine linkers (Figure 4.16a) showed that the alkyl flexible linker failed to promote strong bisintercalation. In contrast, the piperazine linked diquinoline **61** showed highest affinity (K_a 12×10^4 M^{-1}), this bisintercalator confirming that the rigidity of the linker is important.[226] The 4-amino-3-amide diquinolines with different linkers were designed as bisintercalator anticancer drugs[227] are shown in Figure 4.16b. The piperazine containing quinoline was more efficient in the treatment of highly aggressive melanoma cell line A375 than the quinoline containing other linkers. However, the DNA-binding studies were not reported.[227]

The bisthiols **B1**, **B2** and **B3** designed to allow the formation of bisintercalators with quinolines **Q1** and **Q2** were unsuccessful (Figure 4.16c). Bisthiols **B1**, **B2** and **B3** are very flexible, and consistent with reported studies (Figure 4.16a), the entropic penalty of the bisintercalator adopting a conformation suitable for bisintercalation is high. In natural bisintercalators, the distance between the bisintercalated quinoline rings is ~ 10.1 Å. In this study, the distance between the quinoline rings in the potential bis-intercalators formed from the bisthiols **B1**, **B2** and **B3** with quinoline (**Q1** and **Q2**) were estimated to be between approximately 11 Å to 13.5 Å. While the flexibility of the linkers does allow the molecules to adopt a conformation in

which the two quinoline rings can intercalate, the incorporation of the two disulfide functional groups in the linker makes a more precise matching of the distance between the quinoline rings difficult, and the molecules did not overlap with the natural bisintercalators.
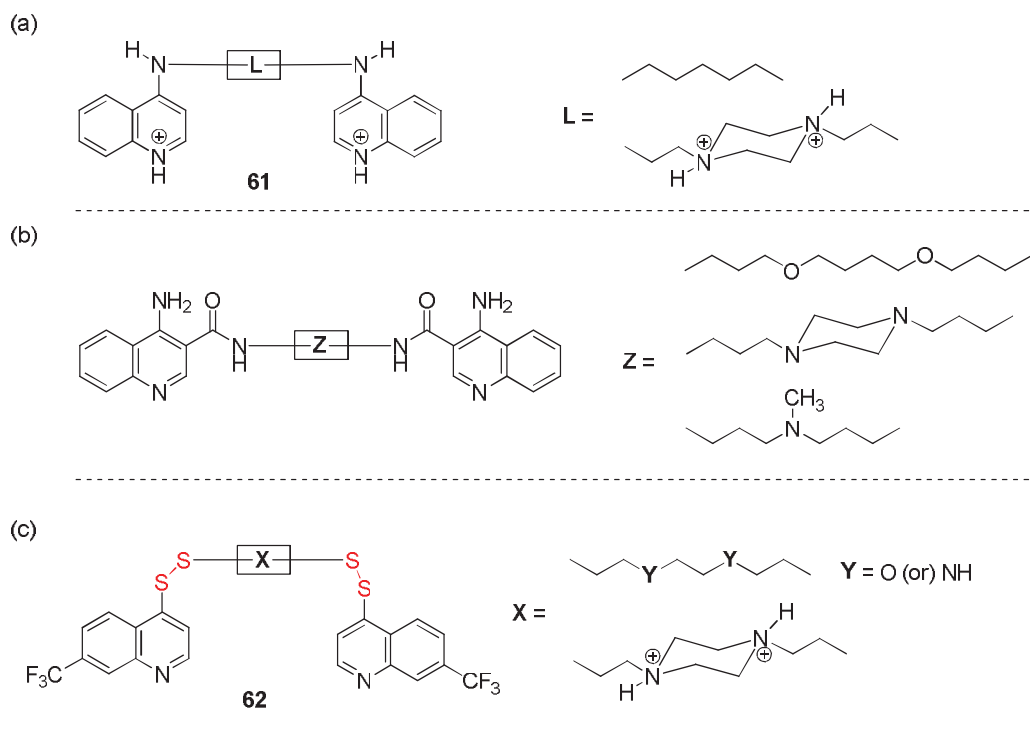
Figure 4.16 (a) Selected literature examples of quinoline DNA-bisintercalators generated by **61**,[226] (b) 4-amino-3-amide diquinoline bisintercalators[227] and (c) comparison of possible bisintercalators formed from quinoline **Q2** derivative **62** and bisthiols **B1/B2/B3**.

4.7 Molecular Visualisation

Molecular visualisation was carried out to understand better the DNA-binding modes of **Q2-S1** and **Q1-S2** and also to assist interpreting the results of the DCL experiments. These studies were carried out in collaboration with James Garner.[228]

The experimental details on the docking protocols are explained under section 4.9.5, however, a brief summary on the methodology to understand the procedure is given below;

- The ligand structures were first sketched in Spartan '04 (Wavefunction Inc) and the geometry optimised using first a MM2 forcefield and then minimised using semi-empirical PM3 forcefield;

- The Protein Data Bank (pdb) was searched for examples of DNA intercalating structures and three representative examples used to build canonical DNA models for DNA intercalating structures;

- The ligands from each pdb files were removed and submitted to 3DDART software and the resulting pdb file was opened in DS Modelling 3.0 and their geometry optimised using CHARMm forcefield;

- The ligand and minimised canonical DNA structure was opened in Autodock Tools and the Gasteiger-Marsili charges were calculated;

- In order to include the DNA fragment, a grid box was created with a defined dimension and resolution, then the grid box was centred on DNA and the grid potential maps were calculated using AutoGrid 4.0.

The molecular visualisation studies were carried out in order to understand the DNA-binding modes of **Q2-S1** and **Q1-S2**. This allows the classification of the resulting binding mode by visual inspection as intercalation, minor groove binding or others such as major groove binding and interaction with phosphate groups. It should be stressed, however, that these docking studies are not accurate enough to produce any realistic quantitative estimation of the DNA-binding affinities of the docked structures.

The interaction of **Q2-S1** and **Q1-S2** with B-DNA sequences d(AT)$_5$ and d(GC)$_8$ were studied in order to rationalise the results obtained in the DCC experiments including why some compounds did not bind to DNA. Overall two binding modes were considered: intercalation using the d(GC)$_8$ rich sequence and minor groove binding using the d(AT)$_5$ rich sequence. It should be noted that these studies were carried out in the absence of water and were used only to provide a useful 3-dimensional picture of possible binding modes. As shown in Figure 4.17, the overall shape, conformational flexibility, steric effects, hydrogen-bonding, π-π interactions, hydrophobic and electrostatic interactions were considered in docking the quinolines with the oligonucleotides.

Q2-S1 and **Q1-S2** were docked into canonical crystal structures of B-DNA d(AT)$_{3-5}$ and d(GC)$_{3-5}$, sourced from the crystal structures with and without intercalation gaps. The DNA structures, where there were 2- and 4-base pair gaps between the intercalation gaps were used to consider the possible intercalation modes of the compounds using both G:C and A:T rich DNA sequences. They were chosen because depending on the linker length and
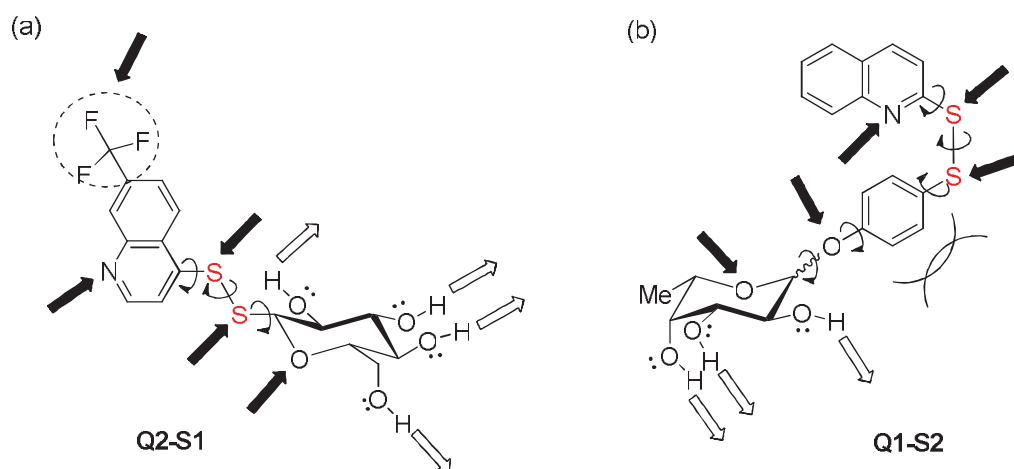
Figure 4.17 shows the hydrogen bond donor (plane arrows) and acceptor (bold arrows) and highlighting the ligand flexibility and steric effects of (a) ligand **Q2**-**S1** and (b) ligand **Q1**-**S2**.

flexibility, the compounds could be mono- or bis-intercalated and/or act as an intercalator-groove binder. However, the majority of the docking compounds with the 2-base pair gap structures, resulted in part of the compound sitting across the second intercalation gap, hence these dockings were discarded.

Visualisation was performed to see if DNA sequence selectivity and probable binding mode would occur under simple models, and whether the non-selectivity of library compounds could be rationalized. For each of **Q2**-**S1** and **Q1**-**S2**, 25 docked conformations were obtained with A:T and G:C DNA sequences, based on the energy calculation, the groove binding mode is favoured for **Q1**-**S2**, whereas the intercalation mode is favoured for **Q2**-**S1**. As none of the library compounds contain a linker, and hence they are unable to bind *via* bis-intercalation.

(a) Docked Q2-S1

While **Q2**-**S1** shows no sequence selectivity that would indicate intercalation *versus* groove-binding ($\Delta \sim 4.18$ kcal mol^{-1} marginal energy difference between lowest energy docked conformations) between A:T or G:C sequences, the results of the DCL experiments suggested that **Q2**-**S1** most likely intercalated with DNA and the marginal difference in energy value (Δ -9.79 kcal mol^{-1}) obtained from the qualitative data also consistent with the experimental results. When **Q2** intercalated with the DNA, the glucose group **S1** located in
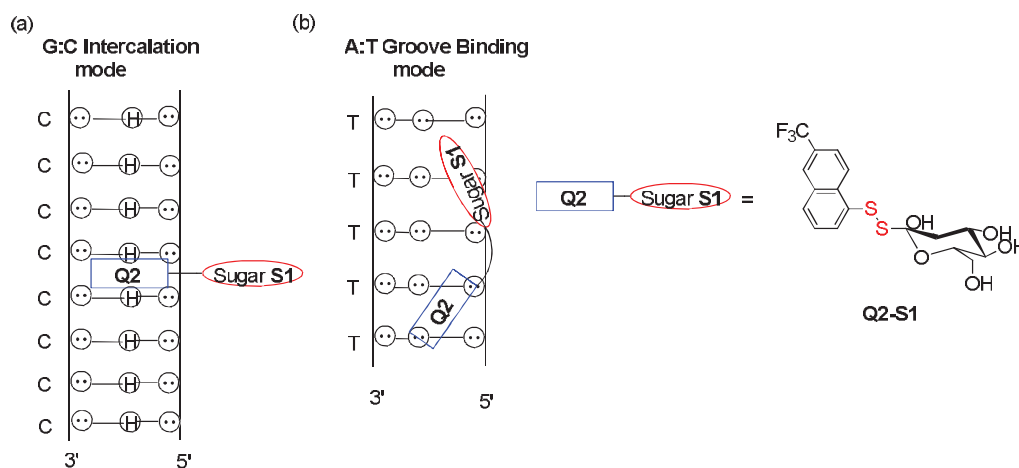
Figure 4.18 shows the DNA-binding modes of **Q2**-**S1** (a) intercalation with G:C sequence and (b) groove binding with A:T sequence.

the groove and not involved in any molecular recognition as shown in Figure 4.18a. There is only one report of DNA-binding studies on glucose derivatives of quinoline.[229] The bis-glucose derivate **63** was designed as a potential DNA-bisintercalator and the mono-glucose derivate **64** was used as a control (Figure 4.19). While the bis-glucose derivative **63** showed some evidence of interaction with DNA, the literature studies showed that the overall weak

binding was attributed to steric hindrance and the linker design. The monoglucose derivative **64** showed the highest affinity with calf thymus-DNA, but no details of the mode of binding were proposed.
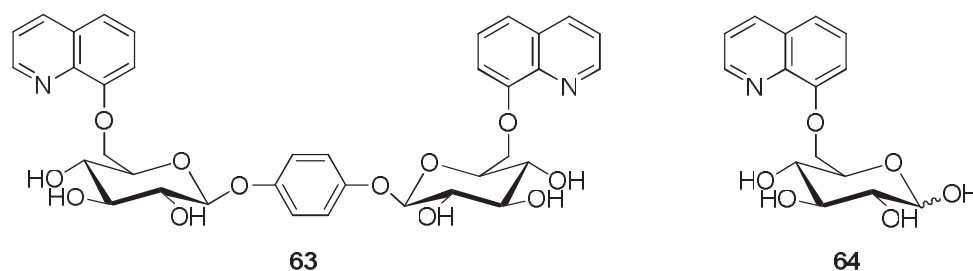
Figure 4.19 DNA intercalator containing quinoline-glucose derivative reported in the literature.[229]

Figure 4.20a shows an image of **Q2-S1** intercalated into a $(GC)_8$ sequence binding. When **Q2** is intercalated, the sugar is located into the groove (see Figure 4.18a), and in this orientation it would be well solvated by water and clearly improve aqueous solubility of **Q2**. The DCL experiments with **Q2** and other water solubilising groups, **A1** and cysteine, supported the conclusions above and the relative order of DNA-binding was **Q2-Cys**>**Q2-A1**>>**Q2-S1** (*APPENDIX*, Figure 3).

Table 4.3 summarises the molecular modelling analysis of the docked structure of **Q2-S1** which showed the presence of hydrogen bonds between one of the F-atom present in the aromatic ring with the G:C sequence and also the hydroxyl group and the sulphur atom present in the sugar moiety (Figure 4.20b). The β-glycoside is shown, and no evidence for preferential interaction of the α- *versus* the β-glycoside could be determined.
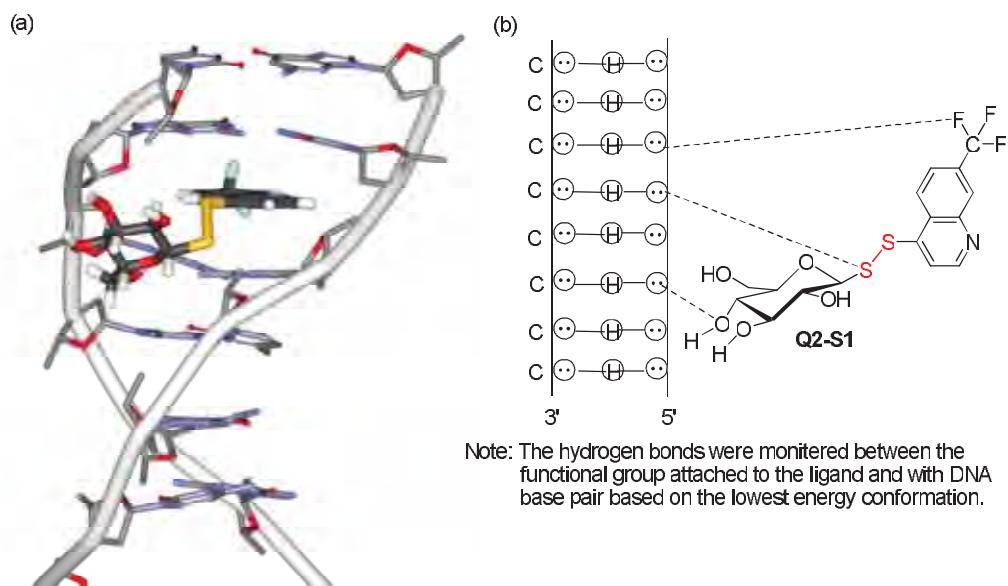
Figure 4.20 (a) Interaction of **Q2-S1** into a (GC)$_8$ sequence as a intercalator 3D image generated by James Garner.[228] DNA (hydrogens and phosphate backbone atoms omitted for clarity), Ligand (dark grey carbon atoms) and (b) formation of hydrogen bonds (indicated by dotted lines) between the ligand **Q2-S1** and G:C sequence.

Figure 4.18b shows the other possible DNA-binding mode of **Q2-S1** in which the sugar could act as a minor groove binder with hydrogen bonds formed between the sugar hydroxyls and the DNA base pairs. While the docked structure shows that formation of hydrogen bonds are possible, there are no examples of glucose as a minor groove binder, and hence the groove binding mode of **Q2-S1** is unlikely to occur.

Intercalation of the quinoline ring also explains why **Q2** does not select the more hydrophobic 2-deoxysugars **S2** and **S3**, which are well-characterised DNA minor-groove binders.[153-154] **Q2-S2** and **Q2-S3**, both contain a bulky aryl group as well as the deoxy sugar. Intercalation of the quinoline ring would

144

result in the aryl sugars located in the groove, and from entropic considerations, as well as possibly some steric interactions, this mode of binding would not be favoured. Intercalation of the quinoline ring **Q2** with **S2** and **S3** side chains as minor groove binders was also considered. However, it was not possible to orient the bulky aryl groups present in **Q2**-**S2** and **Q2**-**S3** in the minor groove without steric collision with the DNA base pairs.

(b) Docked Q1-S2

The DCL experiments suggested that the predominant mode of interaction of **Q1**-**S2** with DNA was as a minor groove binder. Figure 4.21a shows **Q1**-**S2** bound to an $(AT)_5$ sequence in the minor groove. While the α- and β-glycosides have different overall shapes, the α-diastereomer of the sugar of **Q1**-**S2** is predicted to bind better as an A:T groove binder than the β-isomer. However, the visualisation did not provide any rationale for preferred binding mode of only one anomer, and hence the α-glycoside is drawn for simplicity. The molecule fits the curvature of the minor groove and this conformation could potentially be stabilised by an intra-molecular H-bond between the phenyl oxygen and the OH-group of the sugar moiety (Table 4.3) as well also an additional inter molecular H-bond with the A:T base pairs (Figure 4.21b). However, these docking studies are not accurate enough to confirm the sequence selectivity (A:T or G:C) of **Q1**-**S2** molecular level structure.
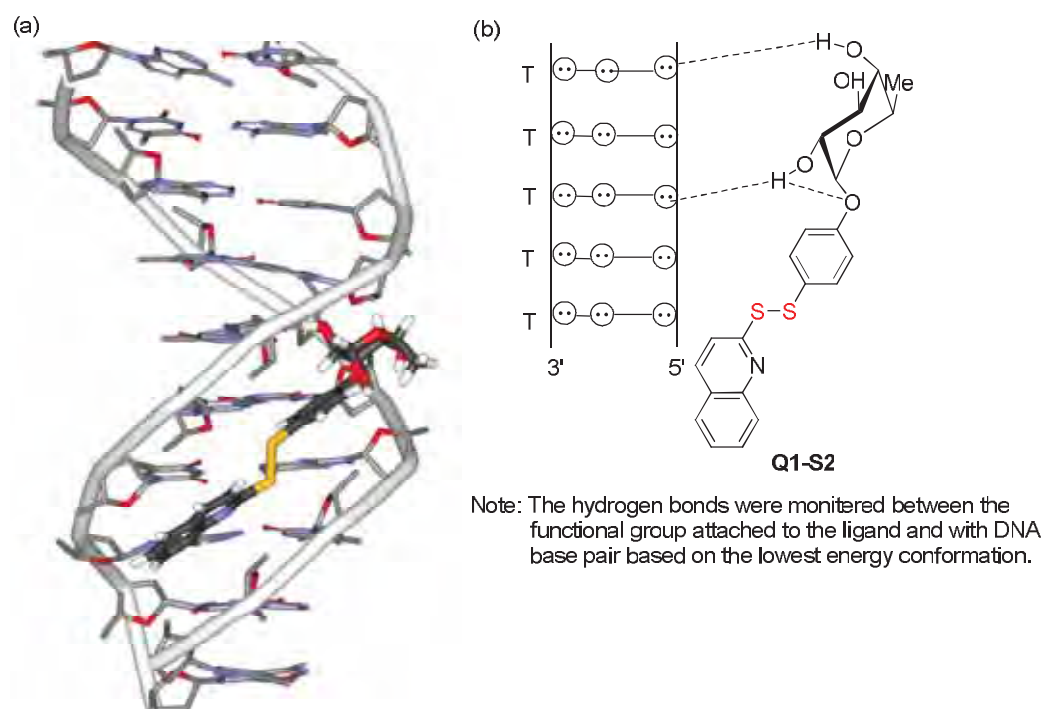
Figure 4.21 (a) Interaction of α-anomer of **Q1-S2** into a (AT)₅ sequence as a minor groove binder 3D image generated by James Garner.[228] DNA (hydrogens and phosphate backbone atoms omitted for clarity), Ligand (dark grey carbon atoms) and (b) formation of hydrogen bonds (indicated by dotted lines) between the ligand **Q1-S2** and A:T sequence.

The interaction of **Q1-S2** with DNA by intercalation of the quinoline chromophore with the aryl sugar acting as a minor groove binder was also considered. Figure 4.22 shows an image of **Q1-S2** intercalated into a (GC)₈ sequence binding. However, in this case, the phenyl ring present in the deoxy sugar presents steric clashes with the base pairs (Figure 4.22) and thus disfavours this binding mode compared with groove binding.
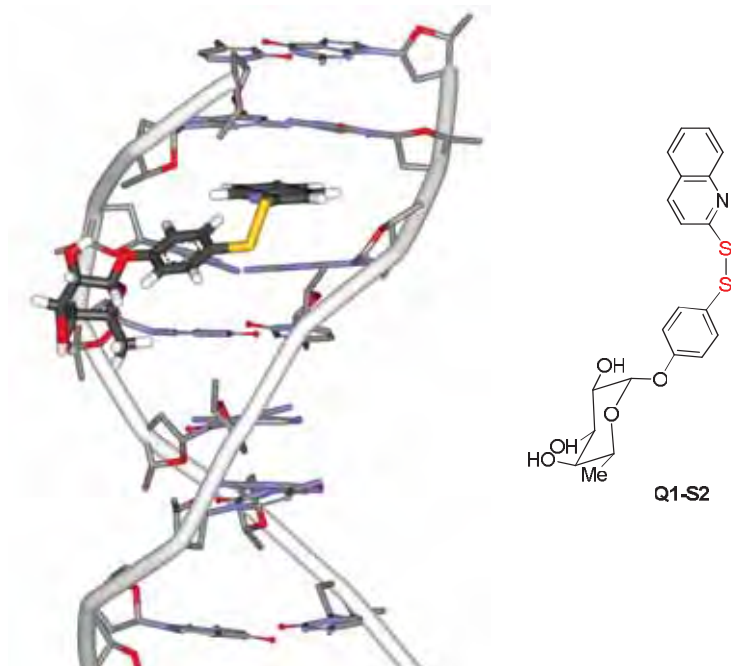
Figure 4.22 Interaction of α-anomer of **Q1-S2** into a (GC)$_8$ sequence as a intercalator 3D image generated by James Garner.[228] DNA (hydrogens and phosphate backbone atoms omitted for clarity), Ligand (dark grey carbon atoms).
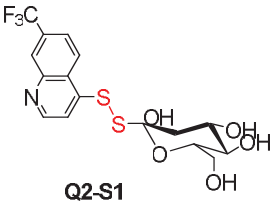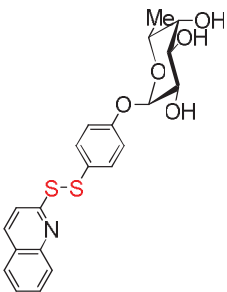
(c) Docked Q1-S3

In contrast, **Q1-S3** contains a benzylic group in place of the aryl group, and this disulfide was not selected by any of the DNA sequences. The introduction of a single methylene group introduces an additional degree of free rotation between the disulfide and the phenyl ring, and as a result the longer length of the molecule and flexibility do not match the curvature of the groove as well as **Q1-S2**.

In summary, the qualitative data obtained from the molecular visualisation studies and also the marginal energy difference calculation predicted that the favoured DNA-binding mode of **Q2-S1** is most likely *via* intercalation with the

G:C rich DNA sequences. Whereas, the quinoline deoxy sugar, **Q1**-**S2** is predicted to bind better as an A:T groove binder. While visualisation was used to understand the DNA-binding modes of **Q2**-**S1** and **Q1**-**S2**, further detailed studies would be required to confirm the molecular level information. The functional groups involved in DNA-binding interactions of **Q2**-**S1** and **Q1**-**S2** are summarised in Table 4.3.

Table 4.3 Summary of Molecular modeling studies of docked **Q2**-**S1** and **Q1**-**S2** with the DNA sequences

Ligand	Mode of Binding	Groups Involving in H-bond	with (GC) DNA	
			BP	PO$_4$
Q2-S1	(GC) Intercalation	Sugar	1	3
		CF$_3$	1	–
		-S-S-	1	–
		Aromatic N-atom	–	–
		O-atom (Sugar)	–	–

Ligand	Mode of Binding	Groups Involving in H-bond	with (AT) DNA	
			BP	PO$_4$
Q1-S2	(AT) Groove Binding	Sugar	2	2
		Phenyl	–	–
		-S-S-	–	–
		Aromatic N-atom	–	–
		O-atom (Sugar)	–	–
		O-atom (Phenyl)	–	–

Note: The hydrogen bonds were monitered between the functional groups attached to the ligand and the DNA base pair (BP), phosphate backbone (PO$_4$) based on lowest energy conformation of the docked structure.

4.8 Summary
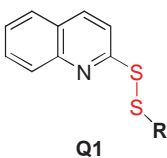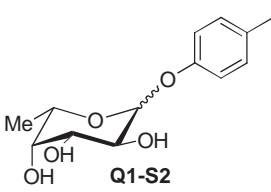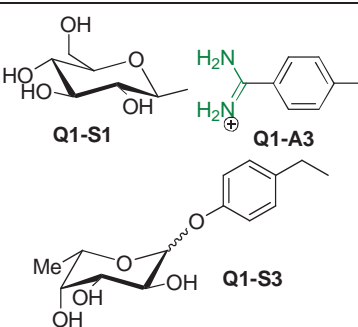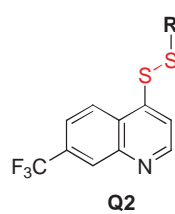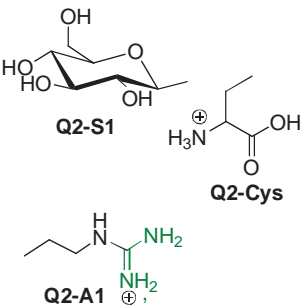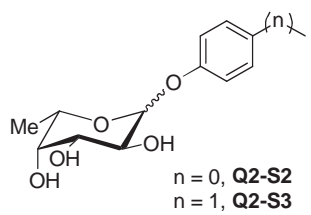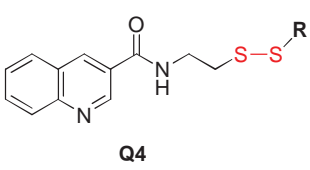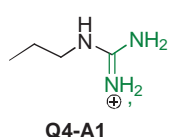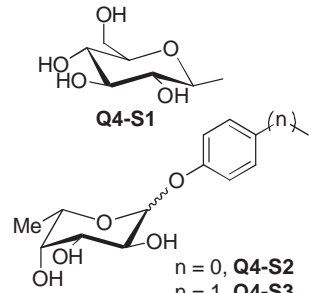
The potential for DCC to identify new DNA-binding molecules was investigated by using three short oligonucleotides, hpGC **O1**, d(GC)$_5$ **O2** and d(AT)$_5$ **O3** as models for DNA and the results were interpreted with the aid of molecular visualisation.

Table 4.4 summarises the key quinoline disulfides that bound to DNA, as well as the disulfides that did not bind (or weakly bound) to DNA, identified using DCC. While the positively charged guanidium quinolines were predicted to bind to DNA, the identification of **Q1**-**S2** and **Q2**-**S1** in particular, was not expected. In the case of **Q1**, DCC allowed direct comparison of the relative affinity of **Q1**-**S2** compared with **Q1**-**S3** for DNA; these two disulfides differ by a single methylene group. In addition, the assay allowed direct comparison of the aryl guanidinium derivative **Q1**-**A3** with the sugar derivatives **Q1**-**S1** and **Q1**-**S3** in the one assay. In the case of the **Q2**, while the charged disulfides and glucose derivatives would be predicted to interact with DNA, neither of the deoxy derivatives **Q2**-**S2** and **Q2**-**S3** were selected by DNA. This result is different to that observed with **Q1**, where one of the deoxy sugar disulfides **Q1**-**S2** was selected by DNA, but the second deoxy sugar disulfide **Q1**-**S3** was not bound to DNA.

Analysis of both the DNA-bound and unbound solutions in the assay provides information on some of the factors that are important for recognition. For example, as shown in Table 4.4, analysis of the DNA-bound fractions using **Q1** showed that none of the selected compounds contained the glucose sugar **S1** or any derivatives of the amidine building blocks. By systematically removing

individual building blocks from the experiment, further information regarding the building block design is possible.

Table 4.4 Summary of DNA-binding studies with quinolines.

It is noted that the assay was carried out under a limited set of reaction conditions that were designed to identify compounds with moderate to strong affinity for DNA and in the absence of salt, which is known to affect DNA-binding constants. Some non-specific binding and multiple binding modes

cannot be ruled out. The relatively minor changes in sequence selectivity are consistent with weak DNA-binding interactions. In the case of **Q2**-**S1** and **Q1**-**S2**, though the compounds bind weakly to DNA, the docking studies predicted that different substitution on the quinoline ring and the hydrogen bond donor/acceptor groups are important DNA-binding characteristics. While visualisation was used to predict the major binding modes of **Q2**-**S1** and **Q1**-**S2** as intercalation and minor groove-binding respectively, further detailed studies would be required to confirm these conclusions. While X-ray crystallography and NMR spectroscopy are typical techniques that provide molecular level information on ligand-DNA interactions, these techniques require a strong binding constant ($K >10^{4-6}$) and strong sequence selectivity in order to form a discrete 1:1 ligand:oligonucleotide complex. The quinolines **Q2**-**S1** and **Q1**-**S2** are weak binders and exhibited only minor variations in binding to different base sequences. Hence X-Ray and NMR spectroscopy are not appropriate techniques for further characterisation.

The flexible bisthiols **B1**, **B2** and **B3**, designed to allow the formation of bis-intercalators with quinolines **Q1** and **Q2**, did not result in the identification of any DNA-binding compounds. The bisthiols **B1**, **B2** and **B3** contain a large number of degrees of rotational freedom and the entropic penalty of bis-intercalators formed from these linkers, adopting a conformation suitable for bis-intercalation is high. The distance between the quinoline rings in the potential bis-intercalators formed from the bisthiols with quinolines **Q1** and **Q2** was comparable to a number of related synthetic bisintercalators reported in the literature.[226-227] The failure of DCC to detect any DNA-binding compounds is consistent with the properties of well-studied DNA bis-intercalators, which are characterised by some rigidity in the linkers and

which contain a shorter distance between the thiol functional groups than was able to be achieved in the design of **B1**-**B3**. The use of aromatic chromophores containing more than two fused rings (i.e. a stronger intercalator) is also preferred with highly flexible linkers for the entropic reasons mentioned above.

4.9 Experimental

4.9.1 Materials and Methods

Commercially available chemicals and solvents (HPLC quality) were used without further purification, unless otherwise stated. All references to water refer to the use of Milli-Q water generated from a Millipore, Milli-Q Bicel A 10 system. The pH was measured on a Beckman Instruments Φ210 pH meter. The mobile phase consisted of eluents A (water containing 0.1% formic acid) and B (acetonitrile) for all HPLC runs.

4.9.2 LC-MS Conditions

The mobile phase consisted of eluents A (water with 0.1% formic acid) and B (acetonitrile) for all runs. LC-MS analysis was performed with a Thermo Scientific 5μm Hypersil GOLD RP C18 column (2.1 × 150 mm column, 5 μm particle size and flow rate 0.2 mL min^{-1}).

The DCL experiment and analysis was performed by LC-MS using a LCQ DECA XP Plus (Thermo Fisher Scientific, San Jose Ca, USA) ion trap mass spectrometer using electrospray ionization (ESI) source, with a Surveyor PDA Plus detector (Thermo Fisher Scientific, San Jose Ca) and was attached to the ESI Source through a divert valve. The instrument conditions were optimised

for sensitivity on both solvent (background) and compound using LC tune software and the data was initially acquired in full scan mode. MS^n product ion scans were performed using collision induced dissociation (30 eV), and the ions generated were measured over the mass range 100-800 Da. Data was analyzed using the Qual Browser feature in Xcalibur 2.1 (Thermo Fisher Scientific, San Jose, Ca, USA).

4.9.3 Normalization with respect to the Internal Standard

The peaks were reported as percentage proportion changes based on peaks areas normalised with respect to an internal standard, 3,5-dihydroxybenzoic acid.[95] The internal standard was prepared as a 2.5 mM stock solution in 5% methanol and water (containing 0.1% formic acid) for untemplated and templated DCL analysis (DNA unbound). For the DNA-bound analysis (templated), 0.5 mM solution in water (containing 0.1% formic acid), diluted from a 2.5 mM stock solution in 5% methanol and water (containing 0.1% formic acid) was used. An equal volume of the internal standard for control (untemplated) and DNA unbound (templated) (10-15 μl, 2.5 mM), DNA-bound (10-15 μl, 0.5 mM) were added (dependent on the area of the largest peak in the DCL) at the time of the analysis.

4.9.4 Selection of Wavelengths for DCL Analysis:

A single wavelength could not be selected for the DCL analysis, as the DCL components have different UV to the starting material. Linear combination of absorbances at two wavelengths near isosbestic points (specific wavelength at which two chemical species have the same molar absorptivity (ε) or –more generally –are linearly related, IUPAC definition), where one set of peaks have

153

high relative absorbance and the others low and *vice versa* at the second wavelength.

There are two wavelengths, 234 and 286 nm, where the relative absorbance of the other two peaks is low and also corresponds to a low standard deviation. The analysis was performed at isosbestic points, to add the peak areas at two distinct wavelengths thereby equalising the sensitivity of peak detection of all components.

4.9.5 Computational methods
(a) Preparation of the Ligands

Quinoline compound structures were created in Spartan '04 (Wavefunction Inc) and geometry optimised with the MM2 forcefield using the default parameters, then minimised using the PM3 forcefield.

(b) Preparation of the Receptors

A search of the Protein Data Bank[230] for intercalating structures suitable to build canonical DNA models yielded 1CX3,[231] with 2-base pairs between the intercalation gaps, and 1AL9,[232] with 4-base pairs between the intercalation gaps. The structure 3EY0[10] was used to build canonical DNA without an intercalation gap. The ligands from each pdb file were removed. The files were then submitted to the 3DDART software portal[233] and the required sequence (A:T or G:C DNA) was entered in the appropriate field. The resulting pdb file was opened in DS Visualiser and the bond order for any of the phosphate groups with missing bonds was corrected. 1CX3 was used to prepare

atDNA_gap and gcDNA_gap. 1AL9 was used to prepare atDNA_4gap and gcDNA_4gap. 3EY0 was used to prepare atDNA_nogap and gcDNA_nogap.

Each canonical structure was minimised prior to docking using DS Modelling 3.0 (Accelrys Inc.). Each structure was typed with the CHARMm forcefield with Momany-Rome partial charges and subjected to a stepwise minimization using the CHARMm (Version 35b3) molecular dynamics module as implemented in DS Modeling 3.0. The system was an in vacuo simulation using default non-bonding parameters (cutoff: 13.50; cutoff-on: 8.00; cutoff-off: 12.00). The structure was minimized with Steepest Descent for 3000 steps (Grad. Tol. 0.1) with all non-hydrogens explicitly fixed (i.e. minimization of all hydrogens). The structure was minimized with Steepest Descent for 3000 steps (Grad. Tol. 0.1) with the aromatic carbons and nitrogens of the base pairs fixed. The structure was minimized with Adopted Basis-set Newton-Raphson (ABNR) for 30,000 steps (Grad. Tol. 0.01) with the aromatic carbons and nitrogens of the base pairs fixed. The output structure was designated the minimised structure. Attempting to minimise the structure without any constraints during the second and third steps resulted in unacceptable distortion of the helix.

(c) Docking Protocol

A quinoline compound and minimised canonical DNA structure was opened in Autodock Tools (ADT) and Gasteiger-Marsili charges were calculated. A grid box was created with 92 x 92 x 92 points and a resolution (i.e. spacing between each point) of 0.375 Å, in order to include the entire DNA fragment. The grid box was centred on the DNA and grid potential maps were calculated using AutoGrid 4.0.

The following parameters were used for each docking; number of runs (25), maximum of energy evaluations (50×10^6), maximum of generations (27,000). Mutation and crossover were applied to the population at rates of 0.02 and 0.80 respectively. Since the grid box enclosed not only the binding site but also the entire DNA fragment, root-mean-square deviation (RMSD) could not be used as an accuracy criterion. A more subjective criterion is to classify the resulting binding mode by visual inspection as intercalation, minor groove binding, or others (major groove binding, interaction with phosphate groups, etc.).

4.9.6 DCC Experimental Procedures

Stock Solutions

(a) Oligonucleotides

5'-Biotinylated oligonucleotides, containing a biotinTEG phosphoramidite (L) linker (Table 4.1), were purchased from Geneworks (0.1-0.3 μmol scale) at HPLC purified quality. The oligonucleotides were annealed in an equal volume (μL equal to nmol of oligos) of TE buffer [10 mM Tris.HCl (pH 8.0 at 25 °C) + 0.1 mM EDTA] by heating at 95 °C for 5 min followed by gradual cooling to room temperature over 6 h.

(b) Building Blocks

Stock solutions of building blocks were prepared in water or 10-20% aqueous methanol solutions at the following concentrations: quinolines **Q1** (1.0 mM), **Q2** (1.0 mM) and **Q4-Y** (1.4 mM); imidazoles **M1** (2.5 mM) and **M2** (2.5 mM); amidines **A1**-**A1** (2.5 mM), **A2**-**A2** (2.5 mM) and **A3**-**A3** (2.5 mM); cysteine (**Cys,**

2.5 mM); thiosugars **S1** (2.0 mM), **S2** (2.1 mM) and **S3-S3** (1.0 mM); and the bisthiols **B1** (1.0 mM), **B2** (1.0 mM) and **B3** (1.0 mM). A 0.5 mM stock solution of the internal standard was prepared by dilution of the 2.5 mM stock solution for DNA-bound (template) samples, thereby accounting for the reduced signal of these samples and ensuring an acceptable baseline in the library region of the spectrum. An equal volume of the internal standard for control (untemplated) and DNA unbound (templated) (10-15 µL, 2.5 mM), DNA-bound (10-15 µL, 0.5 mM) samples were added within each experimental series before LC-MS analysis. All stock solutions were stored in the freezer until required and were allowed to warm to room temperature before use.

(c) Preparation of Dynabeads

Dynabeads® MyOne™ Streptavidin C1 were obtained from Invitrogen and used according to manufacturer's instructions. The Dynabeads are superparamagnetic beads of 1.0 µm in diameter with a streptavidin monolayer covalently coupled to the hydrophilic bead surface. The Dynabeads (10 mg/ml, ~ 7-10×10^9 beads/ml) were in phosphate buffered saline (PBS) pH 7.4, containing 0.01 % Tween®-20 and 0.09% NaN_3 as preservative. The estimated binding capacity of Dynabeads for dsDNA is ~ 20 µg/mg (dependent on the length of the oligomers). The Dynabeads (200 µL, 10 mg/mL suspension) were washed with milliQ water (4×100 µL) followed by 0.1% TFA in water (4×100 µL) and were resuspended in 0.1% TFA in water (50 µL), and were recovered by magnetic separation using Neodymium magnets (Maglab Magnets).

(d) DCL Experiments

Libraries were prepared with the following compositions:

DCL-(a): Quinoline **Q4-Y** (1.4 mM, 75 μL, 105 nmol); amidines **A1-A1** (2.5 mM, 100 μL, 250 nmol), **A2-A2** (2.5 mM, 100 μL, 250 nmol) and **A3-A3** (2.5 mM, 100 μL, 250 nmol); cysteamine (10.0 mM, 100 μL, 1000 nmol) and ammonium hydroxide (10.0 μL, 1.0%) were mixed to give a total volume of 410.0 μL at pH ~ 8.5.

DCL-(b): Quinoline **Q4-Y** (1.4 mM, 75 μL, 105 nmol); amidines **A1-A1** (2.5 mM, 100 μL, 250 nmol), **A2-A2** (2.5 mM, 100 μL, 250 nmol) and **A3-A3** (2.5 mM, 100 μL, 250 nmol); **GSH** (5.0 mM, 75 μL, 375 nmol) and **GSSG** (1.25 mM, 75 μL, 93.75 nmol) and ammonium hydroxide (15.0 μL, 1.0%) were mixed to give a total volume of 540.0 μL at pH ~ 7.4.

DCL-(c): Quinolines **Q1** (1.0 mM, 75 μL, 75 nmol), **Q2** (1.0 mM, 75 μL, 75 nmol) and **Q4-Y** (1.4 mM, 60 μL, 84 nmol); thiosugars **S1** (2.0 mM, 75 μL, 150 nmol), **S2** (2.1 mM, 75 μL, 157 nmol) and **S3-S3** (1.0 mM, 100 μL, 100 nmol), and ammonium hydroxide (10 μL, 1.0%) were mixed to give a total volume of 470.0 μL at pH ~ 8.5.

DCL-(d): Quinolines **Q1** (1.0 mM, 75 μL, 75 nmol) and **Q2** (1.0 mM, 75 μL, 75 nmol); deoxysugars **S2** (2.1 mM, 75 μL, 157 nmol) and **S3-S3** (1.0 mM, 100 μL, 100 nmol) and ammonium hydroxide (5.0 μL, 1.0%) were mixed to give a total volume of 330.0 μL at pH ~ 8.5.

DCL-(e): Quinoline **Q1** (1.0 mM, 75 μL, 75 nmol); imidazoles **M1** (2.5 mM, 100 μL, 250 nmol) and **M2** (2.5 mM, 75 μL, 187.50 nmol); amidines **A1**-**A1** (2.5 mM, 50 μL, 125 nmol), **A2**-**A2** (2.5 mM, 50 μL, 125 nmol) and **A3**-**A3** (2.5 mM, 50 μL, 125 nmol), and ammonium hydroxide (5.0 μL, 1.0%) were mixed to give a total volume of 405.0 μL at pH ~ 8.5.

DCL-(f): Quinoline **Q2** (1.0 mM, 75 μL, 75 nmol); imidazoles **M1** (2.5 mM, 100 μL, 250 nmol) and **M2** (2.5 mM, 75 μL, 187.50 nmol); amidine **A1**-**A1** (2.5 mM, 50 μL, 125 nmol), and ammonium hydroxide (10.0 μL, 1.0%) were mixed to give a total volume of 310.0 μL at pH ~ 8.5.

DCL-(g): Quinolines **Q1** (1.0 mM, 75 μL, 75 nmol) and **Q2** (1.0 mM, 75 μL, 75 nmol); bisthiols **B1** (1.0 mM, 100 μL, 100 nmol), **B2** (1.0 mM, 100 μL, 100 nmol) and **B3** (1.0 mM, 100 μL, 100 nmol) and ammonium hydroxide (5.0 μL, 1.0%) were mixed to give a total volume of 330.0 μL at pH ~ 8.5.

A control experiment was conducted by quenching an aliquot of the equilibrated DCL solution (70 μL) using aqueous TFA (15 μL, 0.1%) to pH ~ 3.0. The internal standard (2.5 mM, 15 μL) was added to the solution to give a total volume of 100 μL, followed by analysis by LC-MS.

(e) Equilibration of DCL Reactions

Disulfide exchange reactions between building blocks were initiated at physiological pH ~ 7.4 or at basic pH ~ 8.5 by the adjustment of the pH with ammonium hydroxide (1%). The libraries were stirred for 3 days with a gentle stream of air bubbled in the solution.

In templated DCL studies, an aliquot of the equilibrated DCL solution (70 μL) was incubated at 30-32 °C with the DNA oligonucleotides (1.0 mM, 9.0 μL) for 3 days. The solutions were agitated manually once daily during the re-equilibration. The experiments were quenched with the addition of aqueous TFA (60-80 μL, 0.1%) to pH ~ 3.0, transferred to a prepared Dynabead sample and was incubated at room temperature for 40 min. The beads were separated from the solution with a magnet and the supernatant (U1) was transferred into an eppendorf tube.

The beads were washed with 0.1% TFA in water (5 × 100 μL) and the washings combined with the supernatant (U1). The solvent was removed *in vacuo* (LABCONCO, CentriVap® Mobile system Model 7812011, 80-90 min at 58 °C) and the resultant residue was diluted with 0.1% TFA in water (85 to 93 μL). To this, the internal standard 3,5-dihydroxybenzoic acid was added (2.5 mM, 7 to 15 μL), to give a total volume of 100 μL, and was analysed by LC-MS.

The DNA was denatured by incubating the Dynabeads with 0.1% TFA in water at 90 °C for 10 minutes (3 × 85 μL). The beads were separated with a magnet and the supernatant (DNA-bound, U2) from each denaturing cycle was combined. The solvent from the combined solutions were removed *in vacuo* and the resultant residue was diluted by 0.1% TFA in water (85 to 93 μL). To this, the internal standard 3,5-dihydroxybenzoic acid was added (0.5 mM, 7 to 15 μL), to give a total volume of 100 μL, and the DNA-bound solution was analysed by LC-MS.

(f) LC-MS Analysis

For each DCL analysis (control, U1 and U2) 40 μL was injected per run. The following elution gradients were employed:

DCL-(a and b): 99% A for 15 min, then gradient was raised to reach 40% B at 110 min, then increased to 100% B at 115 min and held at 100% B for 4 min, then ramped to reach 99% A at 122 min, then held at 99% A until 150 min.

DCL-(c, d and g): 95% A for 5 min, then the gradient was raised to reach 30% B at 35 min, then to 60% B at 65 min, then ramped to reach 100% B at 80 min and held at 100% B for 3 min, then ramped to reach 95% A at 86 min, then held at 95% A until 105 min.

DCL-(e and f):95% A for 5 min, then gradient was raised to reach 60% B at 40 min, then increased to 100% B at 45 min and held at 100% B for 3 min, then ramped to reach 95% A at 51 min, then held at 95% A until 75 min.

The internal standard 3,5-dihydroxybenzoic acid was the largest peak in each spectrum and all peaks were normalised to it. Results are reported as percentage change in composition/percentage composition of each component between the DNA-bound and the control library based on peaks areas of each component (not including the internal standard). The peak areas at 234 nm and 286 nm were summated, equalising the sensitivity of peak detection for all components.

Chapter 5

Biostable Disulfide Mimics

5.1 Biostability of Disulfides

Disulfides are important structures of many proteins and peptides. They have an important function of forming a stabilized and rigidified macrocycle that is often directly relating to biological activity.[234] However, disulfide bonds are generally unstable under biological conditions as they are susceptible to reduction with thiol-containing molecules such as GSH, as well as with redox enzymes.[234-237] Numerous studies of small peptides containing disulfides have confirmed the importance of the disulfide for biological activity.[236,238] As a result, replacement of the disulfide bond with bioisosteric and reductively stable mimics suitable for therapeutic applications have been investigated. These mimics have included more stable linkages such as amides,[239-240] thioethers,[237-238,241] hydrocarbon,[31,242-244] and diselenides.[237,245] Depending on the structure and properties of the peptides, biological activity studies have shown that these strategies can deliver hydrolytically and reductively stable mimics that have the same, or improved biological activities with respect to the parent disulfide.

Very recently, the use of 1,4 and 1,5-disubstituted-1,2,3-triazoles as disulfide mimics has been reported (Figure 5.1).[234,246] Triazoles are readily prepared using "Click" chemistry with the regiochemistry controlled by the choice of catalyst for the reaction.[234] Compared to other mimics, triazoles offer the advantage of directional formation of the disulfide bonds. In the first study, the 1,4-triazole was incorporated into the 17-residue tachyplesin I, that contains two disulfide bonds which maintains a β-hairpin structure. While molecular modeling suggested that there was a high degree of similarity between the disulfide bond in the peptide and the triazole containing analogue, there was a less optimal overlap between the peptide backbones.

162

Nevertheless NMR studies showed highly similar structures and almost identical biological activity of the analogues and the peptide.[234] A second paper compared the effectiveness of the 1,5- versus the 1,4-triazole as a mimic of the 14-residue sunflower trypsin inhibitor. Modeling suggested that the 1,5-triazole would retain the overlap geometry of the peptide whereas the 1,4-triazole would be unable to adopt the same conformation.[246] This hypothesis was confirmed by biological activity results which showed that the 1,5-traizole retained almost full biological activity, whereas the 1,4-triazoles had reduced activity.
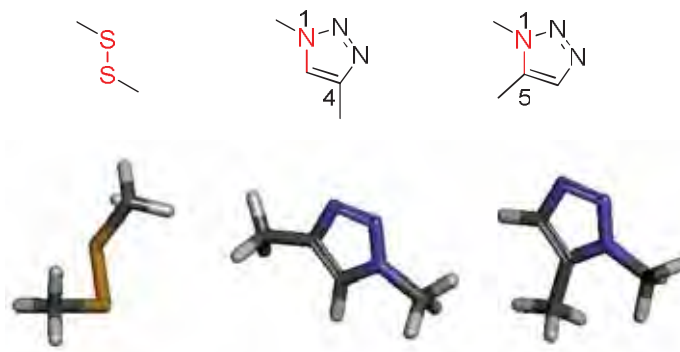
Figure 5.1 Structures of 1,4- and 1,5-disubstituted-1,2,3-triazoles and comparison with the disulfide functional group.

There have been numerous studies on proteins and peptide fragments of proteins, especially α-helices. In many cases, these peptides have contained disulfide bonds. These small peptides have been designed in which the disulfide has been replaced by hydrocarbons and other functional groups, and their DNA-binding characteristics have been reported.[247] In most of the DNA-binding proteins and peptides, the role of the disulfide has been to stabilise the secondary structure of the peptide and reduce the entropic penalty of
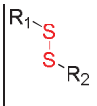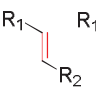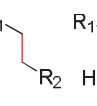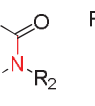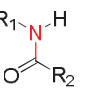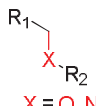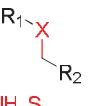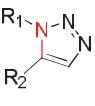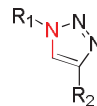
binding to DNA. The amino acid residues involved in DNA recognition are on the opposite surface of the peptide to the disulfide bridge. Hence a wide range of functional groups to substitute for the disulfide bonds are tolerated in the peptide mimics.

In contrast, not surprisingly due to their limited biostability, small molecules that bind to DNA have not been designed to contain disulfide functional groups. Hence there are no reports of successful mimics of disulfides for DNA-binding. The DNA-bisintercalator, triostin A **4** (see Figure 1.2, Chapter 1), which is stabilized by a disulfide bridge,[62,64,248] is a rare example of a stable DNA-binding compound that contains a disulfide bond. However, the disulfide bridge in triostin A **4** has a similar role to disulfides in proteins, and acts to stabilise the cyclic depsipeptide ring and position the amino acid residues in the minor groove of DNA.[61-62] Several acyclic derivatives of triostin A have been reported and hydrocarbon analogues of peptide fragments of triostin A have been shown to retain RNA recognition properties (Figure 1.14, Chapter 1).[31]

5.2 Design of Quinoline Disulfide Mimics

In this work, hydrocarbon, heteroalkane, amides and triazole mimics of **Q1-S2** and **Q2-S1** were designed. The key features in these mimics are summarized in Table 5.1. The design of mimics of **Q1-S2** and **Q2-S1** required consideration of the effect of the mimics on the geometry and distances between R1 and R2, and also whether the disulfide functional group is involved in molecular recognition of DNA.

Table 5.1 Potential stable disulfide mimics and their characteristics; overall bond length data taken from the literature.[249]

Properties	R1–S–S–R2	R1—R2 (alkene)	R1—R2 (alkyl)	R1C(O)–N(H)–C(O)R2	R1–N(H)C(O)R2	R1CH2–X–R2 (X = O, NH, S)	R1–X–CH2R2	triazole	triazole
Rotation & Flexibility	X	—	X	—	—	X	X	—	
H-bond donor / acceptor characteristics	2 [A]	—	—	1 [D] & 1 [A]		1 [A]		2 [A]	
Shape / Steric Effects	—	—	—	—	—	—	—	X	X
Overall bond length	2.05	1.46	1.53	1.23	1.20	1.37 (X=O) 1.43 (X=NH) 1.82 (X=S)		1.27[b]	2.06[b]

[a][A]-Acceptor & [D]-Donor

[b]Distance measured using chemdraw 3D image.

Compared to the parent disulfide, the bond length is reduced in all the mimics and hence the distance between R1 and R2 is reduced. The major conformation of the disulfide bond is *trans*, with the *cis* conformation not energetically favorable (Figure 5.2)[244] In comparison, the alkene and amide mimics have restricted rotation and exist with R1 and R2 *trans* with respect to one another, while in the triazoles, the R1 and R2 groups are *cis* with respect to one another. The heteroatom-containing mimics have similar rotational flexibility to the parent disulfide and the fully saturated hydrocarbon mimic. An important feature for DNA recognition is hydrogen bond donor/acceptor properties and in addition, shape recognition is often important. These properties were also considered in the design. In particular, the alkene and alkyl groups which do not have any H-bond acceptor/donor properties, while the amides, heteroalkyl and triazoles have H-bond acceptor/donor groups at different positions that may alter the DNA-binding properties.

Figure 5.2 Conformational isomers of disulfides.

From the results obtained in Chapter 4 with **Q1-S2**, it was expected that the flexible alkane and heteroalkane analogues would deliver the best mimics as they allowed the formation of the crescent shaped conformation in the minor groove. In contrast, the rigidity of the alkene and amide analogues did not allow overlap of the mimics with the parent disulfide. Similarly both the steric bulk and the distinct geometry of the triazoles were expected to alter the DNA-binding characteristics.

In the case of **Q2-S1**, the intercalation model proposed in Chapter 4 suggested that any disulfide analogue that allowed positioning of glucose into the minor groove would be a good mimic. Hence, it was predicted that all analogues shown in Table 5.1, except possibly the more bulky triazoles, would give similar DNA-binding profiles.

The disulfide mimics shown in Figure 5.3 were synthesized by P. M. Abeysinghe in our group. In the case of **Q1-S2**, thioether β-**65** and amide α-**66** mimics were prepared as pure anomers, while for **Q2-S1**, thioether **67**, 1,4- and 1,5-triazoles **68** and **69** were prepared. Comparison of the properties of the mimics with the parent disulfides **Q1-S2** and **Q2-S1**, was also important. However, only low overall yields and purity **Q1-S2** was obtained. Attempts to prepare the alkene analogues required lengthy procedures and were not pursued.
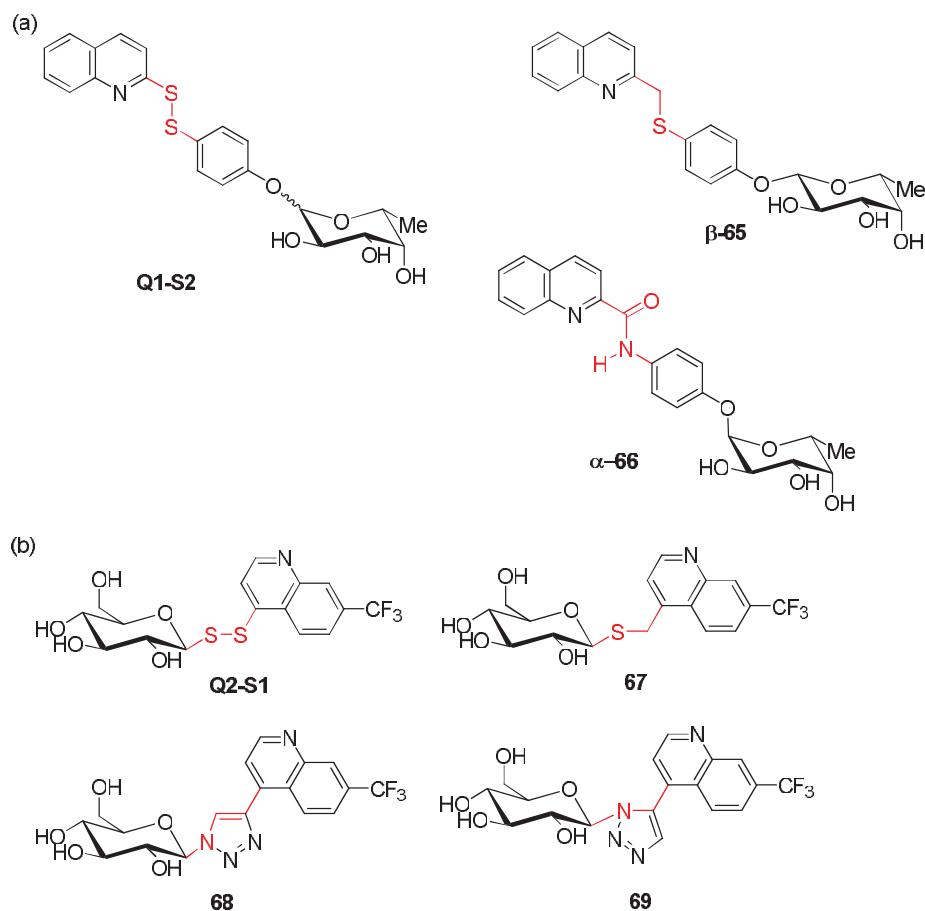
Figure 5.3 Disulfide mimics of (a) parent disulfide **Q1**-**S2** and the mimics thioether β-**65** and the amide α-**66** and (b) the parent disulfide **Q2**-**S1** and the mimics thioether **67**, 1,4- and 1,5-substituted triazoles **68** and **69**.

5.3 DNA-Binding Assay Experiments

The DNA-binding experiments were carried out using the same methods for the DCC experiments presented in Chapter 4, except that the equilibration to form a DCL was not necessary. The parent disulfide and the mimics were incubated in the presence of the oligonucleotides **O1**, **O2** and **O3** for 24 h, and the duplex was denatured and separated, providing solutions of the DNA-

bound material as well as the unbound compounds. Both samples were adjusted to a defined concentration, and an internal standard added to allow LC-MS spectra to be normalised and compared.

It was assumed that allowing a mixture of equimolar amounts of the parent disulfide and the mimics would result in all compounds reversibly binding to DNA with the same affinity and hence the same amount of each compound would be detected in the DNA-bound fraction (corrected for any differences in mole equivalents). If the tested analogues had a different binding mode with DNA compared to the parent disulfide, then this would be reflected in the amounts of bound and unbound compounds measured in solution.

5.3.1 Q2-S1 Mimics

Figure 5.4 shows the DNA-binding experiment of quinoline disulfide **Q2-S1** with the thioether **67**, and triazole click products **68** and **69** (Figure 5.3a) in a ratio of 3:1:2:3 with total scan detection (210-350 nm) as well as a single wavelength at 252 nm (Figure 5.4b and c). The use of equimolar amounts of each of the mimics was not possible due to the solubility issues. Hence, the ratio of 3:1:2:3 was used and the mole ratio was corrected for any differences in mole equivalents. The analysis was carried out using two detection methods, total scan and at a single wavelength. The analysis of the results was not straightforward for several reasons. First, the presence of an impurity from the disulfide **Q2-S1** that co-eluted (as a shoulder peak) with the thioether **67** (Figure 5.4b) meant that accurate measurement of the concentration of **Q2-S1** was not possible. This impurity was not detected in the ^{1}H NMR spectrum of
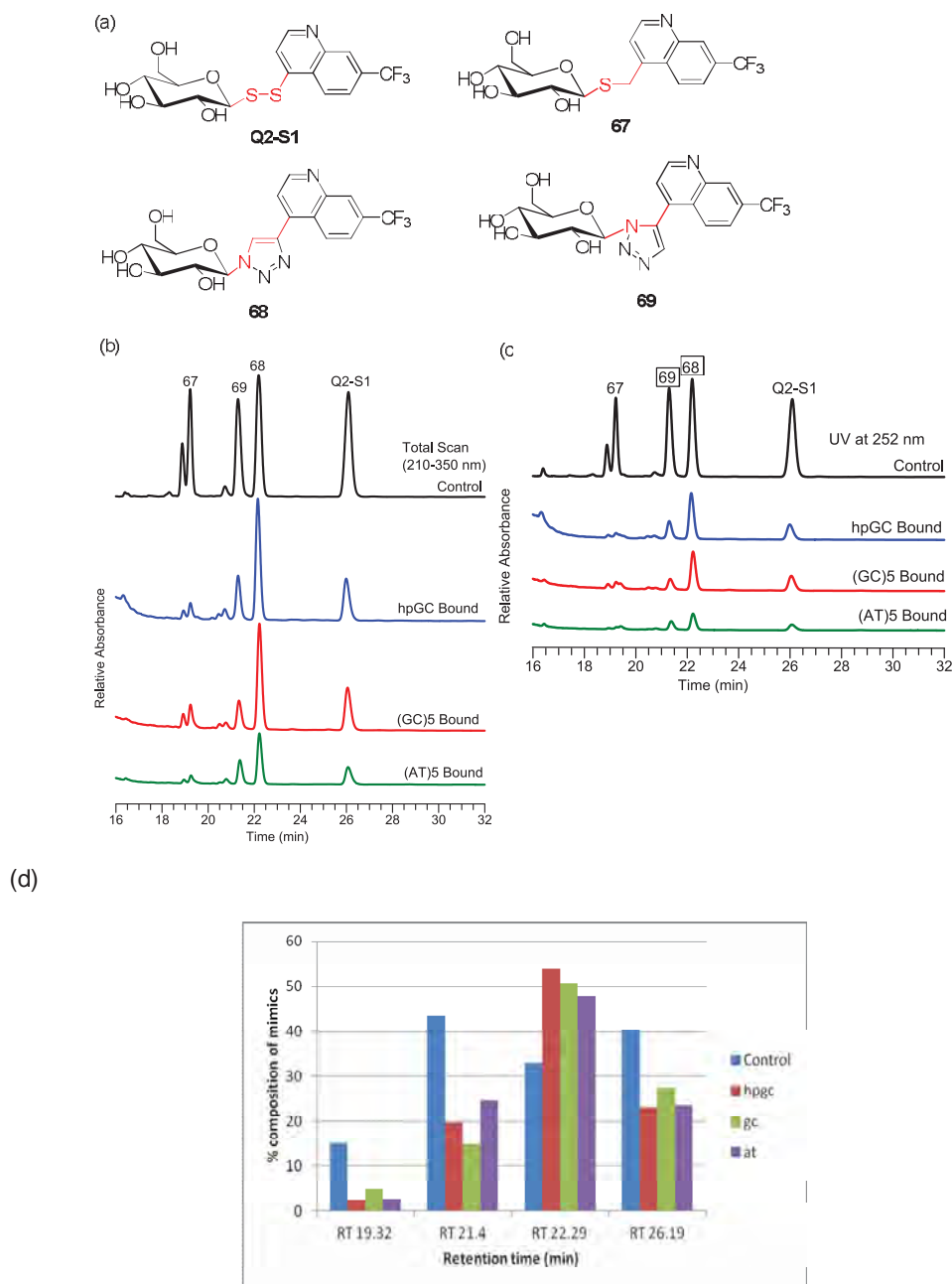
Figure 5.4 DNA-binding assay experiment generated from (a) **Q2-S1** and the mimics **67**, **68** and **69** in a ratio of 3:1:2:3, (b) LC trace with UV total scan detection of the control from 210-350 nm and single wavelength at 252 nm, (c) the control and the DNA bound spectra of **O1**, **O2** and **O3** oligo nucleotides at 252 nm and highlighting the DNA-bound compounds **69** and **68** (box) and (d) percentage composition of the mimics the control experiment.

Q2-**S1**, and was unable to be identified or removed from the sample. While at 270 nm the impurity absorbance was very weak, the analysis was performed at 252 nm, where the relative absorbance of the peaks gave a low standard deviation. Second, the solubility of the mimics was poor, even in 25% methanol/water, and the stock solutions of known concentration could not be prepared. Hence, saturated stock solutions were prepared and the concentrations were calculated using UV-Vis spectroscopy. UV-Vis spectra were measured in methanol in which the compounds showed excellent solubility, as well as 25% methanol/water, which showed some effect on the molar extinction coefficient. For this reason the ratio of the starting mixture of compounds was in the relative molar ratio of 3:1:2:3.

Figure 5.4d shows the percentage composition of the mimics and the disulfide with the respect to the control experiment of the starting compounds, corrected for the different molar ratio of the compounds used in the assay. Relative to the control, the predominant disulfide analogue selected was the 1,4-triazole click product **68** (Figure 5.4c), which was selected in preference to the parent disulfide **Q2**-**S1**. The 1,5-click product **69** had a similar binding profile to the parent disulfide with hairpin **O1** and (GC)$_5$ **O2** oligonucleotides, and showed a slightly enhanced profile with (AT)$_5$ **O3** sequence. In contrast, the thioether **67** showed weak DNA-binding characteristics with all the three sequences, indicating the changing one sulfur atom to a methylene group has a significant effect on DNA-binding (Figure 5.4c). However, given the solubility issues, the use of different solvents in the LC-MS analysis *versus* the UV data, and the errors associated with the sample preparation and concentration

measurements, there is likely to be a significant error associated with the percentage composition figures, and quantitative conclusions cannot be made.

Overall, the 1,4-click compound **68** showed a similar DNA-binding profile to the parent disulfide. The selection of the 1,4-click compound **68** rather than the 1,5-click product **69** suggests that the steric bulk at the 4-position of the quinoline ring is important. While the triazole mimic **69** that has similar geometry to the parent disulfide **Q2-S1**, the slight change in positioning the sugar with respect to the quinoline ring appears to be important. The reduced binding profile of the thioether **67** *versus* the disulfide indicates a more subtle influence of distance between the quinoline and sugar groups, but further experiments on more soluble derivatives are needed to make quantify the DNA-binding affinity of these two compounds.

The intercalation model proposed in Chapter 4 suggested that any disulfide analogue of **Q2-S1** that allowed positioning of glucose into the minor groove would be a good mimic. Hence, it was predicted that thioether **67** and possibly the more bulky triazoles **68** and **69**, would give similar DNA-binding profiles. The results in Figure 5.4 showed that the 1,5-click product **69** showed a similar profile of DNA-binding, while the 1,4-click product **68** showed enhanced DNA-binding and the thioether **67** showed weak affinity for DNA. These are not the results that would be predicted assuming that all analogues bound to DNA by intercalation of the quinoline ring, with the glucose group sitting in the groove and not involved in any molecular recognition. In particular, the thioether analogue **67** shows a very good overlap with the geometry of the disulfide **Q2-S1**, with only a slight reduction on bond length.

Molecular visualisation studies did not predict hydrogen bonding involving the sulfur atoms in **Q2**-**S1** and the reduced affinity of **67** for DNA is not understood. In the case of 1,5-triazole **69**, similar binding profiles were observed, suggesting that the 1,5-triazole is a good mimic of the disulfide. However, no conclusions can be made about the mode of binding with DNA and it is possible that the bulky triazole ring leads to steric clashes, and favours a different binding mode that gives a similar binding profile with the DNA oligonucleotides. The 1,4-triazole has a significant effect on the overall geometry and positioning of the glucose with respect to the quinoline ring. The enhanced DNA-binding was not predicted and suggests that this is most likely interacting with DNA in a different mode.

5.3.2 Q1-S2 Mimics

The effectiveness of **65** and **66** as mimics of **Q1**-**S2** was tested by mixing saturated solutions of **65** and **66** to give an approximately 1:1 ratio of absorbance peaks in the LC-MS of the mixture (Figure 5.5a). As in the case of the analogues of **Q2**-**S1**, the solubility of both mimics was poor, and the sample concentrations were estimated from UV data in methanol, with a slight adjustment for the use of 25% methanol/water in the assay. Both mimics were prepared as a β- and α- anomers of **65** and **66.** A pure sample of the parent disulfide **Q1**-**S2** was unable to be obtained, and hence the experiment was carried out in the absence of the parent disulfide.

Figure 5.5b shows the results of the binding studies with **O1**, **O2** and **O3** oligonucleotides and the percentage composition of the mimics relative to the control experiment corrected for the different mole equivalents in the starting
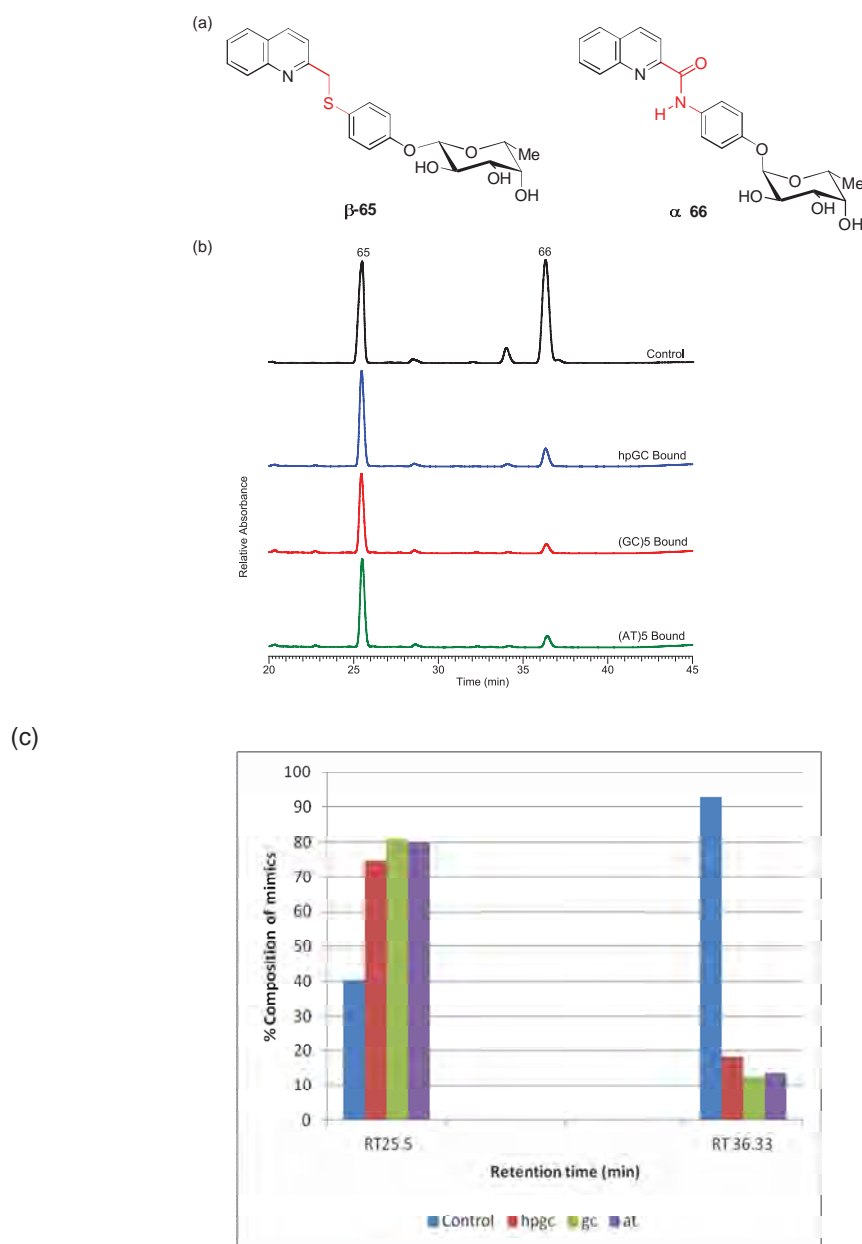
Figure 5.5 DNA-binding assay experiment generated from (a) **Q1-S2** mimics of β-**65** and α-**66** in a ratio of 1:2, (b) LC trace with UV total scan detection (210-350 nm) of the control and the DNA bound spectra of all the **O1**, **O2** and **O3** oligo nucleotides and (c) the percentage composition of the mimics with the control experiment.

173

mixture. Relative to the control, the thioether **65** was selected with all the three sequences and in addition a small amount of the amide **66** was also selected. Unfortunately, in the absence of parent disulfide **Q1**-**S2**, it is not possible to make conclusions regarding the relative amounts of **Q1**-**S2** *versus* **66** that would bind to DNA. However, this experiment clearly showed that the thioether analogue **65** interacts with DNA more strongly than the amide **66**. As predicted, the rigid and planar amide most likely prevents adoption of a crescent shaped conformation that matches the shape of the minor groove.

The DCL experiments suggested that the interaction of **Q1**-**S2** with DNA was as a minor groove binder (Figure 4.16, Chapter 4). The thioether β-**65** that has geometry similar to **Q1**-**S2** bound to an (AT)$_5$ sequence in the minor groove binding similar to the parent disulfide **Q1**-**S2**. The molecule fits the curvature of the minor groove and this conformation could potentially be stabilised by possible H-bonding with the A:T base-pairs.

5.4 Summary

The interaction of the thioether, amide and triazole analogues of the disulfides **Q2**-**S1** and **Q1**-**S2** were studied. The reduced aqueous solubility of these analogues made quantitative comparisons between their relative binding characteristics difficult. In the case of the **Q2**-**S1**, the results of the 1,4-triazole *versus* 1,5-triazole analogues are consistent with independent literature studies with peptides[234,246] which have shown that the 1,5-triazole gives a better geometrical overlap with the parent disulfide. As **Q2**-**S1** was predicted to bind to DNA by intercalation with the disulfide not involved in molecular recognition of DNA, it appears that the bulky triazole group is accommodated.

Surprisingly, the 1,4-triazole showed enhanced DNA-binding and further studies to determine the effect of this structural change on binding would be useful.

In the case of the **Q1**-**S2**, the location of the disulfide in the minor groove proposed by molecular visualisation studies, suggested that changes to this bond would alter DNA-binding significantly. The weak affinity of the amide compared to the thioether confirms this proposed binding mode. However, due to the low overall yields and impurity of **Q1**-**S2,** these analyses could not be compared with the parent disulfide **Q1**-**S2**. Hence, further work is required to confirm the DNA-binding properties of these analogues.

5.5 Experimental

5.5.1 Materials and Methods

The 2-substituted quinoline **Q1** mimics thioether β-**65** and the amide α-**66**, the 4-substituted quinoline disulfide **Q2**-**S1** and the mimics thioether **67**, the triazole click products **68** and **69** were synthesised by P. M. Abeysinghe in our research group.

Commercially available chemicals and solvents (HPLC quality) were used without further purification, unless otherwise stated. All references to water refer to the use of Milli-Q water generated from a Millipore, Milli-Q Bicel A 10 system. The pH was measured on a Beckman Instruments Φ210 pH meter. The mobile phase consisted of eluents A (water containing 0.1% formic acid) and B (acetonitrile) for all LC-MS runs.

The LC-MS condition, solvent preparation, instrumentation and normalisation of the spectra were followed similar to Chapter 4 experimental. The peaks were reported as percentage proportion changes based on peaks areas normalised with respect to the internal standard 3,5-dihydroxybenzoic acid (as per Chapter 4) and also used as an internal standard for the analysis.

For the **Q1** mimics analysis, the total UV scan from 210-350 nm wavelength was selected, whereas for **Q2**-**S1** mimics analysis, due to the overlapping of the impurity presented in the parent disulfide **Q2**-**S1**, the wavelength at 252 nm was considered, where the relative absorbance of the peaks corresponds to low standard deviation.

5.5.2 Experimental Procedures

Stock Solutions

The oligonucleotides stock solutions and the Dynabeads were prepared as mentioned in Chapter 4 experimental.

(a) Q1-**S2** and **Q2**-**S1 Disulfide Mimics**

The disulfide mimics of **Q1**-**S2** and **Q2**-**S1** were not soluble in 25% aqueous methanol and hence the saturated solution was used and the concentrations of the solutions were determined by UV-Vis analysis. The saturated solutions were at the following concentrations: **Q1**-**S2** mimics thioether β-**65** (0.06 mM) and amide α-**66** (0.56 mM); **Q2**-**S1** mimics parent disulfide **Q2**-**S1** (2.0 mM), thioether **67** (1.8 mM), 1,4-click compound **68** (0.46 mM) and 1,5-click compound **69** (0.50 mM).

An equal volume of the internal standard for control (untemplated) and DNA unbound (templated) (25 and 17 µl, 2.5 mM), DNA-bound (14 and 10 µl, 0.5 mM) samples were added within each experimental series before LC-MS analysis. All stock solutions were stored in the freezer until required and were allowed to warm to room temperature before use.

(b) Mimics Assay Experiments

Two experiments were prepared with the following compositions:

Mimic Assay (a): Parent disulfide **Q2**-**S1** (2.0 mM, 15 µL, 30 nmol), thioether **67** (1.8 mM, 5 µL, 9.0 nmol), 1,4-click **68** (0.46 mM, 40 µL, 18.4 nmol) and 1,5-click **69** (0.50 mM, 50 µL, 25.0 nmol) were mixed to give a total volume of 110.0 µL at pH ~ 7.0.

Mimic Assay (b): Thioether β-**65** (0.06 mM, 20 µL, 1.2 nmol), amide α-**66** (0.56 mM, 5 µL, 2.8 nmol) were mixed to give a total volume of 25.0 µL at pH ~ 7.0.

A control experiment was conducted by diluting the aliquot of the **Q1**-**S2** mimics reaction mixture (25 µL, b) using aqueous TFA (50 µL, 0.1%) to pH ~ 3.0 and internal standard (2.5 mM, 25 µL) was added to the solution to give a total volume of 100 µL, followed by analysis by LC-MS. In the case of **Q2**-**S1** mimics (a), which had the original reaction mixture volume of 110 µL, the solvent was removed *in vacuo* (LABCONCO, CentriVap® Mobile system Model 7812011, 40 min at 58 °C) and the resultant residue was diluted with 0.1% TFA in water (83 µL, 0.1%) to pH ~ 3.0 and internal standard (2.5 mM, 17 µL)

was added to the solution to give a total volume of 100 μL, followed by analysis by LC-MS.

(c) Incubation of Reaction Mixtures

The quinoline **Q1-S2** and **Q2-S1** disulfide mimics reactions were carried out at physiological pH ~ 7.0. In templated studies, the reaction mixtures were incubated at 30-32 °C with the DNA oligos **O1**, **O2** and **O3** (1.0 mM, 9.0 μL) for 24 h. The solutions were agitated manually once during the incubation. The experiments were diluted with the addition of aqueous TFA (20-40 μL, 0.1%) to pH ~ 3.0, transferred to a prepared Dynabead sample and was incubated at room temperature for 40 min. The beads were separated from the solution with a magnet and the supernatant (U1) was transferred into an eppendorf tube.

The beads were washed with 0.1% TFA in water (5 × 100 μL) and the washings combined with the supernatant (U1). The solvent was removed *in vacuo* (LABCONCO, CentriVap® Mobile system Model 7812011, 80-90 min at 58 °C) and the resultant residue was diluted with 0.1% TFA in water (75 and 83 μL). To this, the internal standard 3,5-dihydroxybenzoic acid was added (2.5 mM, 25 and 17 μL), to give a total volume of 100 μL, and was analysed by LC-MS.

The DNA was denatured by incubating the Dynabeads with 0.1% TFA in water at 90 °C for 10 minutes (3 × 90 μL). The beads were separated with a magnet and the supernatant (DNA-bound, U2) from each denaturing cycle was combined. The solvent from the combined solutions were removed *in vacuo* and the resultant residue was diluted by 0.1% TFA in water (86 and 90 μL). To

this, the internal standard 3,5-dihydroxybenzoic acid was added (0.5 mM, 14 and 10 μL), to give a total volume of 100 μL, and the DNA-bound solution was analysed by LCMS.

(d) LC-MS Analysis

For each DCL analysis (control, U1 and U2) 40 μL was injected per run. The following elution gradients were employed:

Mimic Assay (a and b): 92% A for 5 min, then gradient was raised to reach 27% B at 12 min, then gradually increased to 35% B at 37 min, then gradient was raised to reach 100% B at 47 min and held at 100% B for 3 min, then ramped to reach 92% A at 53 min, then held at 92% A until 75 min.

The internal standard 3,5-dihydroxybenzoic acid was the largest peak in each spectrum and all peaks were normalised to it. Results are reported as percentage change in composition between the bound sample and the control library based on peaks areas of each component (not including the internal standard). The total UV san from 210-350 nm was considered for **Q1-S2** mimics assay reactions and for **Q2-S1** mimics UV at 252 nm was selected for all components.

Chapter 6

Conclusions

The results in this thesis report the first applications of DCC to identify novel carbohydrate derivatives of quinoline as new DNA-binding compounds. The results extend the initial DCC studies that have been reported to date with quadruplex DNA using peptides and derivatives of simple known DNA-binding compounds to the study of intercalators, sugars and amidines. The work has illustrated the potential and the limitations of DCC to identify intercalator-groove binding molecules that would not be predicted to interact with duplex DNA using the classic rules for DNA recognition.

A new method for the generation of disulfide DCLs under conditions that do not denature or react with DNA was developed using disulfide exchange under mildly basic conditions. This method is a significant improvement on the literature conditions that have used GSSH/GSH in DCC studies with duplex and quadruplex DNA for two reasons. First, Balasubramanian and coworkers[93] have highlighted the different results that can be obtained using disulfide DCLs generated using GSSG/GSH at pH 7.0 and the need to conduct multiple experiments with different concentrations of GSSG/GSH in order to correctly identify the amplified products. Second, while the formation of adducts of GSH in the DCL can provide useful lead compounds, these adducts can also complicate analysis of the results. These conditions should prove useful for future studies with duplex, triplex and quadruplex DNA, as they do not include GSSG/GSH.

Aqueous solubility was a major challenge in the DCL studies, especially with the naphthalimide derivatives **N1-N1** and **N1-Y** and the results suggests that DCC studies to identify new DNA intercalator derivatives must use heterocyclic derivatives that are protonated at physiological pH or incorporate

water soluble functional groups. The DCL studies identified two quinoline-carbohydrate derivatives **Q2**-**S1** and **Q1**-**S2** as new DNA-binding compounds. The initial selection of glucose to give **Q2**-**S1** at the 4-position was not predicted. Similarly, the features present in the semi-rigid **Q1**-**S2** that resulted in DNA-binding compared with the corresponding benzylic derivative **Q1**-**S3** were not readily rationalized using molecular visualisation, and this binding preference was not predicted. While visualisation was used to predict the DNA-binding modes of **Q2**-**S1** and **Q1**-**S2** these docking studies are not accurate enough to produce any realistic quantitative estimation of the DNA-binding affinities of the docked structures. Further detailed studies are required to confirm the proposed molecular level structures, but in the case of the quinolines this will be difficult given their relatively weak interactions with DNA.

The use of the flexible bisthiols **B1**, **B2** and **B3**, to form bisintercalators with quinolines **Q1** and **Q2**, was unsuccessful. A limitation of DCC in this field is the difficulty in designing small building blocks that incorporate the functional groups needed for DCC, are water soluble, and that are compatible with DNA. Typical small molecule DNA binders have MW < 500, and the functional groups that allow equilibration under reversible conditions, in this case the thiol/disulfide interconversion, places restrictions on the distances and spatial orientation of other functionality in the molecule. The development of new reversible reactions, that can be initiated and quenched under conditions compatible with nucleic acids, would expand the opportunities for DNA recognition using DCC, and may allow more precise positioning of functional groups to match the features of a given DNA sequence. Recent studies in our

group have focused on reversible aqueous metathesis reactions for potential applications in DCC as a step towards this goal.[250]

One area where DCC has significant potential is to provide information about the base sequence selectivity of compounds. While only three oligonucleotide sequences were screened in this study, the length of the DNA and base sequence can be changed, and could provide information similar to DNA-footprinting experiments.

Finally, all studies with duplex DNA to date have generated disulfides and imine metal complexes that are not biostable. The results with both the 1,4- and 1,5-triazoles **68** and **69** respectively showed that neither compound gave the same DNA-binding profile as the parent disulfide **Q2-S1**, that would characterize a mimic of the disulfide **Q2-S1**. The triazole rings are quite bulky compared to the disulfide bond, and hence the design of triazole mimics of disulfides would only be appropriate for DNA-binding disulfides identified by DCC in which the sulfur atoms are not involved in hydrogen bonding, or in providing the key conformation and orientation of the functional groups to achieve shape recognition. Surprisingly, the 1,4-triazole showed enhanced DNA-binding and the different binding profiles of the 1,4-triazole (**68**), 1,5-triazole (**69**), suggest that each compound is binding in a different way to DNA and may be interesting DNA binding molecules in their own right.

References

(1)   Blackburn, G. M.; Gait, M. J. *Nucleic Acids in Chemistry and Biology*; Oxford University Press: New York, **1990**.

(2)   Sinden, R. R. *DNA Structure and Function*; Academic Press,INC: London, **1994**.

(3)   Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

(4)   Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 964.

(5)   Wilkins, M. H. F.; Seeds, W. E.; Stokes, A. R.; Wilson, H. R. *Nature* **1953**, *172*, 759.

(6)   Crick, F. H. C. *Proc. Nat. Acad. Sci. USA* **1954**, *40*, 756.

(7)   Rich, A.; Watson, J. D. *Proc. Nat. Acad. Sci. USA* **1954**, *40*, 759.

(8)   Dugas, H. *Bioorganic Chemistry: A Chemical Approach to Enzyme Action*; 3rd ed.; Springer-Verlag New York, Inc.: New York, **1996**.

(9)   Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; *Wiley and Sons, Ltd:England*, **2000**.

(10)   Moreno, T.; Pous, J.; Subirana, J. A.; Campos, J. L. *Acta Crystallogr. Sec.D* **2010**, *66*, 251.

(11)   Rhee, S.; Han, Z.; Liu, K.; Miles, H. T.; Davies, D. R. *Biochemistry* **1999**, *38*, 16810.

(12)   Nguyen, B.; Neidle, S.; Wilson, W. D. *Acc. Chem. Res.* **2008**, *42*, 11.

(13)   Song, G.; Ren, J. *Chem. Commun.* **2010**, *46*, 7283.

(14)   Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Nature* **1980**, *287*, 755.

(15)   Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Principles of Biochemistry: With an Extended Discussion of Oxygen-Binding Proteins*; 2nd ed.; Worth Publishers: New York, **1993**.

(16)   Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *171*, 740.

(17)   Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. *Nature* **1953**, *171*, 738.

(18)   Mamajanov, I.; Engelhart, A. E.; Bean, H. D.; Hud, N. V. *Angew. Chem. Int. Ed.* **2010**, *49*, 6310.

(19)   Thurston, D. E.; Thompson, A. S. *Chemistry in Britain* **1990**, *26*, 767.

(20)   Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, **1984**.

(21)   Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.

(22)   Hunter, C. A. *J. Mol. Biol.* **1993**, *230*, 1025.

(23)   Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Chem. Soc. Perkin Trans. 2* **2001**, 651.

(24)   Privalov, P. L.; Dragan, A. I.; Crane-Robinson, C.; Breslauer, K. J.; Remeta, D. P.; Minetti, C. A. S. A. *J. Mol. Biol.* **2007**, *365*, 1.

(25)   Jantz, D.; Amann, B. T.; Gatto, G. J.; Berg, J. M. *Chem. Rev.* **2003**, *104*, 789.

(26)   Tateno, M.; Yamasaki, K.; Amano, N.; Kakinuma, J.; Koike, H.; Allen, M. D.; Suzuki, M. *Biopolymers* **1997**, *44*, 335.

(27) Fulle, S.; Gohlke, H. *J. Mol. Recognit.* **2010**, *23*, 220.

(28) Ye, X.-S.; Zhang, L.-H. *Curr. Med. Chem.* **2002**, *9*, 929.

(29) Jiang, Y.-L.; Liu, Z.-P. *Mini-Rev. Med. Chem.* **2010**, *10*, 726.

(30) Franceschin, M. *Eur. J. Org. Chem.* **2009**, *2009*, 2225.

(31) Palde, P. B.; Ofori, L. O.; Gareiss, P. C.; Lerea, J.; Miller, B. L. *J. Med. Chem.* **2010**, *53*, 6018.

(32) Wu, Y.; Brosh, R. M. *FEBS J.* **2010**, *277*, 3470.

(33) Kumar, C. V.; Asuncion, E. H. *J. Am. Chem. Soc.* **1993**, *115*, 8547.

(34) Tse, W. C.; Boger, D. L. *Chemistry & Biology* **2004**, *11*, 1607.

(35) Strekowski, L.; Wilson, B. *Mutation Research* **2007**, *623*, 3.

(36) Mitra, S. N.; Wahl, M. C.; Sundaralingam, M. *Acta Crystallogr. Sec. D* **1999**, *55*, 602.

(37) Dervan, P. B.; Poulin-Kerstien, A. T.; Fechter, E. J.; Edelson, B. S. *Top. Curr. Chem.* **2005**, *253*, 1.

(38) Hannon, M. J. *Chem. Soc. Rev.* **2007**, *36*, 280.

(39) Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291.

(40) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284.

(41) Wang, S.; Munde, M.; Wang, S.; Wilson, W. D. *Biochemistry* **2011**, *50*, 7674.

(42) Gallmeier, H.-C.; König, B. *Eur. J. Org. Chem.* **2003**, *2003*, 3473.

(43) Baraldi, P. G.; Bovero, A.; Fruttarolo, F.; Preti, D.; Tabrizi, M. A.; Pavani, M. G.; Romagnoli, R. *Medicinal Research Reviews* **2004**, *24*, 475.

(44) Suckling, C. J. *Expert Opin. Ther. Patents* **2004**, *14*, 1693.

(45) Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 463.

(46) Carrasco, C.; Helissey, P.; Haroun, M.; Baldeyrou, B.; Lansiaux, A.; Colson, P.; Houssier, C.; Giorgi-Renault, S.; Bailly, C. *ChemBioChem* **2003**, *4*, 50.

(47) Tanious, F. A.; Laine, W.; Peixoto, P.; Bailly, C.; Goodwin, K. D.; Lewis, M. A.; Long, E. C.; Georgiadis, M. M.; Tidwell, R. R.; Wilson, W. D. *Biochemistry* **2007**, *46*, 6944.

(48) Lerman, L. S. *J. Mol. Biol.* **1961**, *3*, 18.

(49) Miroshnychenko, K. V.; Shestopalova, A. V. *Int. J. Quantum Chem.* **2010**, *110*, 161.

(50) Zunino, F.; Gambetta, R.; Di, M. A.; Zaccara, A. *Biochem. Biophys. Acta, Nucleic Acids Protein Synth.* **1972**, *277*, 489.

(51) Chaires, J. B.; Dattagupta, N.; Crothers, D. M. *Biochemistry* **1982**, *21*, 3933.

(52) Barthwal, R.; Sharma, U.; Srivastava, N.; Jain, M.; Awasthi, P.; Kaur, M.; Barthwal, S. K.; Govil, G. *Eur. J. Med. Chem.* **2006**, *41*, 27.

(53) Jollès, B.; Laigle, A.; Priebe, W.; Garnier-Suillerot, A. *Chem. Biol. Interact.* **1996**, *100*, 165.

(54) Pratt, W. B.; Ruddon, R. W.; Ensminger, W. D.; Maybaum, J. *The Anticancer Drugs*; 2nd ed.; Oxford University Press, Inc:New York, **1994**.

(55) Dawson, S.; Malkinson, J. P.; Paumier, D.; Searcey, M. *Nat. Prod. Rep.* **2007**, *24*, 109.

(56) Robbel, L.; Hoyer, K. M.; Marahiel, M. A. *FEBS J.* **2009**, *276*, 1641.

(57) Zolova, O. E.; Mady, A. S. A.; Garneau-Tsodikova, S. *Biopolymers* **2010**, *93*, 777.

(58) Yang, P.; Yang, Q.; Qian, X. *Tetrahedron* **2005**, *61*, 11895.

(59) Leng, F.; Priebe, W.; Chaires, J. B. *Biochemistry* **1998**, *37*, 1743.

(60) Leng, F.; Chaires, J. B.; Waring, M. J. *Nucleic Acids Res.* **2003**, *31*, 6191.

(61) Waring, M. J.; Wakelin, L. P. G. *Nature* **1974**, *252*, 653.

(62) Sheldrick, G. M.; Guy, J. J.; Kennard, O.; Rivera, V.; Waring, M. J. *J. Chem. Soc., Perkin Trans. 2* **1984**, 1601.

(63) Takusagawa, F. *J. Antibiotics* **1985**, *38*, 1596.

(64) Ughetto, G.; Wang, A. H. J.; Quigley, G. J.; Marel, G. A.; Boom, J. H.; Rich, A. *Nucleic Acids Res.* **1985**, *13*, 2305.

(65) Doss, R. M.; Marques, M. A.; Foister, S.; Chenoweth, D. M.; Dervan, P. B. *J. Am. Chem. Soc.* **2006**, *128*, 9074.

(66) Fechter, E. J.; Olenyuk, B.; Dervan, P. B. *Angew. Chem. Int. Ed.* **2004**, *43*, 3591.

(67) Portugal, J.; Cashman, D. J.; Trent, J. O.; Ferrer-Miralles, N.; Przewloka, T.; Fokt, I.; Priebe, W.; Chaires, J. B. *J. Med. Chem.* **2005**, *48*, 8209.

(68) Dervan, P. B. *Science* **1986**, *232*, 464.

(69) Nguyen, B.; Lee, M. P. H.; Hamelberg, D.; Joubert, A.; Bailly, C.; Brun, R.; Neidle, S.; Wilson, W. D. *J. Am. Chem. Soc.* **2002**, *124*, 13680.

(70) Rahimian, M.; Kumar, A.; Say, M.; Bakunov, S. A.; Boykin, D. W.; Tidwell, R. R.; Wilson, W. D. *Biochemistry* **2009**, *48*, 1573.

(71) Lanteri, C. A.; Trumpower, B. L.; Tidwell, R. R.; Meshnick, S. R. *Antimicrob. Agents Chemother.* **2004**, *48*, 3968.

(72) Munde, M.; Ismail, M. A.; Arafa, R.; Peixoto, P.; Collar, C. J.; Liu, Y.; Hu, L.; David-Cordonnier, M.-H.; Lansiaux, A.; Bailly, C.; Boykin, D. W.; Wilson, W. D. *J. Am. Chem. Soc.* **2007**, *129*, 13732.

(73) Liu, Y.; Collar, C. J.; Kumar, A.; Stephens, C. E.; Boykin, D. W.; Wilson, W. D. *J. Phys. Chem. B* **2008**, *112*, 11809.

(74) Nagle, P. S.; Quinn, S. J.; Kelly, J. M.; O'Donovan, D. H.; Khan, A. R.; Rodriguez, F.; Nguyen, B.; Wilson, W. D.; Rozas, I. *Org. Biomol. Chem.* **2010**, *8*, 5558.

(75) Bailly, C.; Arafa, R. K.; Tanious, F. A.; Laine, W.; Tardy, C.; Lansiaux, A.; Colson, P.; Boykin, D. W.; Wilson, W. D. *Biochemistry* **2005**, *44*, 1941.

(76)   Lehn, J.-M. *Chem. Eur. J.* **1999**, *5*, 2455.

(77)   Eliseev, A. V.; Lehn, J.-M. *Com. Chem. Biol.* **1999**, *243*, 159.

(78)   Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652.

(79)   Ladame, S. *Org. Biomol. Chem.* **2008**, *6*, 219.

(80)   Corbett, P. T.; Otto, S.; Sanders, J. K. M. *Org. Lett.* **2004**, *6*, 1825.

(81)   Otto, S.; Severin, K. *Top. Curr. Chem.* **2007**, *277*, 267.

(82)   Albrecht, M. *Naturwissenschaften* **2007**, *94*, 951.

(83)   Karan, C.; Miller, B. L. *Drug Discov. Today* **2000**, *5*, 67.

(84)   *Dynamic Combinatorial Chemistry:History and Principles of DCC;Ligands for Biomolecules*; Reek, J. N. H.; Otto, S., Eds.; (wiley-vch:Weinheim), **2010**.

(85)   Hunt, R. A. R.; Otto, S. *Chem. Commun.* **2011**, *47*, 847.

(86)   *Dynamic Combinatorial Chemistry. In Drug Discovery, Bioorganic Chemistry, and Materials Science*; Miller, B. L., Ed.; (wiley-vch Verlag: Weinheim). **2010**.

(87)   Klekota, B.; Miller, B. L. *Trends Biotechnol.* **1999**, *17*, 205.

(88)   Ramstrom, O.; Lehn, J.-M. *Nature Drug Discov.* **2002**, *1*, 26.

(89)   Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 321.

(90)   Otto, S.; Severin, K. In *"Creative Chemical Sensor Systems"*; Schrader, T., Ed.; Springer Berlin / Heidelberg: **2007**; Vol. *277*, p 267.

(91)   Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Drug Discov. Today* **2002**, *7*, 117.

(92)   Whitney, A. M.; Ladame, S.; Balasubramanian, S. *Angew. Chem. Int. Ed.* **2004**, *43*, 1143.

(93)   Ladame, S.; Whitney, A. M.; Balasubramanian, S. *Angew. Chem. Int. Ed.* **2005**, *44*, 5736.

(94)   Tsujita, S.; Tanada, M.; Kataoka, T.; Sasaki, S. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 68.

(95)   Bugaut, A.; Jantos, K.; Wietor, J.-L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. *Angew. Chem. Int. Ed.* **2008**, *47*, 2677.

(96)   Nielsen, M. C.; Ulven, T. *Chem. Eur. J.* **2008**, *14*, 9487.

(97)   Giuseppone, N.; Lehn, J.-M. *Chem. Eur. J.* **2006**, *12*, 1715.

(98)   Hochggurtel, M.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Nicolau, C.; Krause, S.; Schaaf, O.; Sonnenmoser, G.; Eliseev, A. V. *Proc. Nat. Acad. Sci. USA* **2002**, *99*, 3382.

(99)   Reutenauer, P.; Boul, P. J.; Lehn, J.-M. *Eur. J. Org. Chem.* **2009**, *2009*, 1691.

(100)   Boul, P. J.; Reutenauer, P.; Lehn, J.-M. *Org. Lett.* **2004**, *7*, 15.

(101)   Cousins, G. R. L.; Poulsen, S.-A.; Sanders, J. K. M. *Chem. Commun.* **1999**, 1575.

(102)   Simpson, M. G.; Pittelkow, M.; Watson, S. P.; Sanders, J. K. M.

Org. Biomol. Chem. **2010**, *8*, 1173.

(103) Godin, G.; Levrand, B.; Trachsel, A.; Lehn, J.-M.; Herrmann, A. *Chem. Commun.* **2010**, *46*, 3125.

(104) Larsson, R.; Pei, Z.; Ramström, O. *Angew. Chem. Int. Ed.* **2004**, *43*, 3716.

(105) Leclaire, J.; Vial, L.; Otto, S.; Sanders, J. K. M. *Chem. Commun.* **2005**, 1959.

(106) Sarma, R. J.; Otto, S.; Nitschke, J. R. *Chem. Eur. J.* **2007**, *13*, 9542.

(107) Goral, V.; Nelen, M. I.; Eliseev, A. V.; Lehn, J.-M. *Proc. Nat. Acad. Sci. USA* **2001**, *98*, 1347.

(108) Caraballo, R.; Rahm, M.; Vongvilai, P.; Brinck, T.; Ramstrom, O. *Chem. Commun.* **2008**, 6603.

(109) Hutin, M.; Schultz, D.; Nitschke, J. R. *Chimia* **2008**, *62*, 198.

(110) Klekota, B.; Hammond, M. H.; Miller, B. L. *Tetrahedron Lett.* **1997**, *38*, 8639.

(111) Karan, C.; Miller, B. L. *J. Am. Chem. Soc.* **2001**, *123*, 7455.

(112) Chung, M.-K.; Hebling, C. M.; Jorgenson, J. W.; Severin, K.; Lee, S. J.; Gagné, M. R. *J. Am. Chem. Soc.* **2008**, *130*, 11819.

(113) Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Chem. Eur. J.* **2008**, *14*, 2153.

(114) Au-Yeung, H. Y.; Pengo, P.; Pantos, G. D.; Otto, S.; Sanders, J. K. M. *Chem. Commun.* **2009**, 419.

(115) McNaughton, B. R.; Gareiss, P. C.; Miller, B. L. *J. Am. Chem. Soc.* **2007**, *129*, 11306.

(116) Brisig, B.; Sanders, J. K. M.; Otto, S. *Angew. Chem. Int. Ed.* **2003**, *42*, 1270.

(117) Gasparini, G.; Prins, L. J.; Scrimin, P. *Angew. Chem. Int. Ed.* **2008**, *47*, 2475.

(118) Nitschke, J. R. *Acc. Chem. Res.* **2006**, *40*, 103.

(119) Au-Yeung, H. Y.; Dan Pantoş, G.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2009**, *131*, 16030.

(120) Au-Yeung, H. Y.; Pantoş, G. D.; Sanders, J. K. M. *Proc. Nat. Acad. Sci. USA* **2009**, *106*, 10466.

(121) Wong, W.-Y.; Leung, K. C.-F.; Stoddart, J. F. *Org. Biomol. Chem.* **2010**, *8*, 2332.

(122) Rodriguez-Docampo, Z.; Otto, S. *Chem. Commun.* **2008**, 5301.

(123) West, K. R.; Bake, K. D.; Otto, S. *Org. Lett.* **2005**, *7*, 2615.

(124) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2000**, *122*, 12063.

(125) Besenius, P.; Cormack, P. A. G.; Liu, J.; Otto, S.; Sanders, J. K.

M.; Sherrington, D. C. *Chem. Eur. J.* **2008**, *14*, 9006.

(126) Besenius, P.; Cormack, P. A. G.; Ludlow, R. F.; Otto, S.; Sherrington, D. C. *Chem. Commun.* **2008**, 2809.

(127) Besenius, P.; Cormack, P. A. G.; Ludlow, R. F.; Otto, S.; Sherrington, D. C. *Org. Biomol. Chem.* **2010**, *8*, 2414.

(128) Buryak, A.; Severin, K. *Angew. Chem. Int. Ed.* **2005**, *44*, 7935.

(129) Buryak, A.; Severin, K. *J. Comb. Chem.* **2006**, *8*, 540.

(130) Buryak, A.; Pozdnoukhov, A.; Severin, K. *Chem. Commun.* **2007**, 2366.

(131) Ruff, Y.; Lehn, J.-M. *Angew. Chem. Int. Ed.* **2008**, *47*, 3556.

(132) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Science* **2002**, *297*, 590.

(133) Klekota, B.; Miller, B. L. *Tetrahedron* **1999**, *55*, 11687.

(134) McNaughton, B. R.; Miller, B. L. *Org. Lett.* **2006**, *8*, 1803.

(135) Gareiss, P. C.; Sobczak, K.; McNaughton, B. R.; Palde, P. B.; Thornton, C. A.; Miller, B. L. *J. Am. Chem. Soc.* **2008**, *130*, 16254.

(136) Bugaut, A.; Toulmé, J.-J.; Rayner, B. *Angew. Chem. Int. Ed.* **2004**, *43*, 3144.

(137) Bugaut, A.; Bathany, K.; Schmitter, J.-M.; Rayner, B. *Tetrahedron Lett.* **2005**, *46*, 687.

(138) Bugaut, A.; Toulme, J.-J.; Rayner, B. *Org. Biomol. Chem.* **2006**, *4*, 4082.

(139) Portugal, J.; Waring, M. J. *Eur. J. Biochem.* **1987**, *167*, 281.

(140) Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1988**, *27*, 8088.

(141) Teixeira, S. C. M.; Thorpe, J. H.; Todd, A. K.; Powell, H. R.; Adams, A.; Wakelin, L. P. G.; Denny, W. A.; Cardin, C. J. *J. Mol. Biol.* **2002**, *323*, 167.

(142) Rajput, C.; Rutkaite, R.; Swanson, L.; Haq, I.; Thomas, J. A. *Chem. Eur. J.* **2006**, *12*, 4611.

(143) Jantos, K.; Rodriguez, R.; Ladame, S.; Shirude, P. S.; Balasubramanian, S. *J. Am. Chem. Soc.* **2006**, *128*, 13662.

(144) Gooch, B. D.; Beal, P. A. *J. Am. Chem. Soc.* **2004**, *126*, 10603.

(145) Azéma, L.; Bathany, K.; Rayner, B. *ChemBioChem.* **2010**, *11*, 2513.

(146) Bowler, F. R.; Diaz-Mochon, J. J.; Swift, M. D.; Bradley, M. *Angew. Chem. Int. Ed.* **2010**, *49*, 1809.

(147) Boger, D. L.; Chen, J.-H.; Saionz, K. W. *J. Am. Chem. Soc.* **1996**, *118*, 1629.

(148) Gamage, S. A.; Spicer, J. A.; Finlay, G. J.; Stewart, A. J.; Charlton, P.; Baguley, B. C.; Denny, W. A. *J. Med. Chem.* **2001**, *44*, 1407.

(149) Carrasco, C.; Joubert, A.; Tardy, C.; Maestre, N.; Cacho, M.; Brana, M. F.; Bailly, C. *Biochemistry* **2003**, *42*, 11751.

(150) Avendano, C.; Menendez, J. C. *Medicinal Chemistry of Anticancer Drugs*; First ed.; Elsevier, Amsterdam, The Netherlands **2008**.

(151) García-Pérez, M.; Pinto, M.; Subirana, J. A. *Biopolymers* **2003**, *69*, 432.

(152) Frigyes, D.; Alber, F.; Pongor, S.; Carloni, P. *J. Mol. Struct. (Theochem.)* **2001**, *574*, 39.

(153) Walker, S.; Valentine, K. G.; Kahne, D. *J. Am. Chem. Soc.* **1990**, *112*, 6428.

(154) Kahne, D. *Chem. Biol.* **1995**, *2*, 7.

(155) Brana, M. F.; Cacho, M.; Ramos, A.; Teresa Dominguez, M.; Pozuelo, J. M.; Abradelo, C.; Fernanda Rey-Stolle, M.; Yuste, M.; Carrasco, C.; Bailly, C. *Org. Biomol. Chem.* **2003**, *1*, 648.

(156) Antonini, I.; Santoni, G.; Lucciarini, R.; Amantini, C.; Sparapani, S.; Magnano, A. *J. Med. Chem.* **2006**, *49*, 7198.

(157) Kaslow, C. E.; Clark, W. R. *J. Org. Chem.* **1953**, *18*, 55.

(158) Jones, G. *"Heterocyclic Compounds"*; Interscience, **1982**; Vol. *32*.

(159) Mahiou, B.; Gleicher, G. J. *J. Org. Chem.* **1990**, *55*, 4466.

(160) Ferles, M.; Kocian, O. *Collect. Czech. Chem. Commun.* **1981**, *46*, 1518.

(161) Brown, B. R.; Hammick, D. L.; Thewlis, B. H. *J. Chem. Soc.* **1951**, 1145.

(162) Doucet-Personeni, C.; Bentley, P. D.; Fletcher, R. J.; Kinkaid, A.; Kryger, G.; Pirard, B.; Taylor, A.; Taylor, R.; Taylor, J.; Viner, R.; Silman, I.; Sussman, J. L.; Greenblatt, H. M.; Lewis, T. *J. Med. Chem.* **2001**, *44*, 3203.

(163) Kaslow, C. E.; Schlatter, J. M. *J. Am. Chem. Soc.* **1955**, *77*, 1054.

(164) Carter, T. G.; Healey, E. R.; Pitt, M. A.; Johnson, D. W. *Inorg. Chem.* **2005**, *44*, 9634.

(165) Turner, R. A.; Huebner, C. F.; Scholz, C. R. *J. Am. Chem. Soc.* **1949**, *71*, 2801.

(166) Street, J. P.; Skorey, K. I.; Brown, R. S.; Ball, R. G. *J. Am. Chem. Soc.* **1985**, *107*, 7669.

(167) Kaptein, B.; Barf, G.; Kellogg, R. M.; Van Bolhuis, F. *J. Org. Chem.* **1990**, *55*, 1890.

(168) Matziari, M.; Bauer, K.; Dive, V.; Yiotakis, A. *J. Org. Chem.* **2008**, *73*, 8591.

(169) Zheng, T.-C.; Burkart, M.; Richardson, D. E. *Tetrahedron Lett.* **1999**, *40*, 603.

(170) Jacobson, K. A.; Fischer, B.; Ji, X.-d. *Bioconjugate Chem.* **1995**, *6*, 255.

(171) Hernandez-Folgado, L.; Baretić, D.; Piantanida, I.; Marjanović, M.; Kralj, M.; Rehm, T.; Schmuck, C. *Chem. Eur. J.* **2010**, *16*, 3036.

(172) Schmuck, C.; Geiger, L. *J. Am. Chem. Soc.* **2005**, *127*, 10486.

(173) Krishnamurthy, S.; Aimino, D. *J. Org. Chem.* **1989**, *54*, 4458.

(174) DeCollo, T. V.; Lees, W. J. *J. Org. Chem.* **2001**, *66*, 4244.

(175) Miyazaki, K. *Tetrahedron Lett.* **1968**, *9*, 2793.

(176) Kaji, A.; Araki, Y.; Miyazaki, K. *Bull. Chem. Soc. Jpn.* **1971**, *44*,

1393.

(177) Newman, M. S.; Karnes, H. A. *J. Org. Chem.* **1966**, *31*, 3980.

(178) Rogana, E.; Nelson, D. L.; Mares-Guia, M. *J. Am. Chem. Soc.* **1975**, *97*, 6844.

(179) Ashley, J. N.; Barber, H. J.; Ewins, A. J.; Newbery, G.; Self, A. D. H. *J. Chem. Soc.* **1942**, 103.

(180) Bauer, L.; Cymerman, J. *J. Chem. Soc.* **1950**, 109.

(181) Cymerman, J.; Willis, J. B. *J. Chem. Soc.* **1951**, 1332.

(182) Tanizawa, K.; Kanaoka, Y. *J. Biochem.* **1985**, *97*, 275.

(183) Lee, Y.; Mo, H.; Koo, H.; Park, J.-Y.; Cho, M. Y.; Jin, G.-w.; Park, J.-S. *Bioconjugate Chem.* **2006**, *18*, 13.

(184) Pollock, J. R. A.; Stevens, R.; Eyre and Spottiswoode, **1965**; Vol. *5*.

(185) Vaitilingam, B.; Nayyar, A.; Palde, P. B.; Monga, V.; Jain, R.; Kaur, S.; Singh, P. P. *Bioorg. Med. Chem.* **2004**, *12*, 4179.

(186) Ganesan, A. *Angew. Chem. Int. Ed.* **1998**, *37*, 2828.

(187) Lehn, J.-M.; Eliseev, A. V. *Science* **2001**, *291*, 2331.

(188) Rall, T. W.; Lehninger, A. L. *J. Biol. Chem.* **1952**, *194*, 119.

(189) Millis, K. K.; Weaver, K. H.; Rabenstein, D. L. *J. Org. Chem.* **1993**, *58*, 4144.

(190) Fernandes, P. A.; Ramos, M. J. *Chem. Eur. J.* **2004**, *10*, 257.

(191) Schafer, F. Q.; Buettner, G. R. *Free Radical Biol. Med.* **2001**, *30*, 1191.

(192) Cleland, W. W. *Biochemistry* **1964**, *3*, 480.

(193) Lamoureux, G. V.; Whitesides, G. M. *J. Org. Chem.* **1993**, *58*, 633.

(194) Whitesides, G. M.; Lilburn, J. E.; Szajewski, R. P. *J. Org. Chem.* **1977**, *42*, 332.

(195) Xuereb, H.; Maletic, M.; Gildersleeve, J.; Pelczer, I.; Kahne, D. *J. Am. Chem. Soc.* **2000**, *122*, 1883.

(196) Tevis, D. S.; Kumar, A.; Stephens, C. E.; Boykin, D. W.; Wilson, W. D. *Nucleic Acids Res.* **2009**, *37*, 5550.

(197) Hunt, R. A.; Munde, M.; Kumar, A.; Ismail, M. A.; Farahat, A. A.; Arafa, R. K.; Say, M.; Batista-Parra, A.; Tevis, D.; Boykin, D. W.; Wilson, W. D. *Nucleic Acids Res.* **2011**, *39*, 4265.

(198) Chenoweth, D. M.; Dervan, P. B. *Proc. Nat. Acad. Sci. USA* **2009**, *106*, 13175.

(199) Chenoweth, D. M.; Dervan, P. B. *J. Am. Chem. Soc.* **2010**, *132*, 14521.

(200) Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688.

(201) Lown, J. W. *J. Mol. Recognit.* **1994**, *7*, 79.

(202) Pindur, U.; Jansen, M.; Lemster, T. *Curr. Med. Chem.* **2005**, *12*, 2805.

(203) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Nat. Acad. Sci. USA* **1976**, *73*, 804.

(204) Zeglis, B. M.; Pierre, V. C.; Barton, J. K. *Chem. Commun.* **2007**,

4565.

(205) Liu, H.-K.; Sadler, P. J. *Acc. Chem. Res.* **2011**, *44*, 349.

(206) Erkkila, K. E.; Odom, D. T.; Barton, J. K. *Chem. Rev.* **1999**, *99*, 2777.

(207) Kielkopf, C. L.; Erkkila, K. E.; Hudson, B. P.; Barton, J. K.; Rees, D. C. *Nat. Struct. Mol. Biol* **2000**, *7*, 117.

(208) Munde, M.; Lee, M.; Neidle, S.; Arafa, R.; Boykin, D. W.; Liu, Y.; Bailly, C.; Wilson, W. D. *J. Am. Chem. Soc.* **2007**, *129*, 5688.

(209) Athri, P.; Wilson, W. D. *J. Am. Chem. Soc.* **2009**, *131*, 7618.

(210) González-Bulnes, L.; Gallego, J. *J. Am. Chem. Soc.* **2009**, *131*, 7781.

(211) Baraldi, P. G.; Preti, D.; Fruttarolo, F.; Tabrizi, M. A.; Romagnoli, R. *Bioorg. Med. Chem.* **2007**, *15*, 17.

(212) Nguyen, R.; Huc, I. *Angew. Chem. Int. Ed.* **2001**, *40*, 1774.

(213) Huc, I.; Lehn, J.-M. *Proc. Nat. Acad. Sci. USA* **1997**, *94*, 2106.

(214) Capranico, G.; Zunino, F.; Kohn, K. W.; Pommier, Y. *Biochemistry* **1990**, *29*, 562.

(215) Morris, C. R.; Andrew, L. V.; Whichard, L. P.; Holbrook, D. J. *Mol. Pharmacol.* **1970**, *6*, 240.

(216) Grewal, R. S. *Bull World Health Organ.* **1981**, *59*, 397.

(217) Kgokong, J. L.; Wachira, J. M. *Eur. J. Pharm. Sci.* **2001**, *12*, 369.

(218) Cheng, J.; Zeidan, R.; Mishra, S.; Liu, A.; Pun, S. H.; Kulkarni, R. P.; Jensen, G. S.; Bellocq, N. C.; Davis, M. E. *J. Med. Chem.* **2006**, *49*, 6522.

(219) O'Hare, C. C.; Uthe, P.; Mackay, H.; Blackmon, K.; Jones, J.; Brown, T.; Nguyen, B.; Wilson, W. D.; Lee, M.; Hartley, J. A. *Biochemistry* **2007**, *46*, 11661.

(220) Atwell, G. J.; Baguley, B. C.; Denny, W. A. *J. Med. Chem.* **1989**, *32*, 396.

(221) Lutz, R. E.; Patel, A. R.; Ohnmacht, C. J.; Clifford, D. P.; Crosby, A. S. *J. Med. Chem.* **1971**, *14*, 198.

(222) Lutz, R. E.; Ohnmacht, C. J.; Patel, A. R. *J. Med. Chem.* **1971**, *14*, 926.

(223) Cheng, C. C. *J. Pharm. Sci.* **1971**, *60*, 1596.

(224) Allison, J. L.; O'Brien, R. L.; Hahn, F. E. *Science* **1965**, *149*, 1111.

(225) Cohen, S. N.; Yielding, K. L. *J. Biol. Chem.* **1965**, *240*, 3123.

(226) McFadyen, W. D.; Sotirellis, N.; Denny, W. A.; Wakelin, L. P. G. *Biochim. et Biophys. Acta-Gene Expression* **1990**, *1048*, 50.

(227) Szumilak, M.; Szulawska-Mroczek, A.; Koprowska, K.; Stasiak, M.; Lewgowd, W.; Stanczak, A.; Czyz, M. *Eur. J. Med. Chem.* **2010**, *45*, 5744.

(228) Garner, J.; Sherman Durai, C. R.; Harding, M. M. *Unpublished results*.

(229) Li, W.; Zhang, Z.-W.; Wang, S.-X.; Ren, S.-M.; Jiang, T. *Chem. Biol. Drug Des.* **2009**, *74*, 80.

(230) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.

(231) Gallego, J.; Reid, B. R. *Biochemistry* **1999**, *38*, 15104.

(232) Robinson, H.; Priebe, W.; Chaires, J. B.; Wang, A. H. J. *Biochemistry* **1997**, *36*, 8663.

(233) van Dijk, M.; Bonvin, A. M. J. J. *Nucleic Acids Res.* **2009**, *37*, W235.

(234) Holland-Nell, K.; Meldal, M. *Angew. Chem. Int. Ed.* **2011**, *50*, 5204.

(235) Gilbert, H. F. *Methods Enzymol.* **1995**, *251*, 8.

(236) Holmgren, A.; Bjornstedt, M. *Methods Enzymol.* **1995**, *252*, 199.

(237) Muttenthaler, M.; Andersson, A.; de Araujo, A. D.; Dekan, Z.; Lewis, R. J.; Alewood, P. F. *J. Med. Chem.* **2010**, *53*, 8585.

(238) Dekan, Z.; Vetter, I.; Daly, N. L.; Craik, D. J.; Lewis, R. J.; Alewood, P. F. *J. Am. Chem. Soc.* **2011**, *133*, 15866.

(239) Hargittai, B.; Solé, N. A.; Groebe, D. R.; Abramson, S. N.; Barany, G. *J. Med. Chem.* **2000**, *43*, 4787.

(240) Smith, C. W.; Walter, R.; Moore, S.; Makofske, R. C.; Meienhofer, J. *J. Med. Chem.* **1978**, *21*, 117.

(241) Bondebjerg, J.; Grunnet, M.; Jespersen, T.; Meldal, M. *ChemBioChem* **2003**, *4*, 186.

(242) MacRaild, C. A.; Illesinghe, J.; van Lierop, B. J.; Townsend, A. L.; Chebib, M.; Livett, B. G.; Robinson, A. J.; Norton, R. S. *J. Med. Chem.* **2009**, *52*, 755.

(243) Stymiest, J. L.; Mitchell, B. F.; Wong, S.; Vederas, J. C. *Org. Lett.* **2002**, *5*, 47.

(244) Hossain, M. A.; Rosengren, K. J.; Zhang, S.; Bathgate, R. A. D.; Tregear, G. W.; van Lierop, B. J.; Robinson, A. J.; Wade, J. D. *Org. Biomol. Chem.* **2009**, *7*, 1547.

(245) Walewska, A.; Zhang, M.-M.; Skalicky, J. J.; Yoshikami, D.; Olivera, B. M.; Bulaj, G. *Angew. Chem. Int. Ed.* **2009**, *48*, 2221.

(246) Empting, M.; Avrutina, O.; Meusinger, R.; Fabritz, S.; Reinwarth, M.; Biesalski, M.; Voigt, S.; Buntkowsky, G.; Kolmar, H. *Angew. Chem. Int. Ed.* **2011**, *50*, 5207.

(247) Garner, J.; Harding, M. M. *Org. Biomol. Chem.* **2007**, *5*, 3577.

(248) Addess, K. J.; Feigon, J. *Biochemistry* **1994**, *33*, 12386.

(249) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. *J. Chem. Soc. Perkin Trans. 2* **1987**, S1.

(250) Hunter, L.; Condie, G. C.; Harding, M. M. *Tetrahedron Lett.* **2010**, *51*, 5064.

Appendix
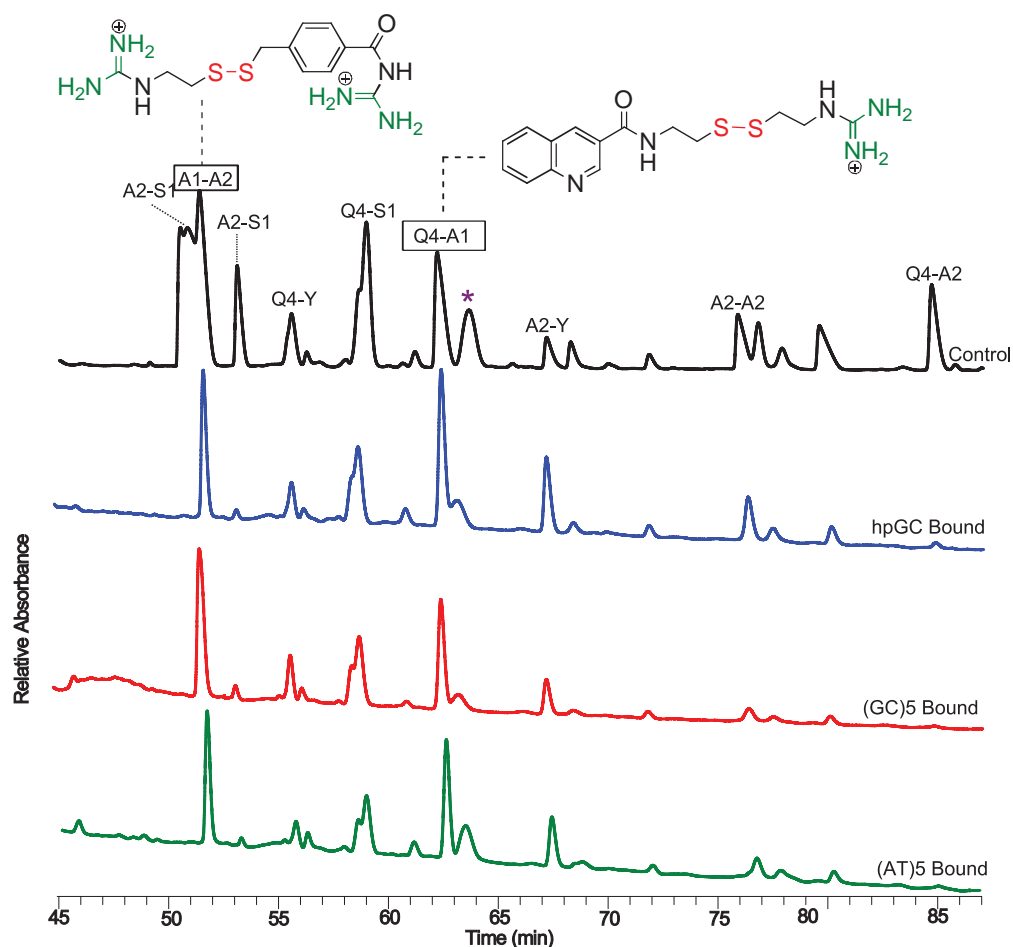
Figure 1 DCL generated from the thiol **S1** and disulfides **Q4**-**Y**, **A1**-**A1** and **A2**-**A2** (Figure 4.5, page 118) LC trace with UV detection of (210-350 nm) for the control DCL and the DNA-bound spectra of hpGC **O1**, (GC)$_5$ **O2** and (AT)$_5$ **O3** oligo sequences highlighting the structures of the DNA-bound compounds

(Note: Benzoic acid (*) was used as an internal standard, however, the internal standard overlapped with the DCL peaks and also peak broadening obtained during the LC-MS experiment and analysis).

193
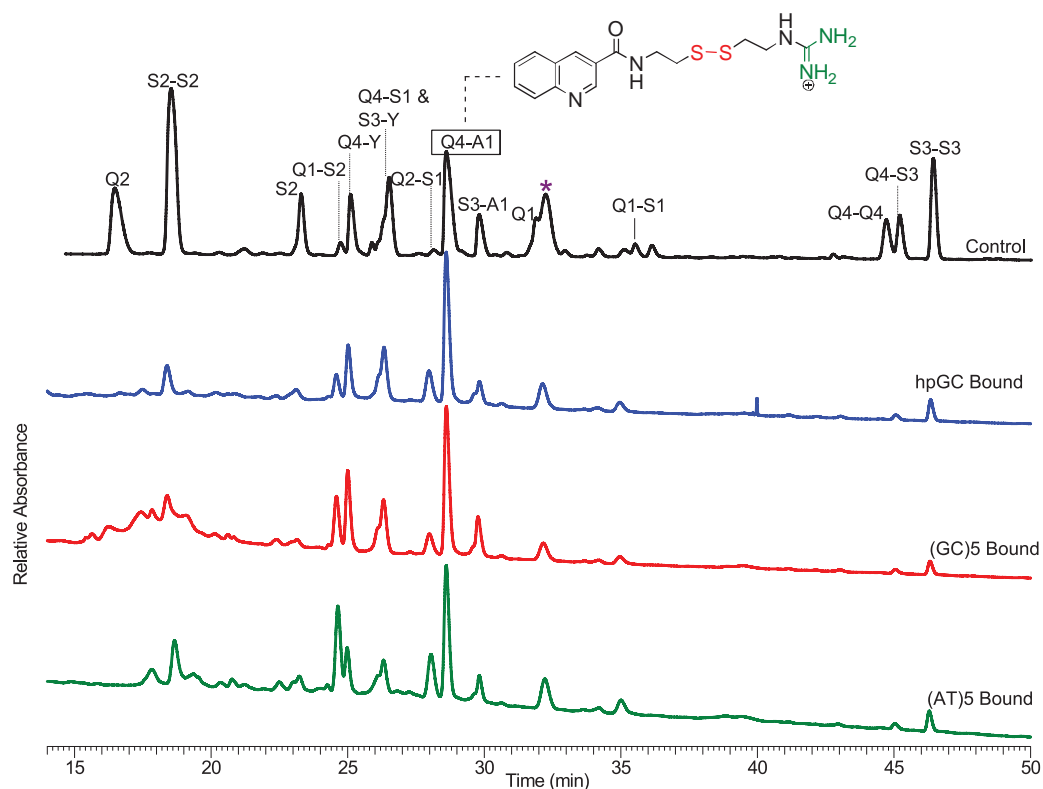
Figure 2 DCL generated from quinolines **Q1**, **Q2** and **Q4-Y**, and thiosugars **S1**, **S2** and **S3-S3** and amidine **A1-A1** (page 129) LC trace with UV detection of (210-350 nm) for the control DCL and the DNA-bound spectra of hpGC **O1**, (GC)₅ **O2** and (AT)₅ **O3** oligo sequences highlighting the structures of the DNA-bound compound

(Note: Benzoic acid (*) was used as an internal standard, however, the internal standard overlapped with the DCL peaks and also peak broadening obtained during the LC-MS experiment and analysis).
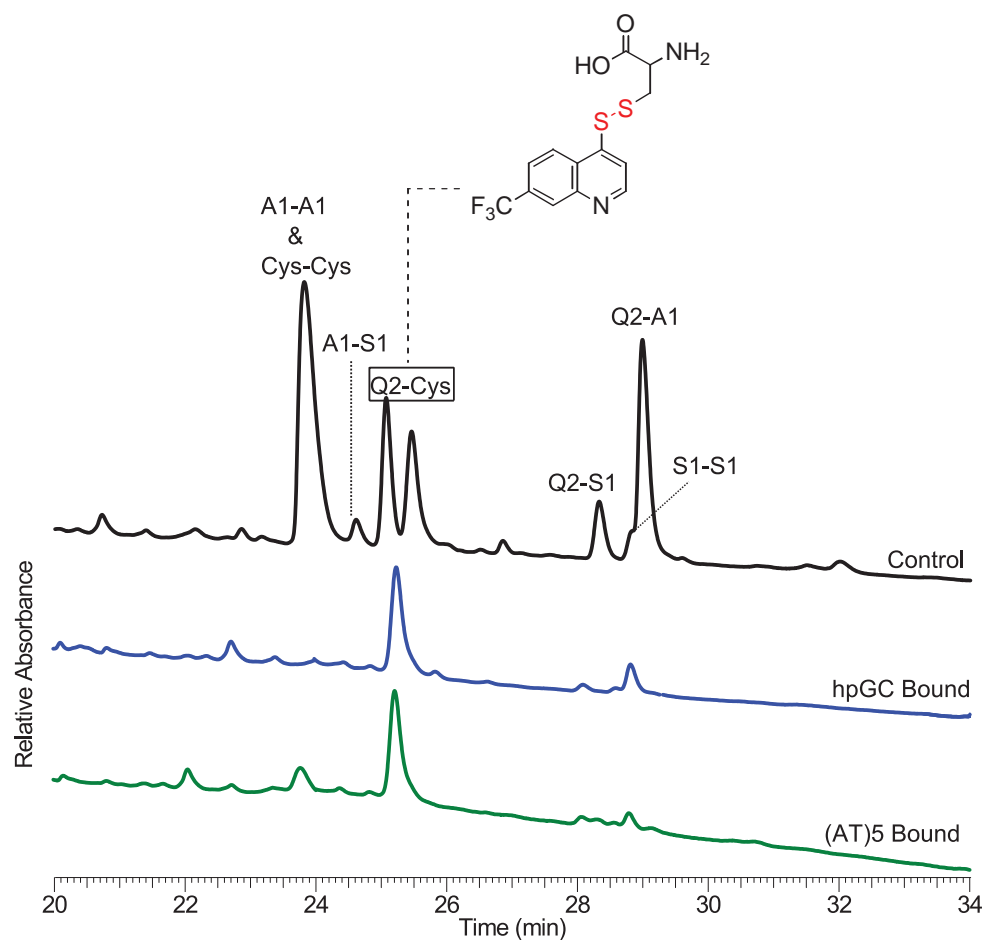
Figure 3 DCL generated from the aqueous soluble building blocks, **A1**-**A1** (charged), cysteine (**cys**, zwitterionic) and **S1** (neutral) with the quinoline **Q2** (page 132), LC trace with UV detection of (210-350 nm) for the control DCL and the DNA-bound spectra of hpGC **O1** and (AT)$_5$ **O3** oligo sequences highlighting the structure of the DNA-bound compound
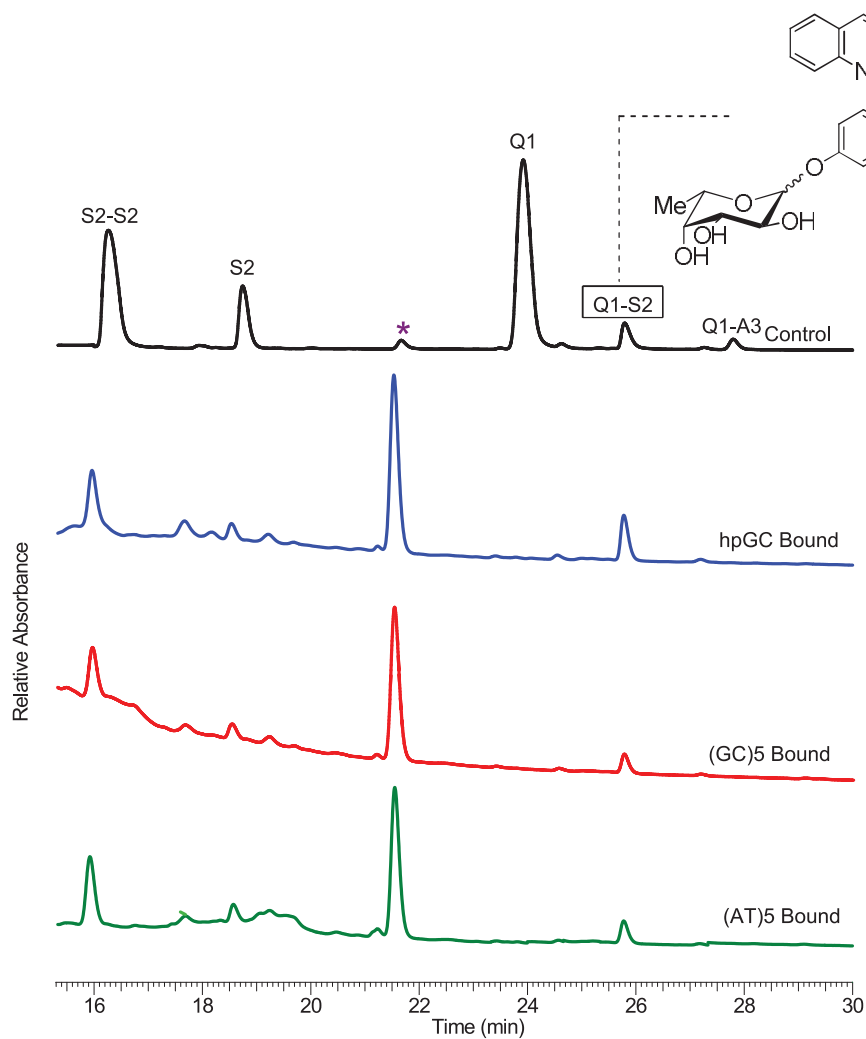
Figure 4 DCL generated from the aqueous soluble building blocks, **A1**-**A1** (charged) and **S2** (neutral) with the quinoline **Q1** (page 132), LC trace with UV detection of (210-350 nm) for the control DCL and the DNA-bound spectra of hpGC **O1,** (GC)$_5$ **O2** and (AT)$_5$ **O3** oligo sequences highlighting the structure of the DNA-bound compound and unidentified (*) compound.
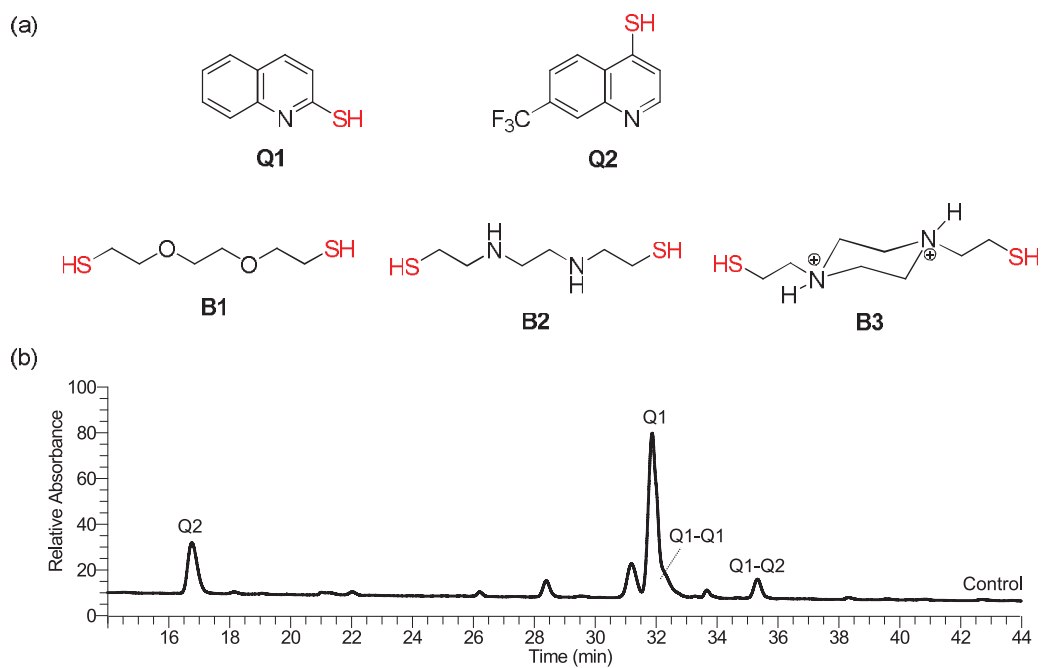
Figure 5 Result of the DCL generated from (a) quinolines **Q1** and **Q2**, and bisthiols **B1**, **B2** and **B3** (page 136) and (b) LC trace with UV detection of (210-350 nm) of the control DCL.