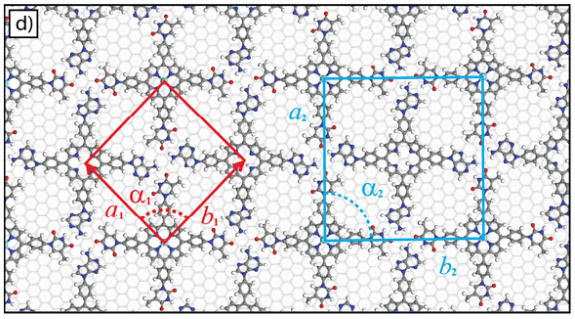
Supramolecular Self-Assembly of DNA and RNA Base Pair Functionalised Porphyrins



Charles W Toft
Prof. Neil Champness

Contents

Acknowledgements

I would firstly like to thank my supervisor, Neil Champness, for first providing me with the opportunity to undertake a summer research project as an undergraduate student and introducing me to the world of supramolecular chemistry and then accepting me to do a PhD in my own area of interest. I am also grateful for his continuous support and advice throughout the project. The group has always extremely welcoming and enjoyable to be a part of and whilst there have been too many to name over the years I wish to give special thanks to a few group members for their contributions, friendship and advice: Nic Pearce, Ben LLewellyn, Tim Hames, Harry Mackilligin, Sarah Haddow, Sam Quinn, Harry Martin, Phil Langer, Hawazen Hassanain and Arjun Patel. I greatly acknowledge the help of a number of people for their contributions to this work: Thank you to Sam Gilson for being an excellent project student and helping with the synthesis of cytosine-functionalised porphyrins. Thank you to Connie Pfeiffer and Will Lewis for their invaluable help with crystallography.Thank you to Matthew Blunt from University College London for running the surface images and for writing our recently submitted paper. I am very grateful to Anna Slater for her foundation work including the synthesis of tetra-thymine porphyrin. Finally, I am grateful to Lisa Hampson for offering advice and support on my work and helping me to improve successive drafts of my writing.

Abstract

The area of supramolecular self-assembly has received significant interest in recent years and a key area in the advancement of this field is the development of supramolecular tectons which are selective and directional in their intermolecular interactions. The four DNA nucleobases; adenine, thymine, guanine and cytosine all interact via complementary, selective and directional hydrogen bonds and have been used extensively in the design of supramolecular systems. These hydrogen bonding interactions and the effect of any substituents are studied further in both the solid and solution phases using an array of functionalised DNA nucleobases. Porphyrins substituted in the meso position by select nucleobases are synthesised and progress has been made towards the synthesis of porphyrins bearing multiple different nucleobases. This process involves either a modified Lindsey porphyrin synthesis route or a series of stoichiometric functionalisations via selective cross-coupling reactions. Finally, the self-assembly of these DNA-functionalised porphyrins is studied. Single-component supramolecular porous networks of Zn-tetra-thymineporphyrin and tetra-adenineporphyrinporphyrin are imaged and hydrogen bonding interactions are found between the molecules and modelled appropriately. These thymine and adenine-functionalised porphyrins are mixed in a 1:1 ratio and a chess-board hyrogen bonding network was formed, based on the complementary A-T hydrogen bonding interaction. This result provides promise that selective supramolecular systems can be studied with guanine and cytosine-functionalised porphryins.

Chapter 1

Introduction

1.1 Supramolecular Chemistry and Self-Assembly

Supramolecular chemistry can be defined as chemistry beyond the molecule.[9][10] That being, it is more than simply the study of the molecule itself and explores the intermolecular interactions between molecules and how that can impact the nanoscale properties of the material. Whereas traditional chemistry primarily focuses on the intramolecular interactions between atoms, such as covalent or organic bonds, supramolecular chemistry examines the intermolecular interactions.[11][12] These include electrostatic forces, hydrogen bonding, $\pi - \pi$ interactions, Van der Waals forces, hydrophobic effects, the mechanical bond and coordination bonds. These interactions and the phenomenon known as cooperativity combine to allow the molecules in supramolecular systems to self-assemble.[12] A key property of these intermolecular interactions is their dynamic nature.[13][14] In contrast to a conventional irreversible formation of a covalent bond the above intermolecular interactions are dynamic. This means that, once formed, they are easily broken and can form again. This low barrier to the formation and breaking of intermolecular interactions means that, under the correct conditions, the molecules can constantly rearrange until the thermodynamic product has formed. This process can be described as the system 'error checking' until a stable system has formed.

Despite the theory of Van der Waals forces being founded in 1873, supramolecular chemistry is a relatively modern area of chemistry and has grown into several key areas of interest since its beginnings as a research topic in 1990.[15][9] There are many fields that are based on the concept of supramolecular chemistry such as molecular self-assembly, molecular recognition, host-guest chemistry and the synthesis of mechanically interlocked molecules.

1.2 Intermolecular Interactions and Cooperativity

The fundamental interactions that are used to design, build and analyse supramolecular systems are intermolecular forces.[12] The intermolecular interactions utilised in supramolecular chemistry include electrostatic forces, hydrogen bonding, $\pi - \pi$ interactions, Van der Waals forces, hydrophobic effects, the mechanical bond and coordination bonds. Each of these interactions has different properties that are the result of different fundamental chemical principles. These different properties can be utilised to allow for a large variation of supramolecular systems to be studied.

1.2.1 Van der Waals Forces

Van der Waals forces are attractive or repulsive intermolecular forces between an electrostatic charge.[16] This electrostatic charge can be due to either a temporary or permanent dipole within

the molecule where a positive charge is attracted to a negative charge. These interactions are relatively weak compared to other intermolecular forces and the strength of the interaction depends on a number of factors including the strength of the dipole, the number of electrons within the molecule and the size and shape of the molecule.[17] Whilst being present in all supramolecular systems Van der Waals are rarely the driving force for the formation of a highly-ordered supramolecular system. This is due to the relatively low strength of Van der Waals forces as well as their lack of directionality and specificity.

1.2.2 $\pi - \pi$ Interactions

π interactions are attractive non-covalent interactions involving the electron-rich π system. Like in Van der Waals forces, a positive charge is attracted to a negative charge. Common π interactions include CH-π interactions, cation-π interactions and $\pi-\pi$ interactions.[18] Cation-π interactions have famously been used by Stoddart *et al* in their synthesis of [2]catenane where the interaction between a pyridinium cation and the electron-rich benzene group is the driving force for its formation. [19] This allowed the position of the interlocked rings to be controlled in relation to each other. Cation-π interactions can be controlled by varying the size of the cation,[20][21] solvation effects,[20][22] and substituents on the π system.[23] Therefore, they are used extensively in supramolecular chemistry as it is possible to tune the strength and geometry of the cation-π interaction.

Another well-known π interaction is the $\pi - \pi$ interaction which is the interaction between two π systems, commonly known as $\pi - \pi$ stacking. These interactions are regularly utilised for the study of planar, π-system-containing species on a surface. Further advancements of this well-known interaction have led to the design and synthesis of a "Buckycatcher" where $\pi - \pi$ interactions between buckminsterfullerene and two corannulene "pincers" have been used to strongly associate C_{60} as shown in Figure 1.1.[24]

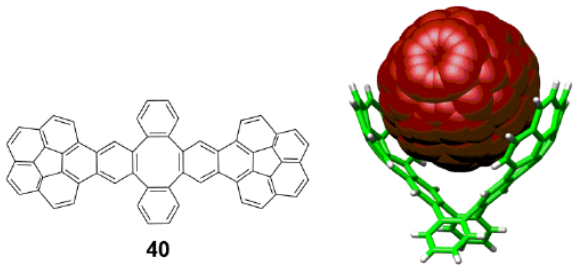


Figure 1.1: The capture of C_{60} with a "Buckycatcher" utilising pi-pi interactions.[24]

Importantly for the scope of this work; pi-pi stacking interactions are also a major interaction responsible for the stacking of nucleobases within DNA and RNA molecules. The DNA nucleobases are planar, aromatic compounds and are therefore suitable for $\pi - \pi$ stacking interactions.[25]

$\pi - \pi$ stacking can occur in a number of different geometries as shown in Figure 1.2. The distance of $\pi - \pi$ interactions is well-defined and this has been extensively studied and observed computationally and crystallographically.[18][26][27]

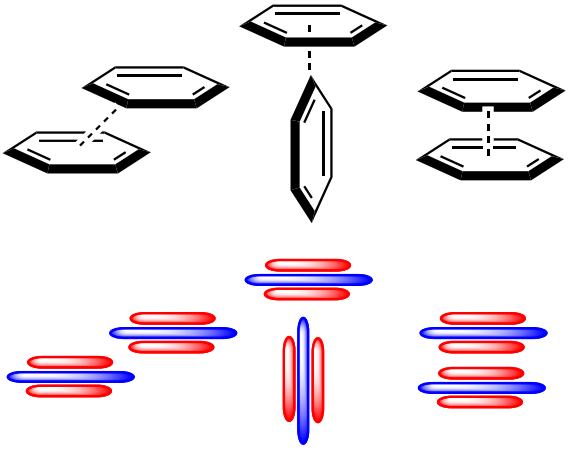


Figure 1.2: The possible geometries of $\pi - \pi$ stacking interactions showing the attractive electrostatic interactions. Modified from [27].

1.2.3 Hydrogen Bonds

Hydrogen bonds are a highly relevant intermolecular interaction in supramolecular self-assembly and have been widely used due their well-defined directional nature. The IUPAC definition defines a hydrogen bond as follows:

"The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment XH in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation."[28]

For example, in water, where X-H = O-H the oxygen atom is more electronegative than hydrogen and the hydrogen can form a bond to another oxygen atom in a different water molecule as shown in Figure 1.3.

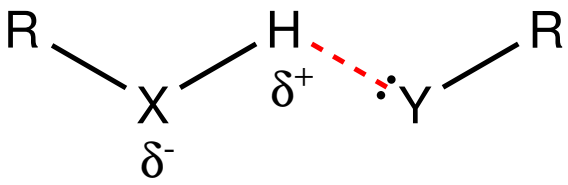


Figure 1.3: Diagram showing the hydrogen bonding interaction between X-H–Y. Hydrogen bond shown by dashed line.

In the above figure the hydrogen attached to the relatively electronegative element X is known as the hydrogen bond *donor*. The atom which forms a hydrogen bond to the *donor* is known as an *acceptor*.

In a hydrogen bond the electronegative atom X draws the electron density from the *donor* H to form a partial positive charge on the H atom. Due to the relatively small size of H atoms this forms a large positive charge density. This positive charge then attracts the lone pair of a hydrogen bond *acceptor* and forms a hydrogen bond. This effect is more pronounced when X is more electronegative as this creates a larger positive charge density on the *donor* H. Therefore, common examples of the

electronegative element in X-H are oxygen, nitrogen and fluorine. Hydrogen bonding and its effects are observed in many everyday phenomena and unquestionably the most well-known of all these phenomena is the relatively high boiling point of water. Water is capable of displaying hydrogen bonding in its simplest form as each molecule is made up entirely of hydrogen bond donors and acceptors. This allows each water molecule to form hydrogen bonds to four adjacent molecules (two as a *donor* and two as an *acceptor*) which creates a hydrogen bonding structure and greatly raises the energy required to melt or boil water as shown in Figure 1.4.

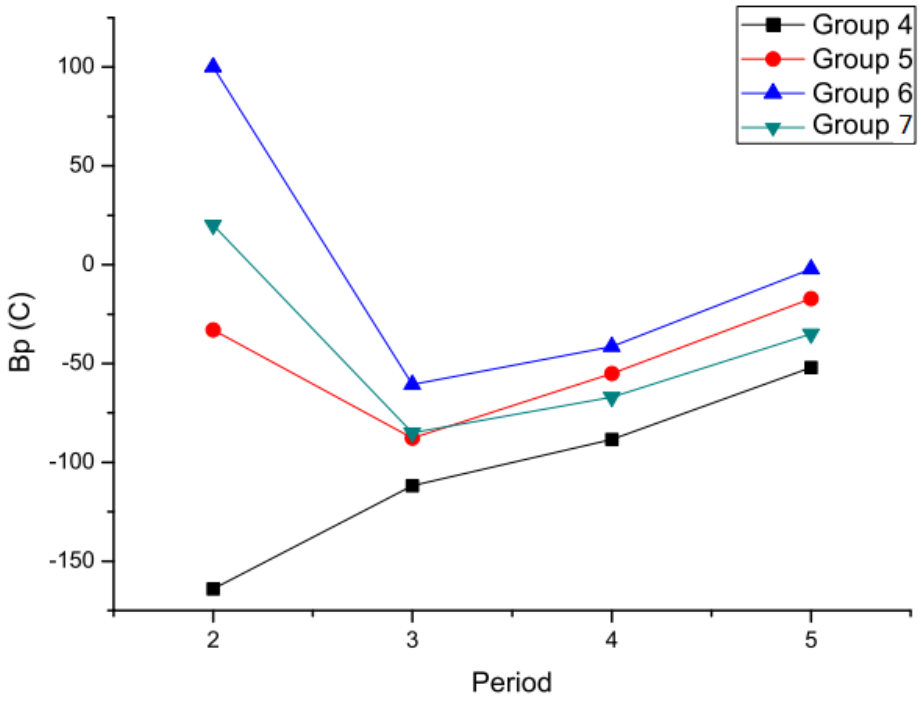


Figure 1.4: Structure of the hydrogen bonding network within water and a plot of the boiling point of the main group hydrides highlighting the elevated boiling points of NH_3, H_2O and HF.

This is also similarly observed in the higher than expected boiling points of the hydrides of nitrogen and fluorine. Going up the periodic table from period 5 to 2 it is expected that the boiling point will decrease and this is indeed the case from period 5 to 3. However, due to the hydrogen bonding displayed by NH_3 H_2O and HF the boiling point is substantially higher than expected. Carbon, an element which is not especially electronegative does not display an elevated boiling point as it does not form any appreciable hydrogen bonding.

A hydrogen bond is most energetically stable at a DAD angle of 180° and therefore are highly directional which leads to the formation of highly ordered structures. The strength of a hydrogen bonding interaction depends on the electronic properties of the donor and acceptor species, the DAD angle, the competitivity of any solvent, the number of hydrogen bonding sites and the sequence of multiple hydrogen bond donors and acceptors.

It has been shown above how the electronic properties of the donor and acceptor affect the strength of hydrogen bonding interaction. A more electronegative atom X will result in a stronger hydrogen bond in the model above. This highlights the effect that substituents can have on hydrogen

bond strength. Electron-withdrawing substituents can increase the electronegative effect of atom X and vice-versa. It can be described that a hydrogen bond will be stronger where the basicity of the H-acceptor and acidity of the H-donor are maximised.

The competitivity of the solvent has a significant effect of on the measured strength of a hydrogen bond interaction. Like all intermolecular interactions, a hydrogen bond is a dynamic interaction and is being formed and broken in solution. In many cases the solvent can also form hydrogen bonds and these compete with the hydrogen bond of interest. A highly competitive solvent such as DMSO or methanol can form hydrogen bonds so the compound has to overcome these for the intermolecular hydrogen bonds to form. Conversely, a non-competitive solvent such as chloroform or hexane will not compete for hydrogen bonds.

The sequence of hydrogen bond donors and acceptors is important in governing the strength of a hydrogen bonding interaction between two molecules due to attractive and repulsive secondary interactions. As expected, as the number of hydrogen bond pairs between two molecules increases so does the strength of the interaction as the forces are additive. There are, however, secondary electrostatic interactions that can have a pronounced effect on the hydrogen bonding pair.

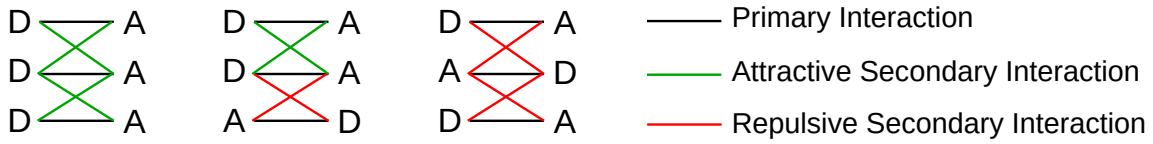


Figure 1.5: The primary and secondary H-bonding interactions of a sequence of H-bond donors (D) and acceptors (A).

As shown in the above diagram, in a hydrogen bond a donor is attracted to an acceptor in a primary interaction. There is also a secondary interaction between the adjacent donors and acceptors. If a donor is paired to an adjacent acceptor then there is an attractive secondary interaction. Conversely, if a donor is paired to an adjacent donor then there is a repulsive secondary interaction. Therefore the most attractive sequence of donors and acceptors is DDD-AAA. In this system all the primary and secondary interactions are attractive. In an alternating sequence of DAD-ADA there are still three primary attractive interactions but all the secondary interactions are repulsive. This results in a weaker interaction between the two molecules.

1.2.4 Cooperativity

Cooperativity is the phenomenon where the binding of an interaction influences the affinity of further interactions.[29][30] Cooperativity can be either positive, where relatively weak non-covalent interactions can become stronger and form more readily once a number of interactions have already formed, or negative, where the affinity of further interactions becomes lower. Positive cooperativity occurs by the formation of an interaction which makes subsequent interactions more prevalent. They may do this by having a 'tether' effect where the first interaction places the rest of the molecule closer so it is more likely to bind a second time. The initial interaction may also have an electronic effect that increases the affinity of further interactions. Two well-known examples of cooperativity in nature exhibit these effects. In haemoglobin the binding of a single oxygen molecule makes further binding events more likely by affecting the electronic properties of the haemoglobin molecule until the full complement of four oxygen molecules have been bound (Figure 1.6 a).[31][32][33] In the DNA double helix the binding of one nucleobase pair 'tethers' the rest of the DNA strand and brings the adjacent pair closer for subsequent binding as shown in Figure 1.6 (b). This results in the accumulation of many relatively weak interactions to form a relatively stable system. This is analogous to a zip on a jacket; each single interlocking interaction is individually weak and relatively

unlikely to form but once the zip is first connected each subsequent link makes the next link easier to form and the whole zip is strong.

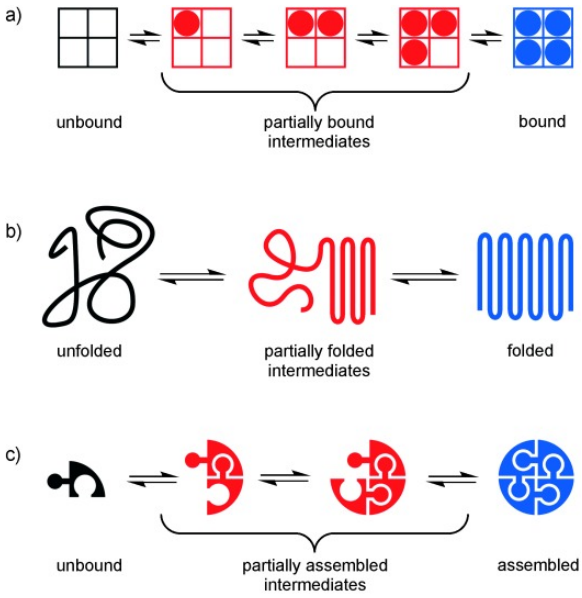


Figure 1.6: Graphical representation of systems displaying cooperativity a) the subsequent binding of oxygen to haemoglobin b) the folding of a protein c) supramolecular self assembly.[29]

Cooperativity is especially prevalent in supramolecular chemistry and is featured in molecular tiling and the synthesis of molecular zippers. Cooperativity is used to the supramolecular chemist's advantage by allowing relatively weak, dynamic, intermolecular interactions to combine to readily form a highly ordered and stable structure in a manner that mimics nature.[34][30]

1.3 The DNA and RNA bases

DNA and RNA are the molecules that contain the genetic information present in all living organisms. DNA is made up of three parts; the nucleobase (adenine, thymine, guanine and cytosine [and uracil in RNA]), the sugar and the phosphate as shown in Figure 1.7. These components form a monomer known as a nucleotide which is repeated to form a polynucleotide. Two polynucleotides are bound together by hydrogen bonds between the nucleobases to form a double helix where eacg nucleobase forms a hydrogen bond to its own complementary pair (adenine to thymine (or uracil) and guanine to cytosine) as published by Watson and Crick in 1953. [35]

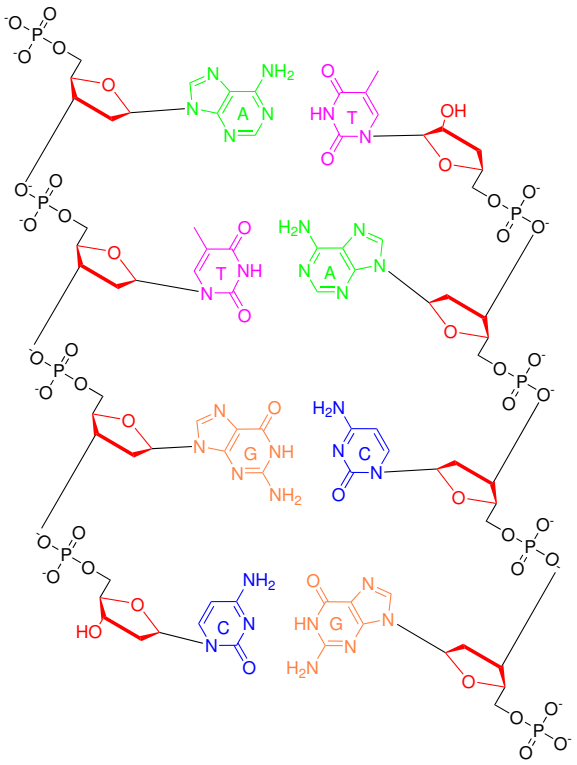


Figure 1.7: Chemical structure of DNA

The DNA nucleobases are based on either the purine or pyrmidine structure. Adenine and guanine are both classified as purines where an imidazole ring is fused to a pyrimidine ring whereas thymine, cytosine and uracil are classified as pyrimidines. It is on these pyrimidine and purine substructures that the numbering of the DNA nucleobases is based as shown in Figure 1.8. The total number of DNA bases on earth are estimated to weigh 50 billion tonnes.[36]

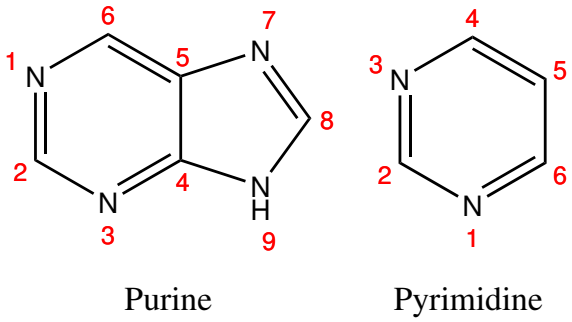


Figure 1.8: Structure and numbering of purine and pyrimidine rings on which the DNA nucleobases are based.

1.4 Hydrogen-Bonding Interactions between the DNA base pairs

The DNA base pairs form complementary interactions due to the favourable interactions between the relevant hydrogen bonding donors and acceptors. In DNA the complementary base pairing that is observed is the well-known Watson-Crick base pairing.[35][37] In this base-pairing adenine is hydrogen bonded to thymine (or uracil) and guanine is hydrogen bonded to cytosine. For a hydrogen bond to form a donor and an acceptor must be presented to one another in a geometry that is energetically favourable. In the Watson-Crick base pairs the structure of each base is such that each base has a donor and acceptor in the right position for its relevant pair. This allows adenine-thymine to form two hydrogen bonds and guanine-cytosine to form three.

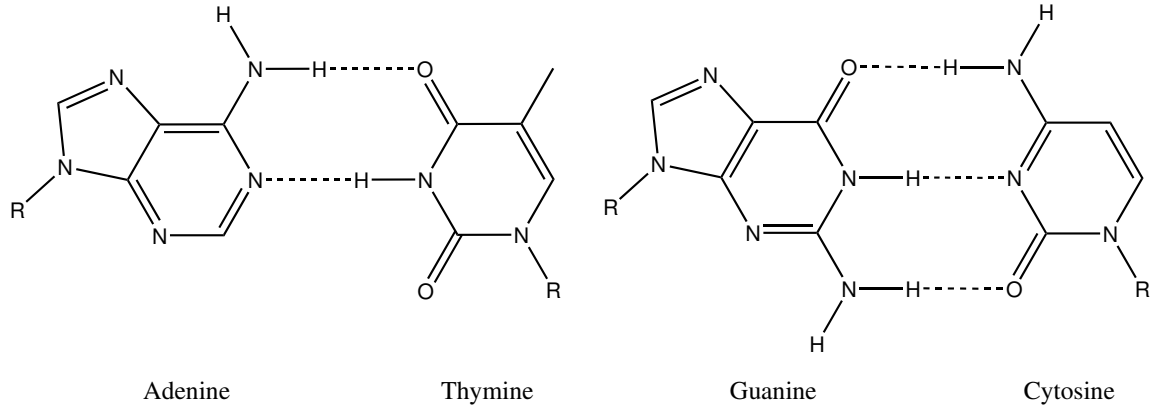


Figure 1.9: The complementary Watson-Crick base pairs.

The hydrogen bonding pairs shown in Figure 1.9 are both purine-pyrimidine pairs as in DNA purine-purine and pyrimidine-pyrimidine pairs are not observed. The reason for this is that there is not enough space within the helix for two purine-based bases to fit as the purine structure is too large. Conversely, the pyrimidine bases are too small to get close enough for hydrogen bonds to form. Adenine-cytosine base pairs are not observed as their structures do not allow the hydrogen bond donors and acceptors to be presented in a complementary manner. Guanine-thymine base pairs may, however, be observed and these are known as wobble base pairs as shown in Figure 1.10. This size selectivity of the base pairs regarding purine-purine and pyrimidine-pyrimidine interactions is due to the overall cooperativity of the system in forming the double helix, where many hydrogen bonding pairs are formed and constrained by the geometry of the sugar-phospate DNA backbone. In the absence of these constraints it is possible for many more hydrogen bonding interactions between the DNA base pairs to be observed. At least 24 possible interactions are possible between the four natural bases.[38]

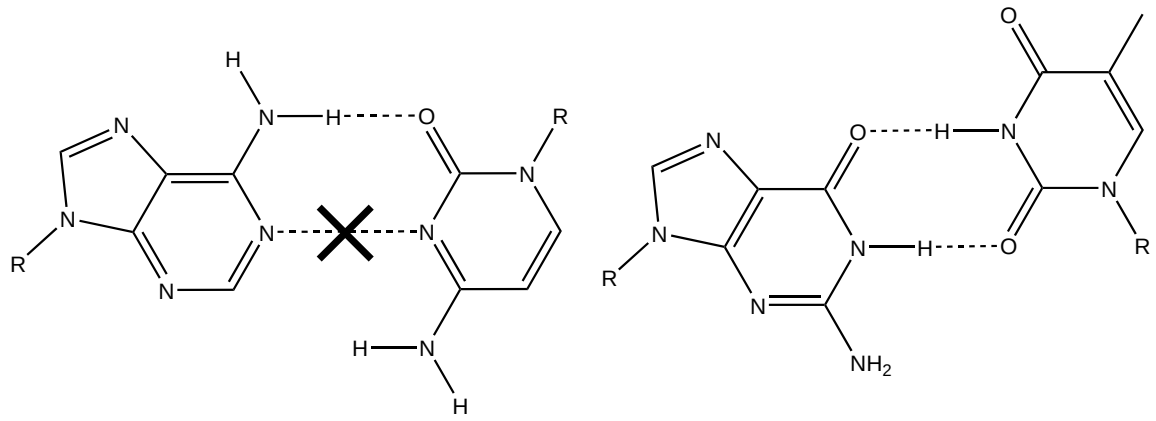


Figure 1.10: The non-complementary adenine-cytosine and the guanine-thymine wobble base pairs.

Another variation of possible hydrogen bonded base pairs are the Hoogsteen base pairs. Hoogsteen base pairs were first reported by Karst Hoogsteen ten years after Watson and Crick first reported their model of the double helix with Watson-Crick pairs.[39] In Hoogsteen base pairs the N-7 nitrogen on the purine base is used to bind the normal Watson-Crick face of the pyrimidine base. Hoogsteen base pairs are rarely observed in DNA as they would require the double helix to adopt a very different shape due to the rotation around the N-9-sugar bond. The hydrogen bond angles are also slightly different to those observed in Watson-Crick hydrogen bonds due to the less-favourable N-7 position.

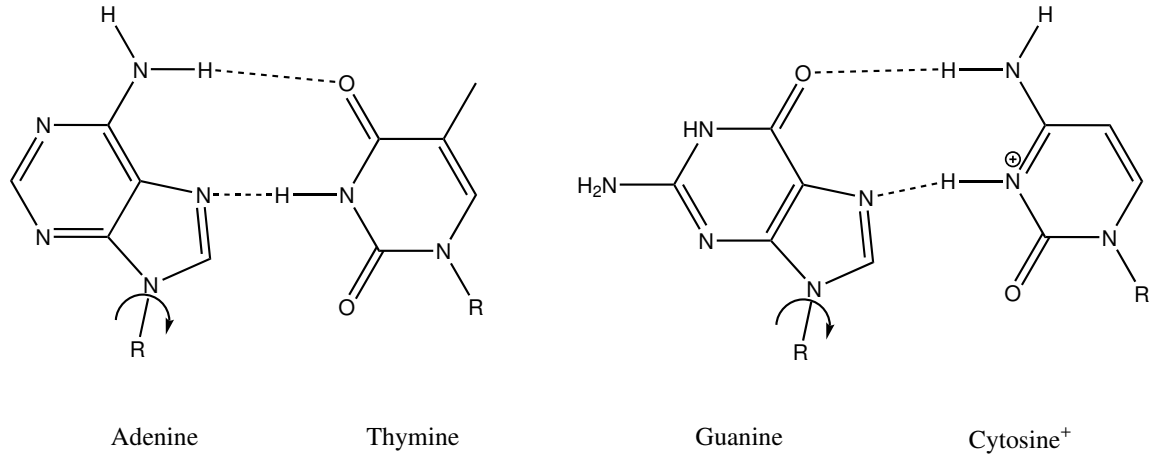


Figure 1.11: The Hoogsteen adenine-thymine and guanine-cytosine base pairs showing the rotation around the N-9-sugar bond.

For the guanine-cytosine Hoogsteen base pair it is necessary for the cytosine nucleobase to be protonated at the N-3 position. With similarly simple modifications it is possible to encounter more nucleobase pairings. One such modification is keto-enol tautomerisation. [40] Keto-enol tautomerisation is the equilibrium between a ketone and an alcohol via the shifting of the bonding electrons.[41] Analogous to keto-enol tautomerism is amino-imino tautomerisation. This is where the transfer of bonding electrons leads to an equilibrium between an amine and an imine functional

group where the imine can also be either *cis* or *trans* across the double bond.[42] All the possible tautomers of the DNA bases via keto-enol and amine-imine tautomerisation are shown in Figure 1.12. In the DNA bases the tautomerisation allows the hydrogen bond donor and acceptor positions to be swapped around as shown in Figure 1.13 and can lead to a number of different possible base-pair arrangements.[43] For all the nucleobases the keto and amino forms dominate and these are the major tautomers which are observed.[40]

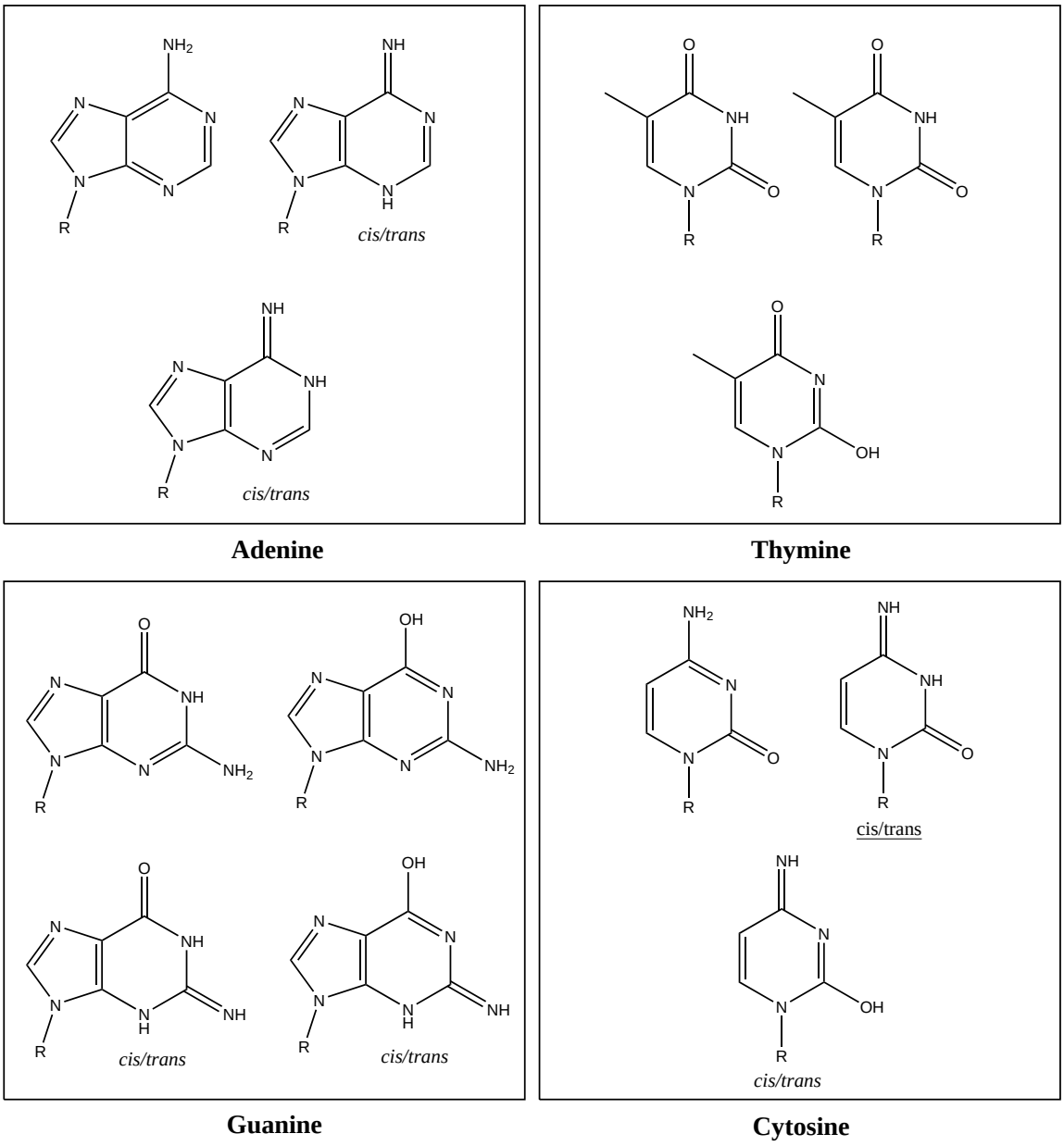


Figure 1.12: Keto-enol and amine-imine tautomers of the DNA bases.

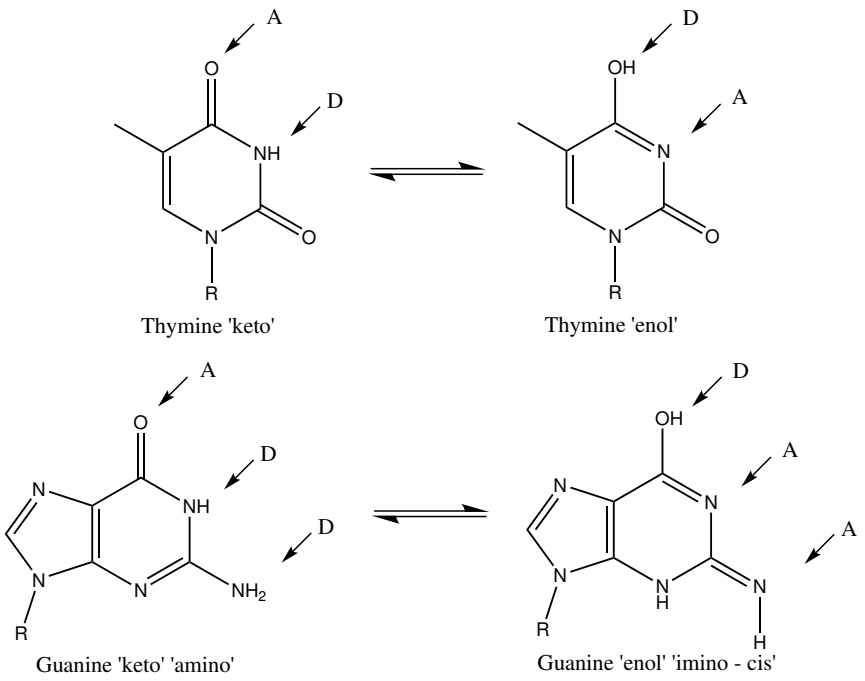


Figure 1.13: The keto-enol and amine-imine tautomerisation of thymine and guanine showing the swapped hydrogen bond donor and acceptor positions.

This tautomerisation and the rearrangement of the donor and acceptor positions causes the classical Watson-Crick base pairing rules to be reversed. Where the keto form of thymine hydrogen bonds to adenine the enol form hydrogen bonds to guanine and the guanine enol form pairs with thymine instead of cytosine. The imino form of adenine also pairs with cytosine instead of thymine and vice versa. It is interesting to note that Watson and Crick first thought that it was the enol form which was prevalent in the structure of DNA. Their colleague, Jerry Donohue, later informed them that the keto form was more likely and this allowed them to deduce the correct complementary base pairs that let to their groundbreaking article.[35][44]

Finally, although improbable in a DNA double helix due to the deformation of the sugar-phosphate backbone required, it is possible for the DNA bases to form hydrogen bonding pairs to themselves along either the Watson-Crick or Hoogsteen faces of the molecule as shown in Figure 1.14.[45][46]

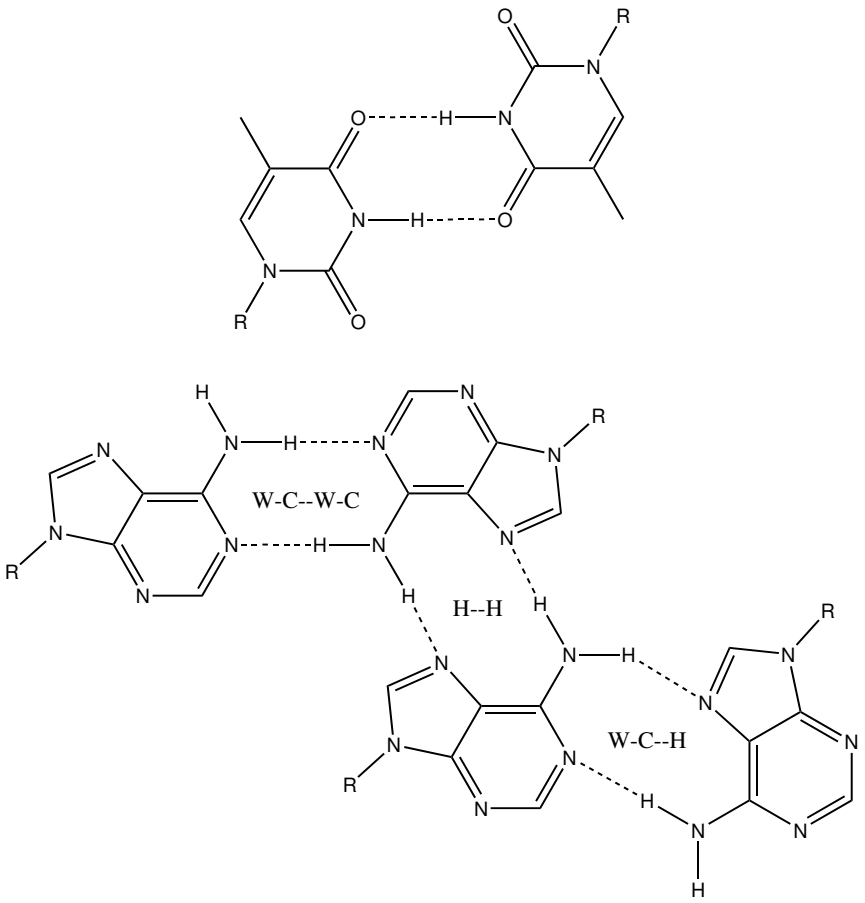


Figure 1.14: The self-dimerisation of thymine and adenine. The adenine structure shows the possibility of both the Watson-Crick (W-C) and Hoogsteen (H) hydrogen bonding faces.

1.5 Hydrogen Bonding in Supramolecular Self-Assembly

Due to their relatively high strength, directionality, selectivity, tunable properties and dynamic nature the hydrogen bond is a intermolecular interaction with well-explored use in supramolecular self-assembly. The synthesis of supramolecular macrocycles is in competition with the formation of linear oligomeric structures. The supramolecular chemist utilises a number of different techniques to control the synthesis of linear species. These include tuning the nucleation-elongation processes (increased nucleation will lead to cyclic species, increased elongation will lead to linear species) and the concentration of the system. The geometric properties of the intermolecular interactions can also be used to preferentially form cyclic species. Finally, templates can be used that lead to ring-closure.

Using intermolecular interactions allows supramolecular systems to form under thermodynamic control due to the dynamic nature of intermolecular interactions. This allows the system to rearrange until the most stable system is formed. The most stable system will be governed by a number of thermodynamic factors including the total number of interactions, the relative strength of those interactions and the entropy of the system and these will govern the geometry of the system and whether a cyclic or linear system is formed and what size of cyclic system is formed as shown in Figure 1.15. The ratio of intramolecular (cyclic) interactions compared to intermolecular (linear) interactions is known as the Effective Molarity (EM) and this ratio determines whether a cyclic or

linear system is formed.

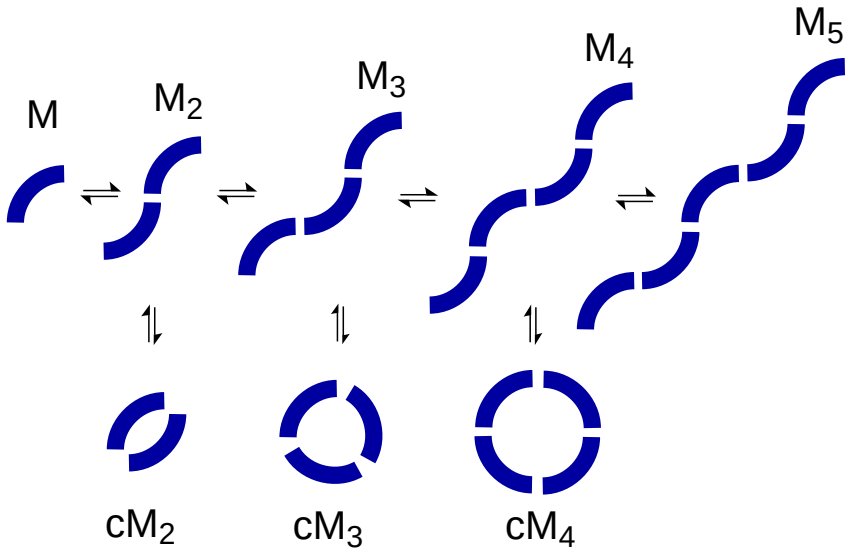


Figure 1.15: Supramolecular equilibrium for the association of a monomer (M) with two binding sites.[47]

In the diagram above the cyclic tetramer (cM_4) is the most stable species. This can be explained by examining the possible relative limitations of the other species present in the above hypothetical equilibrium. (cM_4) forms a cyclic structure with minimal strain on the interactions. The smaller cyclic species have considerable strain on these interactions; the angle is too acute for stable interactions. The cyclic species also forms a higher number of interactions compared to the linear species (4 versus 3 for (cM_4)) and this is thermodynamically favourable. Larger cyclic and linear species are unfavoured due the large decrease in entropy. In summary, the (cM_4) species is the 'goldilocks' species that best combines the effects of conformational arrangement, binding energies and entropy. These effects can be tuned to form the desired supramolecular macrocycle as shown in the examples below.

In 1996 J. Sessler *et al.* synthesised a relatively simple dimer system where two hydrogen bonding molecules associate as shown in Figure 1.16.[48] The hydrogen bonding pairs were based on adenine and uracil nucleobases, highlighting the selectivity of using nucleobases in supramolecular systems. The adenine was found to selectively bind to uracil and vice versa. This formed a strongly bound dimer with a stronger association compared to monomeric nucleobase pairs. This is due to the increased rigidity of the anthracene backbone and cooperative effect of multiple intermolecular interactions in a manner analogous to RNA strands in nature. Once one AU pair is formed a second pair is formed more readily due to the rigidity of the backbone limiting free movement of the free nucleobase pairs. The free pairs are therefore held in close proximity and a hydrogen bond is more readily formed.

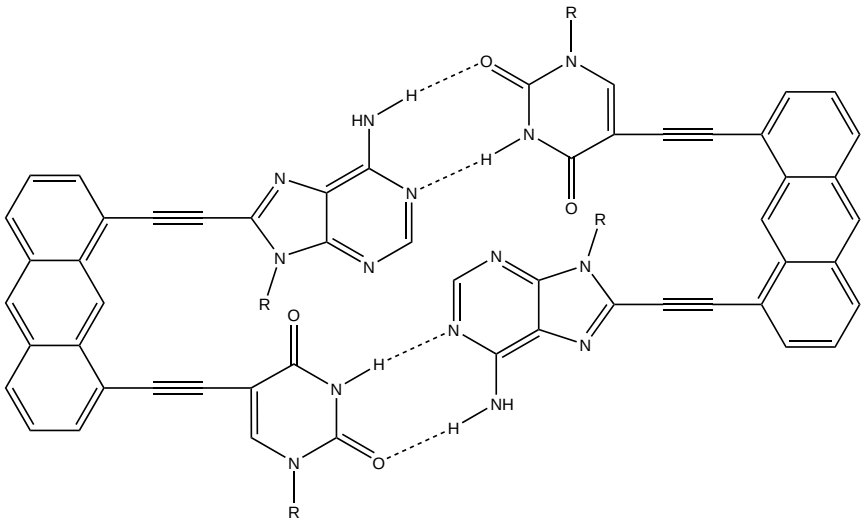


Figure 1.16: The nucleobase based hydrogen bonding dimer synthesised by Sessler *et al.*[48]

The group of Sessler used a similar approach of using DNA nucleobases; this time guanine and cytosine, to form a self-assembled trimeric system (Figure 1.17a). The formation of this trimeric system is driven by the formation of selective hydrogen bonds despite the slight conformational constraint on the system. Perrin *et al.* composed a cyclic tetramer by designing a hydrogen bonding motif bearing both the G and C Watson-Crick hydrogen bonding faces.[49] These faces were placed at an angle of 90° by synthetic design utilising a central pyrrole unit (Figure 1.17b). This has been expanded on further by the groups of Lehn and Mascal to form a cyclic tetramer (Figure 1.17c).[9][50] G and C hydrogen bonding motifs are again used and incorporated into a molecule with a separation angle of 120°. This demonstrates how two similar systems can be tuned to form the desired macrocycle by affecting the structure of the molecule and therefore the relative angle of interaction and then utilising the directional nature of hydrogen bonds.

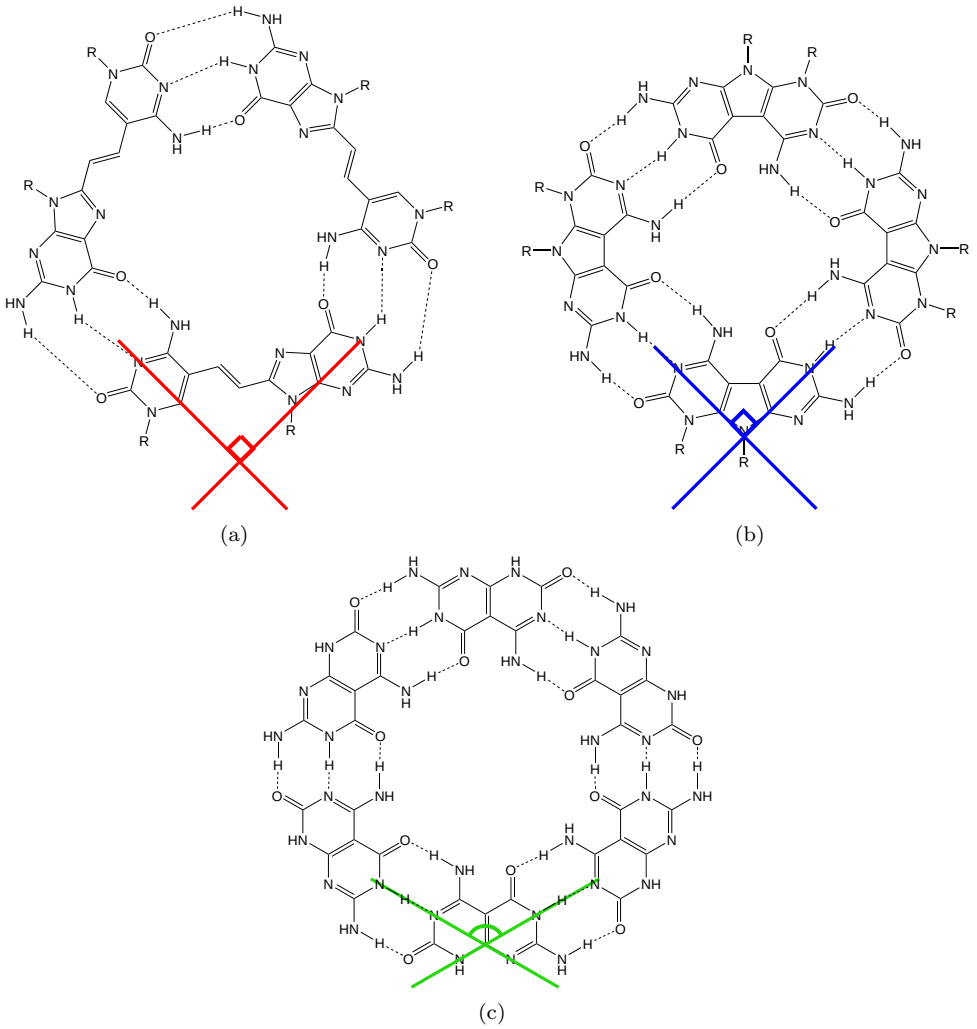


Figure 1.17: Structures of trimeric, tetrameric and hexameric hydrogen bonding macrocycles formed by the interaction between guanine and cytosine nucleobase derivatives.[48],[49],[9]

The group of Gonzalez-Rodriguez have focussed on self-assembly using the DNA nucleobases.[51][47], Recently, the group prepared ditopic monomers bearing the G and C moeity.[52] This structure, with its rigid organic linker, formed unstrained square-shaped assemblies as shown in Figure 1.18. These self-assembled structures were studied by titration experiments using both ^1H NMR spectroscopy and emission spectroscopy. The same group has done analogous work with both adenine and thymine molecules.[53]

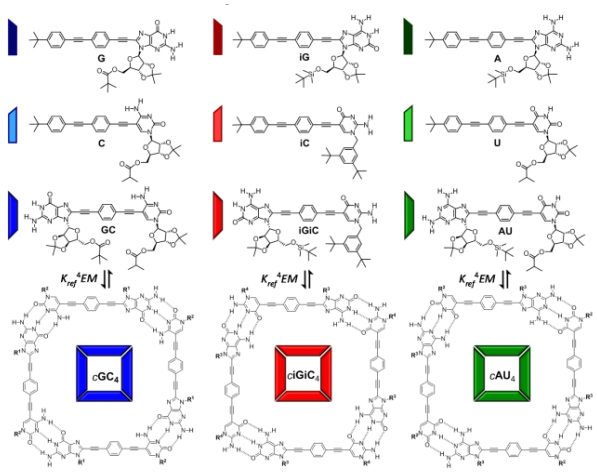


Figure 1.18: Chemical structure of GC monomers and their corresponding cyclic tetramer assembly.[52]

In the work the Gonzalez-Rodriguez group utilised the rigid alkyne group to join the relevant nucleobase to a chosen aromatic linker. This gave the resulting supramolecular structure the desired 90° angle when the Watson-Crick base pair was formed[52][47]. The rigid triple-bond alkyne structure also prevents rotation within the molecule.

1.6 2D Supramolecular Self-Assembly

2D Supramolecular Self-Assembly is the assembly of selective substrates in an ordered arrangement in two dimensions. To allow for the restricted assembly in two dimensions the molecules are usually deposited on a surface such as Highly Ordered Pyrolitic Graphite (HOPG) or Au (111). These surfaces allow for the reproducible assembly of molecules due to their low number of imperfections and high order.[54] These surfaces also allow for convenient use of the chosen surface imaging technique. Surface imaging techniques allow for an atomic level image of the surface to be produced so that the precise molecular arrangement of the self-assembled structure can be observed in real time. Scanning Tunelling Microscopy (STM) is the most-used surface imaging technique in 2D supramolecular self assembly due to its atomic resolution.[55] The scope of 2D supramolecular self-assembly includes the atomic imaging of multicomponent highly-ordered systems. This allows structural properties of the molecules such as bond length and angle to be studied. The self-assembled 2D structure may also produce a porous 2D network where guest molecules can be added and imaged with interesting applications in catalysis, sieves and chemical or biological sensors. The order of the self-assembled structure may also change depending on the conditions. This dependency can be imaged, characterised and utilised for the synthesis of chemical sensors or nano-electronics. Examples of these will be shown below:

The majority of compounds used in 2D supramolecular self-assembly are based on planar aromatic structures. This structure allows for adsorption of the compound onto the surface. An example of this is the adsorption of dehydrobenzo(12)annulene (DBA) and its derivative onto HOPG by De Feyter *et al.*[56] (Figure 1.19).

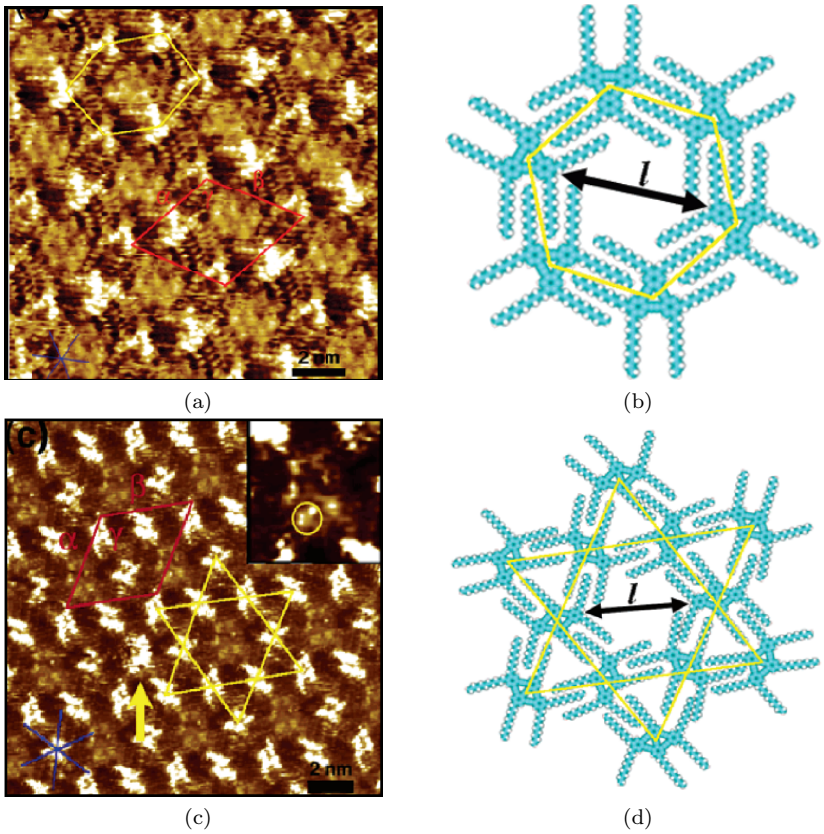
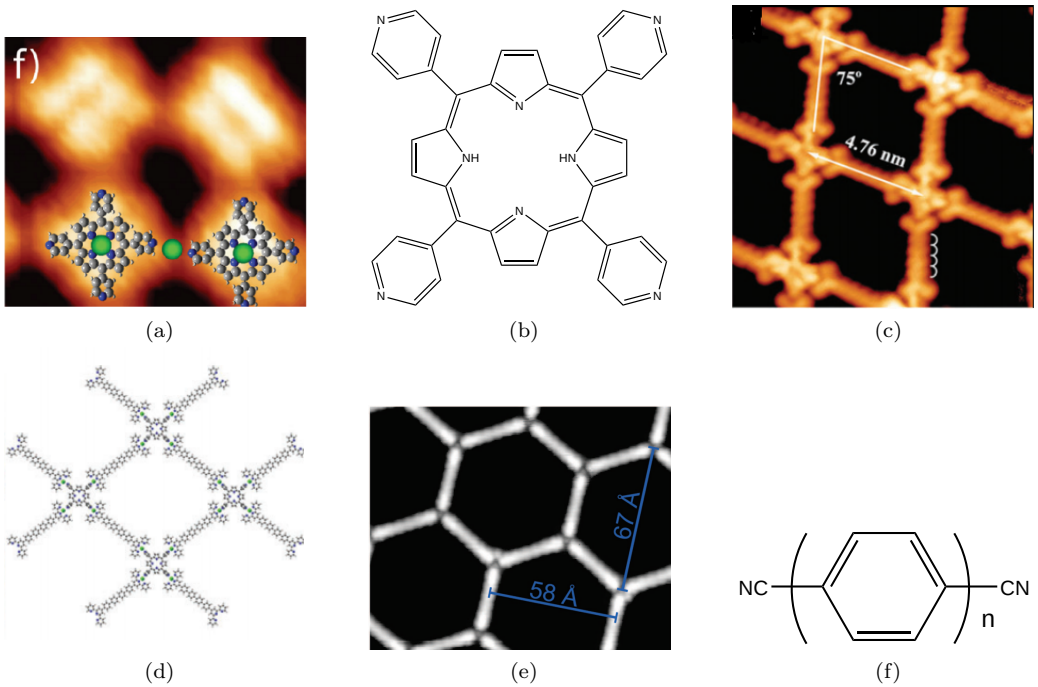


Figure 1.19: STM images of DBA derivatives. a) STM image of DBA physiosorbed at the TCB/HOPG interface (I_{set} = 0,50 nA, V_{bias} = -0.99 V. b) Molecular model of the self assembled DBA structure. c) STM image of DBA derivative physiosorbed at the TCB/HOPG interface (I_{set} = 0,50 nA, V_{bias} = -1.049 V. d) Molecular model of the self assembled DBA derivative structure.[56]

The structures above self assemble via the interdigitation of the alkyl chains which are attracted via relatively weak Van der Waals forces. The formation of these porous architectures is also dependent on the solvent used highlighting the hydrophobic nature of alkyl chains.

Ligand-metal interactions are also used extensively in 2D supramolecular self-assembly due to their relatively high strength, dynamic nature and directionality. Ligands bearing the carbonitrile moeity have been extensively studied with metals such as cobalt, iron and copper as shown by various groups including Li[57], Lin[58] and Schlickum[59] amongst others.[60][61] Various linkers were synthesised bearing the carbonitrile moeity which bonded to the chosen metal. It was possible to tune the surface structure by varying the size of organic linker and utilizing the different ligand-metal chemistry of different metals.



(a) (b) (c)

(d) (e) (f)

Figure 1.20: UHV-STM images of metal-coordinated porous networks. a) High-resolution image of structure (b) on Au (111) with Cu at 300 K (4.5 x 5 nm^2, -0.79 V, 0.29 nA). b) Chemical structure of compound used in image (a)). High-resolution UHV-STM image of the rhombic network formed by (b), Fe and () with a suggested model (d). e) STM image of Co-directed self-assembly of f) (0.1 nA, 0.9 V).[57],[58],[59]

Another ligand used extensively in ligand-metal interactions is the carboxylic acid group. Carboxylic acid groups have been shown to ligate to various metals and form ordered 2D structures. Again, it is possible to design different porous structures by altering the ligand.[62] Stepanow *et al* designed a 2D porous network on the Cu (100) surface with Fe of tunable pore size by utilizing a terepthalic acid-based ligand and increasing the number of phenyl linkers as shown in Figure 1.21.[63]

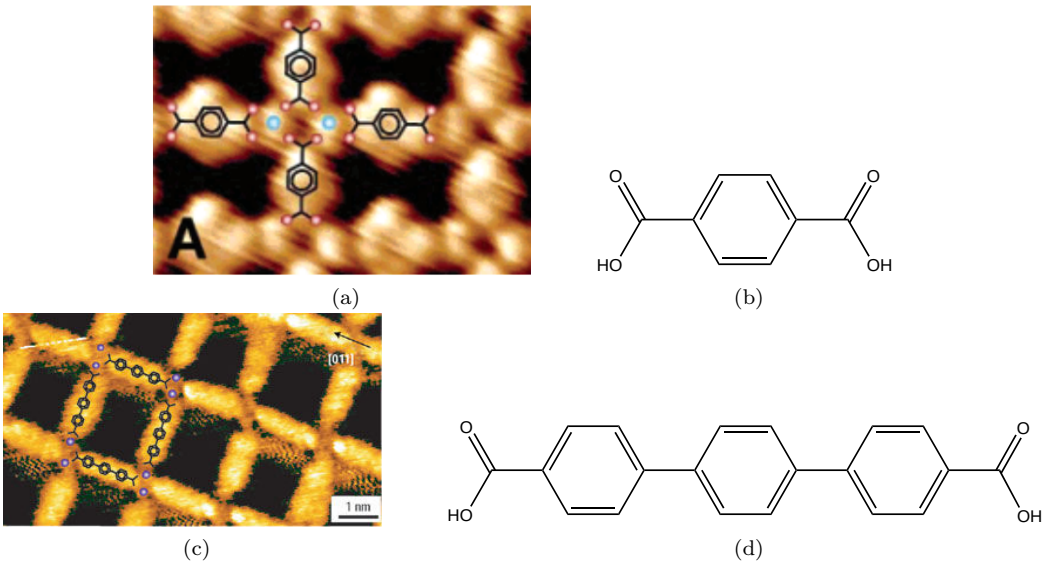
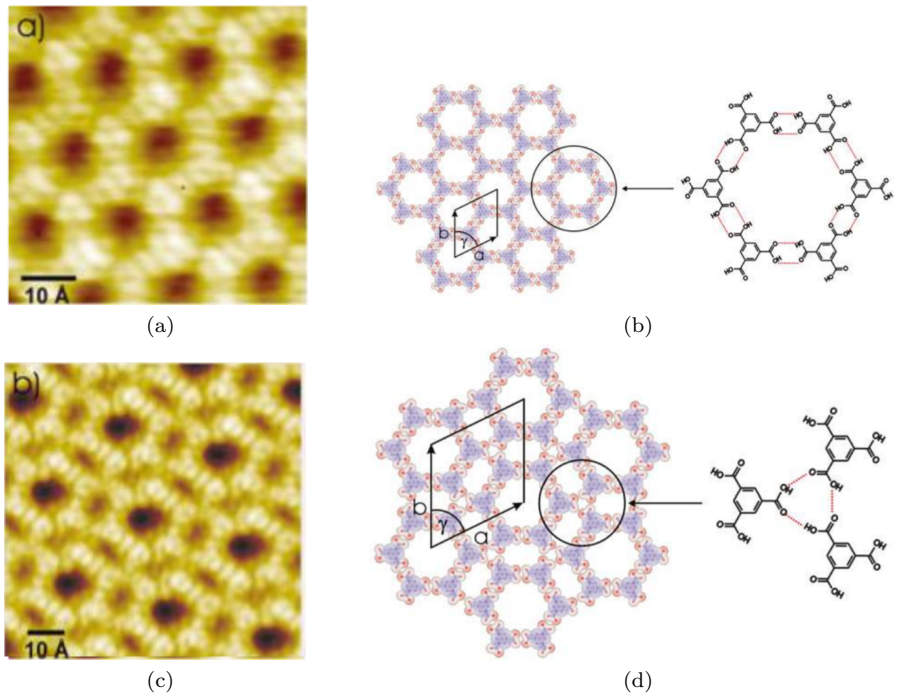


Figure 1.21: UHV-STM images of supramolecular networks based on carboxylic acid-metal interactions. a) STM image of carboxylic acid b) coordinated with Fe on Cu (100). c) STM image of carboxylic acid d) coordinated with Fe on Cu (100). Note the increased pore size with increased linker size.[63]

The carboxylic acid group can also, of course, be utilized in hydrogen bonding 2D supramolecular networks. Carboxylic acids are of particular interest due to their combined donor and acceptor properties which allow for stable, well-ordered and controllable self-assemblies.[64] Early reports demonstrate how compounds such as terepthalic acid and trimesic acid can be used to form 2D surface arrays.[65] Interestingly, the surface self-assembly of these compounds led to the possibility of two polymorphs (Figure 1.22). 'Chicken wire' and 'flower' structures were formed. This shows how a simple component can be used to form controlled complex structures. Each of the two surface polymorphs has a different pore dimension and this can be exploited for host-guest chemistry, catalysis and the design of chemical sensors.



(a) (b)

(c) (d)

Figure 1.22: STM images of supramolecular networks based on carboxylic acid hydrogen bonding interactions. a) UHV-STM image of 1,3,5-benzenetribenzoic acid on HOPG (5.4 nm ^{2}, -1.4 V, 126 pA) with the relevant 'chicken-wire' association mode (b). c) UHV-STM image of 1,3,5-benzenetribenzoic acid on HOPG (8.2 nm ^{2}, -1.4 V, 126 pA) with the relevant 'flower' association mode (d).[65]

By using more complex surface components it is possible to explore a more complex surface polymorphism. This was demonstrated by Beton, Champness and coworkers where terphenyl-3,5,3',5'-tetracarboxylic acid (TPTC) was used to form a 2D supramolecular array where five different patterns were found on the HOPG surface (Figure 1.24 A).[66] The multiple structures observed on the surface interface arises from the use of a rhombus shaped TPTC molecule which can orientate itself in three possible positions on the surface due to the directional hydrogen bonds formed between the carboxylic acid groups. These are represented by the red, blue and green tiles where each of these tiles is the same but rotated by 60°. These red, blue and green tiles can then organise into 5 possible arrangements (Figure 1.24 C-G). No translational symmetry is observed between the arrangement of TPTC molecules.

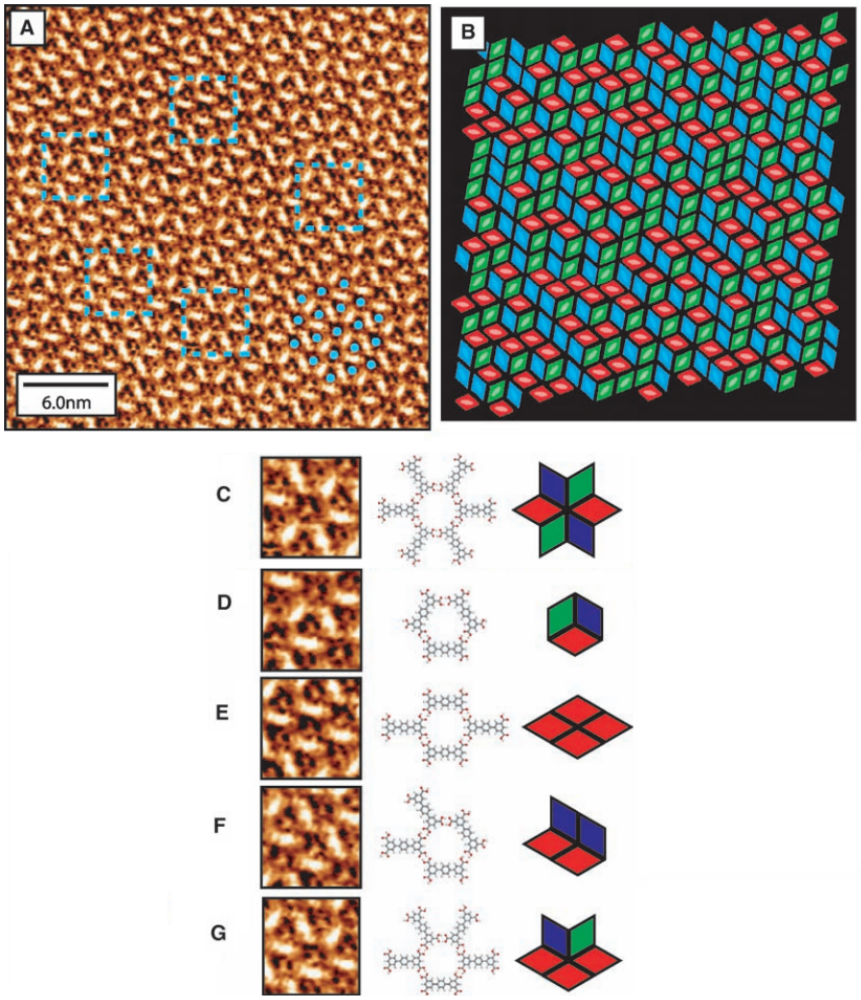


Figure 1.23: A) STM image of TPTC at the nonanoic acid/HOPG interface. B) Image A represented as tiles where the three orientations of TPTC are presented as red, blue and green tiles. C-G) Molecular ball and stick diagrams representing the five possible arrangements of TPTC on the surface.[66]

The supramolecular systems described so far have only made use of a single component. Multicomponent systems have received significant recent interest due to their increased complexity.[67] This increased complexity of supramolecular systems is desired because it allows the individual properties of each component to be combined and tuned to suit the desired application of the system. This can lead to increased selectivity, reactivity or stability of the system with advantages in the field of catalysis, photoelectronic devices, gels,[68] sieves and in the design of systems that mimic those found in nature.[69] Many biological activities such as photosynthesis use complex, multicomponent supramolecular systems to achieve a particular function.

Surface-based chemistry with its associated imaging techniques is an excellent way to study multicomponemt supramolecular systems as it is possible to visualise the exact arrangement of each component in the system.[55] This was demonstrated by Champness and coworkers where a perylenetetracarboxylic diimide (PTCDI) was codeposited with melamine which resulted in a honeycomb structure on the Ag/Si(111) surface.[70]

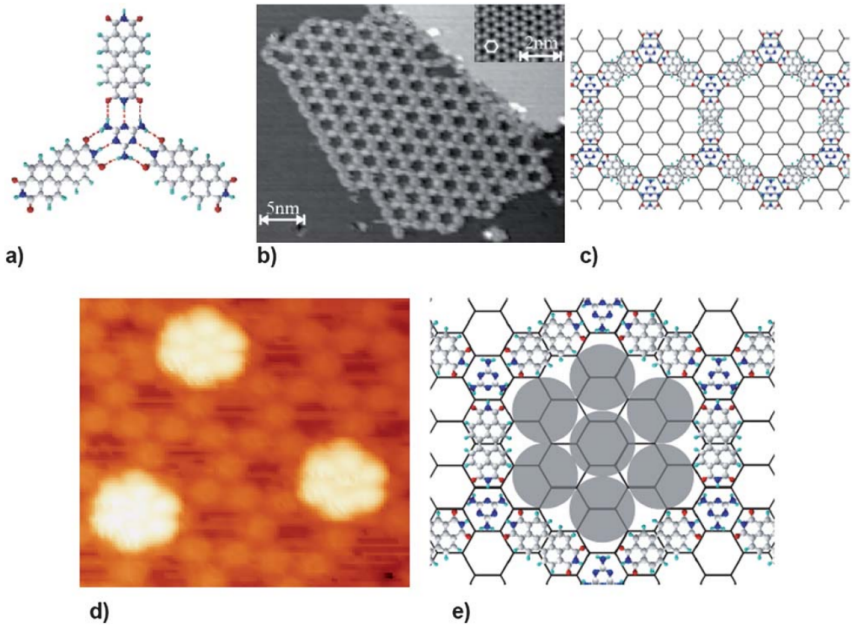


Figure 1.24: Self-assembly of a PTCDI-melamine supramolecular network. a) Chemical structures of PTCDI and melamine and a representation of the PTCDI-melamine junction. Dotted lines represent hydrogen bonds between the molecules. b) STM image of a PTCDI-melamine network. c) Schematic diagram showing the correlation of the network with the Ag/Si(111) surface. d) Image of C_{60} molecules trapped within hydrogen-bonded hosts on Ag/Si(111). e) Model of seven C_{60} molecules trapped within the PTCDI-melamine structure.[70]

The PTCDI selectively hydrogen bonded to melamine due to the three hydrogen bonds that were formed. The honeycomb hydrogen bonding network also formed in correlation with the surface, highlighting the contribution of the surface to the properties of the 2D network. This porous honeycomb structure was also shown to encapsulate C_{60}. For the PTCDI-melamine structure the encapsulation of seven C_{60} molecules was observed. In later studies it was shown that the number of guest molecules could be altered by changing the surface substrate and the structure of the PTCDI molecule.[71][72]

One of the key principles in multicomponent supramolecular systems is the selectivity of each component. To this end the DNA nucleobases have received significant interest due to their selective directional hydrogen bonding nature.[73][74][75][76] The DNA bases themselves are also based on a planar aromatic structure (pyrimidine or purine) and this property makes the DNA nucleobases highly suitable for 2D self assembly on a surface. Early work in this area includes the study of guanine tetramers and the association of $C_{18}H_{37}$-N^9-alkylated guanine on a HOPG surface.[77][78][79] This alkylated guanine species was found to form hydrogen bonded ribbons flanked by interdigitation of the alkylated chains (Figure 1.25 a),c). Addition of potassium cations leads to the rearrangement of the ribbons to form guanine quartets with potassium at the centre of the quartet (Figure 1.25 b),d). This can be reversed by the addition of trifluormethanesulfonic acid which liberates the potassium ions.

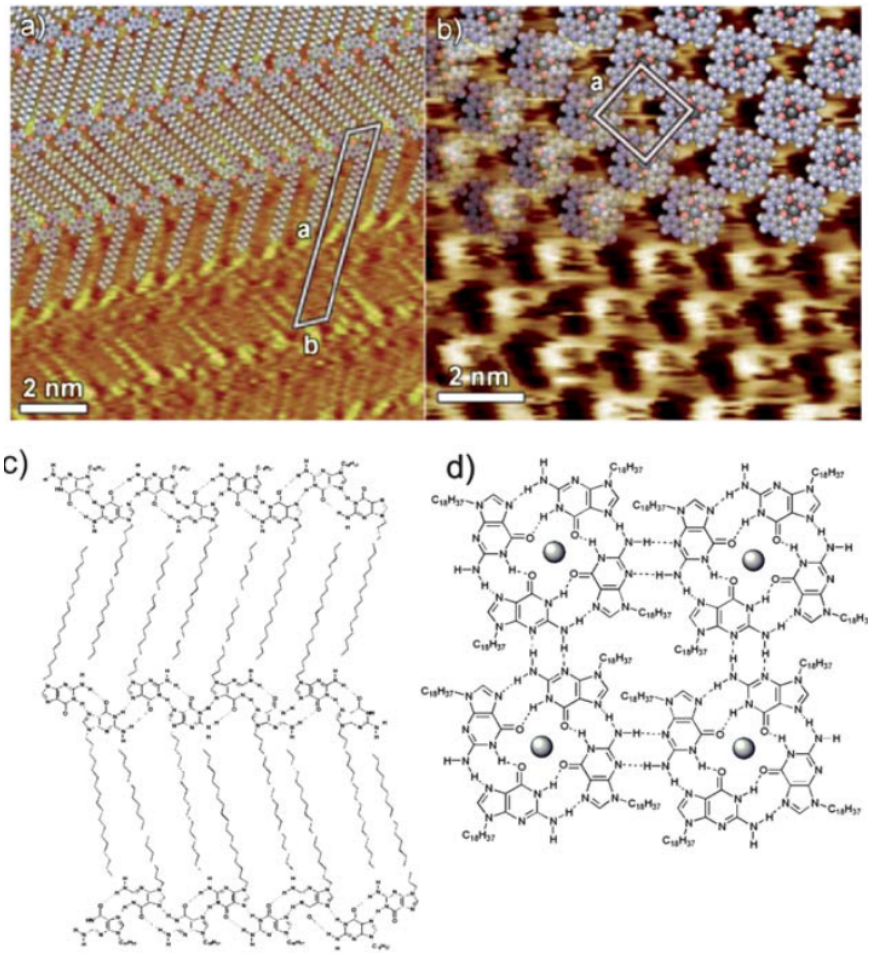


Figure 1.25: STM images of $C_{18}H_{37}$-N^9-alkylated guanine on a HOPG surface. a) STM image of $C_{18}H_{37}$-N^9-alkylated guanine on HOPG with relevant model (c). b) addition of potassium cations leading to rearrangement of the interlocked ribbons to form quartets around the potassium cation with relevant model (d).[77]

Further research into the 2D self-association of each nucleobase has also been carried out by various groups where the nucleobases were found to form cyclic species via interaction at both the Watson-Crick and Hoogsteen faces of each molecule.[45] However, due to the difficulty in imaging and determining the exact identity of each nucleobase mixtures of nucleobases have been less studied on the surface. Mixtures of guanine and cytosine (GC) and guanine and uracil (GU) were studied by Besenbacher *et al.*[76][80][81] where, as expected, in solution the G nucleobase showed strong affinity to C in solution due to the complementary GC hydrogen bonding but binding to U was much weaker. However, when concentrated on the surface binding to both C and U was observed due to high number of H bonding sites that are available. Non-complementary hydrogen bonding pairs were also found between adenine and cytosine which were found to rearrange into A islands and C zigzags upon heating; highlighting the unstable A-C interaction as shown in Figure 1.26.

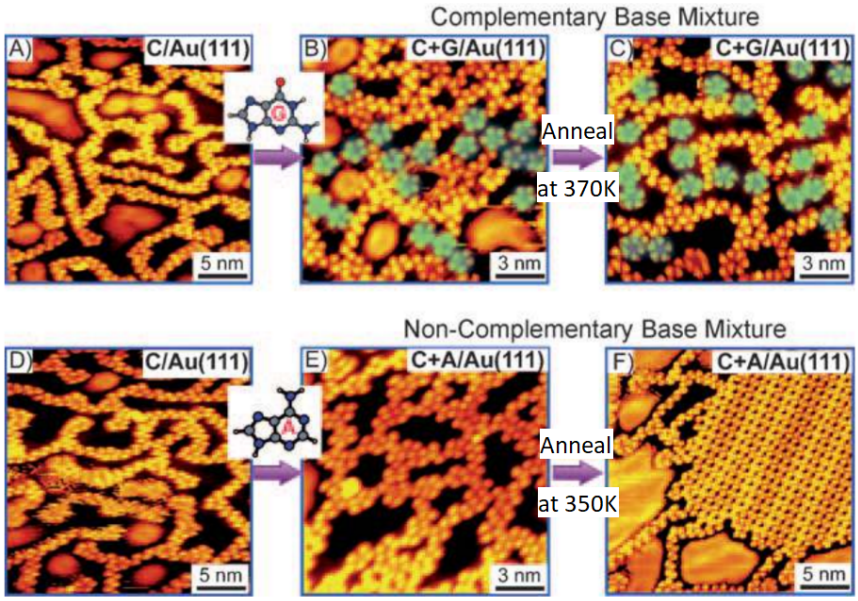


Figure 1.26: STM images of complementary and non-complementary base mixtures on Au (111). a),d) STM image of cytosine on Au (111). b) Addition of guanine to the surface which shows the formation of CG pentamers that are stable upon thermal annealing (c). e) Addition of adenine to the surface forming non-complementary CA pairs that rearrange into A island and C zigzags upon thermal annealing (f).[76]

The same group has also studied the hydrogen bonding surface self-assembly of adenine and thymine and observed a number of different hydrogen bonding assemblies where either Watson-Crick or Hoogsteen hydrogen bonding was present in both dimeric and tetrameric structures.[82]

It is easier to image and differentiate the two components in a 2D supramolecular self-assembled structure by functionalising each component so that each component can be distinguished. The additional functionality of each component may also introduce desirable properties into the supramolecular system that can be taken advantage of. Adenine and thymine have been studied in this manner by Champness *et al.* where a porphyrin was functionalised with thymine in the four meso positions.[83] This thymine functionalised porphyrin was studied on a HOPG surface and was found to self-associate into a highly-ordered structure (Figure 1.27 a),b)) and also to form a multicomponent molecular network when co-deposited with 9-propyladenine (Figure1.27 c),d)). Here, the two components in the molecular network are easier to distinguish due to the porphyrin structure of tetra-thymine porphyrin. This work shows the progress made in the synthesis and study of multicomponent molecular networks and the potential for further advances by utilising the inherent properties of the porphyrin structure including its ability to coordinate metals, the opportunity for differing functionalisation at the *meso* position and finally the electronic properties of the porphyrin structure.

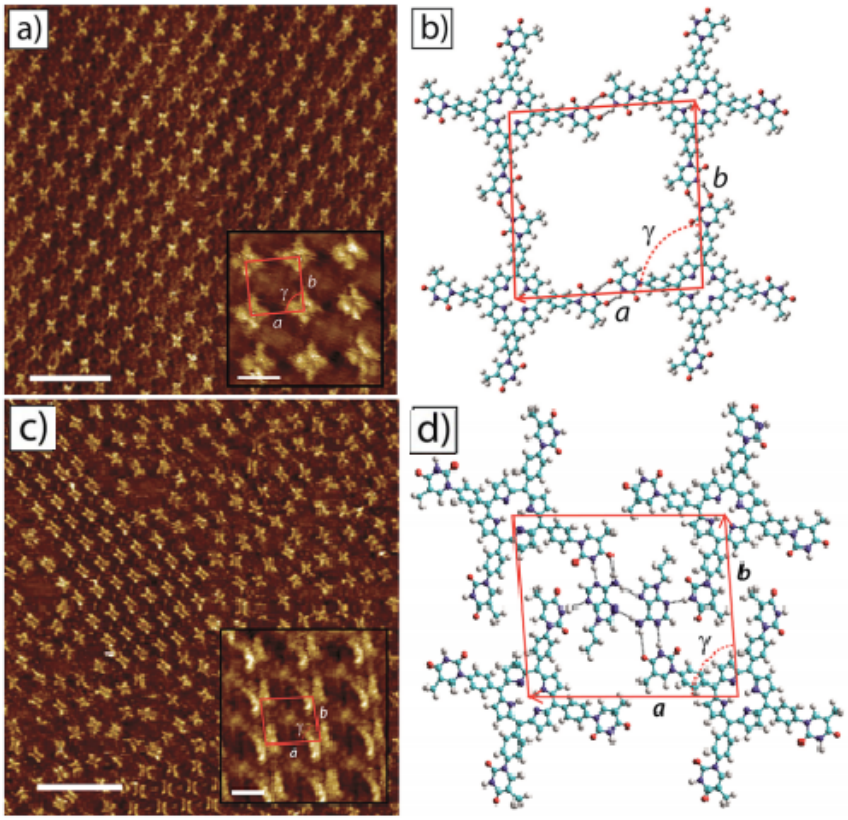


Figure 1.27: a) Large scale STM image of the tetra-Tporphyrin network at the TCBHOPG liquid-solid interface. The insert shows a high resolution, drift corrected STM image of the network with an individual 2D unit cell marked in red. b) Molecular model of the tetra-TP network from MM simulations. c) Large scale STM image of the tetra-TP and 9-propyladenine network at the TCB-HOPG liquidsolid interface. The insert shows a high resolution, drift corrected STM image of the network with an individual 2D unit cell marked in red. (b) Molecular model of the tetra-T-porphyrin network from MM simulations.[83]

1.7 Scanning Tunnelling Microscopy (STM)

Scanning tunnelling microscopy (STM) is a surface imaging technique used to visualise surfaces at an atomic level. It was developed in 1981 and earned its developers, Gerd Bining and Heinrich Rohrer, a Nobel prize in physics.[84] STM allows for 0.1 nm lateral resolution which allows for individual atoms within materials to be visualised.[85]

STM is based on the concept of quantum tunnelling which is the quantum mechanical phenomenon where a subatomic particle passes through a potential barrier that it cannot surpass under the provisions of classical mechanics.[86] In an STM experiment a conducting tip (Figure 1.28) is brought very near (within 1 nm) to the surface that is to be imaged. A voltage is applied and this allows a current to flow between the tip and the surface below. The observed current, known as the tunnelling current, is a function of tip position, applied voltage and the local density of states of the sample.[85] The tunnelling current varies as the tip moves across the surface and encounters different surface heights and density of states and this can be displayed in an image. STM can be measured in constant height mode of constant current mode. In constant height mode the height

of the tip remains constant and the current varies as the tip scans across the surface. In constant current mode the height of the tip is adjusted to provide a constant current by moving the height of the tip with piezo-electric crystals. In this mode, the height of the tip, z, is measured at each position.[87]

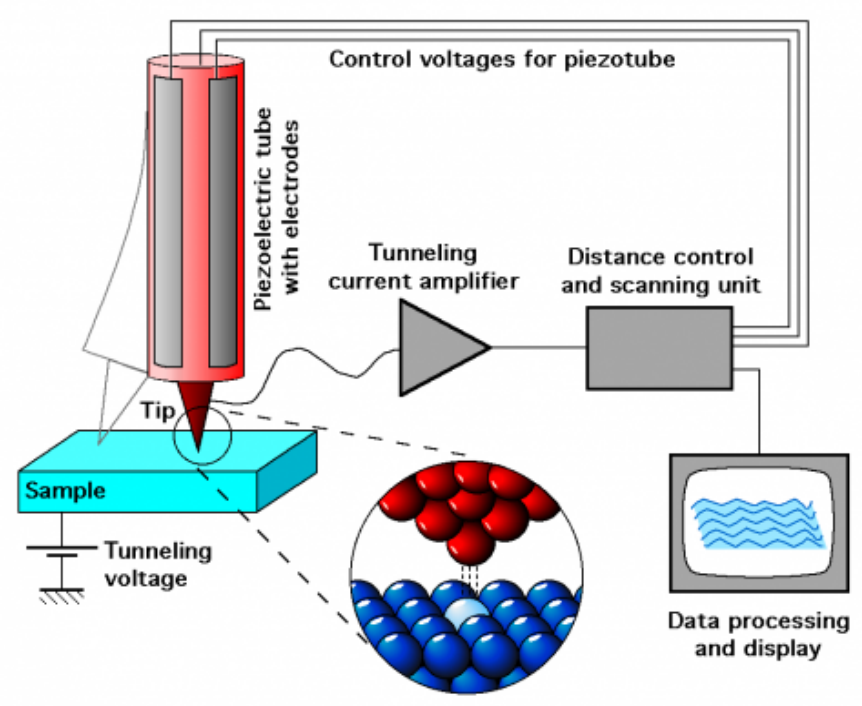


Figure 1.28: Schematic of the STM instrumentation.[88]

STM has been widely used as a surface imaging technique due to its atomic resolution. This allows for the precise location of atoms on the surface to be realized which makes it especially useful for imaging and analysing self-assembled arrangements.[55][89]

1.8 Porphyrin Introduction

Porphyrins are a class of heterocyclic aromatic organic compounds. They consist of four pyrrole subunits connected by a methine bridge. The name "porphyrin" comes from the greek word *"porphyra"* meaning purple and as such most porphyrins are strongly purple or dark red in colour. This strong colouration comes from the strong absorption of light in the visible region. There are many porphyrins which can be found in nature, one of the most well-known of these is heme which is found in red blood cells and is responsible for their characteristic red colour.

Porphyrins are widely used in 2D supramolecular self-assembly due to their planar aromatic core.[90] This allows for them to be readily deposited on a surface and their characteristic 'herringbone' structure can be easily identified via surface imaging techniques as shown below in Figure 1.29.[91]

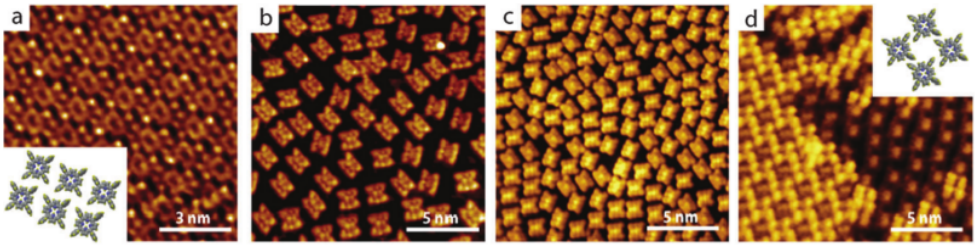


Figure 1.29: STM images of tetraphenylporphyrin on Ag(111) (a) and Cu(111) (b-d).[91]

Porphyrins have particularly interesting characteristics including their redox,[92][93][94] host-guest[95][96] and UV-visible[97] properties allowing for their wide use in solar-cells,[98] chemical sensors[99] and catalysts.[100]

1.9 The Hirshfeld Surface

The Hirshfeld surface of a molecule within a crystal is a computational representation calculated by partitioning spacial regions within the crystal into regions where the electron distribution contribution from the chosen molecule dominates over the electron distribution contribution of the crystal.[101] Where the molecular weight function $w(r)$ is defined as:[101][102]

$$w(r) = \frac{p_{molecule}(r)}{p_{crystal}(r)} \tag{1.1}$$

Where $p_x(r)$ is the electron density of either the molecule or the crystal. The Hirshfeld surface is then defined as the area where $w(r) \geq 0.5$. That being, the contribution of the molecule is greater than that of the crystal.

It is then possible to plot the distance from the surface to nucleus of the nearest atom inside (d_i) or external to (d_e) the surface. If the (d_i) distance is short then this indicates that there is a close contact to an adjacent atom. The distances (d_i) and (d_e) can then be mapped onto the surface to highlight areas where the close contacts of interest are situated on the molecule. They are also plotted on a fingerprint plot which graphically represents the surface against (d_i) and (d_e). An issue with using (d_i) and (d_e) is that they do not account for the size of the atom and this makes it difficult to reliably visualise the close contact between atoms of relatively different sizes. Where the atom is relatively large the distance from the nucleus to the surface will be correspondingly large despite the edges of the atoms being relatively close. To overcome this issue a normalised distance is defined (d_{norm})[103]:

$$d_{norm} = \frac{d_i - r_i^{vdW}}{r_i^{vdW}} + \frac{d_e - r_e^{vdW}}{r_e^{vdW}} \tag{1.2}$$

Where r^{vdW} is the van der Waals (vdW) radius of the appropiate atom internal or external to the surface. Where d_{norm} is shorter than vdW separations this will result in a red area on the Hirshfeld surface, while is used to represent longer contacts.

It is also possible to select areas of the Hirshfeld surface by element to selectively highlight where a certain element is closest inside and outside the surface. This enables particular interactions (such as OH, CC etc) to be highlighted. The contribution to the total surface area of that interaction can also then be calculated to provide information as to what extent a particular interaction is affecting the crystal packing. Combining this with the fingerprint plots can give further information on the interactions.

1.10 Project Scope and Aims

2D supramolecular self-assembly has received significant interest in recent years coinciding with the development of surface imaging techniques such as STM and AFM which allow for real-time images of the molecules on the surface to be visualised. Hydrogen bonding interactions have been used extensively to self-assemble 2D systems due to their selectivity, directionality, dynamic nature and relatively high strength. The four DNA and RNA bases are particularly effective in the field of self-assembly due to the well-understood and selective hydrogen bonds that occur between them. This selectivity allows multicomponent systems to be designed. Multicomponent systems have the advantage of being more complex which allows for more complex and tunable properties to be integrated.

Incorporating the DNA bases into the structure of a porphyrin will allow for multicomponent systems to be designed. A common issue with 2D supramolecular muticomponent systems is differentiating each species within the system. When one component in the system includes a porphyrin ring the two species can be distinguished due to the relatively large size of the porphyrin ring. The porphyrin ring can also be metallated which allows for the differentiation between separate porphyrins on the surface via the AFM imaging technique and the varying electron density observed in different metals.

Members of the group have previously synthesised a porphyrin bearing the thymine DNA base moiety in the meso position(s). The tetrathyminephenylporphyrin was found to associate with itself and its DNA base partner, 9-propyladenine, in a 2D self-assembled structure. This showed how complementary base pairs can self-assemble on a surface and be easily distinguished.

This project aims to build upon this foundation work by utilising the potential of the porphyrin ring to incorporate the other DNA and RNA bases into increasingly complex self-assembled structures. To better understand the interactions between the DNA bases and the effect of substituents on their interactions a series of alkylated and halogenated DNA nucleobases will be synthesised:

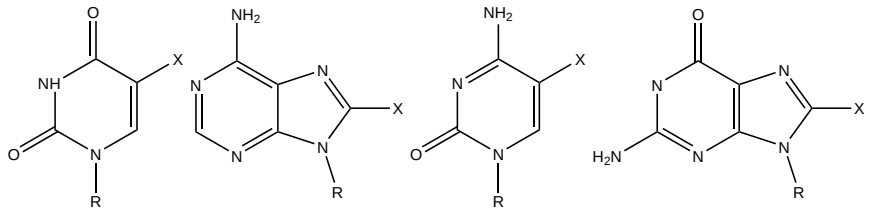


Figure 1.30: Structures of the target N-1 and N-9-alkylated DNA nucleobases. Left - right: N-1-propylthymine, N-9-propyladenine, N-1-propylcytosine, N-9-propylguanine.

The interactions of the above molecules can be studied in solution, as a crystal and on a surface via NMR titration experiments, single crystal X-ray diffractometry and surface imaging techniques respectively.

To further study the interaction of the DNA nucleobases when functionalised with a porphyrin a series of soluble porphyrins which have been functionalised in a single meso position will be functionalised building on the previous work in the group where mono-thymine porphyrin was isolated. These molecules will give a greater understanding of the effect of the porphyrin ring both in solution and as a crystal.

Porphyrins bearing four DNA bases will also be synthesised building on the previously synthesised tetra-thymine porphyrin. The surface self-assembly of these functionalised porphyrins will be studied, including looking the interaction between complementary and non-complementary base pairs. It is hoped that these molecules could self-assemble in organised multicomponent systems such as the one shown below (1.31):

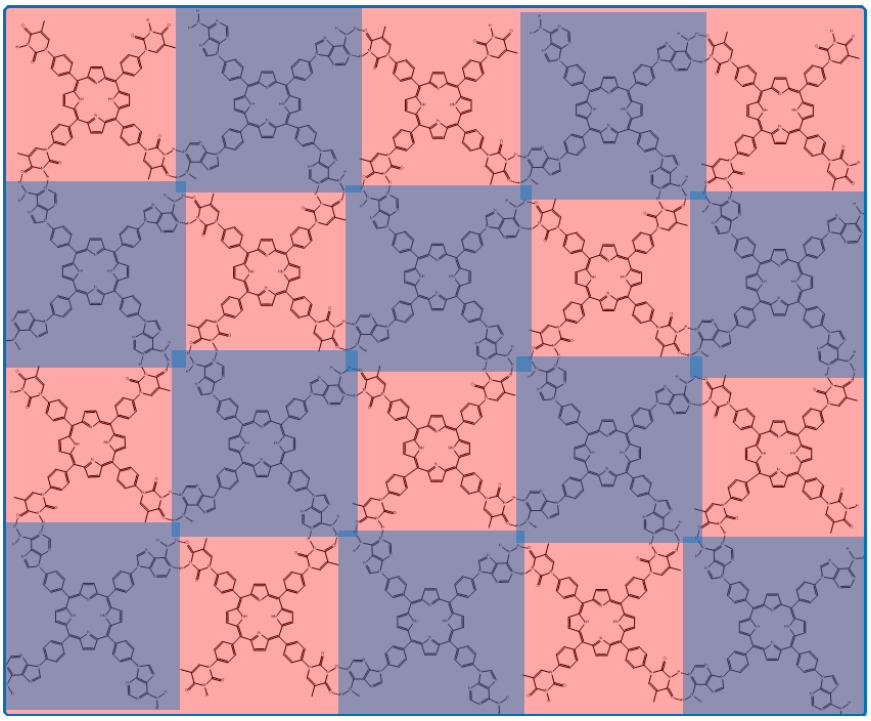


Figure 1.31: Potential 'chessboard' style arrangement between tetra-thymineporphyrin (red) and tetra-adenineporphyrin (blue).

It is important to be able to differentiate each component in a multicomponent system. The porphyrin ring will allow for this to be conveniently done via metallation of the porphyrin ring. These different metals can be differentiated by the surface imaging technique due to changes in the contrast associated with the coordinated metal. Metal coordination also offers the opportunity for additional control over the optical, electronic and catalytic properties of the resulting porphyrin.

Finally, the eventual aim for this project is to synthesise porphryins bearing all four DNA bases. This molecule has the potential to organise into a supramolecular structure utilising the complementary interactions between the nucleobases as shown in the schematic model below (Figure 1.32) where each nucleobase is represented by a red, blue, yellow and green square and the central porphyrin by a purple square.

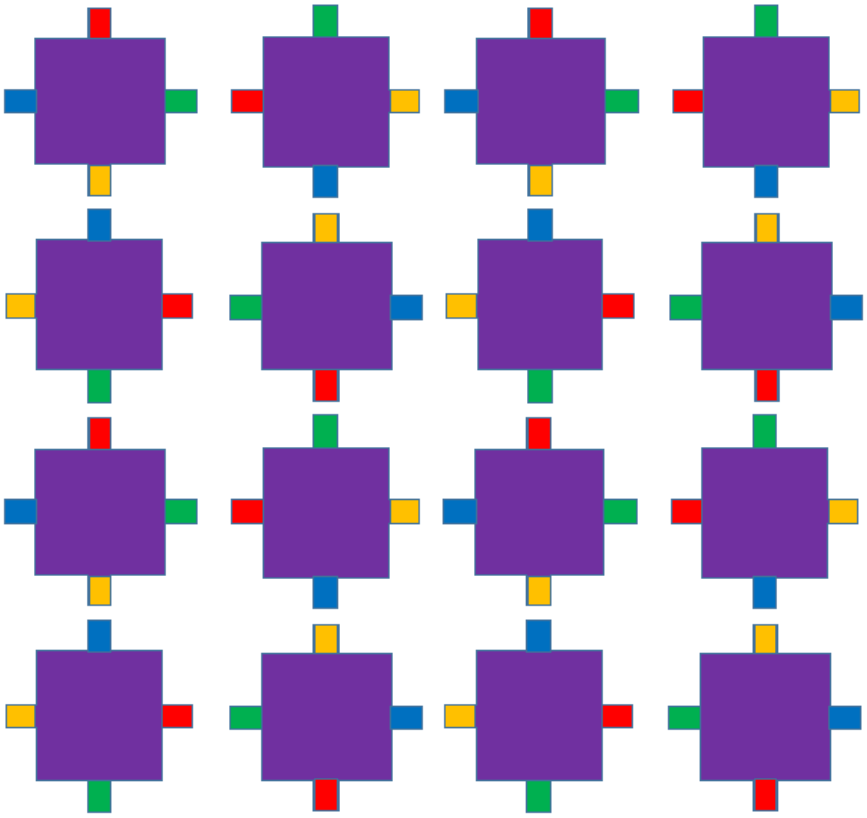


Figure 1.32: Potential supramolecular network formed by a porphyrin functionalised by all four nucleobases.

Despite the above network being formed of a single molecule, each molecule can associate with the surface in four possible relative orinetations each separated by a rotation of 90° in a manner similar to that described by Champness *et al.*[66]. Therefore, a complex supramolecular system of four effectively different units could be designed such as that proposed below; the likes of which have not been previously observed.

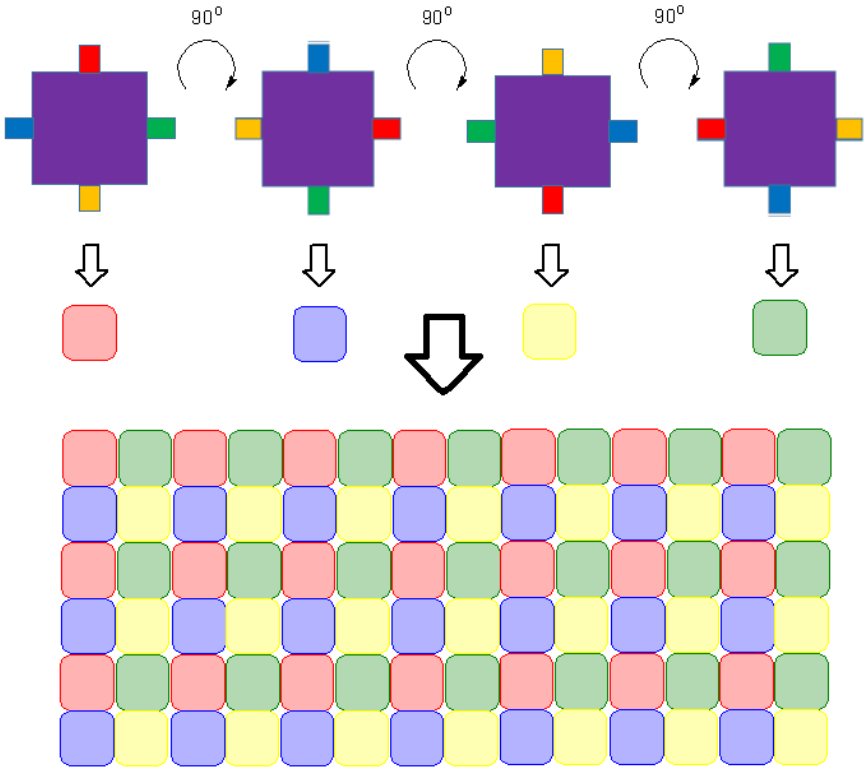


Figure 1.33: Potential supramolecular network formed by a porphyrin functionalised by all four nucleobases.

Chapter 2

The Synthesis of Alkylated and Halogenated DNA Nucleobases

2.1 Introduction

It is well understood that the four DNA nucleobases exhibit complementary hydrogen bonding pairs: thymine pairs with adenine and cytosine with guanine. The reason for these complementary pairs is due to the correct sequence of hydrogen bond donors and acceptors on each pair that allows one base to preferentially bind to another. In the DNA double helix the cooperative nature of multiple interactions occurring simultaneously and the structure of the helix itself with the phospate-sugar backbone also plays a significant role in the complementary binding of the DNA bases. To study the hydrogen bonding interactions of the DNA bases in isolation it is necessary to remove these other contributions from the system.

To study the hydrogen bonding interactions of the DNA bases in detail it is of high importance that the DNA bases are in solution. A known issue with the four DNA bases is their relatively low solubility in common organic solvents. To ensure that the hydrogen bonding interactions being studied are between the DNA bases the solvent must be a non-competitive solvent - the solvent must not also form considerable hydrogen bonds with the DNA bases. Examples of such suitable non-competitive solvents are chloroform, dichloromethane, hexane or benzene. Solvents such as DMSO, methanol, water or acetone are not desirable due to their ability to form hydrogen bonds to the DNA bases themselves. With this in mind a series of DNA bases were synthesised that would be soluble in the previously discussed non-competitive solvents with the aim of studying the hydrogen bonding of these functionalised DNA bases.

2.2 Synthesis of *N*-1 and *N*-9 Alkylated Nucleobases

The most logical position to functionalise the DNA bases is the position that is normally appended to the phosphate in the DNA helix. That being N-1 in the pyrimidine bases and N-9 in the purine bases. This is because this is most-analogous to biological systems and the hydrogen bonding donors and acceptors should be least affected and the effects on the hydrogen bonding should be minimised by steric or electronic effects. It is of added convenience that the N-1 and N-9 positions are relatively reactive and there are many examples in literature of N-alkylation reactions both on the DNA bases and on other purine and pyrimidine systems.[1][104]

The alkylating group of choice was the n-propyl group as it is relatively small so was not expected to change the hydrogen bonding behaviour of the DNA bases but should impart solubility in the

desired non-competitive solvents. The structures of the target N-1 and N-9 propyl DNA nucleobases is shown below in Figure 2.1

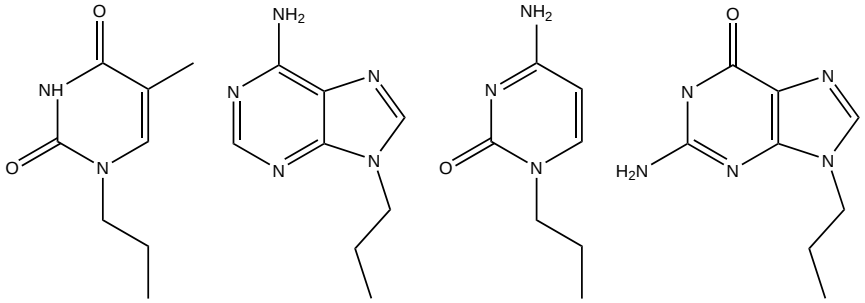


Figure 2.1: Structures of the target N-1 and N-9-propyl DNA nucleobases. Left - right: N-1-propylthymine, N-9-propyladenine, N-1-propylcytosine, N-9-propylguanine.

2.2.1 Propylthymine

The first DNA base for which alkylation was attempted was thymine. The alkylation of thymine with n-bromopropane is shown below in Figure 2.2. These conditions are widely reported in literature where there are numerous reports of alkylation of thymine.[1][5]

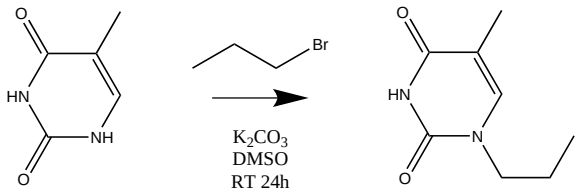


Figure 2.2: The synthesis of N-1-propylthymine via the reaction of thymine with N-bromopropane in DMSO.

This reaction proceeds via nucleophilic alkylation of the N-1 secondary amine. The nucleophilic nitrogen attacks the electrophilic carbon adjacent to the bromide. This carbon is the electrophile as it is δ^+ due to being next to the electronegative bromide. Bromide is also an excellent leaving group which is eliminated from the reaction as a salt. Potassium carbonate then acts as a base to remove a proton and leave the desired alkylated amine as shown in Figure 2.3. From understanding this mechanism it is clear that this reaction has the potential to form a number of undesired products. The first of those is the possibility of the product tertiary amine to react further with N-bromopropane and form an undesired quaternary ammonium salt. This is a well known issue with the alkylation of amines as the product of the alkylation reaction is also nucleophilic. The other possible side product from this reaction is alkylation at the N-3 nitrogen present in thymine. This nitrogen is also nucleophilic and may react with N-bromopropane. Fortunately, this nitrogen is less nucleophilic than N-1 due to the electron-withdrawing effects of the two adjacent carbonyl groups as shown in Figure 2.4.

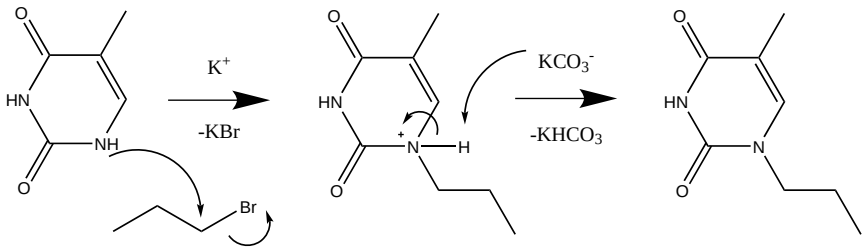


Figure 2.3: Mechanism of the reaction of thymine with N-bromopropane.

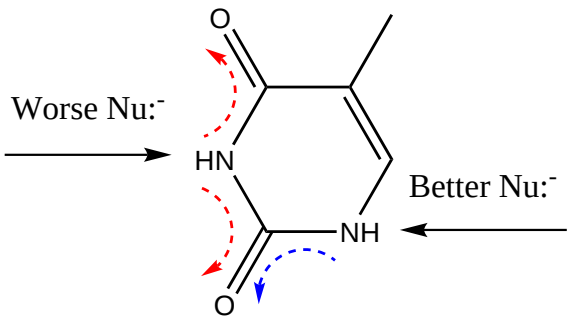


Figure 2.4: Diagram showing how the thymine N-3 is less nucleophilic than N-1 due to the electron-withdrawing effects of the adjacent carbonyl groups.

It is also important to note how the solvent pays a vital role in this reaction. This reaction proceeds via a SN2 mechanism and the nucleophile is much more nucleophilic in polar aprotic solvents than protic solvents. A protic solvent such as methanol would donate a proton which can hydrogen bond to the electron density on the nucleophile and reduce its nucleophilicity. Aprotic solvents such as DMSO, DMF or THF, however, do not donate a proton and the nucleophile is free to react with the electrophile and form the desired product.

This reaction was performed successfully and the desired product was acquired via aqueous work-up to remove the DMSO and column chromatography was used to remove the unwanted N-3 alkylated isomer and dialkylated product which were formed in small amounts. The purity and presence of N-3 propylthymine was ascertained via NMR spectroscopy, Mass spectrometry and single crystal X-ray diffraction. The similarity between the desired N-1 isomer and unwanted N-3 isomer was a particular challenge. Literature reports of the synthesis of N-alkylated thymine compounds report the different Rf values of both isomers. It is often noted that the N-1 isomer has a greater Rf value so is eluted slightly earlier under standard chromatography conditions. To be certain that the correct product has been isolated more quantitative evidence was required. Both isomers have the same mass so mass spectrometry could not be used to distinguish the isomers. The proton and carbon NMR spectra of each compound are also similar with only very slight shifts in the peaks for each isomer. To definitively differentiate between the two isomers 2D HMBC NMR spectroscopy was used. This heteronuclear technique made it possible to easily determine which isomer was present by observing the coupling between the alkyl-group protons and the carbon environments in thymine. In the N-1 isomer the alkyl protons couple to carbons C-2 and C-6. For the N-3 isomer these protons couple to carbons C-2 and C-4. These carbon environments are unambiguously identified as both are in very dissimilar environments (C=O and C=C cf 2x C=O respectively). The HMBC NMR spectra of the N-1 isomer is shown below in Figure 2.5. These different isomers were also confirmed to be correctly assigned by solving the single crystal X-ray diffraction data of each isomer.

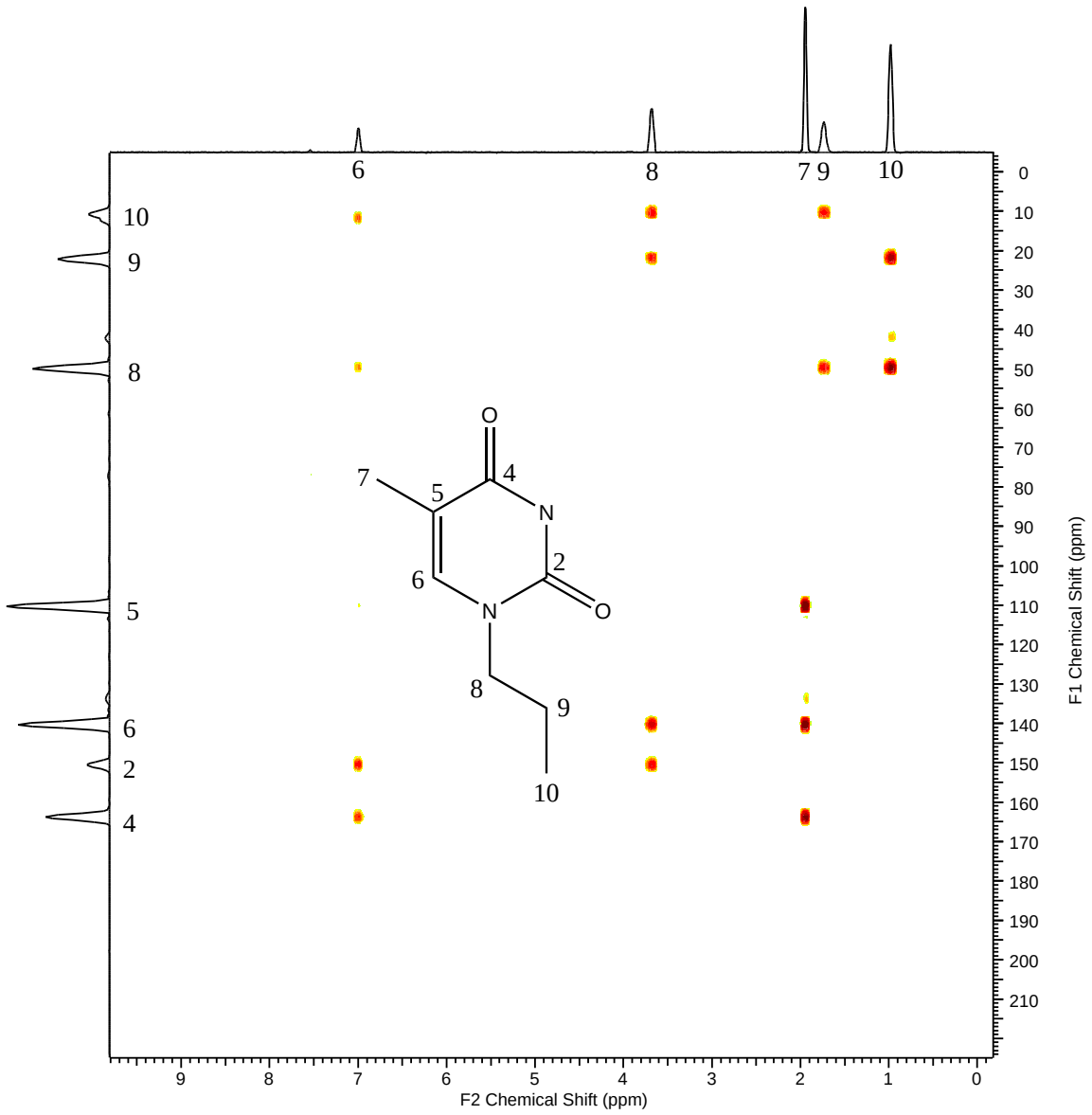
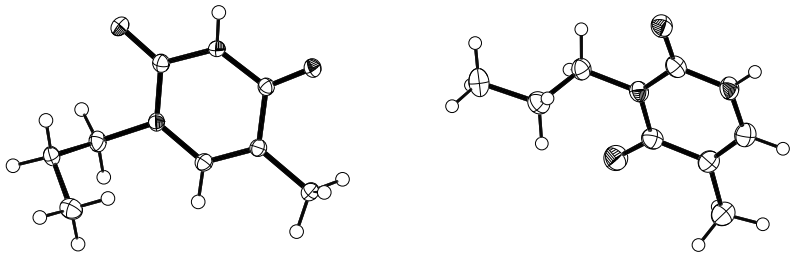


Figure 2.5: HMBC NMR spectrum on *N*-1-propylthymine recorded in $CDCl_3$



(a) (b)

Figure 2.6: Crystal structures of a) *N*-1-propylthymine and b) *N*-3-propylthymine

2.2.2 Propyladenine

The next synthesis of an alkylated DNA nucleobase to be attempted was the synthesis of *N*-9-propyladenine. This reaction proceeds via the same SN2 mechanism as for the previously discussed thymine analogue. However, literature discussions on the alkylation of adenine highlight a number of possible side-products formed during the reaction[105]. The first of these is the unwanted alkylation at the 6-amino position. This unwanted alkylation is overcome with the use of protecting groups. The majority of preparations for *N*-9-alkylated adenine use the Boc protecting group[2]. This protecting group is well known for the protection of amines and was chosen for the synthesis of *N*-9-propyladenine. Other protecting groups used are Fmoc, benzoyl and acetyl, amongst others. The scheme for the synthesis of 6-bis-Boc-amino protected adenine and subsequent alkylation is shown below in Figure 2.7 adapted from literature.[105][2][106][3]

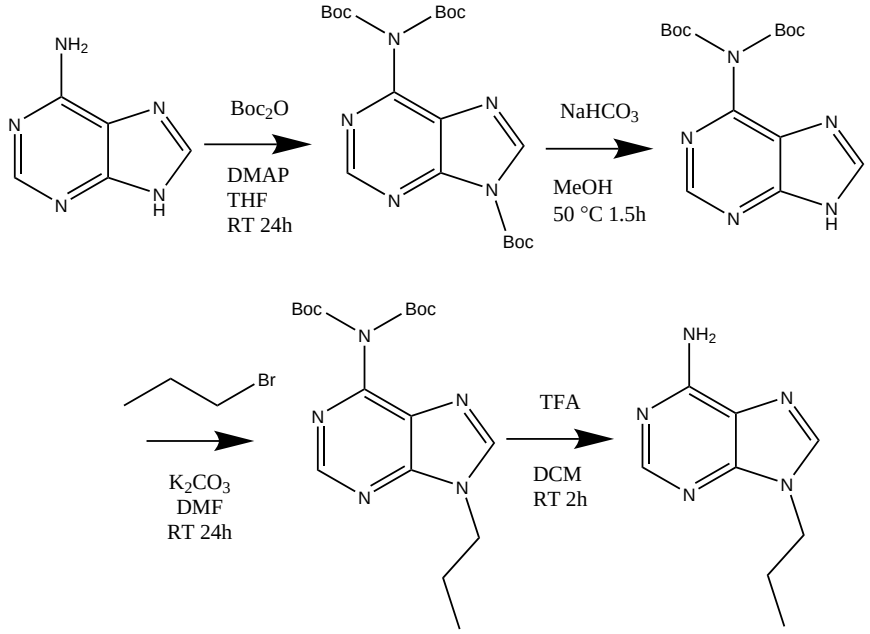


Figure 2.7: Synthesis of *N*-9-propyladenine via the Boc protection route.

The first step of the synthesis of *N*-9-propyladenine was the complete protection of adenine with the Boc protecting group. This was performed by reacting adenine with 3 equivalents of

Boc-anhydride in the presence of 4-(dimethylamino)pyridine (DMAP). Here, DMAP acts as a nucleophilic catalyst by first reacting with Boc-anhydride and creating a more reactive Boc-pyridinium species[3][107]. Adenine then reacts with the Boc-pyridinium species to form the Boc-protected adenine product. The *N*-9-Boc protecting group can then be selectively removed by heating the tris-Boc species in methanol with sodium hydrogen carbonate for 1.5 hours. Care must be taken over the time for this reaction step as leaving the reaction for too long will leave to partial deprotection in the 6-amino position.[2] Once correctly protected and isolated, the 6-bis-Boc aminopurine can be alkylated with *n*-bromopropane as before. It is here that the other side product can be formed. Despite the use of protecting groups alkylation in the unwanted *N*-7 position is observed. The reason for this is due to tautomerisation of the *N*-7, C-8 double bond. This allows *N*-7 to react with n-bromopropane and form the undesired product as shown in Figure 2.8. This tautomerisation and formation of both the *N*-7 and *N*-9 isomers is well-reported in literature. Reported ratios for the formation of the *N*-7 and *N*-9 isomers range between 1:9 and 4:5. In this instance the isomers were formed in a ratio of 1:8 as calculated from the ^{1}H NMR integrals of the crude mixture. The two isomers were not separated at this stage and were both carried onto the final step of the syntheses which was the deprotection of the 6-amino position. The Boc groups were succesfully cleaved via addition of trifluoroacetic acid (TFA) to a solution of the bis-Boc-6-aminopurine in DCM. Upon removal of the Boc groups the *N*-7 and *N*-9 isomers were separated via column chromatography. The *N*-9-propyladenine product was characterised by NMR spectroscopy, mass spectrometry and single crystal X-ray diffraction. The identity of the *N*-7 and *N*-9 isomers was confirmed via HMBC NMR spectroscopy in a similar manner to the *N*-1 and *N*-3 isomers of propylthymine discussed previously as shown below in Figure 2.9.

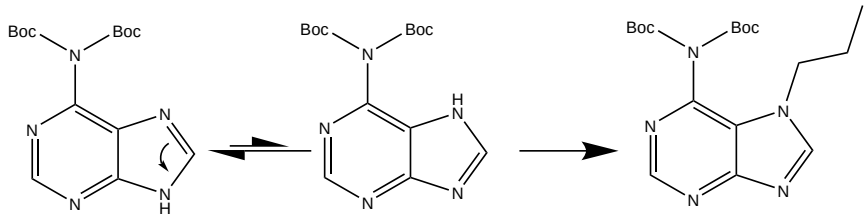


Figure 2.8: The undesired *N*-7 isomer formed via tautomerisation.

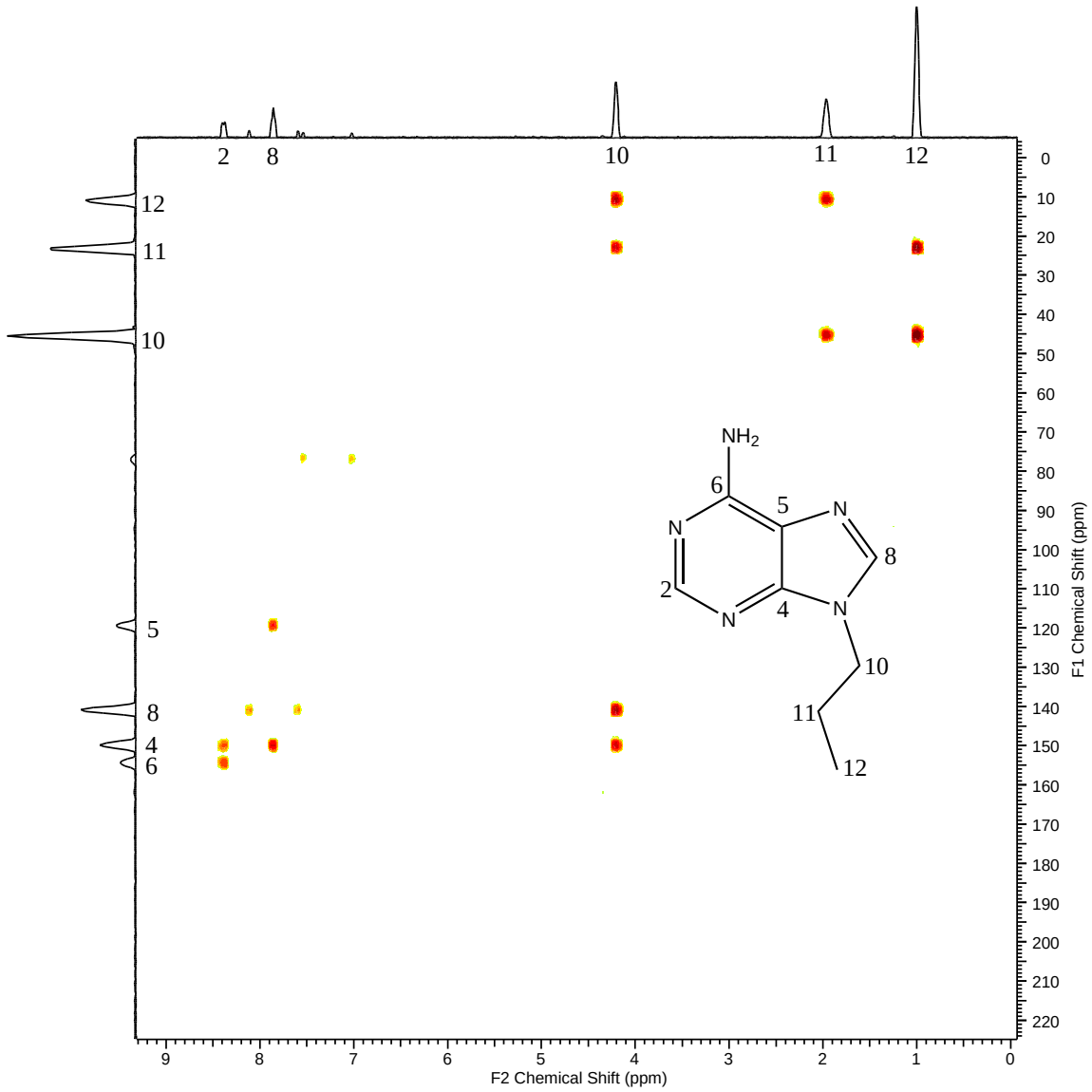


Figure 2.9: The HMBC NMR spectrum of N-9-propyladenine isomer recorded in $CDCl_3$.

2.2.3 Propylcytosine

Literature preparations for the N-1 alkylation of cytosine were particularly difficult to find. With no direct synthetic procedure found the observations from the synthesis of N-1-propylthymine and N-9-propyladenine were applied to the synthesis of N-1-propylcytosine. The 4-amino position was protected with the Boc protecting group in a method adapted from the synthesis of N-9-propyladenine and the N-1 position was subsequently deprotected in the same way. Bis-Boc-4-aminocytosine was then alkylated with n-bromopropane in DMSO and the protected group removed using TFA as shown in Figure 2.10.

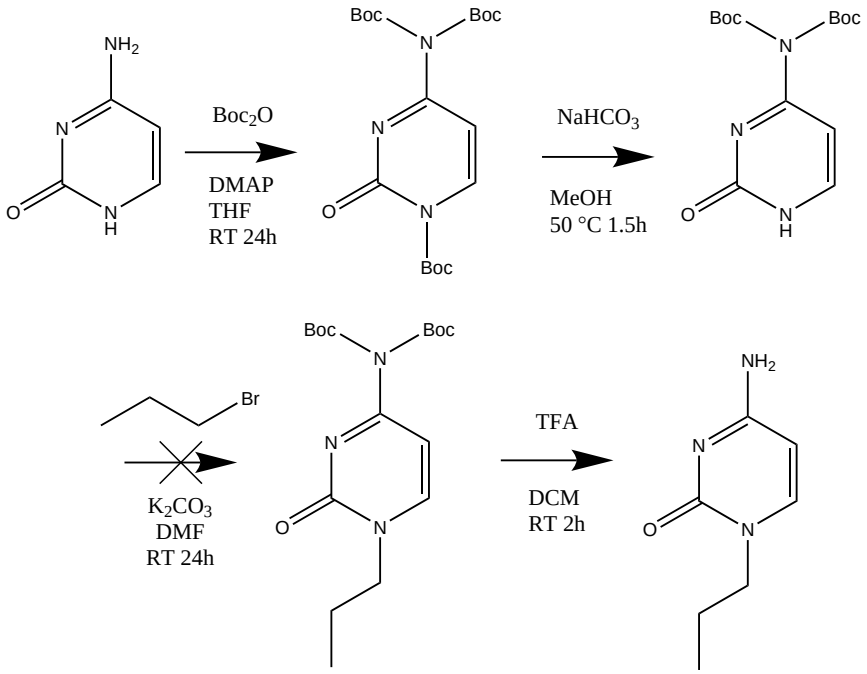


Figure 2.10: Reaction scheme showing the unsuccessful synthetic route to *N*-1-propylcytosine.

Initial observations of the mass spectrum and ^{1}H NMR spectrum suggested the reaction had proceeded as planned as the correct mass was observed and the expected peaks were observed in the ^{1}H NMR spectrum. However, analysis of the HMBC spectrum showed that the correct product had not been formed. For the correct *N*-1 alkylated product it was expected that the primary *N*-1 alkylated CH2 (5.2 ppm) would couple to two carbon environments: the C-2 C=O at ca 160 ppm and the C-6 C=C at ca 130 ppm as shown in Figure 2.12. This was not observed in the HMBC NMR spectrum. Instead the protons were coupled to just a single carbon at ca 160 ppm. This suggested that the an undesired C-2-O-alkylated product had been formed. The formation of this isomer can be explained by considering the keto-enol tautomerisation which may be observed in the DNA nucleobases. This tautomerisation leads to the oxygen becoming the nucleophile which reacts with the *n*-bromopropane and forms the undesired product as shown in Figure 2.11. The identity of this undesired C-2-O isomer was further confirmed by single crystal X-ray diffraction (Figure 2.13).

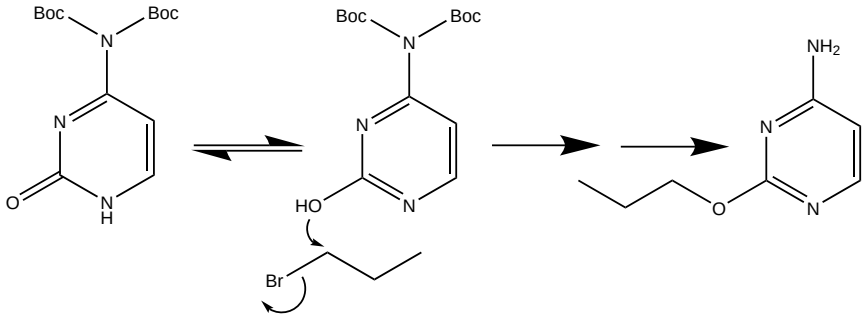


Figure 2.11: The keto-enol tautomerisation of bis-Boc-4-aminocytosine which forms the undesired *C*-2-*O*-propylcytosine.

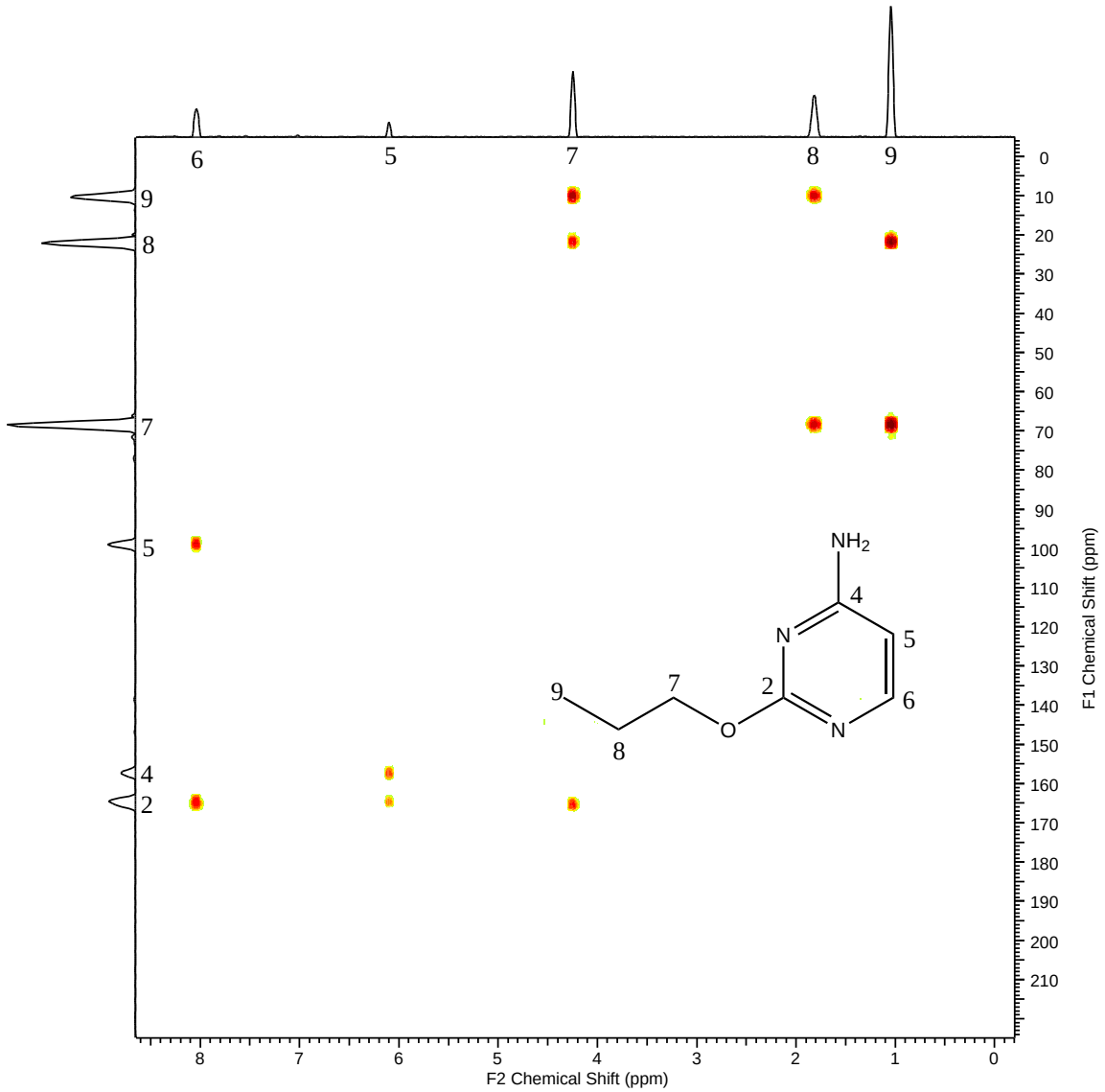


Figure 2.12: HMBC NMR spectrum recorded in $CDCl_3$ showing the undesired *C*-2-*O* alkylated product.

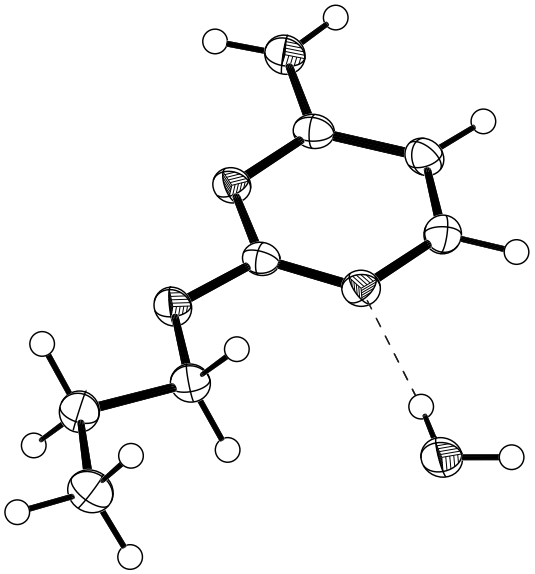


Figure 2.13: Crystal structure of undesired *C*-2-*O* alkylated product.

This reaction was repeated using acetyl-4-aminocytosine[108] and unprotected cytosine and the undesired *O*-alkylated product was formed in each case. The commercially available benzoyl-4-aminocytosine was then used and in this instance the correct *N*-1 alkylated product was formed. The benzoyl protecting group was removed by treatment with ammonium hydroxide to form *N*-1-propylcytosine which was characterised by ^{1}H, ^{13}C and HMBC NMR spectroscopy and mass spectrometry as shown in Figure 2.14 and 2.15.

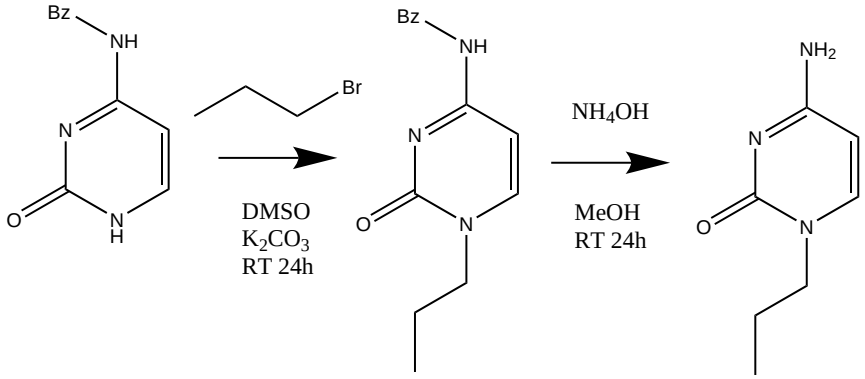


Figure 2.14: The successful synthesis of *N*-1-propylcytosine using the benzoyl protecting group.

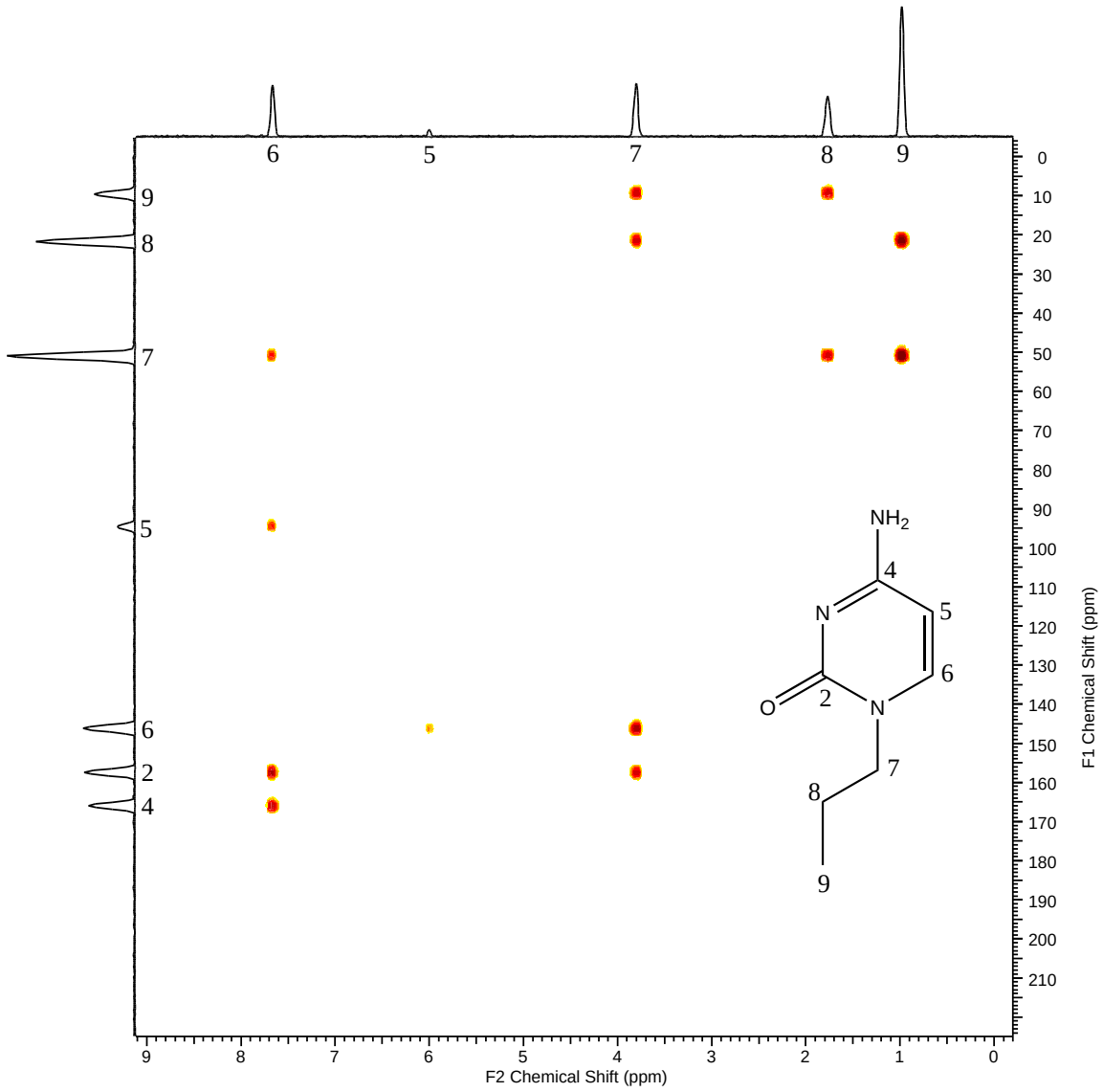


Figure 2.15: HMBC NMR spectrum recorded in $CDCl_3$ showing the desired *N*-1-propylcytosine product.

2.2.4 Propylguanine

The majority of literature reports on the functionalisation of guanine describe the limited solubility of guanine in common laboratory solvents. This makes synthetic procedures more troublesome. It is also clear from the structure of guanine that protecting groups are necessary for the efficient *N*-9-alkylation of guanine. Therefore, 2-amino-6-chloropurine was used as a precursor to guanine and selectively protected with the Boc-protecting group as before to afford 2-(bis-Boc)-amino-6-chloro-9-propylpurine as shown in Figure 2.16.

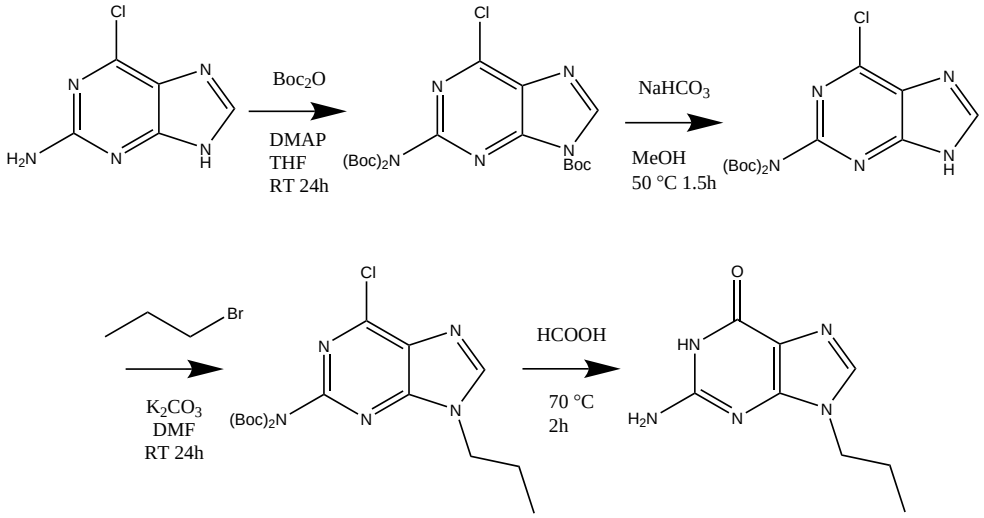


Figure 2.16: The synthesis of N-9-propylguanine via the Boc-protection of 2-amino-6-chloropurine.

The full Boc-protection and selective N-9 deprotection was analogous to the synthesis of propyladenine. The use of the Boc protecting group also has the advantage of vastly increasing the solubility of the substrate in common organic solvents (DCM, EtOAc) due to the bulky t-butyl groups which are part of the Boc protecting group. The alkylation of 2-(bis-Boc)-amino-6-chloropurine with n-bromopropane was also performed successfully and, as expected, both the N-7 and N-9 isomers were formed in a ratio of 1:5 (calculated from the ^{1}H NMR integrals of the crude mixture). These were carried into the next step without separating the two isomers. The final step which involved heating in formic acid served the dual purpose of hydrolising the 6-chloro position and removing the Boc protecting group. It was also possible to isolate 2-amino-6-chloro-9-propylpurine by performing the deprotection reaction using TFA in DCM as these conditions are selective for Boc-deprotection without hydrolising the 6-chloro position. The deprotected N-7 and N-9 isomers were isolated via column chromatography and their identity confirmed by ^{1}H, ^{13}C and HMBC NMR spectroscopy as shown in Figures 2.17, 2.18 and 2.19 as well as mass spectrometry and single crystal X-ray diffraction.

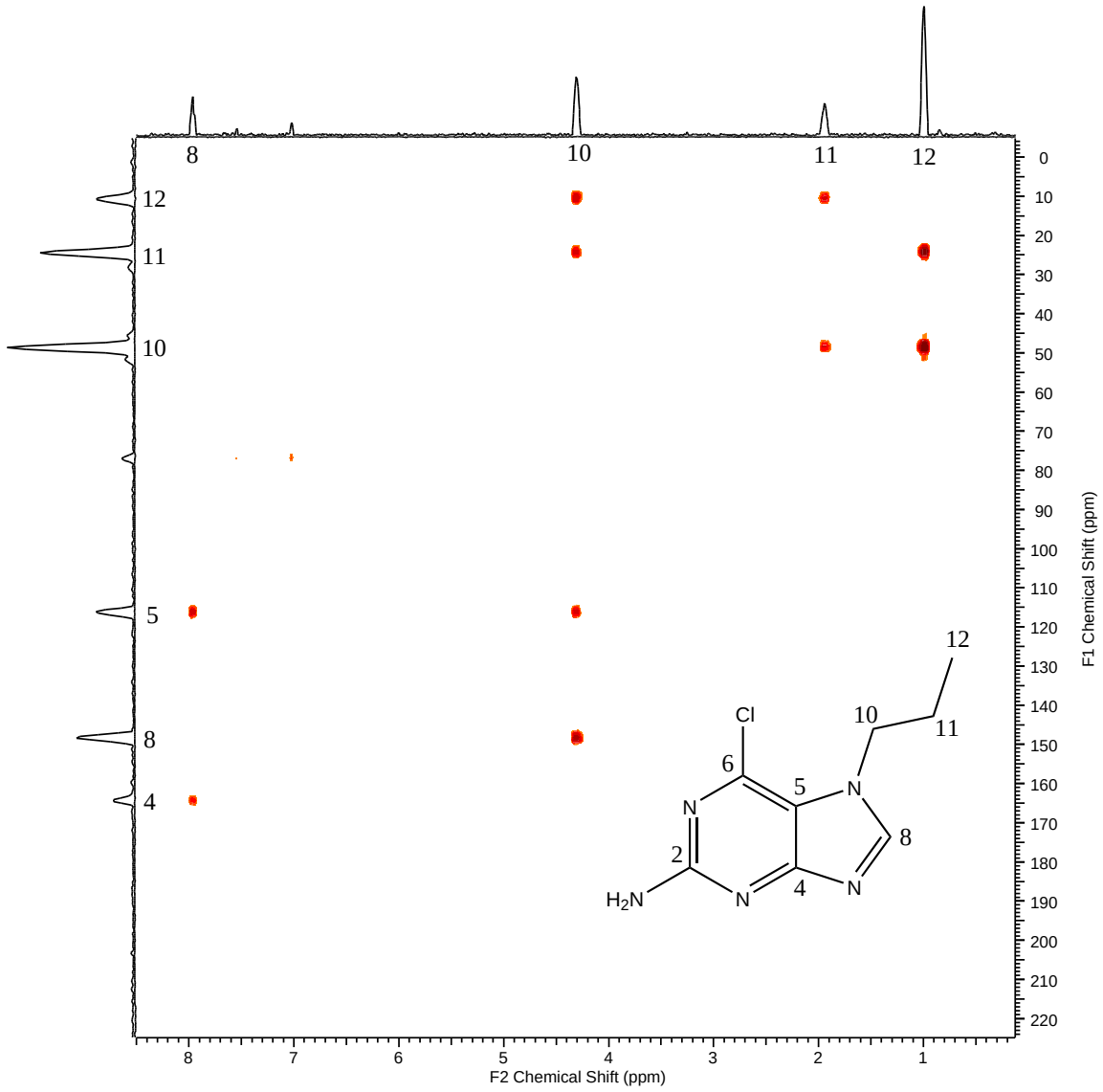


Figure 2.17: HMBC NMR spectrum recorded in $CDCl_3$ showing the *N*-7 2-amino-6-chloropropylpurine isomer.

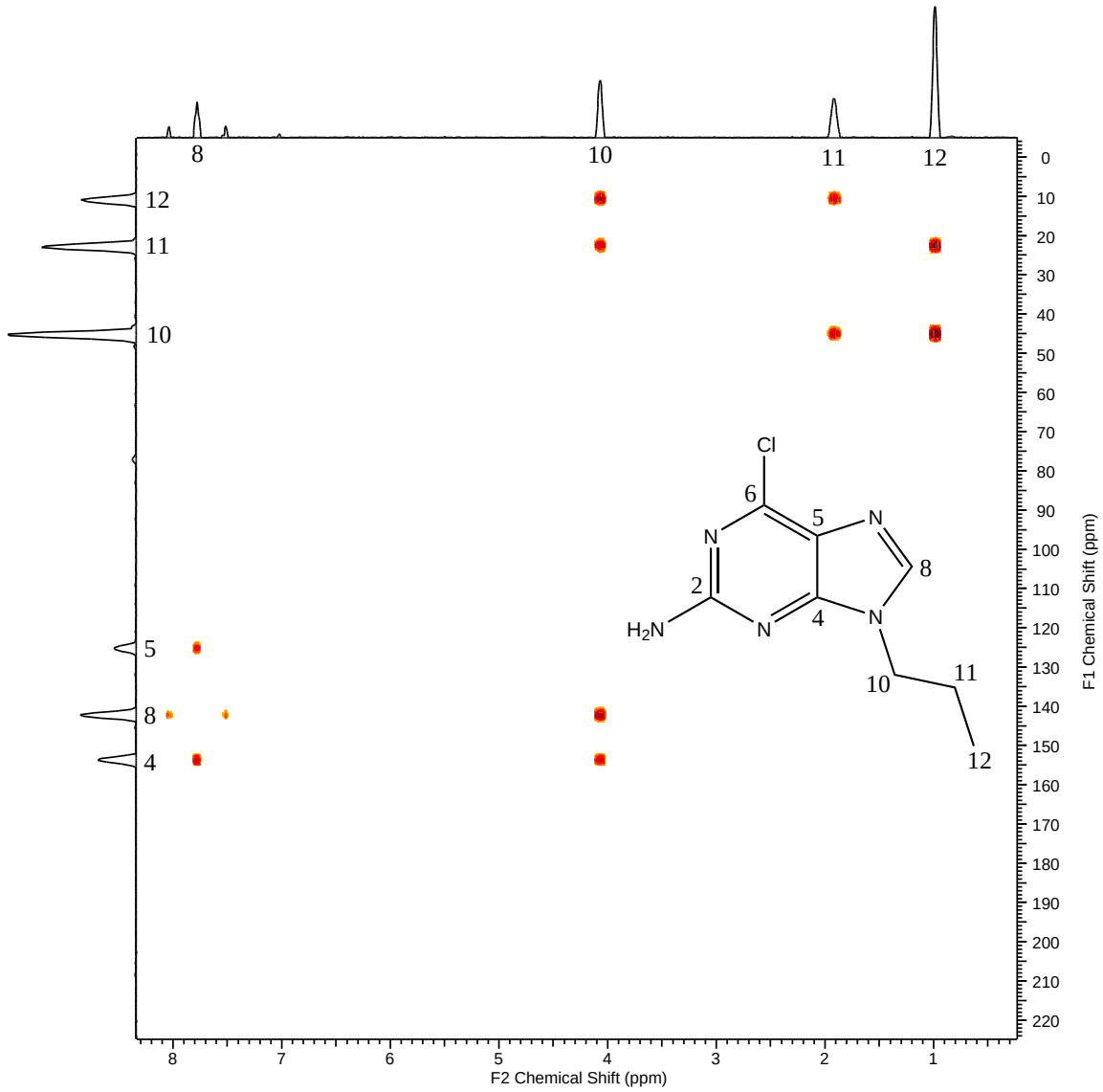


Figure 2.18: HMBC NMR spectrum recorded in $CDCl_3$ showing the *N*-9 2-amino-6-chloropropylpurine isomer.

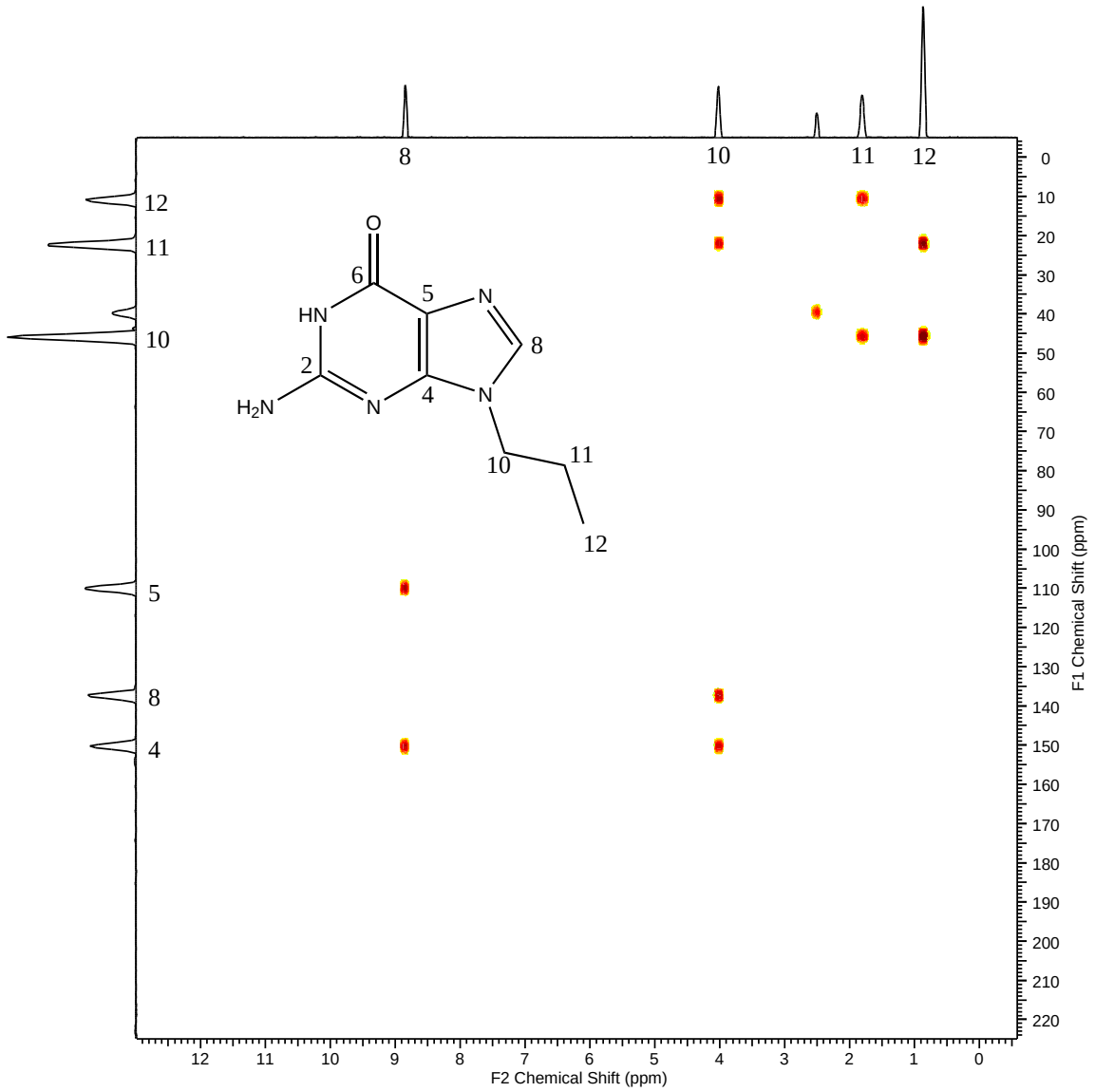


Figure 2.19: HMBC NMR spectrum recorded in DMSO-d showing the *N*-9-propylguanine product.

2.3 Synthesis of Further Functionalised Alkylated Nucleobases

The four DNA nucleobases have been synthesised with a propyl group in the *N*-1 and *N*-9 position. The propyl group was chosen in this position as it was expected to have the least effect on the hydrogen bonding characteristics of the relevant nucleobase. The propyl group is relatively benign with respect to electron donation or withdrawal so should have a minimal effect on the nucleophilicity of the hydrogen bond donors and acceptors (a crucial contributing factor to the strength of a hydrogen bond). The propyl group is also relatively small so should a small steric effect on the formation of the hydrogen bond whilst still imparting sufficient solubility. Alkylation in the *N*-1 and *N*-9 for pyrimidines and purines respectively was also chosen to have a minimal impact on hydrogen bonding due to it not blocking the primary hydrogen bond donors or acceptors and being the same position

where DNA nucleobases are attached to the sugar-phosphate backbone in nature.

With this in mind a number of different substituents were chosen for the four DNA nucleobases which would have differing steric, electronic or positional effects on the molecule. The effect of these substituents were studied by NMR dilution and titration experiments.

Studying the effect of sterics on the hydrogen bonding behaviour of the DNA bases was done by substituting different R-groups in the *N*-1 or *N*-9 positions of the DNA nucleobases. Relatively small groups such as n-ethyl and n-butyl chains through to relatively large groups such as *n*-decyl and benzyl groups were used.

These alkyl and benzyl groups were added to the chosen DNA nucleobase in the same way as described before in Section 2.2 by choosing the relevant alkyl or benzyl halide.

The effect of different electron-withdrawing substituents on the hydrogen bonding behaviour of the DNA bases was done by substituting halogens in various positions of the chosen DNA base. The synthesis of these halogenated DNA bases was more challenging as the addition of halogens also affects the reactivity and regioselectivity of the DNA nucleobase for subsequent reactions so the order and conditions of each reaction had to be carefully considered to achieve the desired target molecules.

2.3.1 Halogenated Thymine

The majority of literature examples where thymine bears a halogen substituent exhibit the halogen in the *C*-5 position[5][6]. In reality this means that the resulting products are in fact halogenated uracil as the halogen is substituted for the *C*-5 methyl group. As uracil and thymine are so similar this was not deemed to be an issue and the ease of adding a halogen in this position made this route the most appealing. For the synthesis of a halogenated thymine derivative the substrate was also alkylated with the *n*-propyl group as before to impart sufficient solubility and for a suitable comparison to the non-halogenated *n*-propylthymine previously studied. Previous literature examples were considered regarding whether to add the halogen first or perform the alkylation reaction first.[6] The bromination of uracil outlined below in Figure 2.20 was the most common procedure found and no examples of the halogenation of *N*-1-alkylated uracil were found.[5] Therefore uracil was first brominated using bromine in DMF at 120 °C (Figure 2.20).

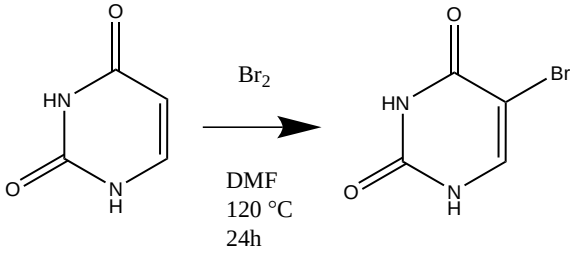


Figure 2.20: The bromination of uracil in the *C*-5 position.

This reaction was highly successful and very simple to perform. The work-up of this reaction was also especially simple and is worth noting. The reaction is simply allowed to cool and chloroform is added to precipitate out the off-white product in excellent yield and purity. The identity of the product was confirmed by mass spectrometry and NMR spectroscopy. HMBC NMR spectroscopy was used to confirm that the desired *C*-5 bromination was achieved and no *C*-6 bromination was observed. This can be explained by studying the mechanism for the reaction as shown in Figure 2.21.

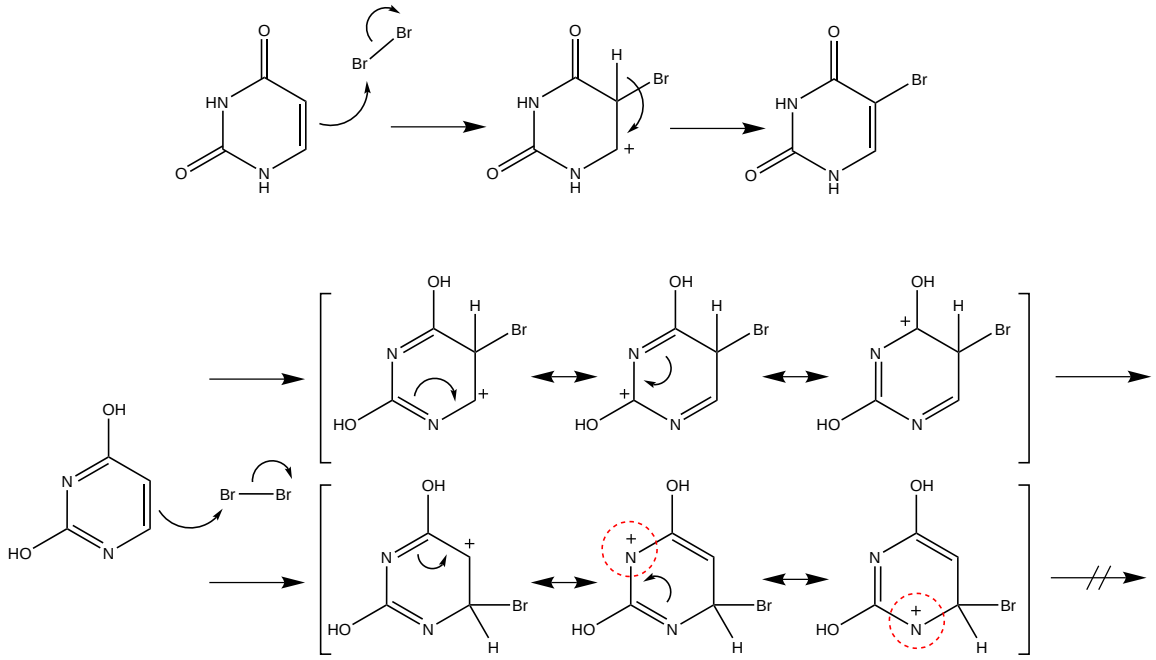


Figure 2.21: The formation of 5-bromouracil via aromatic electrophilic substitution mechanism showing the stable C-5 transition state and the unstable C-6 transition state.

At first glance it would appear that the mechanism for the formation of bromouracil is via an electrophilic substitution of an alkene. However, this does not provide an adequate explanation for the selective regiochemistry of the reaction. It is important to remember the aromatic nature of the pyrimidine ring which can be seen more clearly in the enol tautomer form. The regioselectivity of the electrophilic substitution of pyrimidines in determined by the stability of the transition state [109]. For the C-5 bromination the positive charge resides on C-6. This positive charge can then be stabilised by the conjugated structure. For the C-6 bromination the positive charge on C-5 cannot be stabilised without leading to an unfavoured 6-electron, 2-coordinate nitrogen. The C-5 bromination has the most stable transition state and therefore 5-bromouracil is exclusively formed. Another factor is the well-known ortho-, para-directing property of phenols with respect to electrophilic substitution[109].

The 5-bromouracil formed above was then carried onto the next alkylation step and the previously used alkylation conditions were utilised to attempt to form 5-bromo-N-1-propyluracil. However, this reaction was unsuccessful. The desired N-1-alkylated product was not observed and instead bis-N-1,N-3-propyluracil was exclusively isolated. The reason for this unexpected product is likely to be due to the C-5 bromine having an effect on the reactivity of both the N-1 and N-3 nitrogens. The electron withdrawing bromine may be reducing the nucleophilicity of the N-1 nitrogen and reducing the selectivity between the two positions.

To solve this issue the possibility of performing the alkylation first and then brominating the N-1-alkylated product was considered. Looking at the mechanism for the bromination it can be seen that this route would not be feasible. If the N-1 position was alkylated then the stabilisation of the C-6 positive charge would not be possible and the transition state would no longer be stable. This would greatly limit the success of the bromination reaction. Furthermore, the work-up of precipitating the product out with chloroform would no longer be applicable as the alkylated product would likely be soluble in chloroform. Despite this, uracil was alkylated using standard alkylation conditions nonetheless as N-1-propyluracil is still an interesting product in itself and can be used to study the

effect of the thymine C-5 methyl group.

The alkylation reaction was attempted using a number of different polar aprotic solvents including DMSO, DMF and THF and in all cases the di-alkylated product was formed. Temperatures ranging between -10 and 120 °C were attempted with the same result. Polar protic solvents and non-polar solvents were also used and no reaction was observed as expected. Finally, the base was altered and the use of sodium hydroxide was found to form the desired mono-alkylated 5-bromo-N-1-propyluracil which was identified by ^{1}H, ^{13}C and HMBC NMR spectroscopy (Fig 2.22) as well as mass spectrometry.

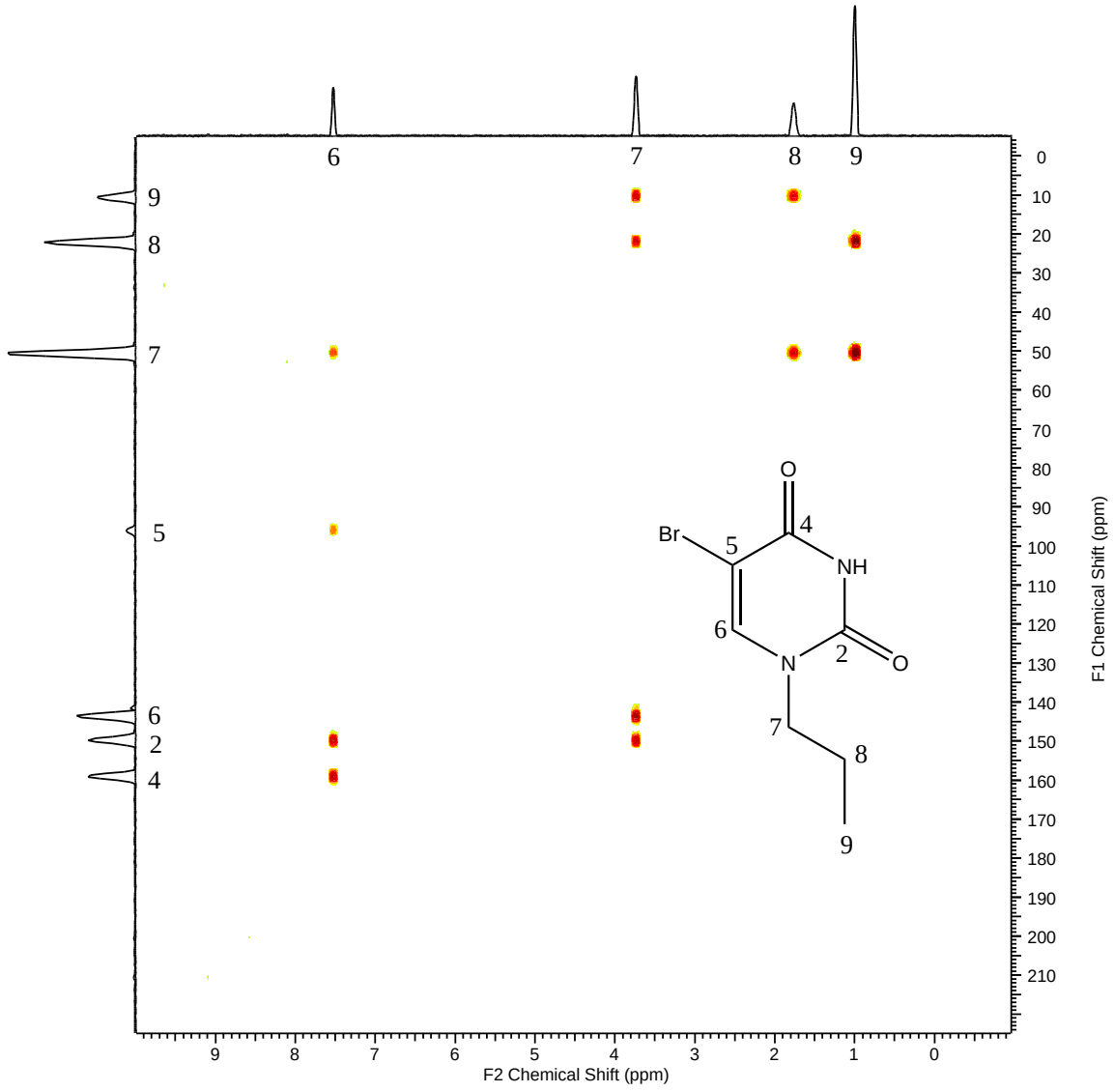


Figure 2.22: HMBC NMR spectrum of 5-bromo-N-1-propyluracil recorded in DMSO-d.

2.3.2 Halogenated Adenine

Due to the different purine substructure of adenine compared to pyrimidine substrucure of thymine it was expected that different reaction conditions would be required for the halogenation of adenine.

Literature examples on the halogenation of adenine report the alkylation reaction being performed first.[6] Therefore, the previously alkylated adenine synthesised in Section 2.2.2 was used. The halogenation of propyladenine was then performed using the conditions shown below in Figure 2.23.

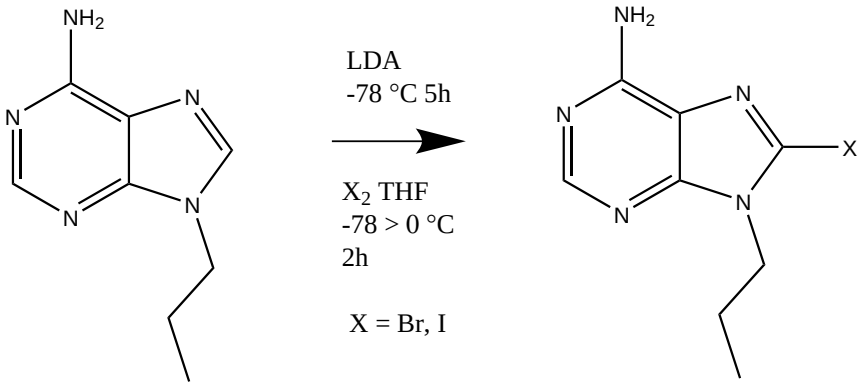


Figure 2.23: The halogenation of adenine via the lithiation of propyladenine with LDA and subsequent reaction with reaction with relevant halogen.[6][8]

In contrast to the halogenation of uracil, which is a direct halogenation, the above reaction proceeds via a lithiated intermediate which then reacts to form the desired halogenated product. The lithiated intermediate is formed by first synthesising lithium diisopropylamide (LDA) from diisopropylamine and n-butyllithium in THF at -78 °C. LDA is a strong, non-nucleophilic base which, upon the addition of adenine readily removes the acidic *C*-8 proton from adenine. The removal of this proton affords the strongly nucleophilic *C*-8 anion which can then react with the electrophilic halogen as shown in the mechanism below (Figure 2.24).

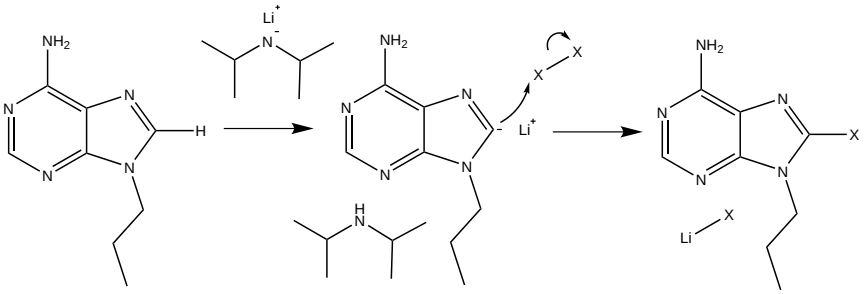


Figure 2.24: The mechanism for the reaction of lithiated propyladenine with the appropriate halogen.
.

This reaction proceeded very efficiently in excellent yield; selectively forming the desired C-8 halogenated adenine. The correct regiochemistry was confirmed by ^{1}H, ^{13}C and HMBC NMR spectroscopy, mass spectrometry and X-ray crystallography.

2.3.3 Halogenated Cytosine

Due to cytosine being based on the same pyrimidine structure as thymine and uracil a similar approach for the synthesis of halogenated cytosine was adopted[2]. Cytosine was first halogenated and then the alkylation was performed. Cytosine was halogenated using iodine and iodic acid in acetic acid as shown in Figure 2.25. This reaction was selective for the *C*-5 position for the

same reason as for 5-bromouracil; that being the stability of the C-6 positive charge. The 5-iodocytosine was then successfully alkylated using n-bromopropane under the same conditions as before. Interestingly, this reaction did not form the undesired O-alkylated product which was observed for the alkylation of cytosine. It is likely that the iodine substituent has an effect on this reactivity.

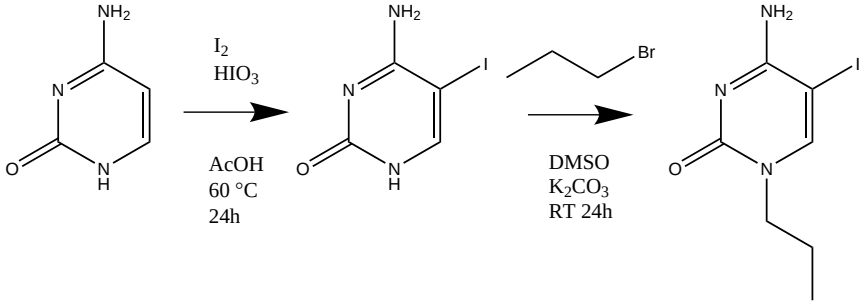


Figure 2.25: The C-5 halogenation of cytosine and subsequent alkylation .

2.3.4 Halogenated Guanine

For halogenated guanine the route previously outlined for halogenated adenine was attempted with the aim of synthesising 6-chloro-2-amino-8-halo-N-9-propylpurine. This route was chosen due to the analogous nature of the guanine precursor 6-chloro-2-aminopurine to adenine. The halogenation of 6-chloro-2-amino-N-9-propylpurine using the previously utilised conditions shown below (Figure 2.26) was attempted.

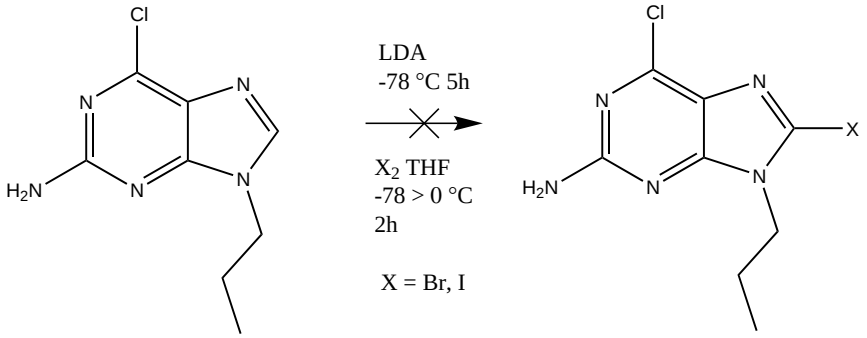


Figure 2.26: The attempted halogenation of 6-chloro-2-aminopurine using LDA which was unsuccessful.

After completing the reaction and work up it was found that all the starting material had been consumed. However, the desired halogenated product was not found. That the halogenation was unsuccessful was initially surprising due to the excellent yield and success of the analogous reaction with adenine. However, a study of the literature and notably a report by Gundersen et al revealed how the subtle differences between 6-chloro-2-aminopurine and adenine can, with an understanding of the reaction mechanism, lead to greatly differing products being formed.[8]

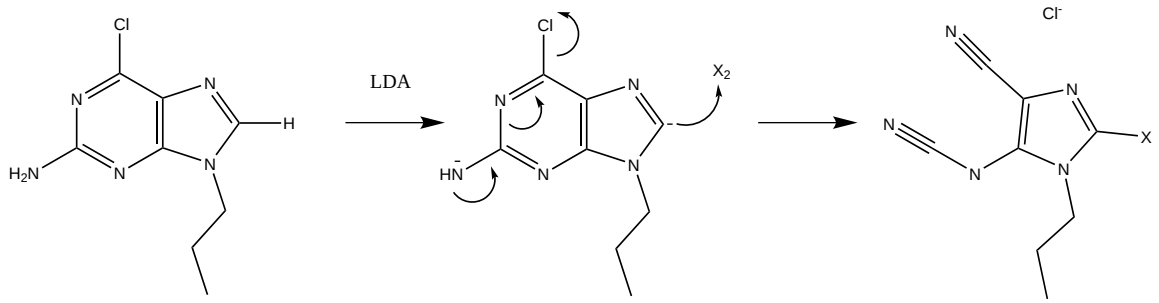


Figure 2.27: Mechanism showing how treatment with LDA leads to the undesired imidazole compound via loss of chloride anion.

In the mechanism above, the very strong LDA base removes any acidic protons. As well as the desired acidic C-8 proton the proton in the 2-amino position is also removed (this happens during the successful adenine halogenation also but the proton is regained during the aqueous work-up). The negative charge on the C-8 anion can attack the electrophilic halogen as desired. However, the negative charge in the 2-amino position can attack through the conjugated system onto the chloride on C-6 as shown in the mechanism above (Figure 2.27. This is possible due to chloride being an excellent leaving group. This breaks the ring-sturucture and forms two nitrile bonds. This is not possible in the case of adenine due to no suitable leaving group being present.

The formation of the undesired imidazole product meant that another method for the halogenation of guanine derivatives was needed. Gundersen *et al.* report a number of different methods for the C-8 halogenation of 2-amino-6-chloropurine derivatives including the protection of the amine group with tetrahydropuran (THP).[8] This technique prevents the anion being formed upon treatment with LDA due to the added steric hindrance of the THP group and the fact that LDA is a relatively bulky base due to the isopropyl groups. Another method reported by Gundersen *et al.* in a separate report is the direct bromination of 2-amino-6-chloropurine derivatives using excess bromine in water at room temperature.[7] This route was chosen as it appeared the most attractive due to its obvious simplicity and the lack of protection and deprotection steps.

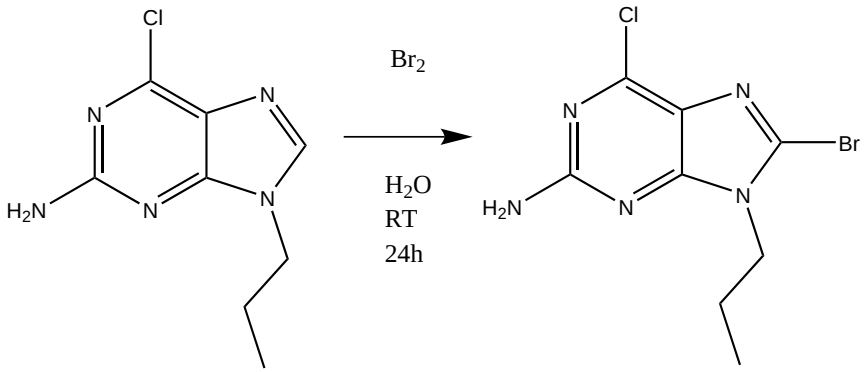


Figure 2.28: The successful conditions for the halogenation of 2-amino-6-chloropurine.

This reaction was performed successfully and the product was easily extracted by allowing excess bromine to evaporate in the fumehood and filtering the suspension to afford the desired product in good yield and purity. The product was characterised by NMR spectroscopy, mass spectrometry and X-ray crystallography.

2.4 Conclusions

In this chapter a large number of alkylated and halogenated nucleobases have been synthesised and characterised as listed in the table below (Table 2.1). Progress has been made towards the efficient synthesis of nucleobases which are functionalised with the chosen alkyl group and halogen and the process has been optimised for each nucleobase. An important conclusion from this work is that the alkylation and halogenation of nucleobases is not trivial and different conditions are required for each base. This includes optimization of variables including the solvent, protecting group, reaction conditions and even order of reaction steps. For each nucleobase a number of isomers are also formed; the identity of which is not easily discernible by simple 1D NMR and mass spectrometry techniques. A reliable and fast method of identifying the correct isomer by 2D HMBC NMR has been outlined, as well as how to separate the desired isomers from the mixture. These alkylated nucleobases display improved solubility in non-polar organic solvents such as DCM as expected and this will allow for their intermolecular interactions to then to be studied in a non-competitive solvent. However, for the cytosine and guanine nucleobases sufficient solubility was not imparted by the alkylation. The synthesis methods outlined in this chapter can be used to synthesise cytosine and guanine nucleobase derivatives with different functionalities in the hope that these will be sufficiently soluble in the chosen solvent. The halogenation of nucleobase derivatives again shows how each nucleobase is remarkably different. Not one method for the halogenation of the these nucleobase derivatives was the same and markedly different synthetic procedures were required for each desired compound. This was explained by understanding the mechanism of each reaction and how the subtle changes in the core pyrimidine and purine structures of the DNA bases affect this. The effect of adding a halogen to the intermolecular interactions of the DNA nucleobases can be studied. Finally, the halogen itself is a useful functional group to allow for further synthetic modifications of the DNA nucleobases to be carried out.

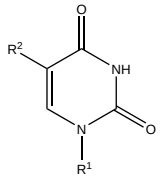
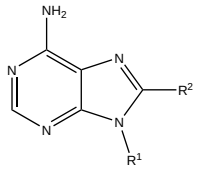
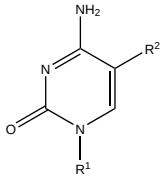
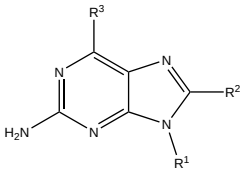
									
	R1	R2	R1	R2	R1	R2	R1	R2	R3
a	*n*-pr	Me							
b	*n*-et	Me							
c	*n*-bu	Me							
d	*n*-pent	Me							
e	*n*-hex	Me							
f	*n*-oct	Me							
g	*n*-dec	Me							
h	benz	Me							
i	*t-bu*-benz	Me							
j	*n*-pr	H							
k	*n*-pr	Br							
l	*t-bu*-benz	Br							
m			*n*-pr	H					
n			*n*-bu	H					
o			benz	H					
p			*t-bu*-benz	H					
q			*n*-pr	Br					
r			*n*-pr	I					
s					*n*-pr	H			
t					*t-bu*-benz	H			
u					*n*-pr	I			
v							*n*-pr	H	OH
w							*t-bu*-benz	H	OH
x							*n*-pr	H	Cl
y							*t-bu*-benz	H	Cl
z							*n*-pr	Br	Cl

Table 2.1: Table of all synthesised DNA base analogues.

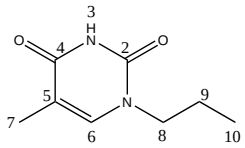
2.5 Experimental

2.5.1 Materials and Equipment

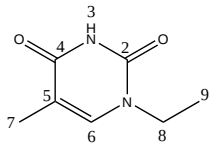
Chemicals were purchased from commercial suppliers and used without further purification. Solid, hygroscopic reagents were dried in an oven before use. Reaction solvents were thoroughly dried before use employing standard methods. Reactions were monitored by thin layer chromatography (TLC) using 0.2 mm aluminium sheets precoated with silica gel 60 F254 (Merck). TLC plates were inspected with a UV lamp featuring both long wave (365 nm) and short wave (254 nm) UV light. Column chromatography was carried out on silica gel Merck 60. Eluent relative vol/vol ratios are indicated in each case. Nuclear magnetic resonance spectroscopy were recorded with a Bruker AV-400 (400 MHz) or Bruker AV-3400 (400 MHz) instrument in the School of Chemistry. The temperature was controlled at 298 K. Chemical shifts (δ) are measured in part per million (ppm) using the signals of the deuterated solvents as the internal standard. EI Mass spectra were taken using a Bruker Apex IV 4.7 T mass spectrometer.

2.5.2 General Procedure for the Alkylation of Pyrimidine and Purine Bases[1]

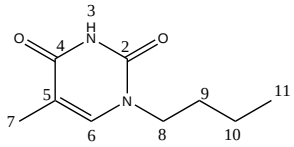
Pyrimidine base (1 g) and potassium carbonate (1.1 eq.) were added to a round bottom flask and dissolved in DMSO (or DMF) (20 ml). Alkyl halide (1.1 eq.) was added via syringe and the solution was stirred overnight at room temperature. The product was washed with water (3x150 ml) and extracted into DCM (3x100 ml). The organic phase was dried over $MgSO_4$, the solvent removed under reduced pressure and the product purified via column chromatography (SiO_2 DCM:EtOAc 10:1 (v:v)) to yield a white powder.



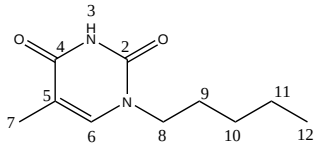
N-1-*n*-propylthymine: (1.1 g, 83 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 9.44 (br. s., 1H, N3), 6.98 (d, J=1.3 Hz, 1H, C6), 3.66 (t, J=7.5 Hz, 2H, C8), 1.92 (d, J=1.1 Hz, 3H, C7), 1.71 (sxt, J=7.4 Hz, 2H, C9), 0.95 (t, J=7.5 Hz, 3H, C10); MS (EI): *m/z+* Calc. = 169.0972 C8H13N2O2 (M+H), Exp. = 169.0978 (M+H), 191.0769 (M+Na)



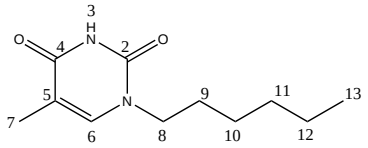
N-1-*n*-ethylthymine: (0.9 g, 74 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.50 (br. s., 1H, N3), 7.00 (d, J=1.1 Hz, 1H, C6), 3.77 (q, J=7.3 Hz, 2H, C8), 1.94 (d, J=1.1 Hz, 2H, C7), 1.31 (t, J=7.2 Hz, 3H, C9); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 164.1 (C4), 150.5 (C2), 140.0 (C6), 110.8 (C5), 43.6 (C8), 14.4 (C7), 12.3 (C9); MS (EI): *m/z+* Calc. = 155.1765 C7H11N2O2 (M+H), Exp = 177.0629 (M+Na)



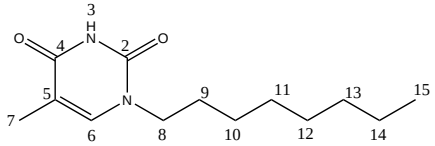
N-1-*n*-butylthymine: (1.2 g, 83 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.65 (br. s., 1H, N3), 6.98 (d, J=1.1 Hz, 1H, C6), 3.70 (t, J=7.4 Hz, 2H, C8), 1.93 (d, J=1.1 Hz, 3H, C7), 1.67 (quin, J=7.4 Hz, 2H, C9), 1.37 (sxt, J=7.4 Hz, 2H, C10), 0.96 (t, J=7.4 Hz, 3H, C11); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 164.1 (C4), 150.7 (C2), 140.4 (C6), 110.5 (C5), 48.3 (C8), 31.1 (C9), 19.7 (C10), 13.6 (C7), 12.3 (C11)



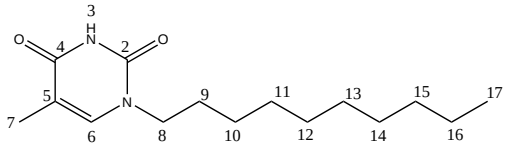
N-1-*n*-pentylthymine: (1.3 g, 83 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 9.07 (br. s., 1H, N3), 6.98 (s, 1H, C6), 3.69 (t, J=7.4 Hz, 2H, C8), 1.93 (s, 3H, C7), 1.68 (quin, J=7.3 Hz, 2H, C9), 1.15 - 1.43 (m, 4H, C10-11), 0.91 (t, J=6.8 Hz, 3H, C12); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 164.3 (C4), 150.9 (C2), 140.4 (C6), 110.5 (C5), 48.5 (C8), 28.7 (C9), 28.5 (C10), 22.3 (C11), 13.9 (C7), 12.3 (C12)



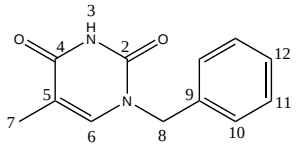
N-1-*n*-hexylthymine: (1.5 g, 90 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.27 (br. s., 1H, N3), 6.98 (d, J=1.3 Hz, 1H, C6)), 3.69 (t, J=7.5 Hz, 2H, C8), 1.93 (d, J=1.1 Hz, 3H, C7), 1.68 (quin, J=7.0 Hz, 2H, C9), 1.22 - 1.42 (m, 6H, C10-12), 0.90 (t, J=6.8 Hz, 3H, C13)



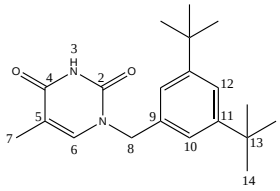
N-1-*n*-octylthymine: (1.6 g, 86 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.55 (br. s., 1H, N3), 6.98 (d, J=1.3 Hz, 1H, C6), 3.69 (t, J=7.4 Hz, 2H, C8), 1.93 (d, J=1.1 Hz, 3H, C7), 1.53 - 1.75 (m, 2H, C9), 1.17 - 1.40 (m, 10H, C10-14), 0.89 (t, J=7.2 Hz, 3H, C15); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 164.0 (C4), 150.7 (C2), 140.4 (C6), 110.5 (C5), 48.6 (C8), 31.7 (C9), 29.1 (C10), 29.1 (C11), 26.4 (C12), 22.6 (C13-14), 14.0 (C7), 12.3 (C15)



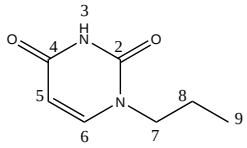
N-1-*n*-decylthymine: (1.5 g, 71 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.25 (br. s., 1H, N3), 6.98 (d, J=1.1 Hz, 1H, C6), 3.69 (t, J=7.4 Hz, 2H, C8), 1.93 (d, J=1.1 Hz, 3H, C7), 1.68 (quin, J=7.1 Hz, 2H, C9), 1.16 - 1.44 (m, 16H, C10-16), 0.89 (t, J=6.7 Hz, 3H, C17); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 163.9 (C4), 150.6 (C2), 140.4 (C6), 110.5 (C5), 48.6 (C8), 31.8 (C9), 29.5 (C10), 29.4 (C11), 29.2 (C12), 29.2 (C13), 29.1 (C14), 26.4 (C15), 22.6 (C16), 14.1 (C7), 12.3 (C17)



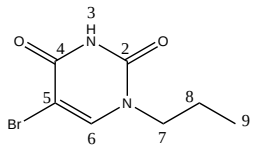
N-1-benzylthymine: (1.5 g, 87 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.57 (br. s., 1H, N3), 7.34 - 7.43 (m, 3H, C11-12), 7.30 (dd, J=7.8, 1.8 Hz, 2H, C10), 6.98 (d, J=1.3 Hz, 1H, C6), 4.90 (s, 2H, C8), 1.89 (d, J=1.3 Hz, 3H, C7); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 163.8 (C4), 151.0 (C4), 139.7 (C6), 135.4 (C9), 129.1 (C11), 128.5 (C10), 127.9 (C12), 111.2 (C5), 50.9 (C8), 12.4 (C7)



N-1-(3,5-di-*t*-butylbenzyl)thymine: (2.3 g, 88 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.71 (br. s., 1H, N3), 7.41 (s, 1H, C12), 7.11 (s, 2H, C10), 7.00 (s, 1H, C6), 4.88 (s, 2H, C8), 1.89 (s, 3H, C7), 1.32 (s, 18H, C14); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 164.0 (C4), 151.7 (C2), 151.0 (C11), 139.8 (C6), 134.5 (C9), 122.5 (C12), 122.2 (C10), 110.9 (C5), 51.3 (C8), 34.8 (C13), 31.4 (C14), 12.4 (C7); MS (EI) *m/z*+ Calc. = 329.2224 C20H29N2O2 (M+H), Exp = 329.2045 (M+H)

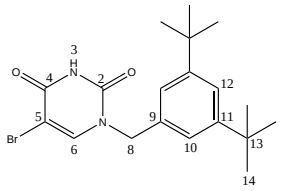


N-1-propyluracil: (1 g, 71 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.51 (br. s., 1H, N3), 7.15 (d, J=7.9 Hz, 1H, C6), 5.70 (dd, J=7.9, 2.4 Hz, 1H, C5), 3.70 (t, J=7.7 Hz, 1H, C7), 1.74 (sxt, J=7.4 Hz, 2H, C8), 0.98 (t, J=7.4 Hz, 3H, C9); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 163.4 (C4), 150.6 (C2), 144.4 (C6), 102.0 (C5), 50.4 (C7), 22.3 (C8), 10.9 (C9); MS (EI) *m/z*+ Calc. = 155.0815 C7H11N2O2 (M+H), Exp. 155.0950 (M+H)

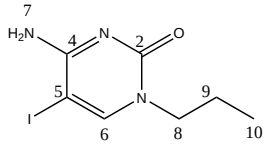


N-1-propyl-5-bromouracil: (0.3 g, 24 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.81 (br. s., 1H, N3), 7.51 (s, 1H, C6), 3.72 (t, J=7.4 Hz, 2H, C7), 1.75 (sxt, J=7.4 Hz, 2H, C8), 0.99 (t, J=7.4 Hz, 3H, C9); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 159.2 (C4), 150.0 (C2), 143.8 (C6), 96.2 (C5),

50.8 (C7), 22.4 (C8), 10.8 (C9); MS (EI) *m/z+* Calc. = 232.9920 C7H10BrN2O2 (M+H), Exp. = 232.9921 (M+H), 254.9744 (M+Na), 250.0188 (M+NH4)



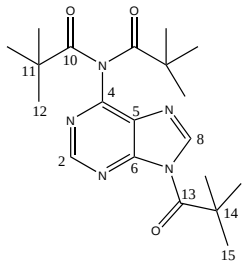
N-1-((3,5-di-*t*-butylbenzyl)-5-bromouracil: (0.9 g, 43 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.39 (br. s., 1H, N3), 7.49 (s, 1H, C12), 7.45 (t, *J*=1.8 Hz, 1H, C6), 7.12 (d, *J*=1.8 Hz, 2H, C10), 4.91 (s, 2H, C8), 1.34 (s, 18H, C14); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 158.9(C4), 152.1 (C2), 150.1 (C11), 143.2 (C6), 133.5 (C9), 123.1 (C12), 122.5 (C10), 96.6 (C5), 52.1 (C8), 34.9 (C13), 31.4 (C14); MS (EI) *m/z+* Calc. = 393.1172 C19H26BrN2O2 (M+H), Exp. = 393.1188 (M+H), 415.0985 (M+Na)



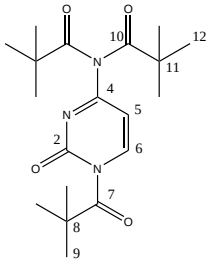
N-1-propyl-5-iodocytosine: (0.7 g, 59 %) ^{1}H NMR (DMSO-d6 ,400MHz): δ (ppm) 8.08 (s, 1H, C6), 6.44 (br. s., 1H, C7), 3.60 (t, *J*=7.2 Hz, 2H, C8), 1.56 (sxt, *J*=7.3 Hz, 2H, C9), 0.81 (t, *J*=7.4 Hz, 3H, C10); ^{13}C NMR (DMSO-d6 ,101MHz): δ (ppm) 163.9 (C4), 154.6 (C2), 152.0 (C6), 55.1 (C5), 48.6 (C8), 21.9 (C9), 10.7 (C10); MS (EI) *m/z+* Calc. = 279.9941 C7H11IN3O (M+H), Exp. = 279.9943 (M+H), 301.9754 (M+Na)

2.5.3 General Procedure for the tris-Boc Protection of Pyrimidine and Purine Bases[2]

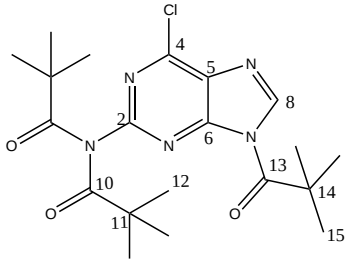
To a N_2 flushed flask containing purine or pyrimidine base (10.0 mmol) and DMAP (0.122 g, 1.00 mmol) was added dry THF (50 ml) via cannula. Boc_2O (9.6 g, 40 mmol)) was added under a N_2 atmosphere and the colourless suspension was stirred overnight at room temperature until a pale yellow solution was formed. The solvent was removed under reduced pressure to give a yellow oil which was used without further purification. Rf (0.45 7:3 hexane:EtOac)



Tris-Boc-adenine: (3.9 g, 92 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.97 (s, 1H, C2), 8.48 (s, 1H, C8), 1.68 (s, 9H, C15), 1.39 (s, 18H, C12); MS (EI) *m/z+* Calc. = 436.2191 C20H30N5O6 (M+H), Exp. = 436.2201 (M+H), 458.2023 (M+Na)



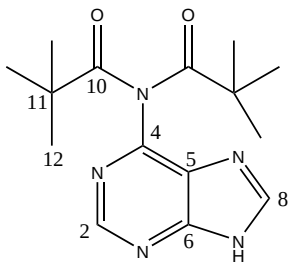
Tris-Boc-cytosine: 3.8 g, 92 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 7.97 (d, J = 7.9 Hz, 1H, C6)), 7.07 (d, J = 7.8 Hz, 1H, C5), 1.60 (s, 9H, C9), 1.56 (s, 18H, C12); ^{13}C NMR ($CDCl_3$, 100MHz): δ (ppm) 162.3 (C10), 160.0 (C7), 150.1 (C4), 149.0 (C2), 143.3 (C6), 96.8 (C5), 85.2 (C8, C11), 27.5 (C9, C12); MS (EI) m/z+ Calc. = 411.2006 C19H29N3O7 (M+), Exp = 411.1907 (M+).



Tris-Boc-2-amino-6-chloropurine: (4.3 g, 91 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.58 (s, 1 H), 1.69 (s, 9 H), 1.47 (s, 18 H), ^{13}C NMR ($CDCl_3$, 100MHz): δ (ppm) 153.0 (C13), 152.7 (C10), 151.2 (C6), 149.9 (C2), 148.7 (C4), 144.3 (C8), 139.6 (C5), 87.4 (C14), 83.2 (C11), 27.2 (C12,C15)

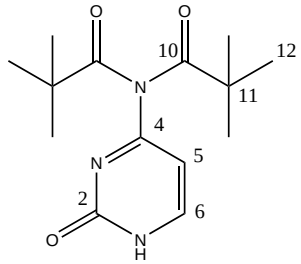
2.5.4 General Procedure for the bis-Boc Protection of Pyrimidine and Purine Bases[2]

The unpurified tris-Boc nucleobase (9.2 mmol) was dissolved in EtOAc (400 ml) and washed with 2M HCl (50 ml), followed by brine (200 ml). The organic phase was dried over $MgSO_4$ and the solvent removed under reduced pressure. The resulting colourless oil was dissolved in MeOH (100 ml) and $NaHCO_3$ sat. (50 ml) was added. The turbid solution was stirred at 50 C for one hour until complete conversion was observed by TLC RF (0.48 EtOAc). The MeOH was removed under reduced pressure and the aqueous layer was diluted with water and extracted with DCM. The organic layer was dried over $MgSO_4$ and evaporated to give a white solid. This crude product was purified by column chromatography (SiO_2) to yield bis-Boc-adenine as a pure product. (2.9 g 8.7 mmol 94 %)

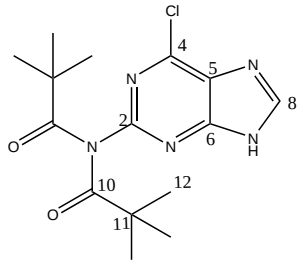


Bis-Boc-adenine: (2.9 g, 94 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.87 (s, 1H, C2), 8.51

(s, 1H, C8), 1.44 (s, 18H, C12); MS (EI) *m/z+* Calc. = 336.1666 C15H22N5O4 (M+H), Exp. = 336.1667 (M+H), 358.1491 (M+Na)



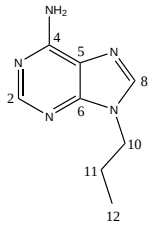
Bis-Boc-cytosine: (2.7 g, 94 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 7.68 (d, J = 7.1 Hz, 1 H), 7.26 (s, 1 H), 7.12 (d, J = 7.1 Hz, 1 H), 1.56 (s, 18 H). ^{13}C NMR ($CDCl_3$, 100MHz): δ (ppm) 163.6 (C10), 158.4 (C2), 149.4 (C4), 145.6 (C6), 96.7 (C5), 84.9 (C11), 27.6 (C12). MS (EI) *m/z+* Calc. = 312.1554 C14H22N3O5 (M + H), Exp = 311.1548.



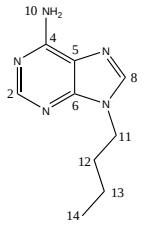
Bis-Boc-2-amino-6-chloropurine: (2.8 g, 83 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.41 (s, 1H, C8), 1.49 (s, 18H, C12), ^{13}C NMR ($CDCl_3$,100MHz): δ (ppm) 155.7 (C10), 150.6 (C6), 150.3 (C2), 148.0 (C4), 147.2 (C8), 127.5 (C5), 83.5 (C11), 27.3 (C12).

2.5.5 General Procedure for the Boc-deprotection of Pyrimidine and Purine Bases[3]

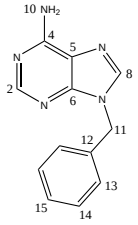
Bis-Boc-alkylnucleobase ($\sim$ 200 mg) was dissolved in DCM (20 ml) and TFA (2 ml) was added. The solution was stirred at room temperature overnight. The reaction was quenched via addition of aq. $NaHCO_3$ sat. and extracted with DCM. The organic layer was dried over $MgSO_4$ and the solvent removed under reduced pressure to yield a white powder.



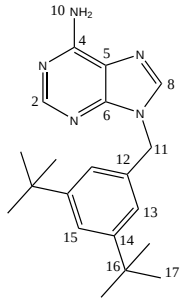
N-9-propyladenine: (80 mg, 85 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.37 (s, 1H, C2), 7.84 (s, 1H, C8), 6.07 (br. s., 2H, N10), 4.19 (t, J=7.3 Hz, 2H, C11), 1.95 (sxt, J=7.3 Hz, 2H, C12), 0.99 (t, J=7.4 Hz, 3H, C13); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 154.6 (C4), 151.2 (C2), 150.0 (C6), 141.0 (C8), 119.5 (C5), 45.7 (C11), 23.4 (C12), 11.2 (C13)



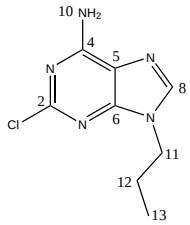
N-9-butyladenine: (81 mg, 83 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.37 (s, 1H, C2), 7.80 (s, 1H, C8), 5.90 (br. s., 2H, N10), 4.20 (t, J=7.2 Hz, 2H, C11), 1.89 (quin, J=7.5 Hz, 2H, C12), 1.38 (sxt, J=7.3 Hz, 2H, C13), 0.97 (t, J=7.4 Hz, 3H, C14)



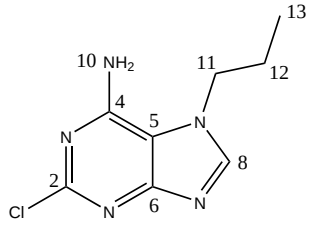
N-9-benzyladenine: (100 mg, 94 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.41 (s, 1H, C2), 7.82 (s, 1H, C8), 7.33 - 7.38 (m, 3H, C14,C15), 7.29 - 7.33 (m, 2H, C13), 6.04 (br. s., 2H, N10), 5.39 (s, 2H, C11); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 154.7 (C4), 151.7 (C2), 150.1 (C6), 140.9 (C8), 135.3 (C12), 129.2 (C14), 128.6 (C15), 127.9 (C13), 119.4 (C5), 47.4 (C11); MS (EI) *m/z+* Calc. = 226.1087 C12H12N5 (M+H), Exp. = 226.1095 (M+H), 248.0910 (M+Na)



N-9-(3,5-di-*t*-butylbenzyl)adenine: (110 mg , 88 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.43 (s, 1H, C2), 7.75 (s, 1H, C8), 7.40 (s, 1H, C15), 7.18 (s, 2H, C13), 5.92 (br. s., 2H, N10), 5.35 (s, 2H, C11), 1.29 (s, 18H, C17); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 155.5 (C4), 153.1 (C14), 151.7 (C2), 150.2 (C6), 140.3 (C8), 134.4 (C12), 122.5 (C15), 122.3 (C13), 119.5 (C5), 47.9 (C11), 34.8 (C16), 31.3 (C17)



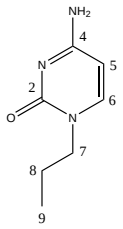
2-amino-6-chloro-*N*-9-propylpurine: (70 mg, 68 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 7.77 (s, 1H, C8), 5.07 (br. s., 2H, N10), 4.06 (t, J=7.2 Hz, 2H, C11), 1.90 (sxt, J=7.3 Hz, 2H, C12), 0.97 (t, J=7.4 Hz, 3H, C13)



2-amino-6-chloro-*N*-7-propylpurine: (32 mg, 73 %)^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 7.95 (s, 1H, C8), 5.03 (br. s., 2H, N10), 4.29 (t, J=7.4 Hz, 2H, C11), 1.93 (sxt, J=7.3 Hz, 2H, C12), 0.98 (t, J=7.4 Hz, 3H, C13)

2.5.6 Synthesis of *N*-1-propylcytosine via Deprotection of Benzoyl Group[4]

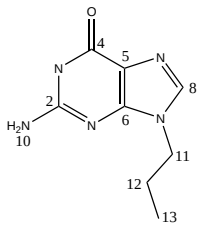
N-1-propyl-*N*-4-benzoylcytosine (257 mg, 1 mmol) was suspended in NH_4OH (4 ml) and methanol (10 ml) and stirred at room temperature overnight. The suspension was then evaporated under reduced pressure and triturated with methanol to afford the pure product as a white powder (120 mg, 78%).



^{1}H NMR (METHANOL-d4 ,400MHz): δ (ppm) 7.61 (d, J=7.3 Hz, 1H, C6), 5.94 (d, J=7.3 Hz, 1H, C5), 3.75 (t, J=7.3 Hz, 1H, C7), 1.71 (sxt, J=7.4 Hz, 1H, C8), 0.93 (t, J=7.5 Hz, 3H, C9); ^{13}C NMR (METHANOL-d4 ,101MHz): δ (ppm) 167.6 (C4), 159.1 (C2), 147.9 (C6), 96.2 (C5), 52.7 (C7), 23.4 (C8), 11.3 (C9)

2.5.7 General Procedure for Hydrolysis and Boc-deprotection of 2-amino-6-chloropurine Derivatives[3]

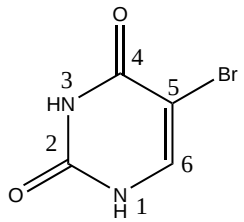
bis-Boc-alkyl-6-chloro-2-aminopurine (206 mg, 0.5 mmol) was added to a round-bottomed flask with formic acid (4 ml) and the yellow solution was stirred at 75 °C for 2 hours. The reaction mixture was concentrated to dryness under reduced pressure and purified by column chromatography (SiO2) to yield an off-white powder (65 mg, 67%).



N-9-propylguanine: ^{1}H NMR (DMSO-d6 ,400MHz): δ (ppm) 8.84 (s, 1H, C8), 7.23 (br. s., 2H, N10), 4.02 (t, J=7.2 Hz, 2H, C11), 1.80 (sxt, J=7.3 Hz, 2H, C12), 0.86 (t, J=7.4 Hz, 3H, C13); ^{13}C NMR (DMSO-d6 ,101MHz): δ (ppm) 155.2 (C4), 154.1 (C2), 150.0 (C6), 137.1 (C8), 109.8 (C5), 45.7 (C11), 22.1 (C12), 10.7 (C13); MS (EI) *m/z+* Calc. = 194.1036 C8H12N5O (M+H), Exp. = 194.1039 (M+H), 216.0851 (M+Na)

2.5.8 Synthesis of 5-bromouracil[5]

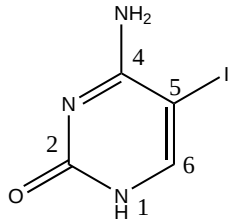
Uracil (1.00 g, 8.92 mmol) was dissolved in DMF (5 mL) and heated to 120C. Bromine (1.57 g, 9.81 mmol) was added via syringe and the solution was stirred at 120 °C for 24h. The reaction was allowed to cool to room temperature and the dark brown solution was poured onto chloroform (500 ml) upon which an off-white precipitate was formed. The solid was collected via suction filtration and washed with chloroform (100 ml) to yield 5-bromouracil which required no further purification (1.50 g, 88 %).



^{1}H NMR (DMSO-d6, 400 MHz): δ (ppm) 11.48 (bs, 1H, N3), 11.20 (bs, 1H, N1), 7.88 (d, J = 6.2 Hz, 1H, C6); ^{13}C NMR (DMSO-d6, 100 MHz): δ (ppm) 160.80 (C4), 151.53 (C2), 142.88 (C6), 95.14 (C5).

2.5.9 Synthesis of 5-iodocytosine[6]

Cytosine (10.0 g, 90.0 mmol), iodine (34.3 g, 135.0 mmol), and iodic acid (22.2 g, 126.0 mmol) were stirred in acetic acid (300 mL) at 40 °C overnight. When the reaction was complete, the mixture was cooled and treated with $Na_2S_2O_3$ (sat. aq.; 200 mL) until a white suspension was obtained. The mixture was then neutralized with NaOH (6 m aq.; 900 mL). The resulting white solid was collected by filtration, and washed with slightly basified water until the filtered water had neutral pH. The solid was dried under reduced pressure to give 5-iodocytosine as a white solid (20.6 g, 96 %).

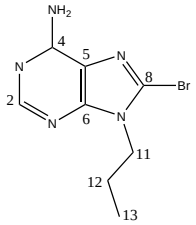


^{1}H NMR (DMSO-d6, 400 MHz): δ (ppm) 10.76 (br. s, 1 H, N1), 7.76 (s, 1 H, C6), 6.48 (br. s, 2 H, NH2); ^{13}C NMR (DMSO-d6, 100 MHz): δ (ppm) 164.4 (C4), 155.8 (C2), 149.4 (C6), 55.2 (C5); MS (EI) *m/z+* Calc = 237.9471 C4H5N3OI (M+H)+, Exp = 237.9464

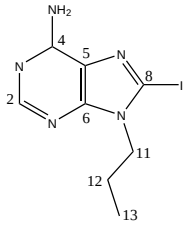
2.5.10 General Procedure for *C*-8 Halogenation of Purine Bases[7][8]

Procedure a) Treatment with LDA:

To a solution of diisopropylamine (0.145 mL, 1.03 mmol) in dry THF (3 mL) n-BuLi (0.536 mL, 1.00 mmol, 1.87 M in hexane) was added dropwise at -78 °C under N_2. After stirring for 30 minutes a solution of alkyl nucleobase (0.500 mmol) in THF (1.5 mL) was added and stirred for a further 60 minutes at -78 °C. A solution of halogen (1.00 mmol) in THF (1.5 mL) was added dropwise and the resulting dark mixture was stirred for 60 minutes at -78 °C and allowed to warm to room temperature. Aqueous $NHCl_4$ sat. (5 mL) was added and the resulting mixture was extracted with EtOAc (3 × 50 mL). The organic extracts were washed with brine (100 mL), dried over $MgSO_4$ and evaporated under reduced pressure to yield pure compound as an off-white solid.



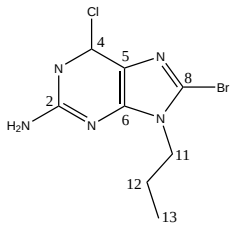
8-bromo-*N*-9-propyladenine: (110 mg, 85 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.33 (s, 1H, C2), 5.81 (br. s., 2H, NH2), 4.19 (t, J=7.3 Hz, 2H, C11), 1.90 (sxt, J=7.4 Hz, 2H, C12), 0.99 (t, J=7.4 Hz, 3H, C13); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 153.7 (C4), 152.4 (C2), 151.3 (C6), 127.8 (C8), 119.9 (C5), 46.1 (C11), 22.8 (C12), 11.0 (C13); MS (EI) *m/z*+ Calc. = 256.0192 C8H11BrN5 (M+H), Exp = 256.0186 (M+H)



8-iodo-*N*-9-propyladenine: (140 mg, 91 %) ^{1}H NMR ($CDCl_3$,500MHz): δ (ppm) 8.29 (s, 1H, C2), 5.78 (br. s., 2H, NH2), 4.15 (t, J=7.5 Hz, 2H, C11), 1.89 (sxt, J=7.4 Hz, 2H, C12), 1.00 (t, J=7.5 Hz, 3H); ^{13}C NMR ($CDCl_3$,126MHz): δ (ppm) 153.7 (C4), 152.6 (C2), 151.4 (C6), 122.4 (C8), 100.2 (C5), 47.6 (C11), 23.1 (C12), 11.1 (C13); MS (EI) *m/z*+ Calc. = 304.0054 C8H11IN5 (M+H), Exp = 304.0052 (M+H)

Procedure b) Direct halogenation:

Bromine (aq. sat. 5 ml) was added to a suspension of purine base (200 mg) in water (2ml). The flask was closed stirred at room temperature for 24 hours. The flask was opened and left until excess bromine had evaporated. The solvent was then evaporated under reduced pressure and the solid purified by column chromatography (SiO_2 DCM:MeOH 98:2 v:v).



2-amino-6-chloro-8-bromo-*N*-9-propylpurine: (220 mg, 80 %) ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 5.08 (br. s., 2H, NH2), 4.08 (t, J=7.4 Hz, 2H, C11), 1.86 (sxt, J=7.4 Hz, 2H, C12), 0.98 (t, J=7.5 Hz, 3H, C13);

Chapter 3

Intermolecular Interactions of Alkylated and Halogenated DNA Nucleobases

3.1 Hydrogen Bonding Interactions of *N*-1 and *N*-9 Alkylated Nucleobases

Following on from the synthesis of the alkylated DNA nucleobases the hydrogen bonding interactions of these nucleobases can now be studied. The propyl group has been chosen with a view to increase the solubility of the DNA bases whilst having as little impact as possible on the hydrogen bonding interactions. The propyl group itself has no significant hydrogen bond donors or acceptors and is relatively small so should have minimal steric effect. The group is also relatively benign with respect to electron donation or withdrawal and finally, the propyl group was added in the N-1 and N-9 positions of the DNA pyrimidines and purines respectively so as to closely mimic the structure of the DNA bases found in nature and not block any of the hydrogen bonding sites. The hydrogen bonding interactions of the alkylated nucleobases were studied via a series of NMR titrations and by studying the crystal structures.

3.1.1 The Theory and Practices of NMR Titration

The principle of an NMR titration is to observe the changes in the NMR spectrum of the sample when the concentration is changed or an additional analyte is added to the sample. These changes in the NMR spectrum provide information on how the chemical environments are being affected by the change in concentration or addition of another analyte. This can be used to study hydrogen bonding interactions as the chemical shifts of the functional groups within a hydrogen bonding dimer are different to those within a monomer. For example, in a simple carboxylic acid the acid proton is in a different chemical environment in the monomer to when it is a hydrogen bond donor in a dimer as illustrated in Figure 3.1. The chemical shift observed in the NMR spectrum provides information on whether the sample is, on average, in the free state (monomer) or bound state (dimer). This, coupled with the known changes in concentration across multiple data points can allow the binding constant to be calculated and the stoichiometry of the interaction to be deduced.

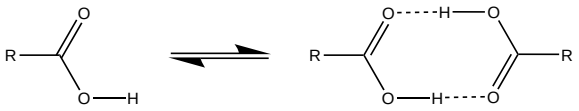


Figure 3.1: Illustration showing the different carboxylic acid proton environments in both the monomer and dimer which can be observed by NMR spectroscopy.

In the above Figure 3.1 the carboxylic acid proton is in a different chemical environment for both the free state and the bound state. This results in a different chemical shift being observed in the proton NMR spectrum. The chemical shift for the monomer is expected to be upfield relative to the dimer as the hydrogen bonding within the dimer further deshields the carboxylic acid proton. Hydrogen bonding is more prevalent at higher concentrations as at lower concentrations the molecules are surrounded by solvent molecules. Therefore, as concentration is increased more hydrogen bonding molecules are in close proximity and can interact and the resonant proton is shifted downfield.

For an NMR titration experiment studying hydrogen bonding it is important that the solvent itself does not form appreciable hydrogen bonds to the analyte as this would produce inaccurate results as the solvent will be competing with the sample to form hydrogen bonds. Therefore, it is preferred to perform the titration experiment in a non-competitive solvent such as DCM, chloroform or benzene. It is however, still possible to use competitive solvents such as acetone, methanol or DMSO if at some point the analyte can 'out-compete' the solvent and form hydrogen bonds. Care must be taken when using protic solvents such as methanol as these may exhibit a H-D exchange with the sample and cause the disappearance of the resonant H signal.

3.1.2 NMR Titration Studies of *N*-1 and *N*-9 Alkylated Nucleobases

Self-Association Studies

The hydrogen bonding interactions of N-1 and N-9 alkylated nucleobases were studied by performing various NMR titrations. The self-association binding constants for each alkylated nucleobase were calculated by performing a dilution experiment on each sample. For the dilution experiments a high concentration of the sample was prepared to represent the sample in the 'bound (AA)' state and this was sequentially diluted until a dilute sample representing the 'unbound (A)' state was formed. ^{1}H NMR spectra were recorded for each titre and the NH chemical shift was recorded for each concentration. The data was fitted to the equations below using non-linear curve fitting procedure. This data was then fitted to the following equations.[110]

$$[AA] = \frac{1 + 4K_d[A]_0 - \sqrt{1 + 8K_d[A]_0}}{8K_d} \tag{3.1}$$

$$[A] = [A]_0 - 2[AA] \tag{3.2}$$

$$\delta_{obs} = \frac{2[AA]}{[A]_0}\delta_d + \frac{[A]}{[A]_0}\delta_f \tag{3.3}$$

Where $[AA]$ is the concentration of the bound species, K_d is the association constant, $[A]_0$ is the total concentation, $[A]$ is the concentration of the free species, δ_d is the limiting bound chemical shift of the dimer and δ_f is the free chemical shift.

The dilution experiment for each alkylated nucleobase was carried out and the data from the analytical hydrogen bond donor NH chemical shift(s) fitted to the equations 3.4-3.6 above and solved using purpose-written software on a computer. During the experiment the NH peaks were found to

shift upfield as the sample was diluted as shown in Figure 3.2. This indicates that at as the sample is diluted the presence of bound species decreases and unbound species increases. The hydrogen bonded bound species exhibits a downfield chemical shift as the hydrogen bond has the effect of deshielding the H nucleus. The free species, on the other hand, exhibits an upfield chemical shift as the H nucleus is no longer deshielded by the electron-withdrawing hydrogen bond acceptor. Whilst a shift in the hydrogen bonding NH peak was observed, as expected, for *N*-1-propylthymine no shift was observed in any of the other peaks in the NMR. This indicates that no other parts of the molecule are displaying any appreciable interactions. This is also indirect evidence that the correct *N*-1 alkylated isomer is present. In the N-1 isomer the hydrogen bonding NH proton is a relatively large distance from any other proton in the molecule. This means that any inductive effect of the presence of hydrogen bonding would not be observed in any adjacent proton environments. (It is likely that the *C*-2 and *C*-4 carbonyl environments would be affected but it is not practical to study carbon environments in a NMR dilution experiment due to the low concentrations which would require long collection times.)

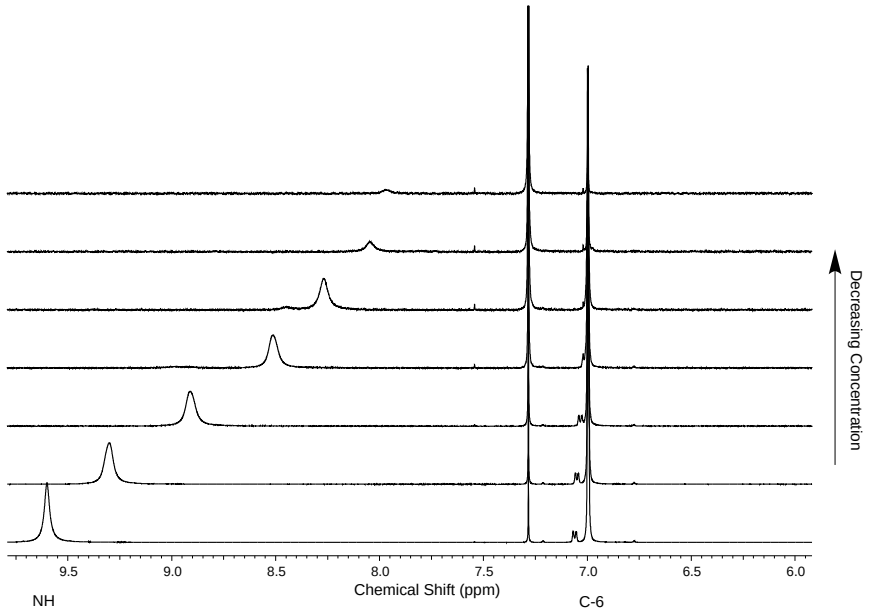


Figure 3.2: Overlayed ^{1}H NMR spectra of the dilution experiment of *N*-1-propylthymine in $CDCl_3$ at 298 K.

The relative amounts of monomer and dimer in the hydrogen bonding system at a given concentration can be represented by a speciation plot which plots the percentage population of [A] and [AA] in solution against concentration as shown in Figure 3.3. As expected, at low concentrations the population is effectively exclusively [A] as the sample is so dilute then dimers are highly unlikely to form. As concentration increases the concentration of [AA] increases until the dimer is in the majority. In solution it is not feasible to achieve a population of 100% dimer as this would be a solid crystal structure. In solution there is a dynamic exchange of hydrogen bonding pairs so even in a saturated solution there is likely to always be unbound species.

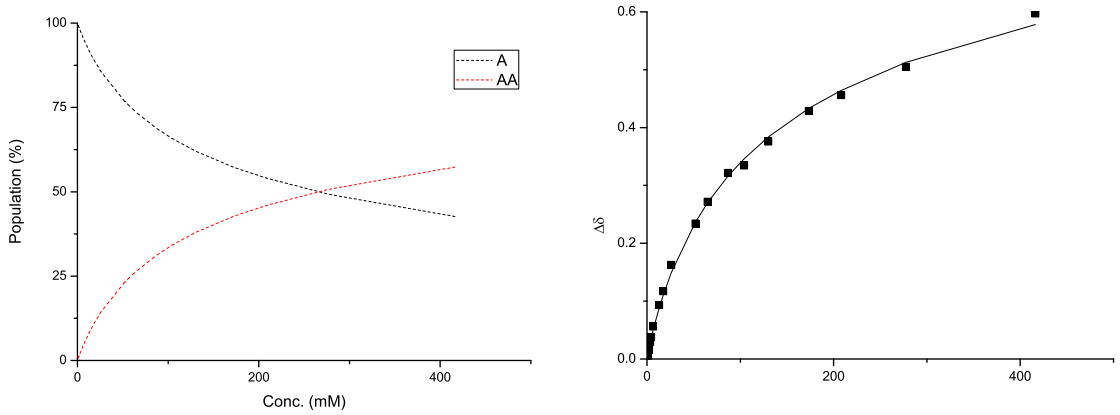


Figure 3.3: a) Speciation chart of population of [A] and [AA] against concentration for the self-association of *N*-1-propylthymine. b) Plot showing the change in chemical shift of the thymine NH against concentration for the self-association of *N*-1-propylthymine.

The solution of the equations 3.4-3.6 also importantly yields the association constant for the self-association of the sample. This gives an indication of the strength of the hydrogen bonding interaction. The self-association constant for propylthymine was calculated to be 1.79 M^{-1} These value is in accordance to literature values for the association of similar alkylated nucleobases analogues. This is a relatively weak interaction, as expected, as thymine does not preferentially form a complementary pair with itself.

Host-Guest Binding Studies

To study the hydrogen bonding interactions of the synthesised alkylated nucleobases with their respective pairs host-guest titration experiments were carried out. It was expected that the association constants for the complementary DNA base pairs would be considerably higher than those found for the self-association studies. For the host-guest titrations a Job plot was first plotted for each host-guest mixture. A Job plot is a method of continuous variation used for calculating the stoichiometry of a binding event. The stoichiometry of binding is required so that the correct equations can be applied for the calculation of the association constant. In a binding event of A with B it is possible that a single A will bind with a single B; this is a 1:1 isotherm. However, a single A may bind with more than one B; a 1:2, 1:3 etc. isotherm. For a Job plot the total molar concentration of each species are held constant and their mole fractions are varied between 0 and 1. The observable (for ^{1}H NMR titration the change in chemical shift ($\Delta\delta$)) is plotted against the mole fraction. For a 1:1 isotherm the Job plot will be symmetrical with the peak in $\Delta\delta$ at the 0.5:0.5 mole fraction. At this mole fraction the ratio of host and guest are equal and each species can bind to one of the other. At all other mole fractions one species is in excess and a lesser degree of binding is observed resulting in a lower $\Delta\delta$.

For the binding of alkylated DNA nucleobases a 1:1 binding isotherm is expected as each nucleobase has one primary selective hydrogen bonding site (the Watson-Crick face). Although it is possible for a nucleobase to bind on the Watson-Crick and Hoogsteen face simultaneously, common sense suggests that at the equimolar mole fraction 1:1 binding will be observed. Below are the Job plots recorded for the complementary base pair of *N*-1-propylthymine and *N*-9-propyladenine (Figure 3.4).

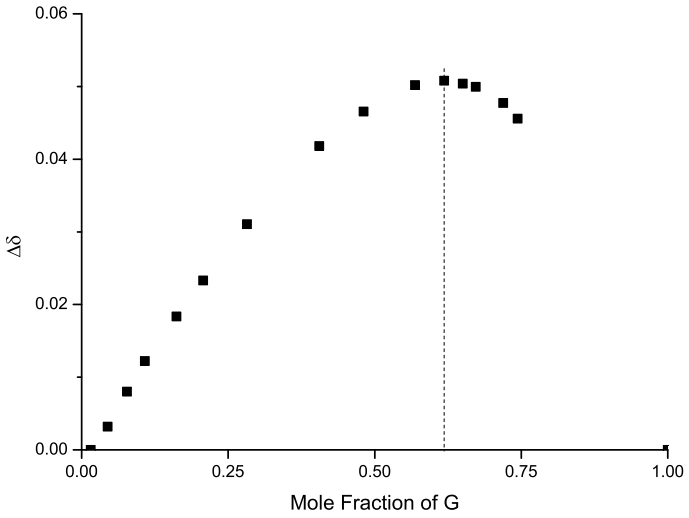


Figure 3.4: Job plot of propylthymine-propyladenine titration where propylthymine was the host (H) and propyladenine was the guest (G). Recorded in $CDCl_3$ at 298 K.

The Job plot above shows a maximum $\Delta\delta$ at a guest mole fraction of 0.62. This suggests a 1:2 binding stoichiometry which is different to the 1:1 stoichiometry that was expected. This can be explained by the possibility for both Watson-Crick and Hoogsteen hydrogen bonding. If both these hydrogen bonding modes are observed it is possible for each adenine molecule to bind to two thymine molecules. These hydrogen bonding pairs would both result in a shift in the relevant NH proton peak in the NMR spectra. To confirm whether Watson-Crick or Hoogsteen hydrogen bonding present in the host-guest system it is possible to monitor chosen CH shifts of the host and guest. The hydrogen bonding in each position will affect the CH shift of protons which are adjacent to the hydrogen bonding donors differently as shown in Figure 3.5. For thymine-adenine Watson-Crick binding a shift in the C-2 CH proton will be observed whereas for Hoogsteen binding a shift in the C-8 CH proton will be observed. For the cytosine-guanine base pair it is somewhat simpler as for Watson-Crick binding both the N-1 and C-2 NH hydrogen bond donors are involved in hydrogen bonding but for Hoogsteen binding these donors are not involved so any shift in the N-1 and C-2 NH protons must be due to Watson-Crick binding. If Hoogsteen binding is present the C-8 CH will be shifted.

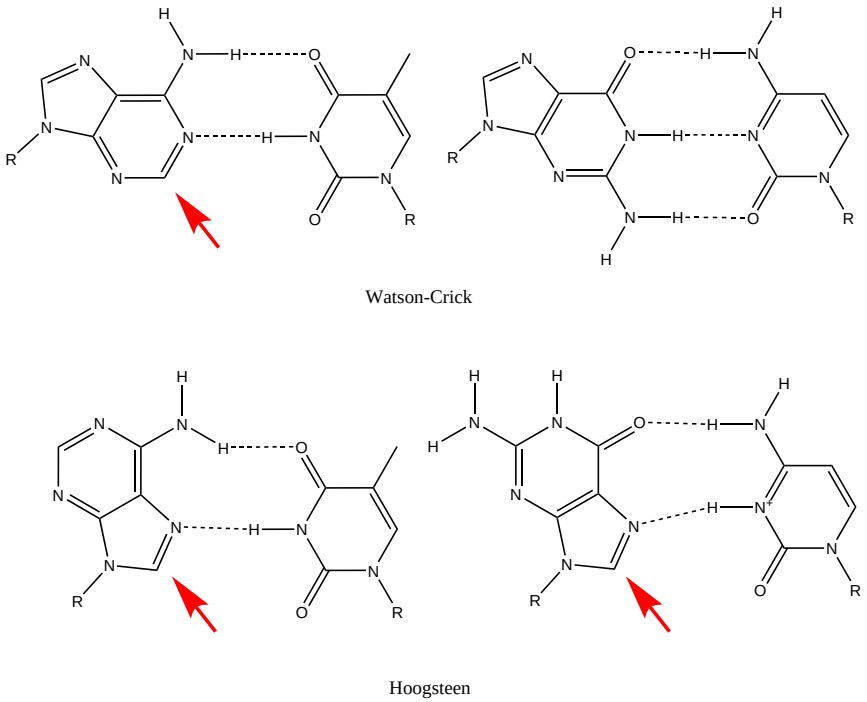


Figure 3.5: The different geometries of Watson-Crick and Hoogsteen binding between the complementary base pairs highlighting the environments that are shifted depending on which hydrogen bonding mode is observed in red.

For the thymine-adenine host-guest titration shifts in both the C-2 and C-8 protons were observed as shown in Figure 3.6. This indicates that both Watson-Crick and Hoogsteen hydrogen bonding are present in the host-guest system. Both of these binding modes are shown in Figure 3.7.

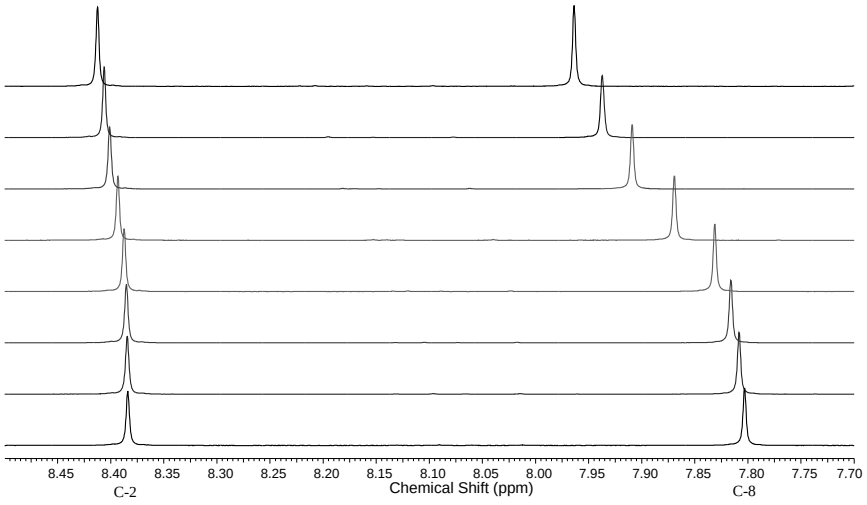


Figure 3.6: ^{1}H NMR spectra showing the shift in the C-2 and C-8 signals of propyladenine as the concentration of propylthymine is increased. Recorded in $CDCl_3$ at 298 K.

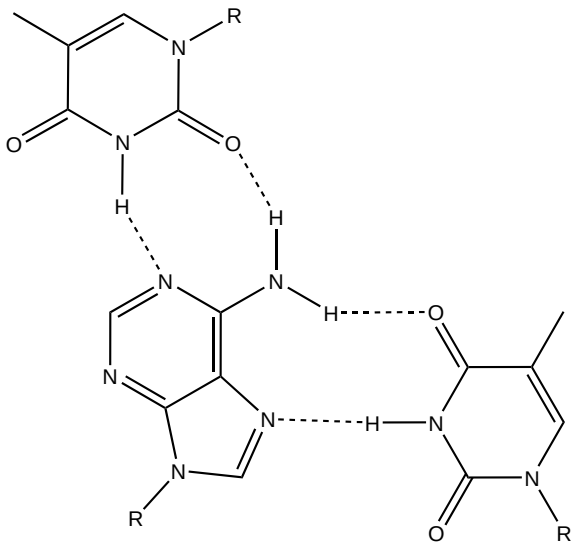


Figure 3.7: Structure of the 1:2 thymine:adenine hydrogen bonding stoichiometry as observed in the Job plot showing the Watson-Crick and Hoogsteen binding as confirmed by the ^{1}H NMR spectra above (3.6).

For the host-guest NMR titrations of the complementary base pairs a solution of guest was titrated into a solution of host of constant concentration and ^{1}H NMR spectra were recorded for each titre. The relevant NH shift of the host was used for fitting to the following equations. For each complementary pair the host-guest experiment was repeated with the host and guest roles reversed each time.

$$[AA] = \frac{1 + 4K_d[A]_0 - \sqrt{1 + 8K_d[A]_0}}{8K_d} \tag{3.4}$$

$$[A] = [A]_0 - 2[AA] \tag{3.5}$$

$$\delta_{obs} = \frac{2[AA]}{[A]_0}\delta_d + \frac{[A]}{[A]_0}\delta_f \tag{3.6}$$

The speciation and association constants for the binding of the complementary base pairs were calculated and plotted (Figure 3.8) and, as expected, a larger association constant is observed than for the dilution experiments. For the AT base pair a binding constant of 70.4 M^{-1} was calculated.

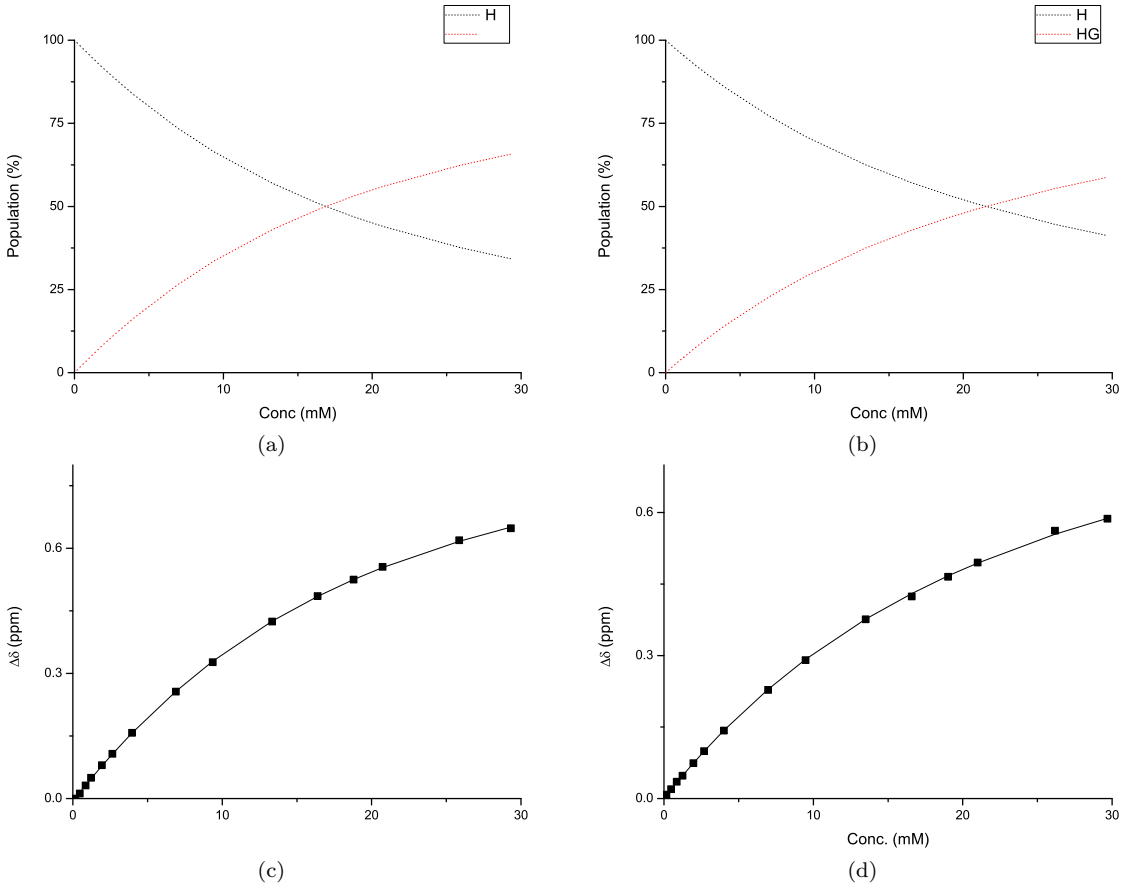


Figure 3.8: a),b) Speciation chart of the PTPA and PAPT NMR titration experiments. c),d) Plot of the change in chemical shift against concentration for the PTPA and PAPT NMR titration experiments.

3.1.3 Crystallographic Studies of *N*-1 and *N*-9 alkylated Bases

The solid-phase crystal structure of the alkylated DNA bases was determined by single-crystal X-ray crystallography. This allowed the solid-state interactions between the bases to be studied. In solution it is easy for the molecules to move freely and orientate into a favourable position for hydrogen bonding interactions. In the solid state other interactions such as $\pi - \pi$ stacking may take priority in the formation of a crystal structure.

A single crystal of *N*-1-propylthymine was grown by slow evaporation of $CDCl_3$ and the crystal structure was solved for the crystal structure shown below (Figure 3.9):

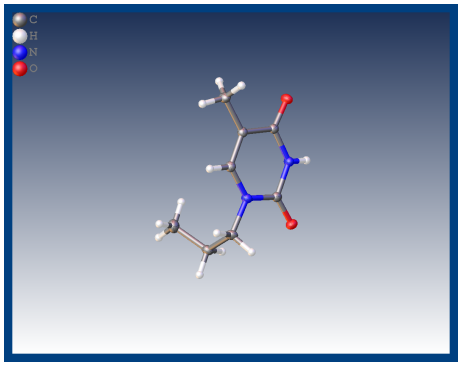
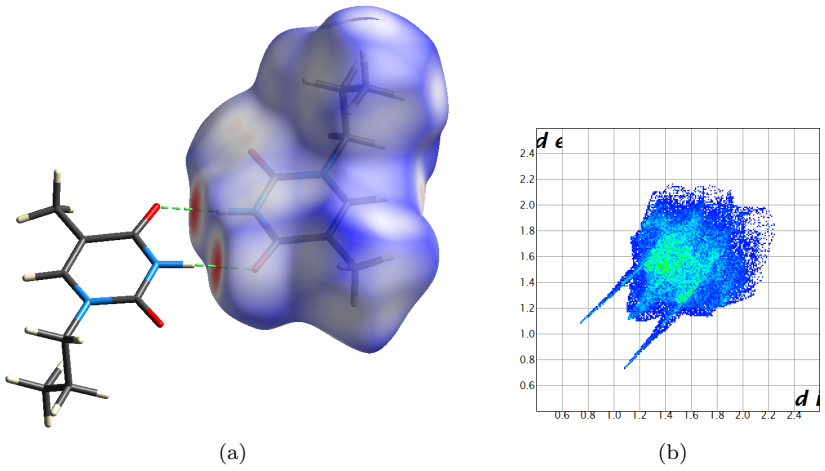


Figure 3.9: Crystal structure of N-1-propylthymine.

The crystal structure of N-1-propylthymine clearly shows the molecules arranged in the position for Watson-Crick hydrogen bonding. The distances and angles between the relevant hydrogen bond donors and acceptors also indicates that hydrogen bonding is present. To further study the intermolecular hydrogen bonds between the molecules the Hirschfeld surface of the N-1-propylthymine crystal was plotted. The Hirshfeld surface plots the electron-distribution around a chosen portion of the crystal structure. Different coloured areas of d_{norm} plotted on the Hirshfeld surface indicate that two atoms are in contact at a normalised distance below the Van der Waals radius of the relevant atoms internal and external to the surface [103]. The close contact of atoms is a fundamental feature in the formation of bonds and intermolecular interactions so a peak in the Hirshfeld surface indicates that an intermolecular interaction is present between the molecule within the surface and an external molecule. The Hirshfeld surface can be displayed visually and superimposed on the crystal structure to show the locations of any peaks or troughs in d_{norm}. The surface can also be shown graphically to plot the distance between the atoms within the surface (d_i) and outside the surface (d_e) to the surface.



(a) (b)

Figure 3.10: a) Hirshfeld surface and b) finger plot of N-1-propylthymine crystal structure.

For the crystal structure of N-1-propylthymine the Hirshfeld surface clearly shows two areas of short d_{norm} in red (Figure 3.10). These are where the hydrogen bonds are expected to be over the C=O hydrogen bond acceptor and N-H hydrogen bond donor. These areas of short d_{norm} indicate that there is an attractive intermolecular interaction between these parts of the molecule and the

adjacent molecule external to the Hishfeld surface. Hirshfeld surface analysis of the adjacent molecule also shows peaks in the positions over the complementary hydrogen bond donors and acceptors (not shown for clarity). The relatively large peak shown by the strong red contrasting colour and sharp peaks in the finger plot suggest that the interaction is a hydrogen bond due to the well known directional nature of a hydrogen bond and how a hydrogen bond is similar to a covalent bond in how it shares electrons. To be sure that short d_{norm} contacts are due to electron density provided by hydrogen bond donors and acceptors it is possible to selectively show the Hirshfeld surface by internal and external element as shown in Figure 3.11. Looking at the Hirshfeld surface it can be seen that there are other red highlighted areas which indicates that other interactions are present in the crystal. There is also an interaction between the C-6 proton and the C-2 carbonyl. Interestingly, these interactions were not observed in the NMR dilution experiment as no other peaks were found to be shifted. This shows a difference between the interactions in the crystal and solution phases.

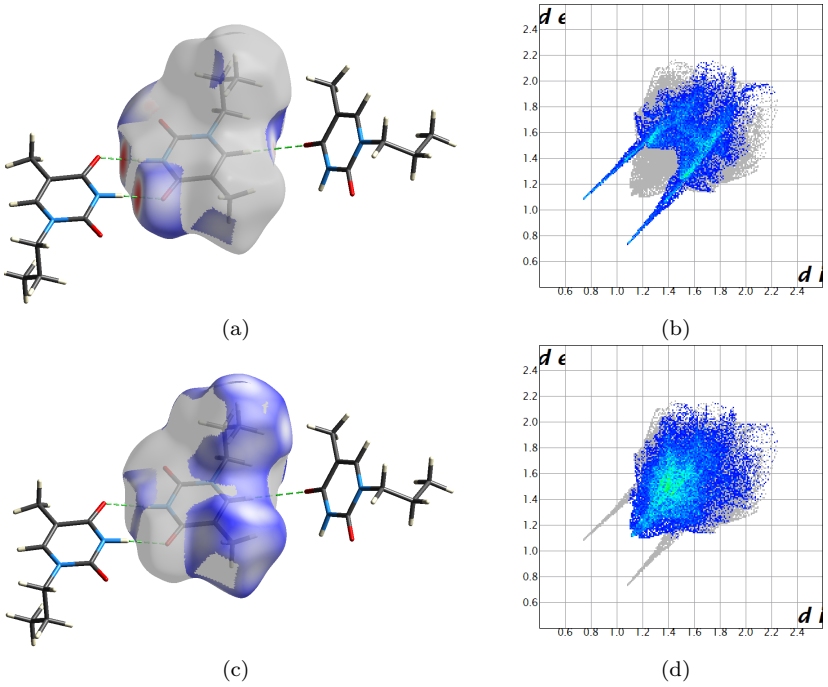


Figure 3.11: Hirshfeld surface and finger plot of propylthymine crystal structure by internal and external element a),b) = Oi He; c),d) = Hi He

For the selected internal and external element Hirshfeld plots above it is clear that the peaks in the Hirshfeld surface are due to hydrogen bond donors and acceptors. When O is chosen for the internal element and H for the external element a clear peak is seen in the Hirscheld surface. This shows that these two elements are close as would be the case in a hydrogen bond. For the reciprocal of this interaction it can be seen that Hi is interacting with Oe as expected for the other part of the hydrogen bonding pair. For the secondary HO interaction between the C-6 proton and C-2 carbonyl it can be seen by the intensity and size of the red area of the Hirshfeld surface that this is not as strong an interaction. The distance of this interaction is 2.618 Å compared to 1.959 Å for the Watson-Crick paired hydrogen bond. This is expected as the CH hydrogen bond donor is less effective than NH as carbon is less electronegative than nitrogen.

When other elements are chosen no particularly short d_{norm} distances are observed. The other large contribution to the Hirshfeld surface comes from Hi He. These close contacts are much more diffuse and not seen in a sharp peak as shown by the absence of red areas and the broad peak in

the finger plot which indicates much weaker electrostatic interactions. It is not surprising that these electrostatic interactions are on the area of the molecule bearing the alkyl chain and methyl group.

The surface area and volume of the Hirschfeld surface was calculated at 203.05 $Å^2$ and 201.13 $Å^3$. Hydrogen bonding interactions (OH and HO) cover 30.8 % of the total Hirschfeld surface and electrostatic interactions (HH) cover 50.8 % of the surface.

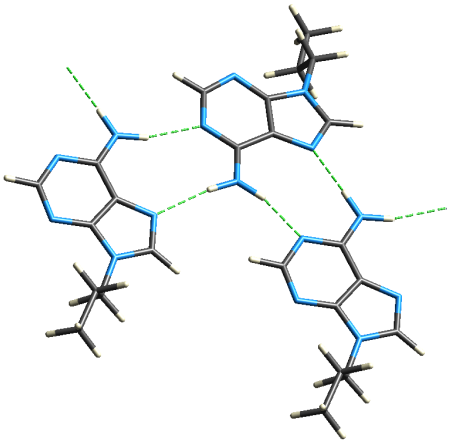
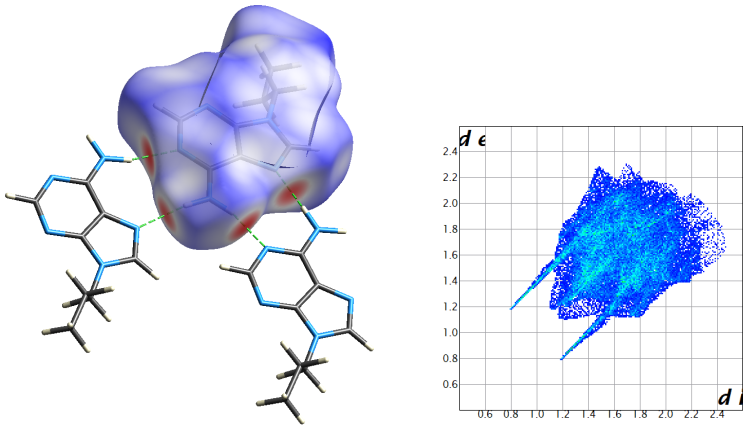


Figure 3.12: A view of the hydrogen bonded network in the crystal structure of *N*-9-propyladenine.

For the single crystal structure of adenine above (Figure 3.15) both Watson-Crick (W-C) and Hoogsteen (H) hydrogen bonding modes can be seen. In the crystal lattice each molecule binds on both hydrogen bonding face and the Watson-Crick face binds to the Hoogsteen face. This result was not expected as the Watson-Crick binding is known to be the mode selected in nature and W-C-W-C hydrogen bonding allows for the most ideal hydrogen bonding angles. However, in a crystal lattice the interactions that favour the highest stability of the entire lattice are favoured. This suggests that W-C-W-C hydrogen bonding then leads to less stable interactions throughout the rest of the lattice (such as less stable H-H interactions or unfavourable steric interactions). Therefore, the W-C-H bonding allows the total lattice energy to be minimised.

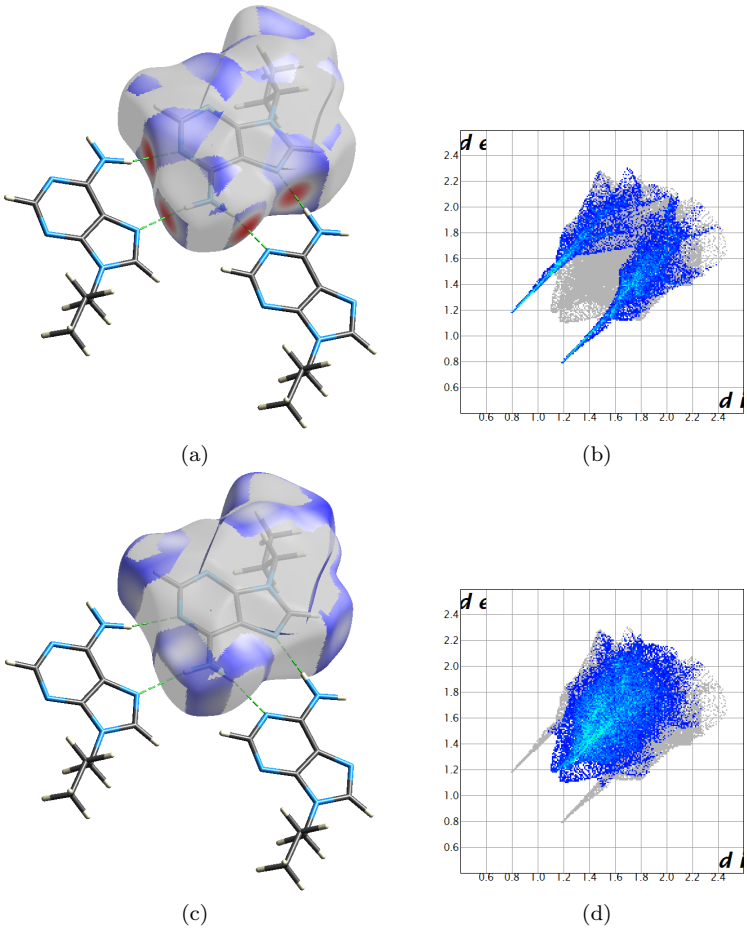


(a) (b)

Figure 3.13: a) Hirschfeld surface and b) finger plot of N-9-propyladenine crystal structure.

The Hirshfeld surface for N-9-propyladenine shows that there is intermolecular hydrogen bonding

present between each propyladenine molecule at the W-C-H site. This is shown by the red regions in the Hirschfeld surface in areas where the hydrogen bond donors and acceptors are close. The Hirshfeld surface was calculated to have a total area of 217.94 Å^2 and volume of 219.04 Å^3.



(a) (b)

(c) (d)

Figure 3.14: Hirshfeld surface and finger plot of propyladenine crystal structure by internal and external element a),b) = Ni He (reciprocal); c),d) = Hi He

The Hirshfeld surfaces of *N*-9-propyladenine when selected by element also show how the intermolecular interactions are due to hydrogen bonding donors and acceptors. There is also a relatively large electrostatic interaction but it neither as strong as the hydrogen bonds nor is it especially directional and has minimal effect on the orientation of the molecules within the crystal lattice.

Close contacts between N and H contribute 30.1% of the total surface area whereas 54.1% of the total surface is due to H-H interactions.

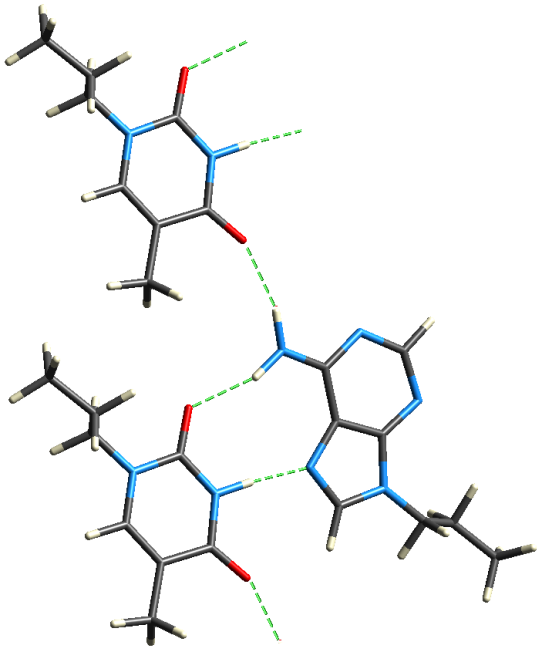


Figure 3.15: Cocrystal structure of N-1-propylthymine and N-9-propyladenine.

It was also possible to grow a cocrystal of thymine and adenine. It was interesting to discover that in this cocrystal the expected Watson-Crick hydrogen bonding was not observed between the complementary base pair. The thymine molecule binds to the Hoogsteen face of adenine via the reverse Watson-Crick hydrogen bond mode. It can be seen from the crystal structure that the hydrogen bond lengths and angles are not ideal; particularly the C-6-amino proton on adenine. It is possible that this intermolecular lattice structure is the most stable when all the other factors of crystal packing are considered. Alternatively, this could be just one of many polymorphs that form when a crystal grows and other arrangements are also possible.

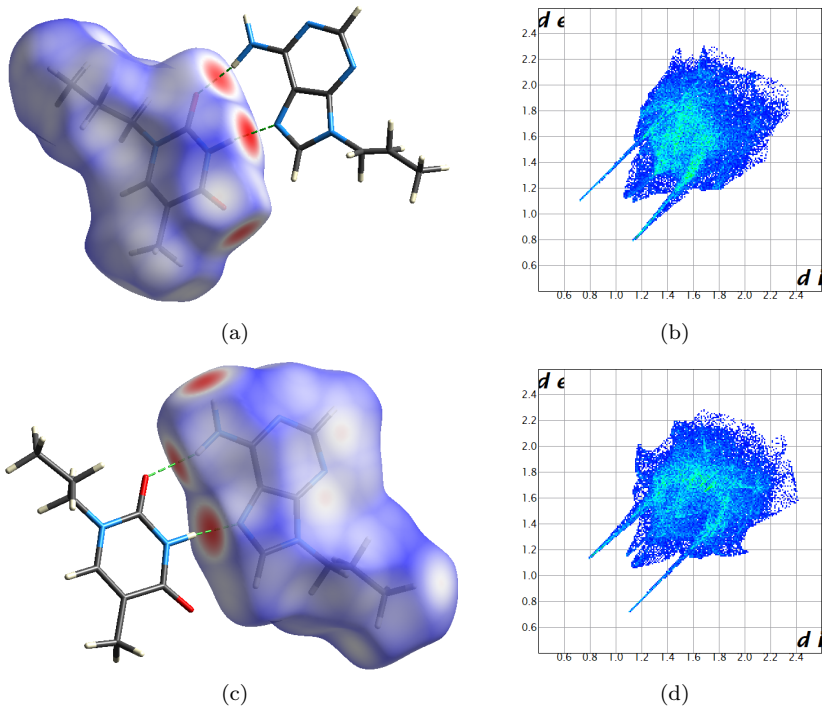


Figure 3.16: Hirshfeld surface and finger plot of propylthymine and propyladenine cocrystal. a) b) = propylthymine; c) d) = propyladenine

The Hirshfeld surface of both components of the cocrystal highlights the hydrogen bonding between the two molecules as there are areas of high electron density where the two surfaces meet between the relevant hydrogen bond donors and acceptors. It can be seen that there is some interaction on both adenine C-6-amino protons suggesting that Watson-Crick interactions are also present. However, there is no peak over the N-1 nitrogen on adenine which shows that Watson-Crick hydrogen bonding is not present.

As expected, the surface areas and volumes of propylthymine and propyladenine surfaces are comparable to those calculated for the self-association crystals. The propylthymine crystal had a surface area of 209.11 $Å^2$ and volume of 206.99 $Å^3$. Propyladenine had a surface area and volume of 212.96 $Å^2$ and 212.05 $Å^3$.

The contribution to the total surface area by element is also comparable for both the complementary cocrystal and the self-association crystal. For thymine H-H interactions contribute 52.1 % and OH and HN interactions contribute 24.9 %. For adenine the H-H interactions contribute 50.4 % and the hydrogen bonding interactions (NH and HO) contribute 25.9 %.

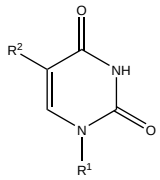
3.2 Effect of Substituents on the Self-Assembly of Alkylated DNA Nucleobases

The four DNA nucleobases have been synthesised with a range of different alkyl, benzyl and halogen groups to study the steric and electronic effect of these groups on hydrogen bonding self-assembly.

Firstly, the effect of different steric groups on the hydrogen bonding interactions was studied. This was done by studying the hydrogen bonding interactions between the nucleobases in both solution and crystal phases with different alkyl and benzyl substituents. It was expected that as the steric bulk increased the strength of the hydrogen bonding interaction would go down.

3.2.1 Alkylated Thymine

A number of thymine analogues were analysed by NMR dilution experiments to study the effect of the alkyl group and any substituents added to the core thymine pyrimidine structure. The association constants for the analysed thymine analogues are shown below in Table 3.1.



	R^1	R^2	K_d M^{-1}	r D-A (Å)	θ D-H-A
a	*n*-pr	Me	1.79	1.959 H1-O1	169.746 N1-H1-O1
b	*n*-et	Me			
c	*n*-bu	Me	1.38		
d	*n*-pent	Me	1.49	1.980 H1A-O1B, 1.974 H1B-O1A	171.325 N1A-H1A-O1B, 168.008 N1B-H1B-O1A
e	*n*-hex	Me			
f	*n*-oct	Me	3.93		
g	*n*-dec	Me	3.78		
h	benz	Me	4.28	1.991 H1-O2	172.079 N1-H1-O2
i	*t-bu*-benz	Me	2.94	1.977 H1-O1	175.511 N1-H1-O1
j	*n*-pr	H	4.41		
k	*n*-pr	Br	10.30		
l	*t-bu*-benz	Br			

Table 3.1: Table of association constants for the synthesised thymine analogues as calculated by NMR dilution and hydrogen bonding lengths (r) and angles (θ) as calculated from the relevant crystal structure.

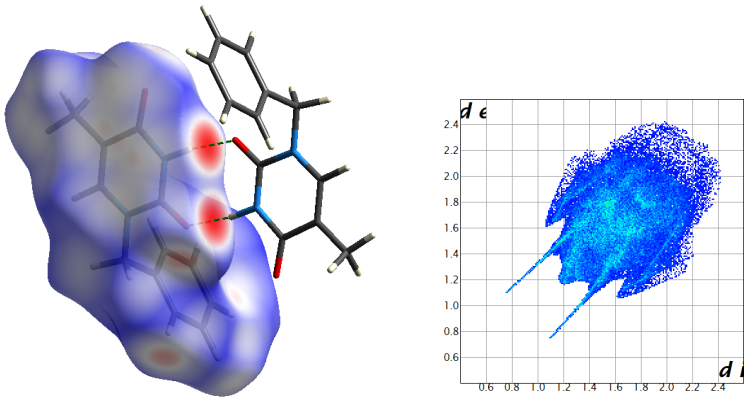
It can be seen that the alkyl or benzyl group in the N-1 position of thymine does not have a significant effect on the association constant. The association constant is slightly higher (indicating a stronger interaction) for the longer *n*-oct and *n*-dec chains. This may be due to the interdigitation of the long alkyl chains providing an extra stabilising effect. It was expected that the long alkyl chains may provide a steric barrier to association by wrapping around the molecule and possibly blocking the hydrogen bonding part of the molecule but this does not appear to be the case. For the benzyl group the association constant is slightly higher. This may be due to additional π interactions caused by the aromatic benzyl group. This can be investigated by looking at the NMR spectra. That the association constant is lower for the t-butylbenzyl group supports this theory as the *t*-butyl groups would interfere with the π interactions and lower the association constant. The group with the greatest effect on the association constant was the bromine substituted in the *C*-5 position. This was found to increase the association constant by ca five times (although the association constant is still relatively small). This suggests that bromine is either having an effect on the electronic structure of the thymine analogue which, in turn, affects its hydrogen bonding. Alternatively, the bromine may provide additional halogen bonds that stabilise the hydrogen bonding structure.

For the crystal structures of DNA bases with different alkyl groups it was expected that the total Hirshfeld surface area and volume would increase with increasing size of alkyl or benzyl group. It was also expected that the contribution of electrostatic interactions to the surface would increase relative to the contribution of hydrogen bonds.

Entry	Surface Area ($Å^2$)	Surface Volume ($Å^3$)
a	203.06	201.15
d	256.56	271.22
h	246.07	255.08
i	389.93	466.45
m	219.04	217.94
q	240.63	247.06
r	247.49	248.02
x	230.88	229.67

Table 3.2: Table of Hirschfeld surface areas and volumes calculated from the crystal structure of DNA base analogues.

For the alkyl chains the Hirshfeld surface area and volume increased as expected as the chain length was increased. This is purely because the crystal is larger as the chain is lengthened by an increasing number of carbons. These alkyl chains increase the non-polar, electrostatic interactions between the surfaces. However, for the benzyl groups some other interactions were observed that were of note.



(a) (b)

Figure 3.17: Hirschfeld surface and finger plot of benzylthymine crystal structure

The hydrogen bonding between the expected donors and acceptors was seen at the reverse-Watson-Crick site as evidenced by the strong red areas on the Hirshfeld surface as shown in Figure 3.17. In addition to these further red areas can be seen on the Hirshfeld surface and there is evidence of some secondary peaks in the fingerprint plot. This shows that some other interactions are present in the crystal structure. Growing the crystal structure to show more adjacent molecules and analysing the Hirshfeld surface by element showed that further hydrogen bonding was present as well as π interactions.

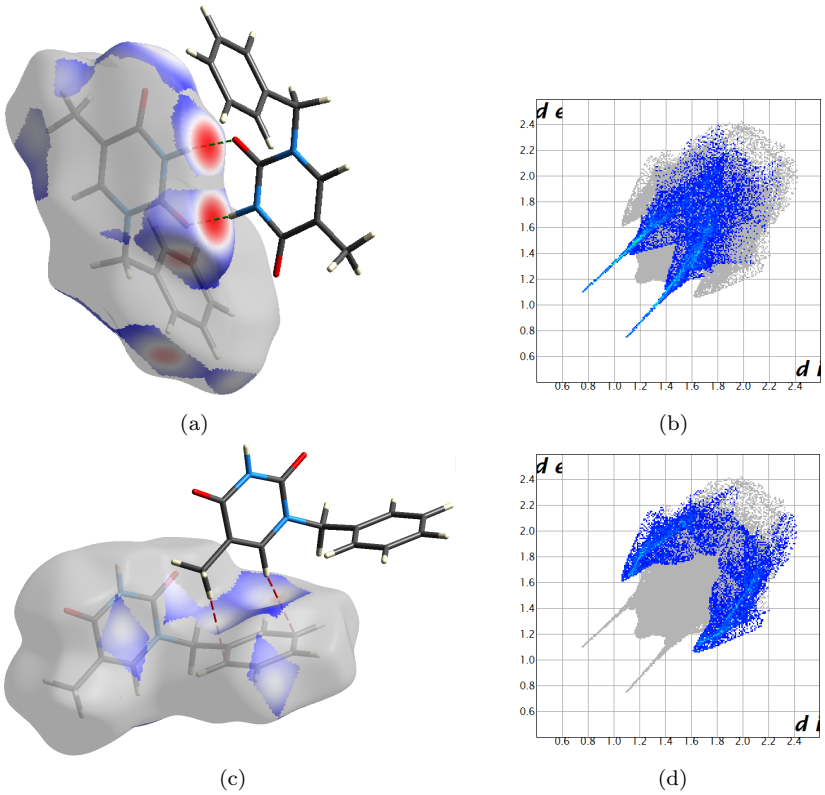


Figure 3.18: Hirschfeld surface and finger plot of benzylthymine crystal structure by element also showing the reciprocal contact between internal and external element. a), b) = Oi He; c), d) = Ci He

The further hydrogen bonding interactions were present between the C=O acceptors on thymine and the C-H donors on the benzyl group. CH protons are not as effective hydrogen bond donors as the more commonly observed NH or OH donors. This is due to the lower electronegativity of carbon compared to nitrogen or oxygen which does not induce as great a positive charge density on the proton. This decreased strength of hydrogen bond between CO and CH can be observed by much smaller red areas on the Hirshfeld surface.

For *N*-1-propylthymine any secondary interactions observed in the crystal were not observed in the NMR dilution. However, the other hydrogen bonding interactions between thymine and the benzyl group in *N*-1-benzylthymine can be observed in the NMR dilution spectra. A small shift was observed in both the benzyl, *C*-5 methyl and *C*-6 thymine protons as shown in figure (3.19). These shifts are not likely to be due to inductive effects of the hydrogen bond as they are relatively far from the site of the hydrogen bond so this suggests that for *N*-1-benzylthymine there are secondary interactions observed in the solution phase.

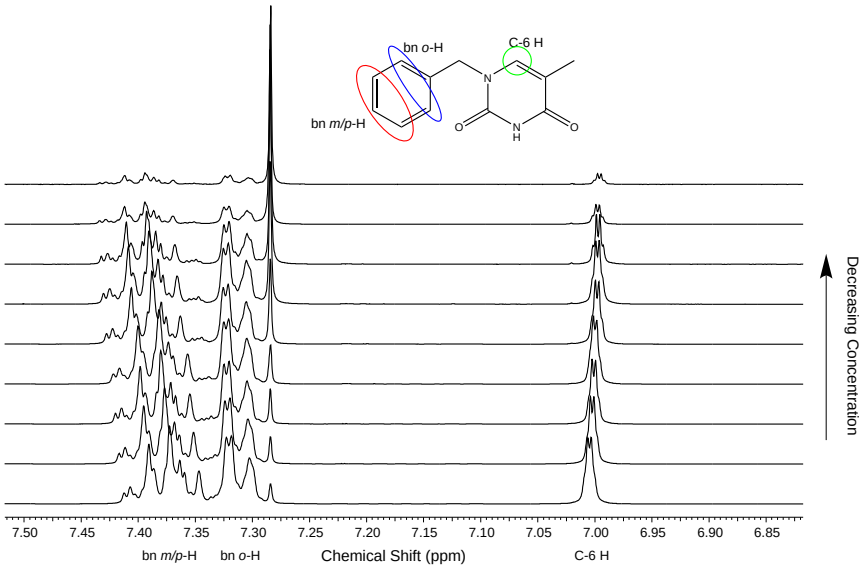


Figure 3.19: Overlay of the ^{1}H NMR dilution experiment of *N*-1-benzylthymine showing the shift in benzyl and *C*-6 protons of *N*-1-benzylthymine performed in $CDCl_3$ at 298 K.

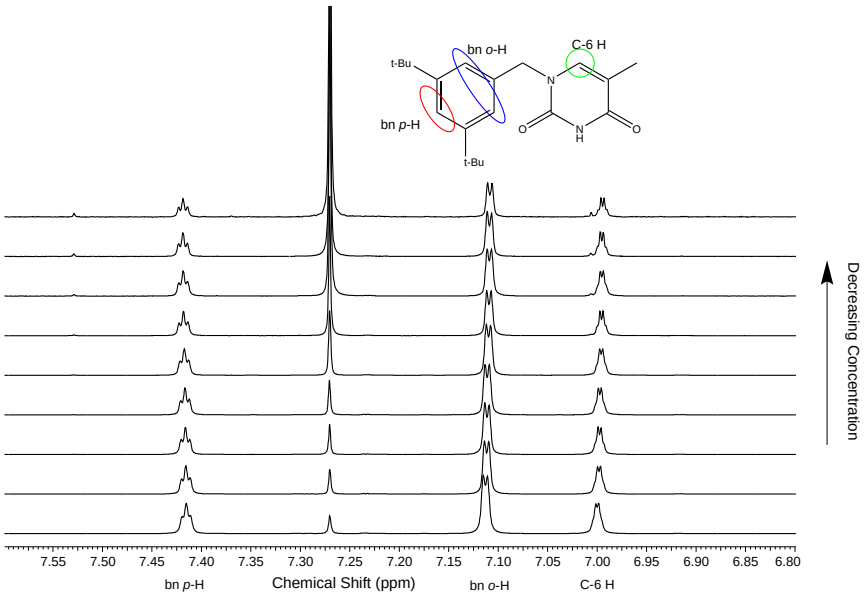
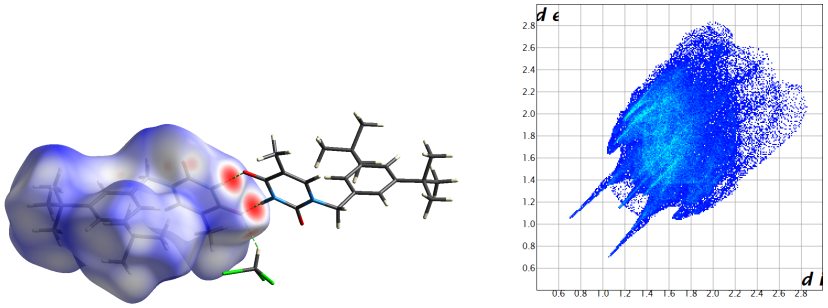


Figure 3.20: Overlay of the ^{1}H NMR dilution experiment of *N*-1-benzylthymine showing the lack of shift in benzyl and *C*-6 protons in *N*-1-(3,5-di-*t*-butyl)benzylthymine performed in $CDCl_3$ at 298 K.

The π interactions were present between the benzyl group carbons and the protons on the methyl and C6 group on thymine. These impact of these interactions on the crystal structure can be seen by observing the angle of the benzyl group. It was expected that the benzyl group would orthogonal to the thymine as this would minimise any steric interactions. The total surface area of the Hirschfeld surface was 246.07 Å^2 with a volume of 255.08 Å^3. Of this 23.3 % of the total surface were OH

interactions and 47.5 % were HH interactions. These were both slightly lower than expected due to the relatively large 18.8 % contribution of CH interactions which was not seen for the other crystal structures.

For the 3,5-(di-t-butyl)-benzyl group it was expected that these secondary hydrogen bonding interactions and π interactions would not be as prevalent due to the added steric bulk of the t-butyl groups. As can be seen from the NMR spectra above (Figure 3.20), this is indeed the case. No significant shift was observed in the benzyl, C-6 or C-5 methyl peak during the NMR dilution experiment. This suggests that the t-butyl groups inhibit these interactions. However, they were still observed in the crystal structure as can be seen in Figure 3.21 and 3.22.



(a) (b)

Figure 3.21: Hirschfeld surface and finger plot of 3,5-(di-t-butyl)-benzylthymine crystal structure.

For this crystal structure the hydrogen bonding between the thymine nucleobases was observed as expected as shown by the large red areas between the thymine hydrogen bond donors and acceptors. A number of smaller interactions were also present in the crystal structure including the interaction between the C-2 C=O and C-5 methyl group on adjacent molecules. These interactions are relatively weak due to the weaker hydrogen bonding between CH donors. A considerably larger hydrogen bonding interaction was observed between the C-4 C=O and the chloroform solvent molecule. For this hydrogen bond the donor is also a CH proton. However, it is considerably larger due to the inductive electronegativity of the three chlorine atoms. It can be seen in Figure 3.22 (c), d)) that Cl-H interactions are also present in the crystal and it these that are responsible for the non-symmetrical shape of the Hirshfeld finger plot as Cl is not present within the surface.

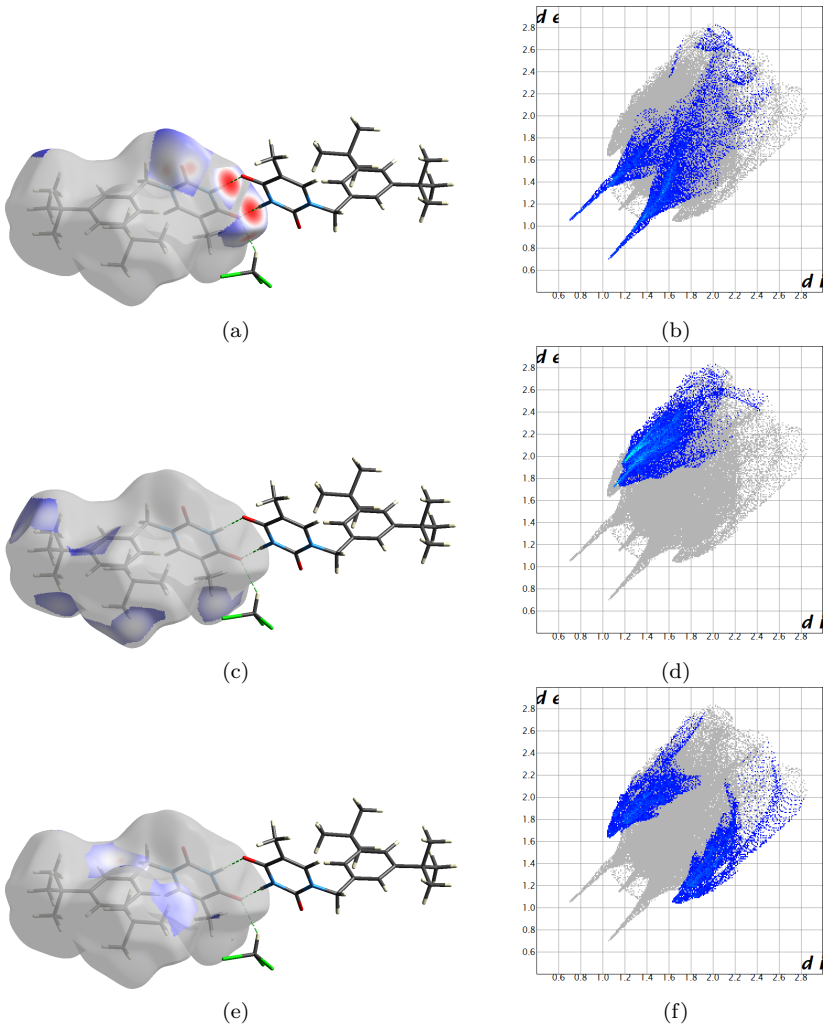
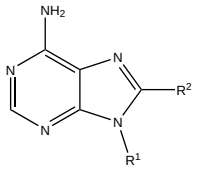


Figure 3.22: Hirshfeld surface and finger plot of 3,5-(di-*t*-butyl)-benzylthymine crystal structure by element also showing the reciprocal contact between internal and external element. a), b) = Oi He; c), d) = Hi Cle; e), f) = Ci He

It was not a surprise that 3,5-(di-*t*-butyl)-benzylthymine was the largest of the thymine crystal structures due to the bulky *t*-butyl groups. The surface had an area of 389.93 $Å^2$ and volume of 466.45 $Å^3$. With 54.6 % of this surface comprising HH interactions this crystal also had the largest contribution from electrostatic interactions.

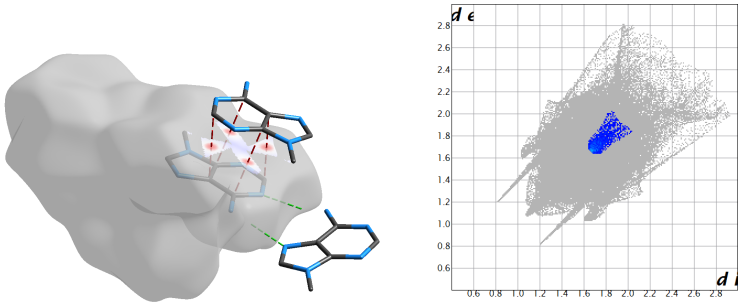
3.2.2 Alkylated Adenine



	R^1	R^2	K_d M^{-1}	r D-A (Å)	θ D-H-A
	n-pr	H	3.07	2.106 H5Bb-N2A,	163.193 N5B-H5Bb-N2A,
				2.207 H5Aa-N3B	169.828 N5A-H5Aa-N3B
				2.119 H5Ab-N2B,	161.122 N5A-H5Ab-N2B,
				2.226 H5Ba-N3A	165.883 N5B-H5Ba-N3A
c	*n*-bu	H			
h	benz	H		2.171 H00a-N002	174.939 N003-H00a-N002
				2.200 H00b-N005	161.704 N003-H00b-N005
i	*t-bu*-benz	H	0.29		
	n-pr	Br		2.390 H1Ab-N2B	174.012 N1A-H1Ab-N2B
				2.235 H1Ba-N4A	166.027 N1B-H1Ba-N4A
				2.161 H1Bb-N2A	169.851 N1B-H1Bb-N2A
				2.288 H1Aa-N4B	166.777 N1A-H1Aa-N4B
l	*n*-pr	I		2.098 H1Ab-N2B	155.241 N1A-H1Ab-N2B
				2.160 H1Ba-N4A	171.104 N1B-H1Ba-N4A

Table 3.3: Table of association constants for the synthesised adenine analogues as calculated by NMR dilution and hydrogen bonding lengths (r) and angles (θ) as calculated from the relevant crystal structure.

The crystal structure of 3,5-(di-*t*-butyl)-benzyladenine was of interest due to the stacking interactions observed between the adenine molecules. π stacking interactions were observed between the benzyl-functionalised thymine molecule but these interactions were between the benzyl group and not the nucleobase itself. π stacking of the nucleobases is of greater interest as it is also observed within a biological DNA helix and could suggest some cooperativity of the intermolecular interactions. π stacking interactions between the nucleobases will induce the close proximity of hydrogen bonding pairs and may make further hydrogen bonds more likely. The directional nature of π stacking and hydrogen bonds could work together to increase the likelihood of further interactions as shown in Figure 3.23.



(a) (b)

Figure 3.23: Simplified crystal structure of 3,5-(di-*t*-butyl)-benzyladenine showing the π stacking interactions by analysing the Hirschfeld surface by element Ci Ce.

As well as the C-C π stacking interactions shown above, the adenine molecules were found to hydrogen bond alternatively on the Hoogsteen and Watson-Crick faces as seen for other adenine analogues. Of particular interest here was the slight twist in the hydrogen bond between the Watson-Crick face of the highlighted molecule and the Hoogsteen face of the adjacent partner. It is likely that this twist is due to the steric hindrance of the relatively large 3,5-(di-t-butyl) benzyl group as shown by the interactions observed on this part of the molecule in Figure 3.24.

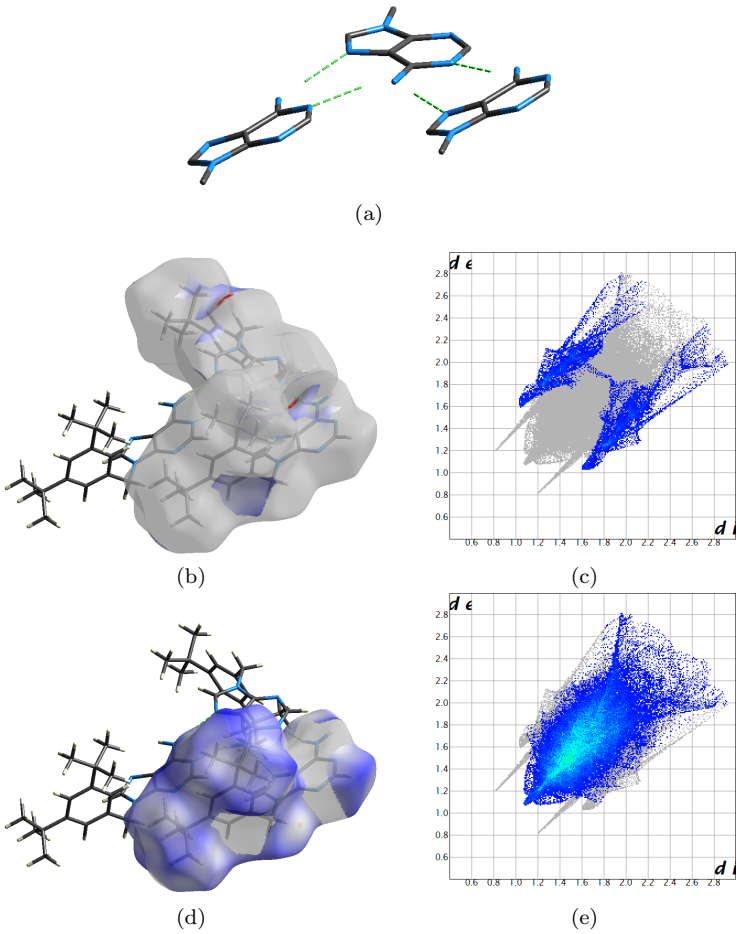
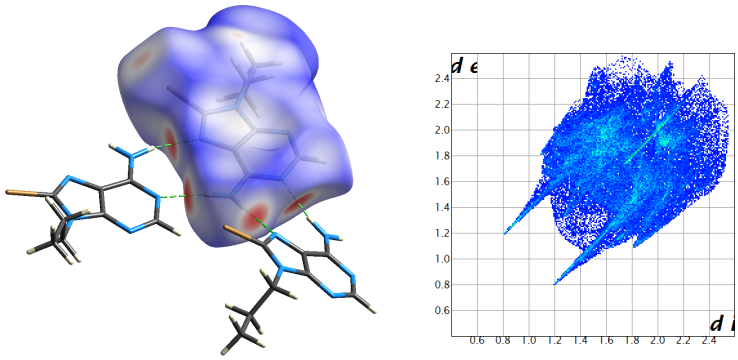


Figure 3.24: a) A simplified structure 3,5-(di-*t*-butyl)-benzyladenine showing the twist in the WC-H hydrogen bond caused by; b) c) = Ci He (reciprocal) and d) e) = Hi He close contacts.

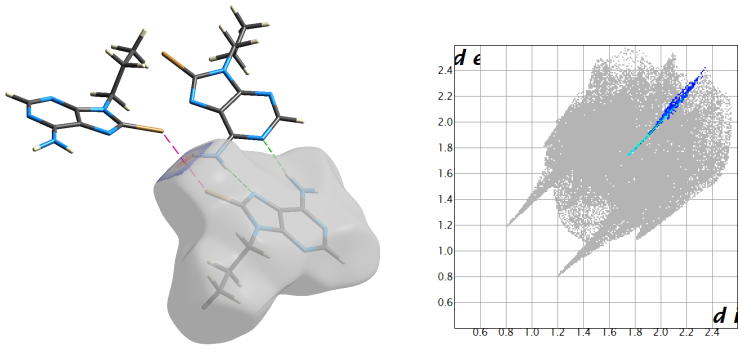
For the crystal structure of 8-bromo-*N*-9-propyladenine it was expected that the Watson-Crick binding face would be preferred compared to the Hoogsteen face due to the electron-withdrawing effects of the bromine. This would make the *N*-7 nitrogen a less effective hydrogen bond acceptor. However, as can be seen in Figure 3.25 both Watson-Crick and Hoogsteen faces bind alternatively in a manner similar to that which was observed for *N*-9-propyladenine.



(a) (b)

Figure 3.25: Crystal structure and Hirshfeld surface and finger plot of 8-bromo-*N*-9-propyladenine.

The binding of the Watson-Crick face to the Hoogsteen face suggests that, although less favourable than pure Watson-Crick binding, it is favourable in this crystal structure due to the other interactions present in the crystal lattice. It is noticeable that a considerable twist can be seen in the hydrogen bonding between the Watson-Crick face of the surface highlighted molecule and the Hoogsteen face of its hydrogen bond partner. This is in contrast to what was observed for N-9-propyladenine where the molecules were arranged in a planar arrangement. This twisted hydrogen bond is not the most favourable orientation of a hydrogen bond due to the strained bond angles present and it suggests that other interactions are forcing the molecules to twist to allow for the total energy to be lower in the system. Analysing the Hirshfeld surface by element shows some Br-Br interactions and it is likely that it is these that cause the twist in the hydrogen bond as shown in Figure 3.26. The Hirschfeld surface is also considerably larger for 8-Bromo-*N*-9-propyladenine than *N*-9-propyladenine as expected due to the relatively large bromine atom. From the crystal structure of 8-iodo-*N*-9-propyladenine the same Watson-Crick to Hoogsteen hydrogen bonding was observed as for the bromo and unhalogenated analogue. However, the presence of water molecules hydrogen bonding to the adenine structure made further analysis of the crystal structure difficult.

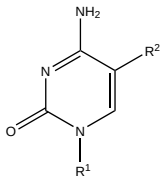


(a) (b)

Figure 3.26: Crystal structure and Hirshfeld surface and finger plot of 8-bromo-*N*-9-propyladenine by element showing the Br-Br interaction (magenta) which causes the twist in the hydrogen bond (lime).

3.2.3 Alkylated Cytosine

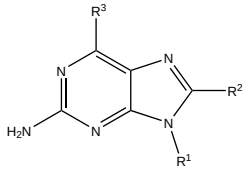
It was not possible to study the interactions of the alkylated cytosine analogues. This was due to the insufficient solubility of the molecules and a failure to find crystallisation conditions that led to successful isolation of a single crystal.



	R^1	R^2	K_d M^{-1}
a	*n*-pr	H	
i	*t-bu*-benz	H	
l	*N*-pr	I	

Table 3.4: Table of the synthesised cytosine analogues. No dilution experiments were performed on these molecules.

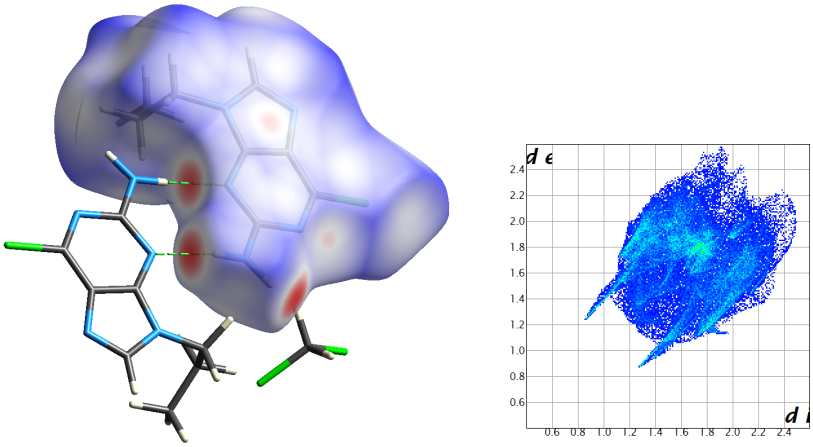
3.2.4 Alkylated Guanine



	R^1	R^2	R^3	K_d M^{-1}	r D-A (Å)	θ D-H-A
a	*n*-pr	H	OH			
	t-bu-benz	H	OH		2.037 H2Cb-N4B	168.152 N2C-H2Cb-N4B
					1.959 H1Ca-O1B	165.887 N1C-H1Ca-O1B
					1.943 H1Ba-N4D	173.425 N1B-H1Ba-N4D
					2.008 H2Bb-O1D	157.573 N2B-H2Bb-O1D
					2.033 H2Db-N4A	165.117 N2D-H2Db-N4A
					1.992 H1D-O1A	164.504 N1D-H1D-O1A
h	*N*-pr	H	Cl	0.01		
i	*t-bu*-benz	H	Cl			
k	*n*-pr	Br	Cl		2.281 H2b-N1	163.526 N2-H2b-N1

Table 3.5: Table of association constants for the synthesised guanine analogues as calculated by NMR dilution and hydrogen bonding lengths (r) and angles (θ) as calculated from the relevant crystal structure.

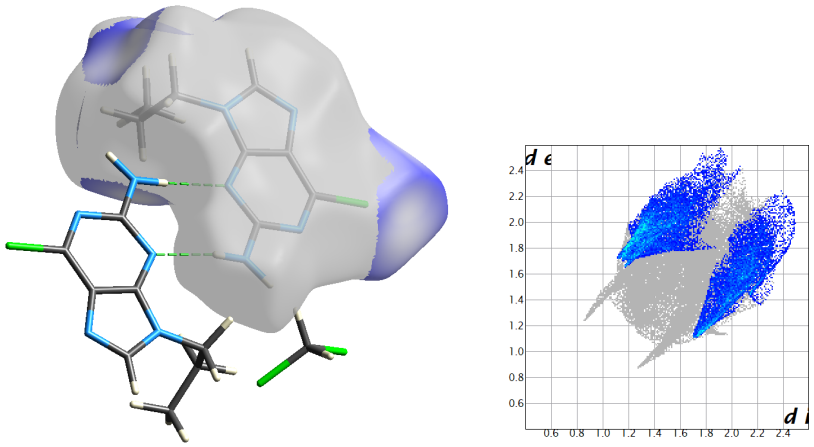
The crystal structure of 2-amino-6-chloro-*N*-9-propylpurine was particularly interesting due to the two molecules engaging in hydrogen bonding through neither the Watson-Crick or Hoogsteen faces of the molecule as seen in Figure 3.27



(a) (b)

Figure 3.27: Crystal structure and Hirshfeld surface and finger plot of 2-amino-6-chloro-*N*-9-propylpurine.

Instead of hydrogen bonding through the Watson-Crick or Hoogsteen face of the molecule the hydrogen bonding was observed between the 2-amino protons and the *N*-3 nitrogen. This site is not normally observed in nature due to the phosphate-sugar backbone limiting its access to other nucleobases within the DNA helix. In the alkylated structures this position is also the most sterically hindered. It could be suggested that the chlorine atom is having an effect on the hydrogen bonding of the purine molecule and making the Watson-Crick and Hoogsteen faces less favourable. Chlorine is relatively electronegative element and can withdraw electrons from the *N*-1 nitrogen making it a less effective hydrogen bond acceptor. The chlorine atom also makes the Hoogsteen face unavailable quite simply as there are no hydrogen atoms to act as hydrogen bond donors. However, it is possible that the Watson-Crick face hydrogen bonding is not observed in the crystal structure due to the other factors that govern the growth of a crystal such as sterics. Observing the Hirschfeld surface by element also shows some halogen bonding interactions between the chlorine atom and the propyl chain (Figure 3.28) so it is possible that the effect of these could also make the *N*-3 nitrogen hydrogen bonding site more favourable.



(a) (b)

Figure 3.28: Hirschfeld surface and finger plot of 2-amino-6-chloro-*N*-9-propylpurine showing the halogen bonding between Cli and He.

3.3 Conclusions

The hydrogen bonding interactions of a number of different DNA base analogues were studied. Analysis by NMR dilution and titration showed that, as expected, a complementary base pair of adenine and thymine associates more strongly than the relevant self-associative dimer. A co-crystal of propyladenine and propylthymine was also analysed and a W-C-H hydrogen bonding mode observed. The effect of alkyl chain length and size on the hydrogen bonding interactions of DNA nucleobase analogues was also investigated. NMR dilution experiments showed that the alkyl chain has a minor effect on the strength of hydrogen bonding interaction. However, when aryl groups are introduced into the chain evidence of secondary π interactions are observed in the NMR spectra. These π interactions were also observed via Hirshfeld surface analysis of the crystal structure of arylated nucleobases. It was found that the introduction of bulky steric groups can limit these π interactions as shown in the NMR spectra and Hirshfeld surface. Halogen groups can also have an effect on the intermolecular interactions of DNA nucleobases due to the formation of secondary interactions and the electronic effect of the halogen on the hydrogen bond donors and acceptors.

3.4 Experimental

3.4.1 Binding Studies

Deuterated chloroform was purchased from Sigma Aldrich and filtered through basic alumina and dried over 4Åmolecular sieves. Basic alumina was purchased from Sigma Aldrich and used as supplied.

Self-association

A high concentration of analyte was prepared in $CDCl_3$ and 450 μl of this solution was transferred to a capped NMR tube. A small volume of pure $CDCl_3$ was added successively to dilute the sample and a ^{1}H NMR spectrum recorded following each dilution. The data was fitted to a dimerisation model by solving the equations outlined previously using non-linear curve fitting procedure.

Host-Guest association

A chosen analyte was employed as the host of which the solution concentration was kept constant. The host was dissolved in $CDCl_3$ to make a 2 ml solution (4 5 mM). 500 μl of the host solution was transferred to a capped NMR tube and a ^{1}H NMR spectrum was recorded. The guest compound was dissolved in the remaining host solution and titrated into the NMR tube. ^{1}H NMR spectra were recorded following addition of each titre. The data was fitted to 1:1 binding isotherm by solving the equation outlined previously. The chemical shift of NH on host was used for fitting.

Crystallography

Single crystal X-ray diffraction experiments were performed on a Rigaku Saturn724+ diffractometer equipped with a rotating anode using monochromated Cu-Kα radiation ($\lambda = 1.5418$ Å) at 120 K. The structures were solved by direct methods using either SHELXS or SHELXT[111] and refined with SHELXL[112] using a least squares method. OLEX2 software was used as the solution, refinement and analysis program.[113] All hydrogen atoms were placed in geometrically calculated positions; non-hydrogen atoms were refined with anisotropic displacement parameters.

Chapter 4

The Synthesis of DNA-Functionalised Porphyrins

To study the surface self-assembly of the DNA nucleobases it was necessary to synthesise a supramolecular tecton bearing at least one DNA nucleobase. For surface self-assembly studies it was important that the tecton would readily associate on a surface. It was also important that the tecton could bear multiple DNA nucleobases if possible. Finally, a well-defined molecular structure that would present the DNA nucleobases to each other in a controllable, predictable manner was required. To this end, a porphyrin was chosen as functionalisation is possible in multiple positions, it is planar and its surface behaviours are well studied. A porphyrin also bears four-fold symmetry and as there are four DNA bases this will allow for the possibility of evenly distributing four different DNA bases on the molecule. Thymine functionalised porphyrins have also been synthesised previously in the group so there is a proven synthetic route which can be adapted and comparative studies can be performed[83].

There are numerous methods and procedures for synthesising functionalised porphyrins as outlined in Figure 4.1. Porphyrins are synthesised from the condensation of an aldehyde with pyrrole[114]. Therefore, one method for synthesising functionalised porphyrins is by synthesising functionalised aldehydes. This functionalisation will then be carried onto the final product. This method was employed previously in the group for the synthesis of thymine functionalised porphyrins[83]. In a similar manner it is also possible to functionalise the pyrrole moeity. Both these methods impart the functionalisation before the synthesis of the porphyrin ring. Another method involves the synthesis of a porphyrin with reactive groups and then adding the desired functionalisation onto the porphyrin once it has been formed. Both these methods have their own advantages and disadvantages which will be discussed. Furthermore, each method presents different features which may be utilised to form the desired product with tuned properties in an efficient a manner as possible.

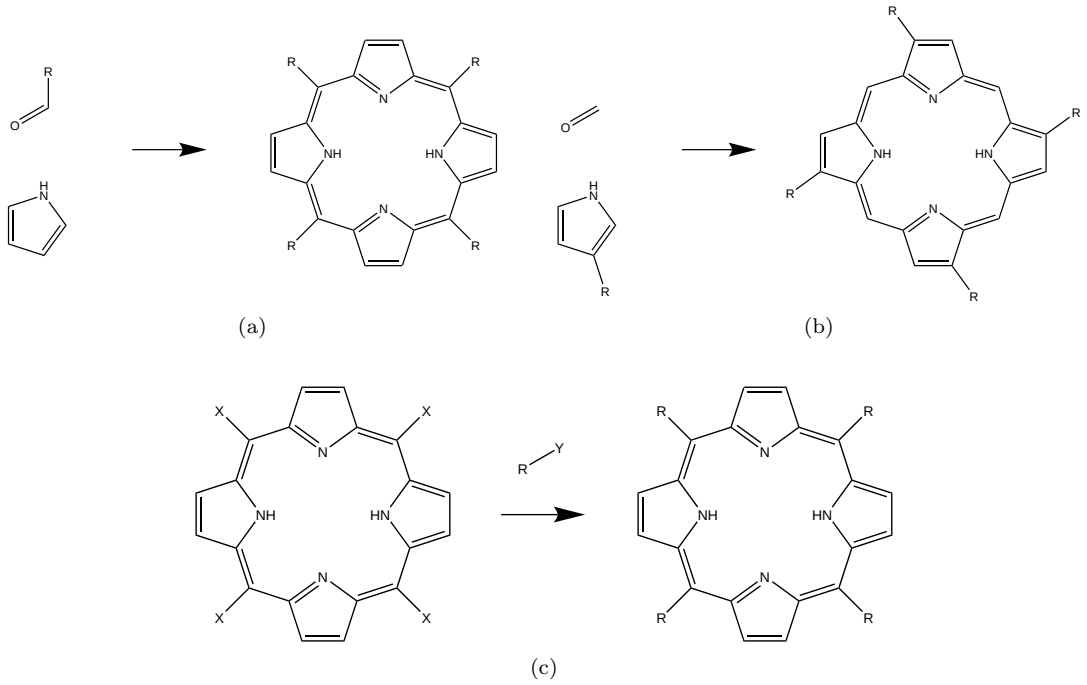


Figure 4.1: Simplified schemes of different methods for the production of functionalised porphyrins via; a) functionalised aldehydes b) functionalised pyrrole c) post-porphyrin-synthesis functionalisation.

The method of using functionalised aldehydes was utilised for the synthesis of thymine functionalised porphyrins[83]. This method adds the functionalisation at the meso position on the porphyrin which is particularly useful as it reliably places each functionalised base 90 degrees apart from each other in the (5,10,15,20) positions which is important for control of the self-assembly of the compounds. This technique also introduces the functionalisation early on in the synthetic route meaning that the functionalisation is likely to be done using simpler and cheaper starting materials. If the functionalisation step is particularly difficult or poor yielding it may be beneficial to perform the reaction will less valuable starting materials without any possible hindrance that using a more complex molecule may bring (solubility, reactivity, selectivity etc.).

The method of functionalising the pyrrole moeity positions the functionalisation in a different position to that achieved when functionalising the aldehyde. A clear disadvantage of using this method is the lack of control over the regiochemistry of the final product. The functionalisation could occur in the positions shown in Figure 4.1 where the functionalisation is evenly spaced (2,7,12,17) or, alternatively, in the (2,7,12,18) or (2,8,12,18) positions due to the two different possible orientations of the pyrrole. This lack of control makes this method undesirable for the synthesis of supramolecular tectons as the products are often difficult to separate.

Post-synthetic functionalisation of the porphyrin can be performed in either the meso position or on any other position on the porphyrin where the reaction conditions suit. In some cases the functionalisation reaction may be selective which allows for the efficient synthesis of porphyrins where multiple functionalisations are present. This is especially applicable in this work where the functionalisation of a porphyrin with multiple different nucleobases is desired. This technique also allows for the synthesis of a much simpler porphyrin; at least at first. Porphyrin synthesis is typically poor yielding and porphyrins are often not trivial to separate. Therefore, synthesising a porphyrin from simple starting materials in large quantities and then adding the functionalisation after may be a more efficient route. If the functionalisation is added before the porphyrin is synthesised it is

likely that at least 70% of the functionalised starting material will be lost.

For the study of more complex self-assembled systems bearing multiple selective hydrogen bonding sites it is necessary to synthesise a porphyrin bearing multiple different DNA bases. Ideally, a porphyrin bearing one of each of the four DNA bases would be synthesised. The synthesis of porphyrins bearing different meso functionalities is not trivial and there are a number of different methods for their synthesis, as detailed in the Introduction and outlined in Figure 4.2. In brief, there is the Adler-Longo modified Rothemund method where the different functionalised aldehydes are simply mixed in the stoichiometric quantities with pyrrole and the statistical product is formed[115][116][117]. There is the more considered approach first described by Lindsey where each 'side' of the porphyrin is assembled separately and then selectively reacted together in a multi-step synthesis[114][118]. It is also possible to synthesise a symmetric porphyrin and then selectively add each independent unit.

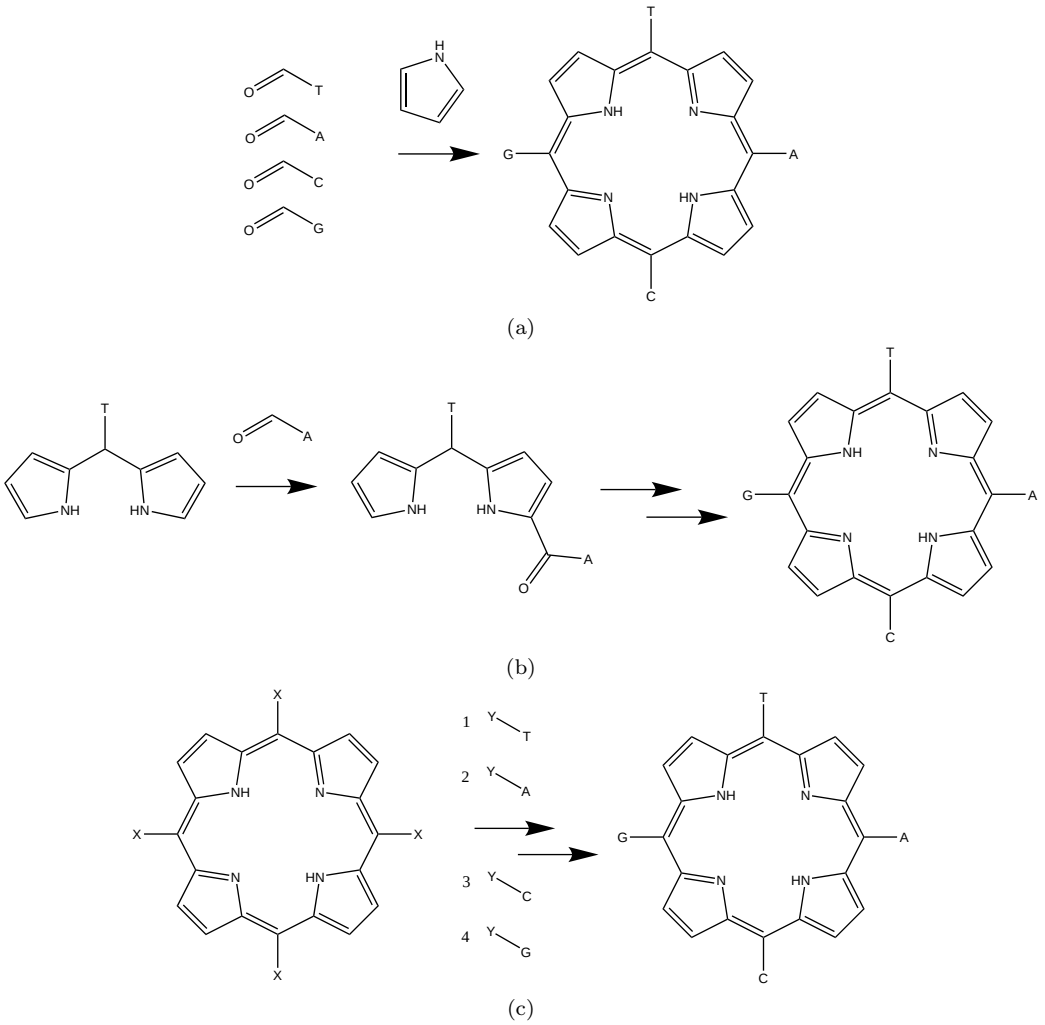


Figure 4.2: Simplified schemes of different methods for the production of functionalised porphyrins via; a) mixed functionalised aldehydes b) Lindsey multi-step synthesis c) post-porphyrin-synthesis selective functionalisation.

Each of the methods for synthesising porphyrins with different meso functionality has advantages

and disadvantages. The mixed aldehyde method, while being synthetically the most simple, has the major disadvantage of being almost entirely unselective. It is a relatively simple method as in most cases it is possible to synthesise large amounts of each aldehyde with relative ease and then the starting materials are simply mixed together. However, as shown in Table 4.1 below, this statistical approach will yield 59 different possible products from one reaction (assuming each aldehyde has equal reactivity). Some of these products will be formed in very small amounts but the obvious difficulty in purification of the mixture demonstrates how the mixed aldehyde method is unsuitable for the synthesis of a porphyrin where the meso functionality is not all the same. Even the TACG desired product is formed with three possible positional isomers.

4:0	3:1	2:2*	2:1:1*	1:1:1:1**
TTTT	TTTA	TTAA	TTAC	TACG
AAAA	TTTG	TTCC	TTAG	
CCCC	TTTC	TTGG	TTCG	
GGGG	AAAT	AACC	AATC	
	AAAG	AAGG	AATG	
	AAAC	CCGG	AACG	
	CCCT		CCTA	
	CCCA		CCTG	
	CCCG		CCAG	
	GGGT		GGTA	
	GGGA		GGTC	
	GGGC		GGAC	

Table 4.1: Table of different possible products formed when 4 aldehydes are statistically mixed in an Adler-Longo method. * = 2 positonal isomers; ** = 3 positional isomers

The synthesis route described by Lindsey or adaptations thereof are the most common route used for the synthesis of porphyrins where differing meso functionalities are present.[114][118] This is because it overcomes the problem of statistical mixtures. In this method the porphyrin is sequentially assembled in a multistep synthesis. First, a functionalised dipyrromethane is synthesised which is selectively reacted with another aldehyde (or carboxylic acid, acid chloride etc.) and then another dipyrromethane and so on until a porphyrin ring is formed. This method has the clear advantage of selectively forming the desired product. It is also possible to control the positional isomerisation of the final product. However, this method requires a large number of steps and the use of challenging reaction techniques due to the lack of stability of the intermediates.

The post-synthetic functionalisation route can be made selective for the synthesis of porphyrins by utilising selective functionalisation techniques[119]. These may include the use of protecting groups or by manipulating the stoichiometry of each functionalisation step so that only the desired product is formed. For example, Y-T may be reacted with a stoichiometric excess of tetra-X-porphyrin so that only tri-X-mono-T-porphyrin is formed. This technique may also produce positional isomers and unwanted products where the functionalisation occurs twice on the same molecule. However, many fewer isomers will be formed with this technique than the statistical mixed isomer approach as shown in the flowchart below (Figure 4.3). A limitation of this approach is the difficulty in finding functionalisation steps that are selective. Successive functionalisation steps may also lead to low yields of the final product once any side products and the culmination of imperfect yields are taken into account.

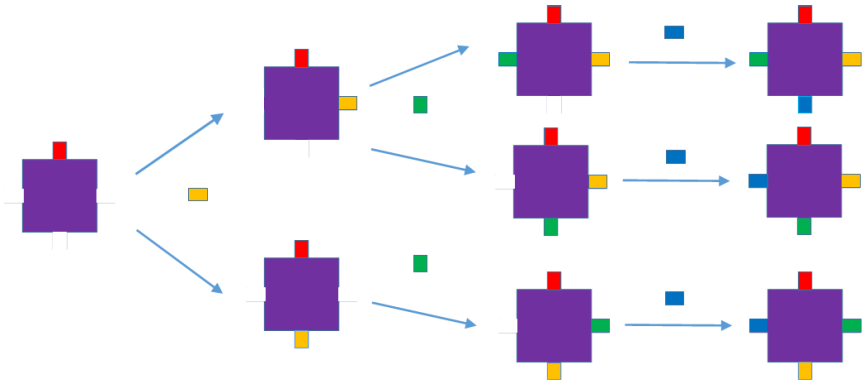


Figure 4.3: Flowchart showing the three positional isomers formed via successive selective functionalisations of a porphyrin.

4.1 Synthesis of DNA-Base Functionalised Aldehydes

4.1.1 *N*-1 and *N*-9 arylated Nucleobases

Previous work in the group synthesised a thymine-functionalised porphyrin by synthesising an aldehyde bearing the thymine moiety and performing the relevant condensation reaction to form the desired porphyrin[83]. This approach was successful so it was adapted to synthesise analogous porphyrins bearing the other DNA bases. Firstly, the DNA functionalised aldehydes in Figure 4.4 were synthesised. This process was largely adapted from previous work in the group or the excellent report by Gothelf *et al.*[120][121].

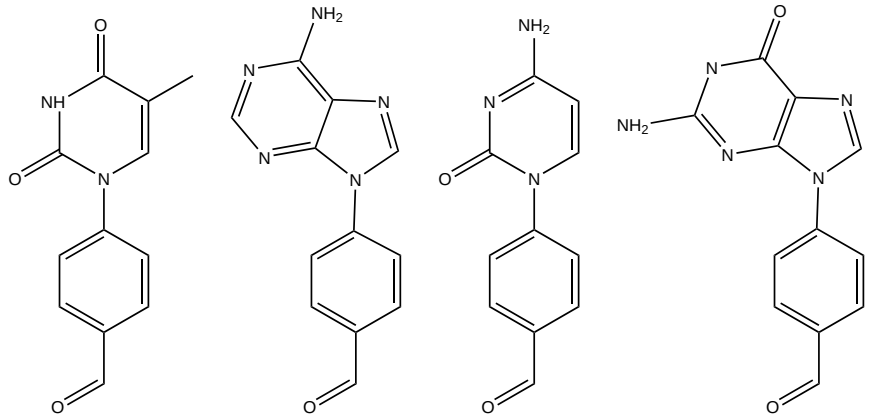


Figure 4.4: The DNA nucleobase functionalised aldehydes synthesised from the adapted procedure by Gothelf *et al.*

The DNA base functionalised aldehydes above all feature the para-formylphenyl group in the *N*-1 or *N*-9 position for pyrimidines and purines respectively. The goal was to functionalise the DNA nucleobases at the same nitrogen that is normally attached to the sugar-phospate backbone in DNA. This was with the aim of allowing the final products to perform in a similar way to the *N*-1 and *N*-9 functionalised aldehydes synthesised previously. The aldehyde functional group is separated from the DNA nucleobase by a phenyl group. This is to sufficiently space the nucleobases out in the final product. The phenyl group is also planar and relatively rigid compared to an alkyl chain. This will

facilitate control of the position of the DNA base in the final product and allow for predictable and reliable self-assembly. The nucleobase N-C bond can also rotate to provide sufficient movement of the DNA nucleobase which will allow for facile hydrogen-bonding to occur.

For the synthesis of N-1 or N-9 aryl functionalised DNA bases it was necessary to form a C-N bond. For this the previously utilised SN2 nucleophilic substitution reaction was not suitable. The SN2 reaction for the formation of alkyl DNA bases was successfully performed using an alkyl halide and this reaction is not possible using an aryl halide. Aryl halides are relatively unreactive to nucleophilic substitution due to the steric hindrance of the aryl group and the increased C-X bond strength in aryl halides compared to alkyl halides[109]. Gothelf *et al.* report the synthesis of N functionalised aryl DNA nucleobases via a Chan-Lam Evans modified Ullmann coupling[122][123]. This is a copper-mediated coupling reaction where an aryl boronic acid is oxidatively coupled to NH or OH compounds in air[124]. This coupling has advantages over the comparable Buchwald-Hartwig coupling as it may be performed in air at room temperature and a relatively inexpensive copper catalyst is used in contrast to a palladium catalyst. The synthesis of N-aryl nucleobases has classically been performed via heterocyclization of suitable precursors[125][126]. This technique, however, has often led to low yields and a new synthetic route is required for each nucleobase. Considering the compelling disadvantages of other routes and the prior success in synthesising N-1 aryl thymine the Chan-Lam Evans coupling route was chosen for the synthesis of N-1 or N-9 para-formylphenyl nucleobases using the conditions shown below (Figure4.5):

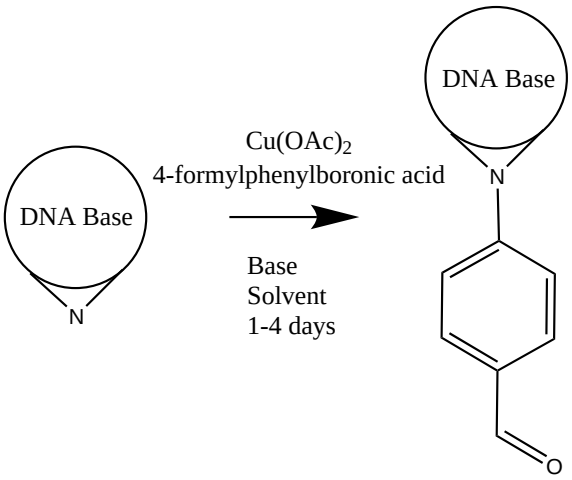


Figure 4.5: Scheme for the Chan-Lam Evans Ullmann coupling used for the synthesis of N-1 or N-9 paraformylphenyl nucleobases

Gothelf *et al.* reported the requirement for protecting groups in the coupling reaction[120]. Protecting groups are required to eliminate possible N-arylation at undesired positions. A further benefit of using protecting groups is to increase the solubility of the DNA bases in the reaction solvent and to allow for more convenient purification by column chromatography. A number of different protecting groups and solvents were investigated by Gothelf and those which gave the best results in each case were used[120]. It was interesting to note that different protecting groups and solvents were found to be optimal for each base. Different bases were also tested with similar results. For N-1-paraformylphenylthymine the benzoyl group was used. For this the synthesis of N-3-benzoylthymine was required. This was synthesised in two steps from thymine as described by Reese *et al.*[4]. N-3-benzoylthymine was then carried forward into the synthesis of N-3-benzoyl-N-1-paraformylphenylthymine using the conditions shown below (Figure 4.6). Pyridine was used as the base and DCM was used as the solvent.

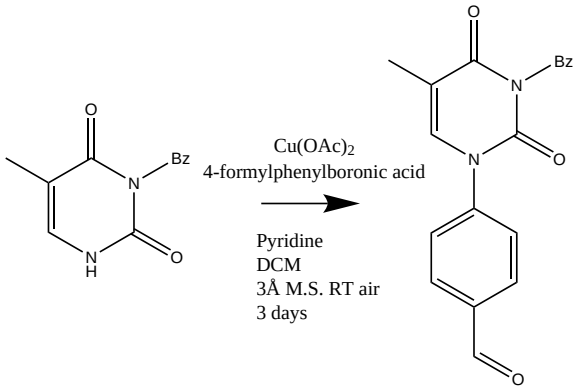


Figure 4.6: Scheme for the Chan-Lam Evans Ullmann coupling of 4-formylphenylboronic acid and *N*-3-benzoylthymine.

The reaction was carried out for three days until no further presence of starting *N*-3-benzoylthymine could be observed by TLC or mass spectrometry. The product was then isolated via aqueous work up, using EDTA as a chelating agent to remove copper, and purified by column chromatography. The identity of the final product was confirmed by mass spectrometry and proton and carbon NMR spectroscopy.

For the synthesis of *N*-9-aryl adenine derivatives it was described in the literature how the Boc protecting group gave optimal results under Chan-Lam Evans coupling conditions[120]. Therefore 6-bis-Boc-aminopurine was synthesised in two steps as previously described in the synthesis of propyladenine (Section 2.2.2)[3]. The 6-bis-Boc-aminiopurine was then reacted with 4-formylphenylboronic acid under the Chan-Lam-Evans coupling conditions shown below (Figure 4.7).

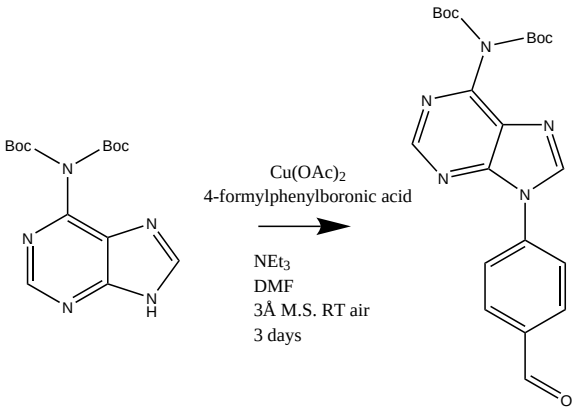


Figure 4.7: Scheme for the Chan-Lam Evans Ullmann coupling of 4-formylphenylboronic acid and 6-bis-Boc-aminopurine.

For adenine it was reported by Gothelf that improved yields were observed when triethylamine was used as a base and DMF was used as a solvent. It is suggested that is due to the increased strength of triethylamine as a base and the increased solubility of the products in DMF. This reaction was not successful when 6-mono-Boc-aminopurine was used which is likely due to the reduced solubility of the mono-Boc starting material. Due to the fact that arylated purines can chelate to Cu(II) it was especially important to remove the copper with the strongly chelating EDTA during the work up[127]. The isolation of pure product from the reaction was more challenging than

for the thymine species for a number or reasons. Firstly, the DMF solvent proved more challenging to remove compared to DCM due to is miscibility with water and high boiling point. It was necessary to thoroughly remove all traces of DMF as even small traces would lead to complications when column chromatography was used. Small amounts of partial product deprotection were also observed during the purification. When column chromatography was performed both bis and mono-Boc products were eluted. This is likely due to the mildly acidic properties of the silica which was used. A combination of these factors led to comparatively low yields for the desired product. It should be noted that the mono-Boc product, whilst not the desired product, is still of some use as the Boc-protecting groups will have to be removed from the final porphyrin product.

The synthesis of N-1-paraformylphenylcytosine was reported by Gothelf *et al.* via a 4-allythiocytosine derivative[120]. This was synthesised from the commercially available 4-thiouracil. This route has the advantage that 4-allylthiocytosine derivatives can readily be converted to cytosine by methanolic ammonia or converted to uracil upon treatment with strong aqueous acid or base[128]. The synthesis of N-1-paraformylphenylcytosine under the conditions described by Gothelf *et al.* was attempted by Sam Gilson as shown below (Figure 4.8).

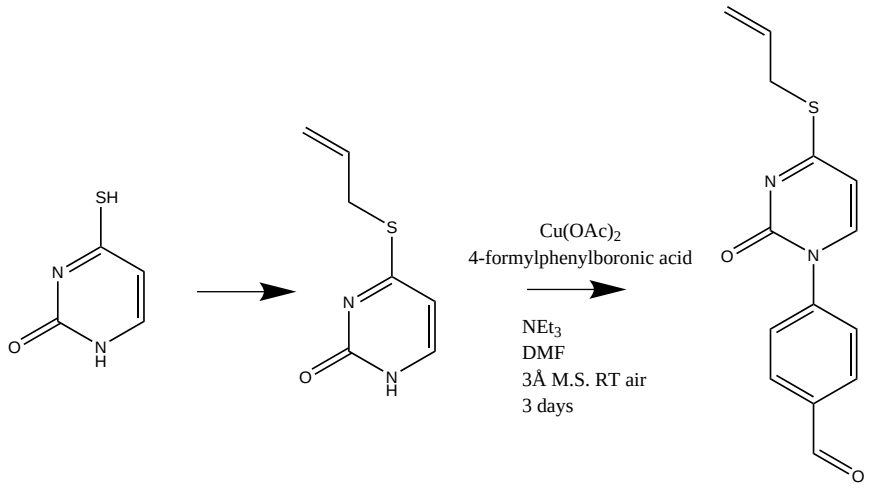


Figure 4.8: Scheme for the Chan-Lam Evans Ullmann coupling of 4-formylphenylboronic acid and 4-vinylthiouracil.

Whilst this synthesis was successful, only low yields were obtained and although 4-thiouracil is commercially available, it is prohibitively expensive. Therefore, alternative routes for the synthesis of *N*-1-paraformylphenylcytosine were investigated. Reactions using the commercially available and inexpensive *N*-4-benzoylcytosine as used for the synthesis of alkylcytosine derivatives were attempted but these were unsuccessful. The attachment of the Boc protecting group was then attempted using *N*-4-bis-Boc-cytosine as synthesised in Section 2.2.3. The Chan-Lam-Evans coupling of this starting material was then performed successfully using the condtions shown below (Figure 4.9. The correct regiochemistry of this cytosine aldehyde was confirmed by the crystal structure.

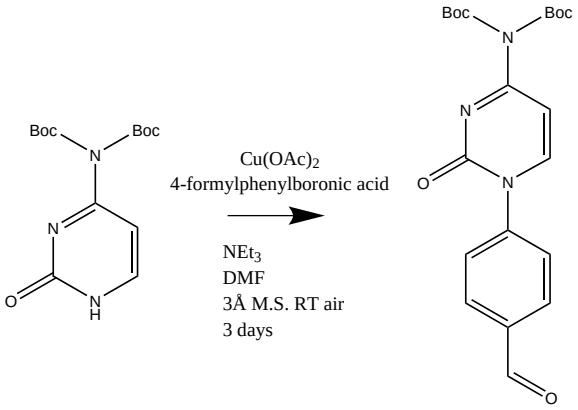


Figure 4.9: Scheme for the Chan-Lam Evans Ullmann coupling of 4-formylphenylboronic acid and *N*-4-bis-Boc-cytosine.

For a guanine functionalised phenyl aldehyde a 2-amino-6-chloropurine derivative was used. This was due to the notoriously low solubility of guanine compounds in common organic solvents. The 2-amino-6-chloropurine derivative was protected using the Boc protecting group as described earlier for the synthesis of alkylguanine (Section 2.2.4) and the coupling reaction with 4-formylphenylboronic acid was performed using the conditions below (Figure 4.10). Hydrolysis and Boc-deprotection can be performed simultaneously by heating the product in aqueous acid to afford the guanine derivative.

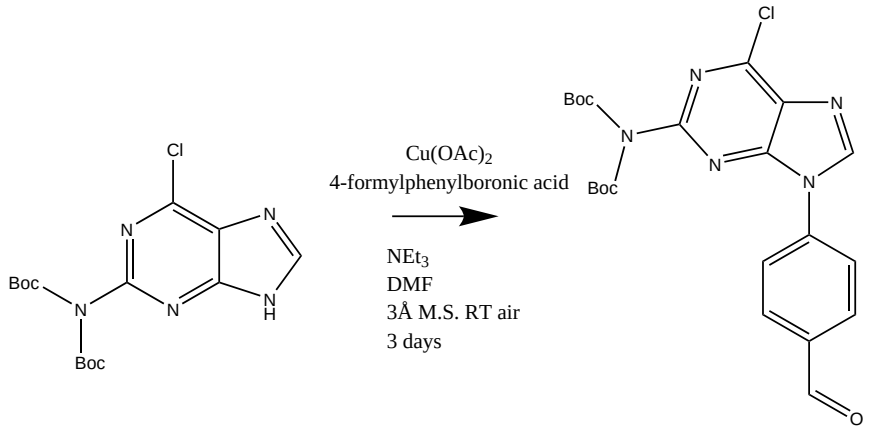


Figure 4.10: Scheme for the synthesis of a guanine functionalised aldehyde precursor via the Chan-Lam Evans Ullmann coupling 4-formylphenylboronic acid and bis-Boc-2-amino-6-chloropurine.

In the report by Gothelf *et al.* a number of different methods for the deprotection of the coupled aryl nucleobases are described[121][121]. In this work it was decided that the protecting group would be retained until after the porphyrin step. This is because the protecting groups aid in the solubility of the DNA nucleobases in common organic solvents. Therefore, it was expected that porphyrin synthesis and purification would be easier with the protecting groups present. Once the DNA functionalised porphyrin had been successfully synthesised and isolated then the protecting group could be removed. Deprotection steps are generally highly yielding with easily separable side products so it was not expected that these would present many significant challenges.

4.1.2 *C*-5 and *C*-8 Arylated Nucleobases

The functionalisation of DNA aldehydes in the *N*-1 and *N*-9 positions as shown above was chosen as it resembles the functionalisation site of the DNA sugar backbone. This should allow the DNA nucleobases to hydrogen bond and interact in a similar way. Changing the position of the functionalisation may have an impact on the angle at which the DNA base pairs will approach each other when engaged in hydrogen bonding. This may allow for different self-assembled structures to be assembled. Changing the functionalisation site may also have an impact on the hydrogen bonding due to steric or electronic effects of the pendant group. To this end, *C*-5 and *C*-8 functionalised arylated nucleobases were synthesised. The *C*-5 and *C*-8 positions for pyrimidines and purines were chosen due to the availability of previously synthesised *C*-5 and *C*-8 halogenated nucleobases. A study of the literature showed that these are the most commonly substituted sites[129][5]. There are many examples of *C*-5 and *C*-8 functionalisation of nucleobases and many of these rely on metal-catalysed cross-coupling reactions[106]. There are a large number of examples of cross-coupling reactions where a halogen-containing component is coupled to an aryl group including the Suzuki, Sonogashira and Heck reactions[130][131][132][133][134][135][136]. An advantage of these cross-coupling reactions on the *C*-5 and *C*-8 halogenated DNA bases is that they are more selective compared to the Chan-Lam-Evans coupling as there is only one halogen group present on each DNA base compared to multiple nitrogen sites. This enhanced selectivity means that protecting groups are not required. The DNA bases possess, however, notoriously poor solubility in common organic solvents so solubility can be enhanced by utilising an alkyl chain in the *N*-1 or *N*-9 positions. An advantage of not using a protecting group is that it reduces the number of steps in the synthetic route and no further deprotection reactions are required once the porphyrin has been synthesised. Conveniently, the required starting materials have already been synthesised in Section 2.3 and these can be reacted with the relevant aryl derivative.

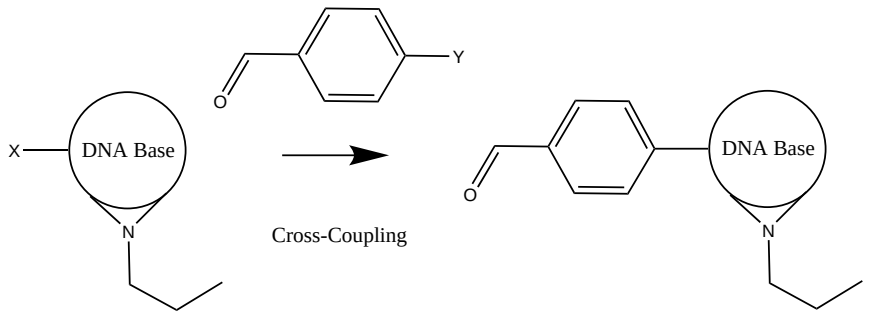


Figure 4.11: Scheme for the synthesis *C*-5 or *C*-8 arylated nucleobases via a relevant cross-coupling reaction of the *C*-5 and *C*-8 halogenated nucleobases synthesised in Section 2.3.

There are numerous examples of cross-coupling reactions of halogenated nucleobases and due to the wide scope of cross-coupling reactions in synthetic chemistry there are often many possible solvents, catalysts, bases and reaction conditions to consider[137]. Therefore, a number of screening reactions were set up to investigate a suitable set of reaction conditions that yielded the desired product in appreciable yield. The aim of these screening reactions was not to thoroughly investigate the best and most efficient route possible with an emphasis on the effect of each and every aspect of the reaction conditions. Only the desired arylated nucleobases in acceptable yield to carry into further reactions were required. Due to its simplicity and the affordability and availability of the starting materials, the Suzuki reactions was chosen first for screening. Here, the readily available 4-formylphenylboronic acid could be used and a number of different solvents, bases and temperatures were screened. The conditions were taken from examples in literature where analogous halogenated nucleobases were reacted with aryl boronic acids[106][137]. However, in almost all cases the starting

materials were different in at least one factor (halogenated nucleoside, different aryl functionality) and this almost certainly had some effect on the results of the screening reactions performed.

A number of different solvents were screened with various reaction conditions and no product was observed. In the majority of cases only the starting halogenated DNA base was observed. This was unexpected as the reaction conditions were taken from comparable examples in literature. It is likely that the reason for the failure of these reactions was due to the lack of solubility of the starting materials. The majority of literature examples include a halogenated DNA nucleoside[137]. Compared to a DNA nucleobase these compounds have a 5-carbon sugar and this sugar group will likely change the solubility of the starting material compared to an alkyl group. After testing a number of solvents a mixture of dimethoxyethane (DME) and water (4:1 v:v) was found to be successful[106]. Unlike previous solvents, this mixture readily dissolved all components of the starting reaction mixture and it is likely that this can be attributed to its success. These successful reaction conditions shown below (Figure 4.12) were successfully utilised for the synthesis of C-5 and C-8 paraformylphenyl nucleobases.

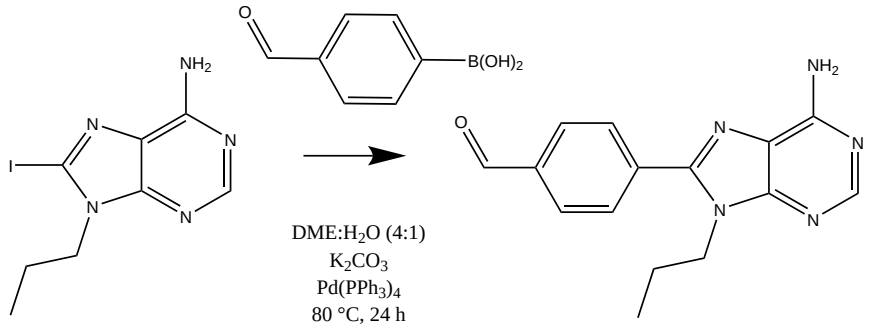


Figure 4.12: The synthesis of C-8 arylated adenine via a Suzuki reaction of 8-iodo-N-9-propyladenine synthesised in section 2.3 with 4-formylphenylboronic acid. These conditions were successfully used for C-5 arylated thymine and cytosine derivatives.

4.2 Synthesis of DNA Functionalised Porphyrins

4.2.1 Adler-Longo Method

A number of different DNA functionalised aldehydes have been synthesised where the paraformylphenyl group is in either the N-1 or N-9 position or in the C-5 or C-8 position for pyrimidines and purines respectively. For the synthesis of a symmetric porphyrin where all the meso substituents are the same the chosen aldehyde is simply reacted with pyrrole to form a porphyrin. This was done previously in the group with N-3-benzoyl-N-1-paraformylphenylthymine to form tetra-N-3-(benzoylthymine)phenylporphyrin (TNbzTPP)[83]. The benzoyl protecting groups were readily removed using ammonium hydroxide as shown below to provide tetra-N-3-thyminephenylporphyrin (TNTPP).

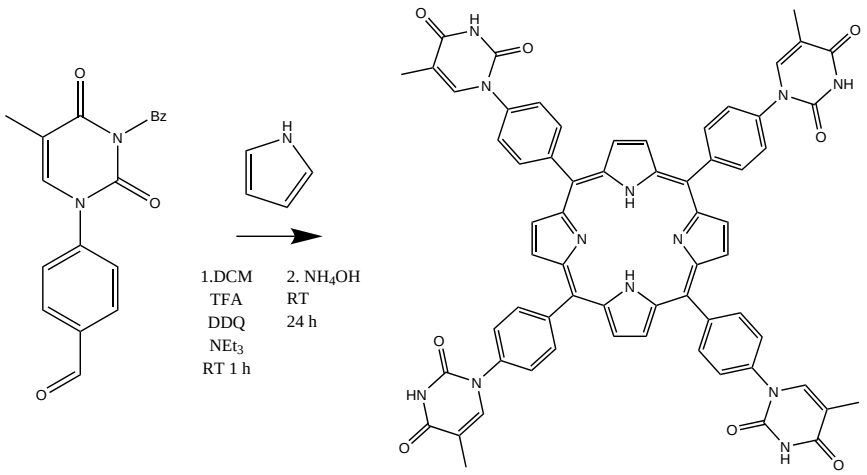


Figure 4.13: The synthesis of TNTPP via the reaction of *N*-3-benzoyl-*N*-1-paraformylphenylthymine with pyrrole and subsequent benzoyl deprotection with ammonium hyrdroxide as previously described by Champness *et al.*[83]

The final TNTPP product was isolated in 8 % yield over two steps as a dark-purple powder. The identity of the product was confirmed by ^{1}H and ^{13}C NMR spectroscopy as well as MALDI mass spectrometry with an exact mass of 1110.36. As expected, the final product was not readily soluble in non-polar organic solvents such as DCM or chloroform. Solubility could be achieved by addition of a couple of drops of TFA to the NMR sample in $CDCl_3$. This is due to the twist in the protonated porphyrin structure afforded by the addition of acid; a change from dark purple to green is also observed. The solubility of TNTPP was not sufficient for NMR titration studies due to the high concentrations required for these experiments.

This same synthetic route was also attempted using the *N*-9-adenine aldehyde. *N*-9-formylphenyl-*N*-6-bis-Boc-adenine was reacted with pyrrole under the conditions shown below to form tetra-*N*-9-(bis-Boc-adenine)phenylporphyin (TNBocAPP).

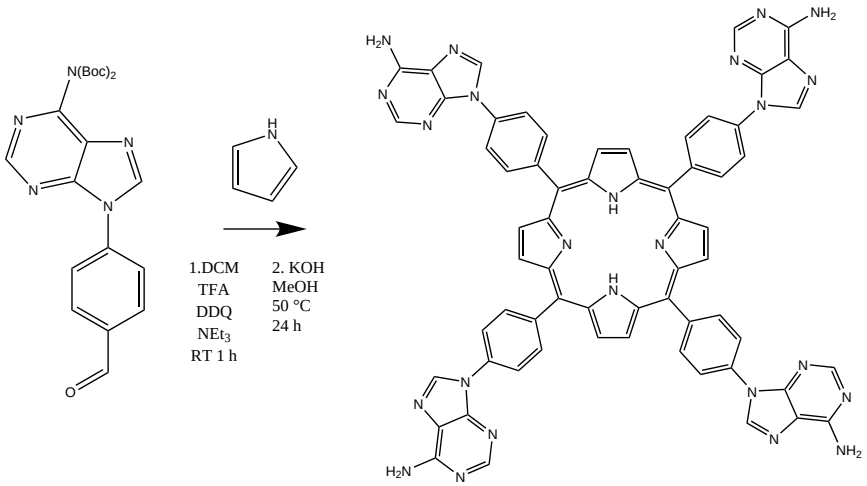


Figure 4.14: The synthesis of TNAPP via the reaction of *N*-9-formylphenyl-*N*-6-bis-Boc-adenine with pyrrole and subsequent Boc deprotection with potassium hyrdroxide.

Isolation of this product was much more challenging than for the thymine functionalised porphyrin. The reaction yielded a mix of adenine-containing porphyrins with a loss of any number of 8 possible Boc-protecting groups. The presence of these products was confirmed by MALDI mass spectrometry. The loss of each successive Boc group resulted in a reduction by 100 mass units as shown below (Figure 4.16. This mixture of products made purification and isolation of a single product difficult. This is due to the considerable effect that the Boc groups have on the solubility and Rf of the compound. Where a large number of Boc groups are present the compound is readily soluble in chloroform and is eluted by column chromatography at an appreciable rate using DCM:EtOAc (90:10) as eluent. However, the solubility and Rf of the compound drops rapidly where fewer Boc groups are present. For the mono-Boc and Boc-deprotected product the compound is sparingly soluble in DMSO and requires the addition of methanol and triethylamine to the chromatography eluent. Despite all of these Boc-protected adenine porphyrin compounds being useful it was challenging to isolate an acceptable amount of each compound due to the vastly differing properties of each compound. The loss of the Boc protecting groups is likely due to the TFA used in the porphyrin synthesis. TFA is a known reagent used for the removal of Boc groups.

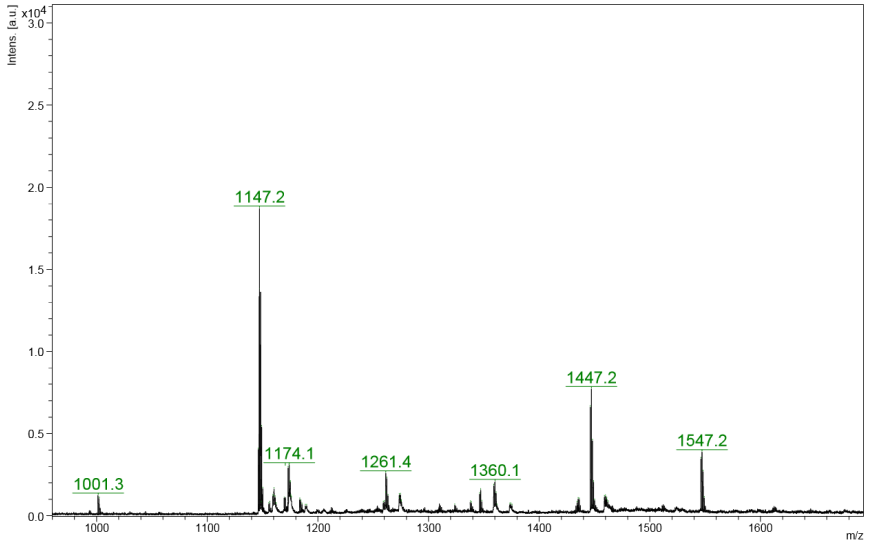


Figure 4.15: MALDI mass spectrum of crude TNAPP showing partially deprotected products. m/z 1447 = 3xBoc, m/z 1547 = 4xBoc .

To overcome the issue of having a mixture of compounds where the number of Boc groups n=0-8 it was suggested that the reaction mixture be combined and fully deprotected to afford a single product. The Boc deprotection was achieved by treatment of the mixture with potassium hydroxide in methanol at 50 ° C. These conditions did afford the desired deprotected product. However, the reaction produced an appreciable amount of incomplete deprotection. This is likely due to the low solubility of the deprotected product. The final tetra-*N*-9-adeninephenylporphyrin (TNAPP) was characterised by NMR spectrosopy and MALDI mass spectrometry with a mass of 1146.

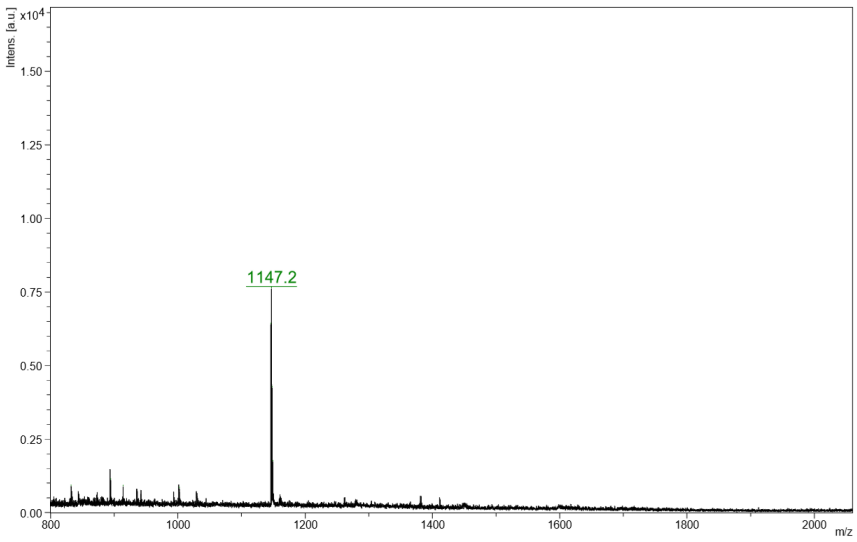


Figure 4.16: MALDI mass spectrum of deprotected and purified TNAPP. m/z Calc. 1147.41 (M+H); Exp. 1147.27 .

The synthesis of tetra-*N*-1-cytosinephenylporphyrin (TNCPP) was also attempted using *N*-1-formylphenyl-*N*-4-bis-Boc-cytosine. The Boc-protected tetra cytosine porphyrin was synthesised using the same conditions as used previously for porphyrin synthesis. The fully protected product was isolated and the troublesome issue of undesired Boc-group removal was not observed. This, coupled with the high solubility of the fully-Boc-protected product, allowed for a pure product to be isolated and an NMR spectrum of high quality was collected as shown below (Figure 4.17).

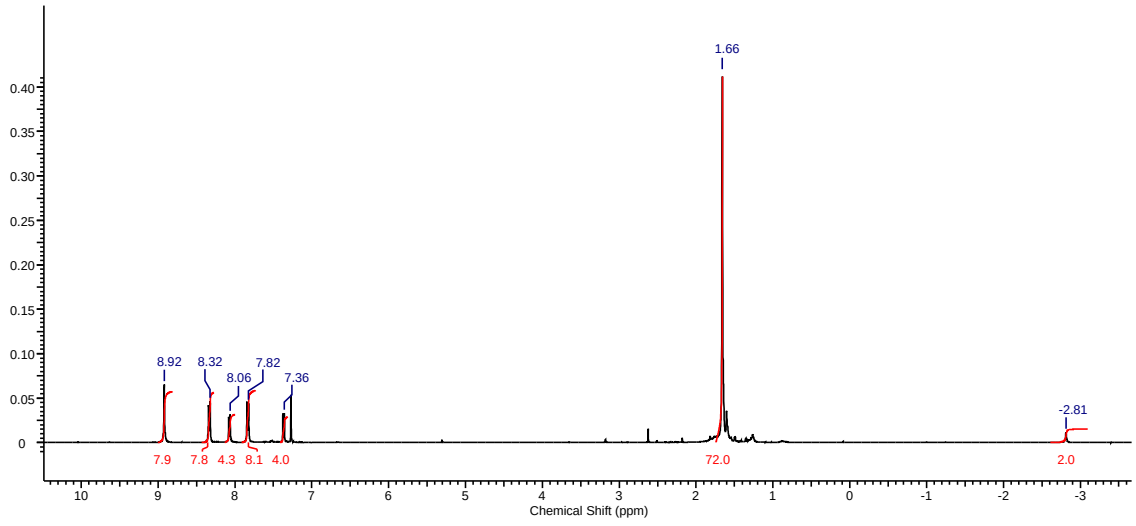


Figure 4.17: Proton NMR spectrum of TBocNCPP recorded in $CDCl_3$.

The synthesis of porphyrins where the DNA nucleobase is functionalized in the *C*-5 or *C*-8 position was attempted using the *C*-5 and *C*-8 functionalized aldehydes synthesised earlier. Unfortunately, all attempts using these aldehydes were unsuccessful. The *C*-5-formylphenyl-*N*-1-propylcytosine was reacted with pyrrole under the same conditions which were used to synthesise

the tetra-DNA-nucleobase substituted porphyrins previously. The reaction started as a colourless solution which darkened upon addition of TFA as seen previously. This darkening indicated that the reaction had proceeded successfully. However, upon quenching the reaction the desired product was not observed in the MALDI mass spectrum. This same negative result was seen when *C*-8-formylphenyl-*N*-9-propyladenine was used. It is suggested that these negative results are due to insufficient solubility of the starting material preventing the reaction from completely forming the necessary cyclic porphyrin structure.

4.2.2 Lindsey Method

The DNA functionalised porphyrins synthesised in the previous section are all symmetric porphyrins where the same functionalisation is present in each of the four meso positions. To compare the hydrogen bonding properties of the DNA functionalised porphyrins to the previously studied alkyl nucleobases it is necessary to study the compounds in NMR titration experiments. For these experiments the compounds must be readily soluble in a suitable solvent. All of the tetra-DNA base functionalised porphyrin do not display sufficient solubility to be analysed by NMR titration. For a model that is easily comparable to the previous studies of alkylated DNA bases it is necessary to synthesise a porphyrin bearing a single DNA base. This will allow the same 1:1 binding model to be studied as before, whereas, a tetra-DNA base functionalised porphyrin could display a 4:1 binding model. For a mono-DNA base functionalised porphyrin it is also possible to increase the solubility of the compound by adding solubilising groups in the other three meso positions. This approach was taken previously in the group for the thymine functionalised porphyrin: the tetra thymine porphyrin was not sufficiently soluble for NMR titration studies so a mono thymine porphyrin was synthesised. The solubility of this compound was increased by the addition of *t*-butyl groups.

The synthesis of a mono-DNA base porphyrin requires the aforementioned type of porphyrin as there are two different components in the meso position. For the synthesis of mono-thymine porphyrin Champness *et al.* report the synthesis of a porphyrin in a method adapted from that pioneered by Lindsey[83][114]. Porphyrin synthesis by simple mixing of different aldehydes would produce a statistical mixture of compounds as explained earlier. Instead, the approach used uses a multi step synthesis where each 'side' of the porphyrin is sequentially and selectively assembled and added.

First, the synthesis of 3,5-di-tert-butylbenzyldipyrromethane was performed. This was achieved by first synthesising 3,5-di-tert-butylbenzaldehyde from 1-bromo-3,5-di-tert-butylbenzene which was then reacted with excess pyrrole as shown below (Figure 4.18).

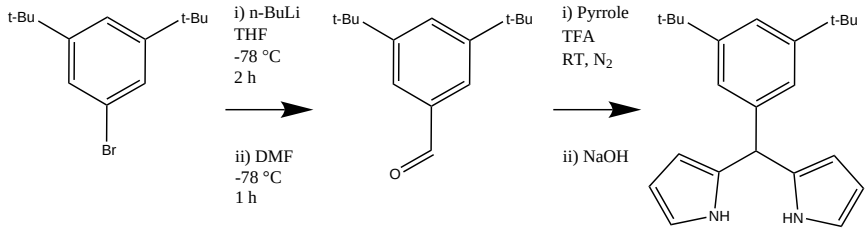


Figure 4.18: Reaction scheme for the synthesis of 3,5-di-tert-butylbenzyldipyrromethane via 3,5-di-tert-butylbenzaldehyde.

The bulky aldehyde 3,5-di-tert-butylbenzaldehyde was synthesised by first reacting 1-bromo-3,5-di-tert-butylbenzene with n-butyllithium[138]. The starting material undergoes a metal-halogen exchange and forms the highly reactive lithiated intermediate. To this intermediate DMF is added. The highly polar character of the lithiated species makes it an excellent nucleophile and it readily reacts with the electrophilic DMF to afford the desired aldehyde in excellent yield. Formation of

the lithiated species is required as aryl halides are known to be poor nucleophiles with respect to nucleophilic substitution[109].

This aldehyde was then carried on to the formation of 3,5-di-tert-butylbenzyldipyrromethane[139]. The formation of a dipyrromethane is similar to the formation of a porphyrin where an aldehyde undergoes a condensation reaction with pyrrole under acidic conditions. The essential condition of this reaction that ensures a dipyrromethane is formed is the stoichiometry of the aldehyde and pyrrole components[140]. For a porphyrin the aldehyde and pyrrole are reacted in a 1:1 ratio and this allows the ring to form. For a dipyrromethane the ring formation is prevented by using an excess of pyrrole (in this case pyrrole was used as the solvent). This ensures that each aldehyde molecule reacts with two pyrrole species and then cannot react further as all the aldehyde has been consumed. An important distinction is also the lack of oxidation step present in the synthesis of a dipyrromethane. The addition of an oxidising reagent such as p-chloranil or DDQ would form an unsaturated aromatic dipyrromethene. Due to their aromatic character dipyrromethenes are relatively stable and not suitable as a building block for the synthesis of porphyrins. Therefore, the dipyrromethane reaction is quenched by addition of base. Due to the unstable nature of the dipyrromethane product the synthesis, work up, purification and storage of this compound must be carried out in the dark to avoid decomposition into undesired polymeric species. The desired product was synthesised in 81% yield and characterised by NMR spectroscopy and mass spectrometry.

This dipyrromethane compound represents one 'side' of the porphyrin. It was then necessary to synthesise the other 'sides' which can be selectively added onto this building block. This required the synthesis of a protected 3,5-di-tert-butylbenzyl carbonyl species via the reaction schemes (Figure 4.19) shown below. A protected carbonyl species is required for the selective and controlled addition to the dipyrromethane compound.

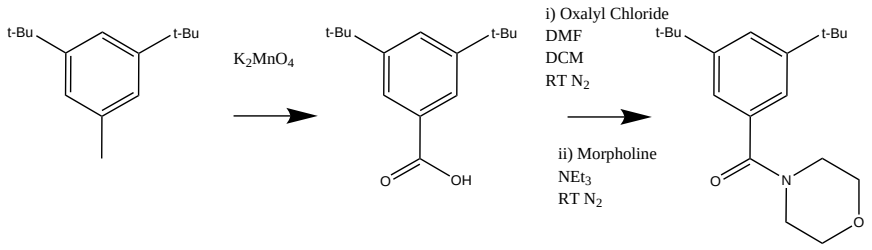


Figure 4.19: Reaction scheme for the synthesis of morpholine-3,5-di-tert-butylbenzylamide via 3,5-di-tert-butylbenzoic acid.

For the synthesis of 3,5-di-tert-butylbenzoic acid 3,5-di-tert-butyltoluene was oxidised by treatment with potassium permanganate[141]. This carboxylic acid was then converted to the morpholine-amide in two steps. First, the carboxylic acid was reacted with oxalyl chloride to form an acid chloride. This highly reactive acid chloride was then rapidly reacted with morpholine in the presence of catalytic triethylamine to form the desired amide. This chosen route for the amide synthesis was chosen due its relatively high yield and the availability of starting materials and reagents. There are numerous other possible routes for amide coupling due to its relevance in peptide synthesis. These often include coupling reagents such as DCC, HOBt and HATU, amongst others.[142],[143]

The amide-protected 3,5-di-tert-butylcarbonyl species could then be reacted with the previously synthesised dipyrromethane to synthesise three 'sides' of the desired porphyrin in the reaction shown below (Figure 4.20).

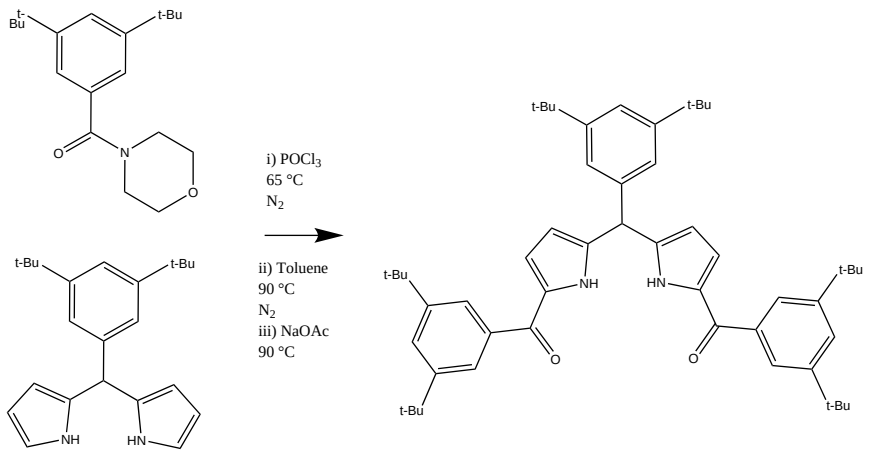


Figure 4.20: Reaction scheme for the synthesis of morpholine-3,5-di-tert-butylbenzylamide via 3,5-di-tert-butylbenzoic acid.

The reaction above is an acylation of an electron rich arene[109][144][145]. Under the conditions above the reaction is known as the Vilsmeier Haack reaction. For this reaction the first step is the formation of the formylating agent (or Vilsmeier Reagent). This is formed from the morpholine-amide and phosphorus oxychloride as shown in the mechanism below (Figure 4.21).

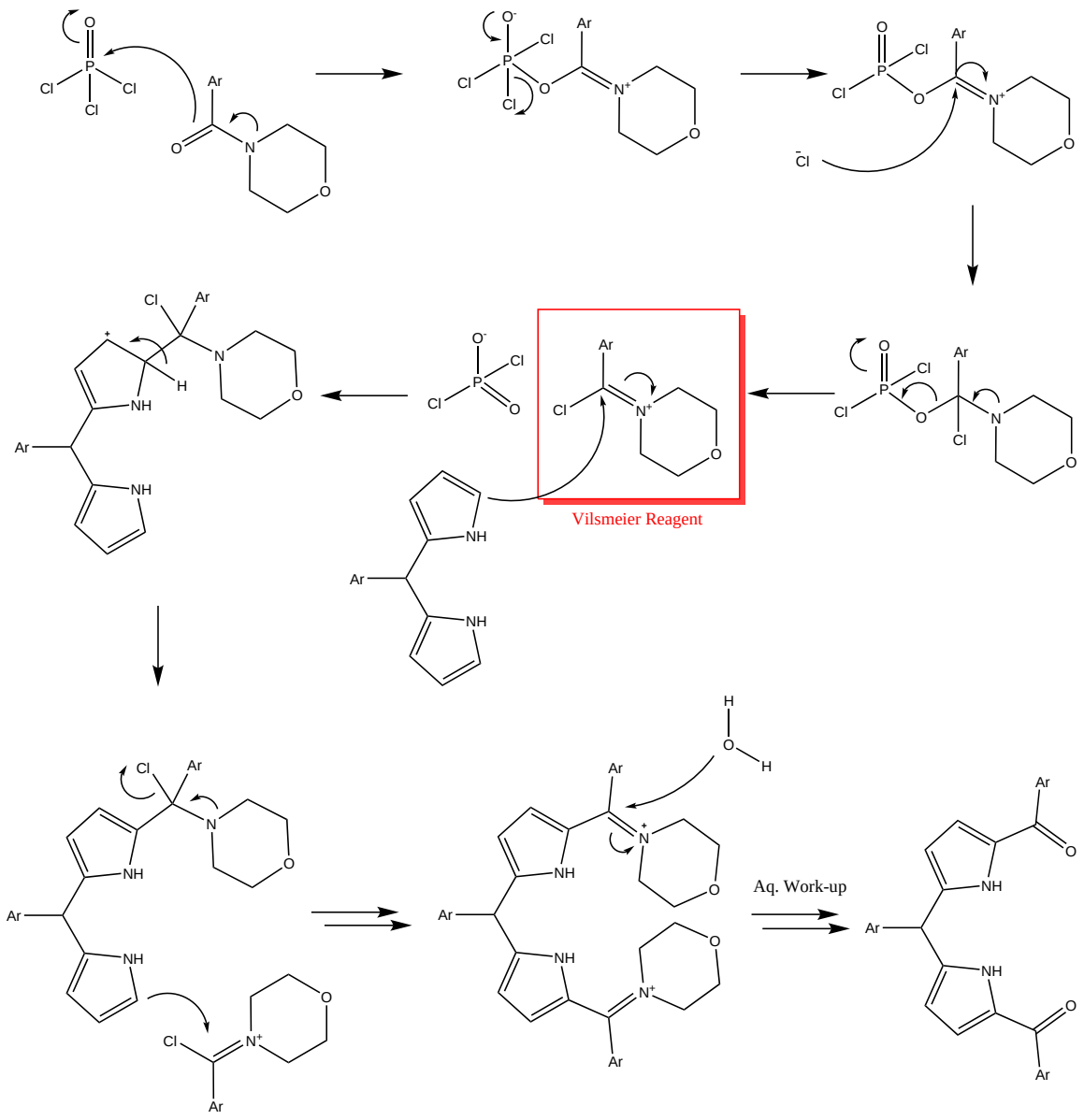


Figure 4.21: Mechanism for the Vilsmeier Haack reaction.

The Vilsmeier Reagent is then attacked by the electron rich dipyrromethane in an electrophilic substitution reaction. The electrophilic substitution occurs selectively at the α position of the pyrrole due to the stability of the intermediate. When electrophilic substitution occurs at the α position there are three resonance forms that stabilise the positive charge. For substitution at the β position there are only two resonance forms as shown below (Figure 4.22) .

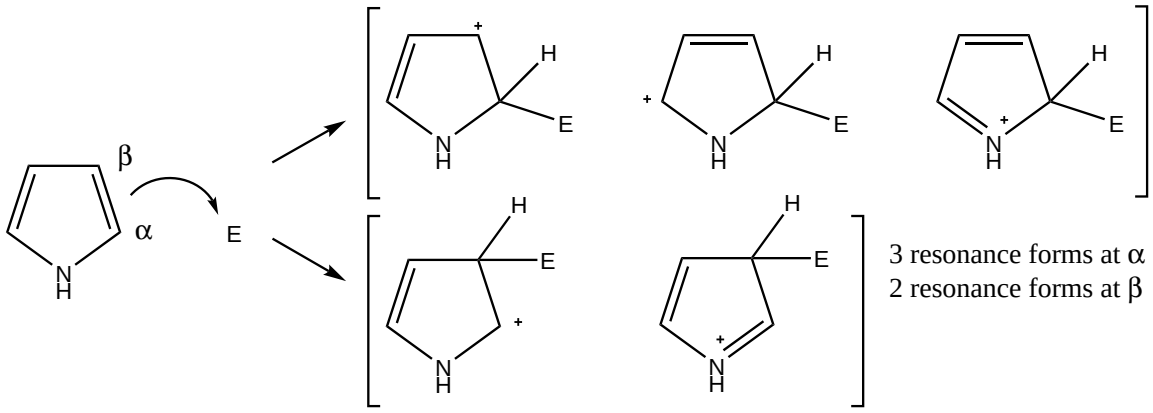


Figure 4.22: The resonance forms of the intermediate formed during electrophilic substitution of pyrrole showing the increased stability of the intermediate upon substitution at the α position.

After electrophilic substitution the aromaticity of the arene is restored by removal of the proton by another molecule of amide, and another iminium ion is generated (not shown for clarity). Chloride is then lost as a suitable leaving group. The iminium ion formed in this step is then converted to a carbonyl by an aqueous work up. Importantly, this final iminium ion cannot undergo further electrophilic substitution due to the lack of suitable leaving group. This ensures the reaction stops after a single substitution of each amide and the desired product is formed.

The desired product was isolated as a dark purple powder in 86% yield after purification by column chromatography and the identity was confirmed by NMR spectroscopy.

For the fourth and final 'side' of the porphyrin it was necessary to synthesise a DNA base functionalised dipyrromethane. This dipyrromethane would react with the other three 'sides' and form the desired porphyrin. A thymine functionalised dipyrromethane was previously synthesised in the group and the same approach was used with an adenine aldehyde as shown below (Figure 4.23).

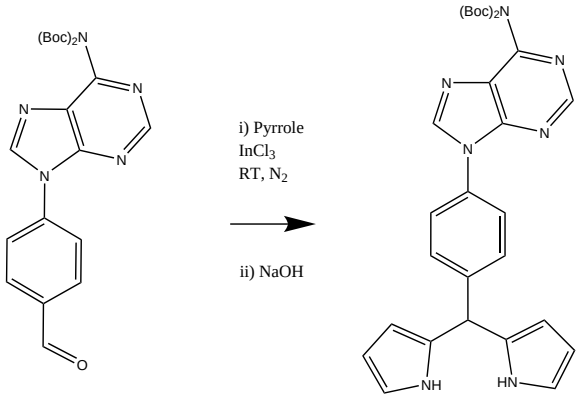


Figure 4.23: The synthesis of bis-Boc-adenine-N-9-phenyl-(meso)dipyrromethane.

The synthesis of the adenine functionalised dipyrromethane was similar to that of 3,5-di-tert-butylphenyl-dipyrromethane where the chosen aldehyde was reacted with excess pyrrole. For the adenine dipyrromethane synthesis the acid was changed to the milder indium chloride. It was found that, whilst this makes the reaction proceed at a slower rate, fewer side products were observed. This made isolation of the desired dipyrromethane easier. The work up of the reaction was also modified due the lower solubility of the product in hexane compared to the 3,5-di-tert-butyl analogue. Once

the reaction was complete the reaction was quenched with the addition of aqueous sodium hydroxide (1 M) and the product extracted with DCM. After evaporation of the solvent the crude product was purified by column chromatography to afford the pale-yellow product in 26 % yield. This protected dipyrromethane was then combined with the acylated tert-butyldipyrromethane in the next step to synthesise the desired porphyrin.

The synthesis of the final mono-adenine-tri(di-tert-butyl)phenylporphyrin was achieved by combining the two previously synthesised dipyrromethanes in the method outlined by Lindsey as shown below (Figure 4.24).

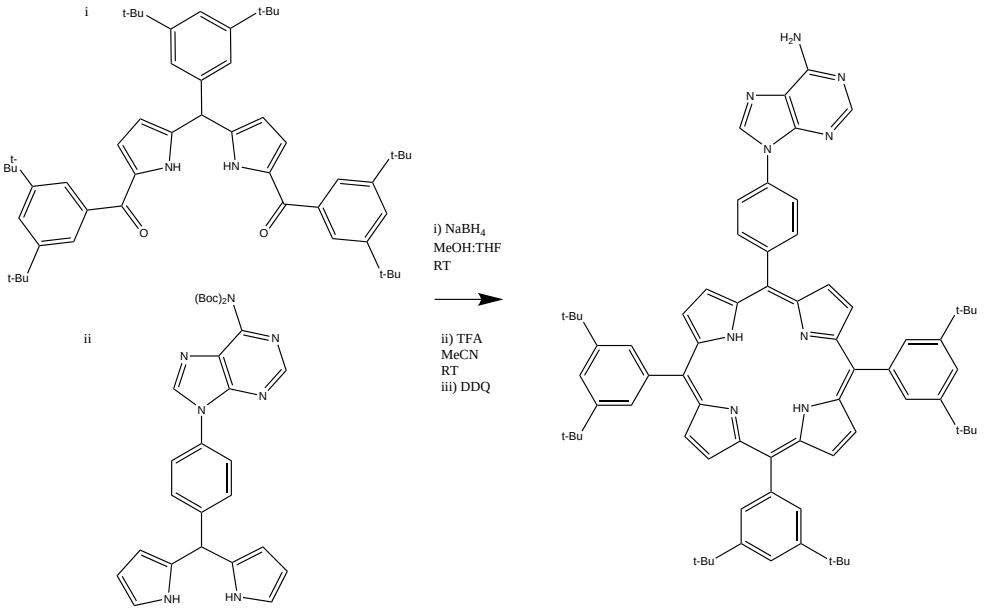


Figure 4.24: The synthesis of mono-adenine-tri(di-tert-butyl)phenylporphyrin.

First, the acylated dipyrromethane was reduced to the corresponding carbinol by treatment with sodium borohydride ($NaBH_4$) in mixture of THF and methanol (1:1 v:v). Sodium borohydride is a well known reagent used for the reduction of ketones to alcohols and is commonly used in preference over the more powerful lithium aluminium hydride ($LiAlH_4$)[109]. Both reagents will selectively afford the desired ketone product and $NaBH_4$ does not violently react with water meaning that water-free conditions are not required. Upon the addition of $NaBH_4$ significant effervescence was observed. Therefore, the reagent was added slowly over the course of 20 minutes until the dark purple starting material was no longer present and a pale orange solution was observed. The resulting carbinol species was isolated by careful evaporation of the solvent and dried under vacuum. The dicarbonyl species is now more susceptible to electrophilic attack upon addition of the acid catalyst and the adenine functionalised dipyrromethane. The dicarbinol species is protonated by the acid and water is released resulting in the highly electrophilic carbocation as shown in Figure 4.25. This carbocation then rapidly reacts with the nucleophilic dipyrromethane and, after oxidation with DDQ, the desired mono-adenine-tri-(di-tert-butyl)phenylporphyrin (MATtBPP) is formed.

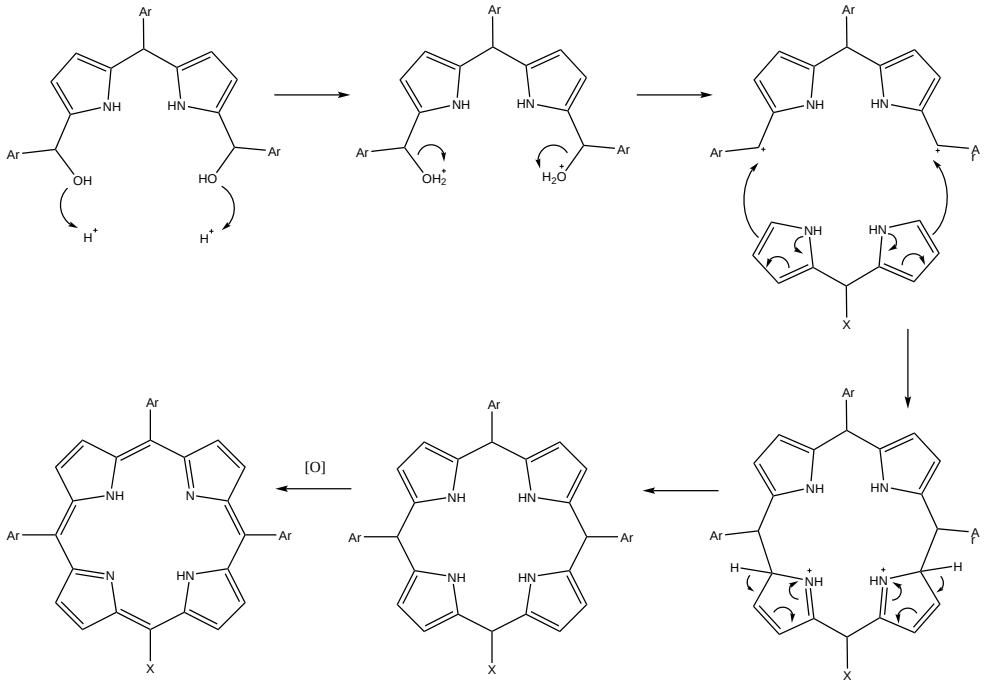


Figure 4.25: The mechanism for the acid catalysed condensation reaction between the dicarbinol species and the nucleophilic dipyrromethane. (It is worth noting that both aromatic substitution reactions do not in fact occur simultaneously.)

The product was isolated by column chromatography and the pure product was characterised by NMR spectroscopy, MALDI mass spectrometry and X-ray crystallography.

The multi-step synthesis detailed above did successfully afford the desired porphyrin. Synthesis of this compound using the mixed aldehyde method would afford a large quantity of side products and this would increase the purification steps and limit the yield of the desired product. Conversely, the multi-step synthesis afforded only the desired product and this avoided excessive purification. Despite this synthesis method requiring a large number of steps (7 in total) and some complicated experimental procedures most steps proceeded in excellent yield with minimal purification. The 3,5-di-tert-butylphenyl aldehyde and carboxylic acid were both synthesised in quantitative yield with trivial purification. These were both transformed into the relevant dipyrromethane and amide in 81 % and 85 % yield respectively. The diacylated dipyrromethane was then formed from these components in 86 % yield resulting in a net yield of 59 % over 5 steps. This yield demonstrates how even minor inefficiencies in yield can lead to a considerable loss of product once combined over the course of a multi-step synthesis. However, the acylated dipyrromethane is synthesised from largely inexpensive materials (with the exception of pyrrole which is essential in any porphyrin synthesis) and a total yield of 59 % is highly acceptable over the course of 5 steps. It is also worth noting that, once synthesised, the acylated dipyrromethane can be used as a building block for the synthesis of a number of porphyrins where the meso functionality is different. A different dipyrromethane with the desired functionality can be reacted with the acylated species.

The yield for the synthesis of the adenine functionalised dipyrromethane was relatively low at 26 %. This is due to the lack of stability of dipyrromethane compounds in general, in addition to the poor solubility of the compound which made purification difficult. The final porphyrin-forming step was also low yielding, but not especially so when compared to many other examples of porphyrin synthesis[146]. Once the final step was completed the overall combined yield was roughly 1 % (59 x 22 x 12) over all 7 steps. This yield would be improved greatly if the synthesis of the functionalised

dipyrromethane could be bettered. However, the synthesis of porphyrins is well known for its low yields and the yield of a multi-step synthesis will always be low where the final step has a yield not exceeding 30 %. It is worth remembering that high yields are not the focus of this work where the synthesis of complex, novel compounds is the main aim.

4.3 Post-Synthesis Functionalisation

It is also possible to synthesise a functionalised porphyrin by introducing the desired functionality after the porphyrin has been synthesised. Benefits of this method include the synthesis of a porphyrin from readily available starting materials and the opportunity to exploit the selectivity of cross-coupling reactions to efficiently synthesise the desired porphyrin. For this it is necessary to first synthesise a porphyrin with the relevant reactive sites to which the functionality can be added. It has been shown earlier in this work and in the literature that cross-coupling reactions can be used to successfully functionalise aromatic compounds with DNA bases. In these reactions halogens, boronic acids and alkyne functionalities have been used as the reactive site for cross coupling. Therefore, the tetraphenylporphyrins bearing these functionalities shown in Figure 4.26 were synthesised.

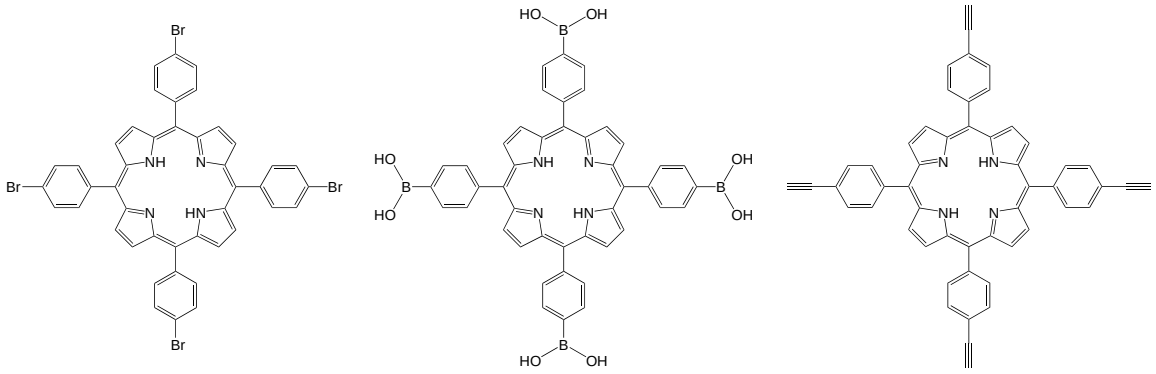


Figure 4.26: The porphyrins used for post-synthetic functionalisation.

4.3.1 Porphyrin Synthesis

Tetra-(4-bromophenyl)porphyrin was synthesised from the condensation reaction of commercially available 4-bromobenzaldehyde with pyrrole in propionic acid as shown below (Figure 4.27). Propionic acid is a relatively mild acid with a high boiling point and is used as the solvent for the reaction. The use of a relatively mild acid is required so as to effectively protonate the aldehyde without leading to excessive polymerisation of the pyrrole starting material. The product is conveniently insoluble in cold propionic acid so upon completion the reaction can be cooled and filtered which, after washing with methanol, yielded the product as bright purple crystals. It was possible to synthesise this porphyrin in relatively large quantities despite the modest yield due to the commercial availability and affordability of the starting material. Using the method where refluxing propionic acid is the solvent also allows much greater concentrations of starting materials to be used which allowed large quantities to be synthesised without using excessively large quantities of solvent, column chromatography or unwieldy large reaction vessels. The method where DCM is used as the solvent with an acid catalyst although typically slightly higher yielding, performed at room temperature and suitable for a higher range of substrates require low concentrations (500 mM) and this was not deemed practical for larger scale synthesis. The bromine group could then be used as a suitable reactive substrate for cross-coupling reactions.

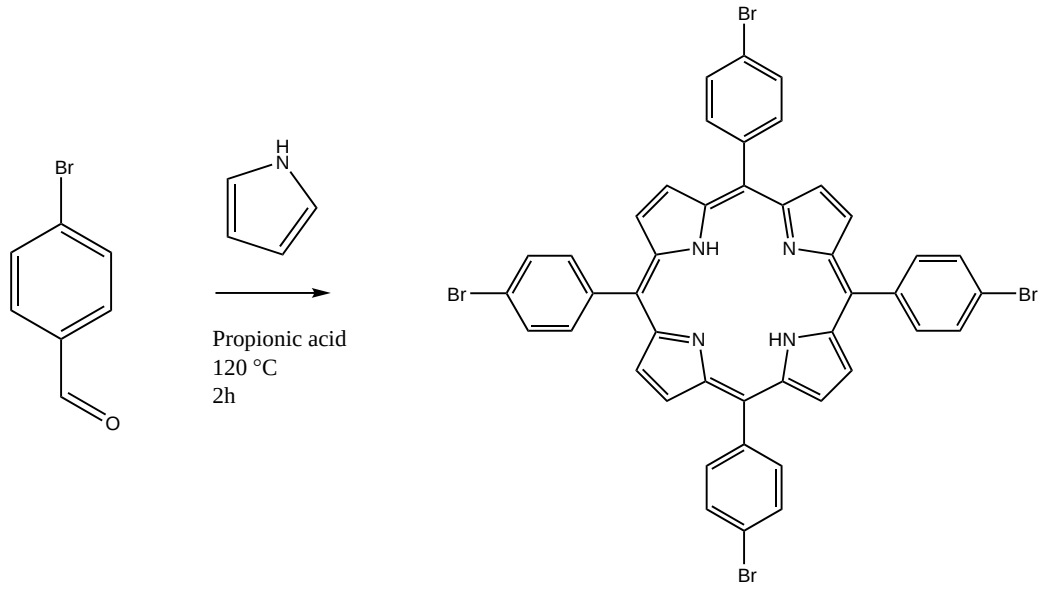


Figure 4.27: The synthesis of tetra-(4-bromophenyl)porphyrin by reacting 4-bromobenzaldehyde with pyrrole in refluxing propionic acid.

The synthesis of tetra-(4-boronic acidphenyl)porphyrin was first attempted using the same conditions as above but with the commercially available 4-formylphenylboronic acid. This method was not successful. This was not surprising and the failure of this reaction was likely due to the low solubility and stability of 4-formylphenylboronic acid. To overcome this the pinacol ester derivative was used in a method outlined by Stoddart *et al.*[147]. Boronic esters are a common alternative to boronic acids due to their increased solubility in non-polar solvents and chiefly for their increased stability. The boronic ester aldehyde derivative was successfully reacted with pyrrole to afford a boronic ester functionalised porphyrin which could then be deprotected to afford tetra-(4-boronic acid phenyl)porphyrin.

Finally, a tetra-(4-ethynylphenyl)porphyrin was synthesised from 4-ethynylbenzaldehyde. This ethynyl group functionalised porphyrin could be used for Sonogashira cross-coupling reactions. 4-ethynylbenzaldehyde was first synthesised from 4-bromobenzaldehyde in a Sonogashira cross-coupling reaction. In a Sonogashira cross-coupling a terminal alkyne is coupled to an aryl or vinyl halide with a palladium catalyst, a copper co-catalyst and an amine base. Here, 4-bromobenzaldehyde, the aryl halide was reacted with TMS-acetylene, the terminal alkyne, in the mechanism shown below (Figure 4.29)[134].

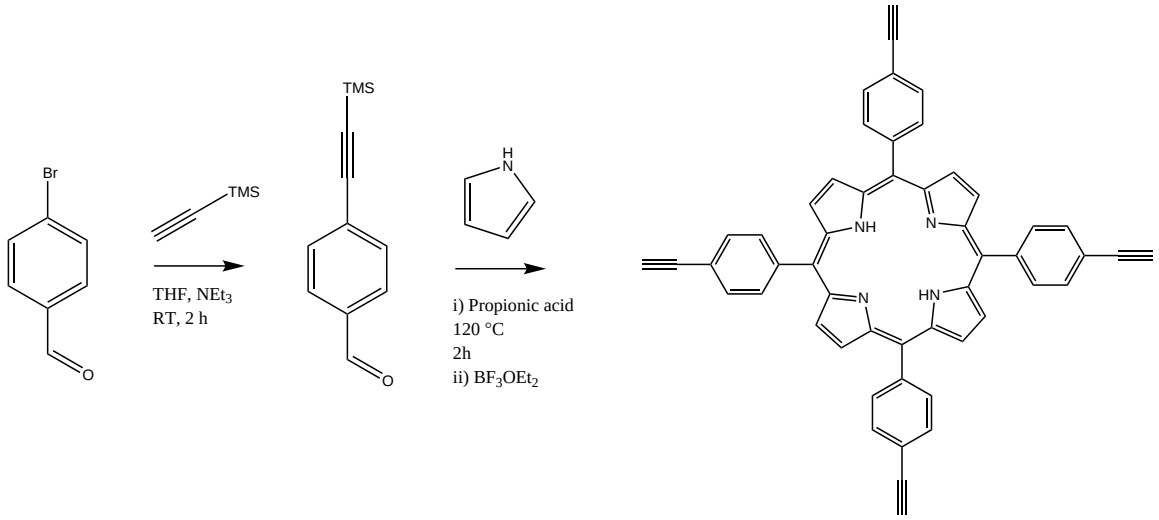


Figure 4.28: The mechanism of the Sonogashira cross-coupling reaction.

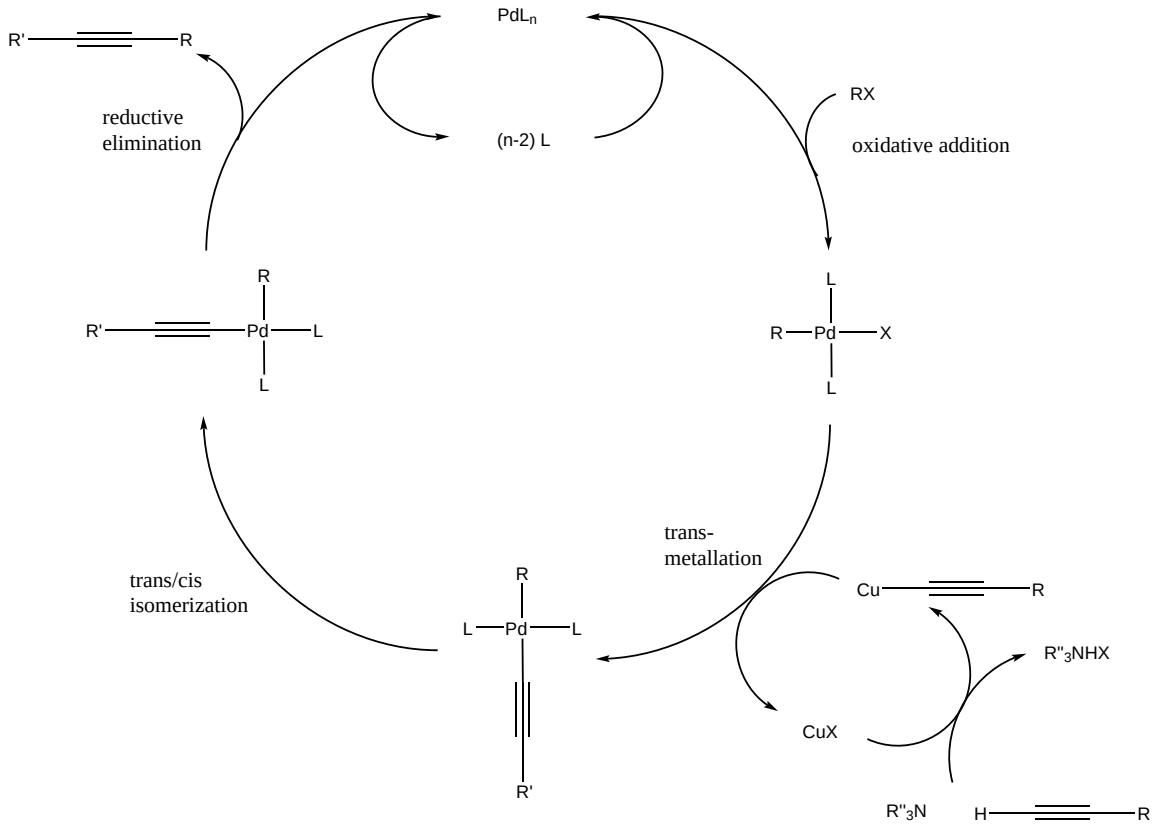


Figure 4.29: The mechanism of the Sonogashira cross-coupling reaction.

First, two successive ligands are dissociated from the palladium catalyst. In all Sonogashira reactions either $PdPh_4$ or $PdCl_2Ph_2$ was used ($PdCl_2Ph_2$ is first reduced from Pd^{II} to Pd^0 in situ).

Oxidative addition of the aryl halide then coordinated the aryl group to the palladium catalyst. The terminal alkyne, which is first coordinated to the copper cocatalyst with the aid of amine base is then trans-metallated to the palladium catalyst. In this process the alkyne group exchanges with the halogen, in turn regenerating the copper halide cocatalyst. Isomerization of the alkyne-palladium species then forms the cis isomer. Reductive elimination of this complex yields the cross-coupled alkyne product whilst regenerating the palladium catalyst.

In the reaction above, TMS-acetylene was used as the terminal alkyne as TMS is a common protecting group for alkyne groups. A protecting group is required to prevent Sonogashira coupling from occurring on both sides of the alkyne. This would form the unwanted 4,4-ethynyl dibenzaldehyde species. The TMS protecting group can also be readily removed when desired. The 4-ethynylbenzaldehyde could be isolated by simple filtration of the reaction mixture and washing with hexane to afford the crude product in excellent yield. Any 4-bromobenzaldehyde starting material could be removed with the use of a cold-finger. This separation technique exploits the different sublimation temperatures of two compounds. The impure compound is placed in a vessel and heated under vacuum. The low pressure greatly reduces the boiling point of 4-bromobenzaldehyde. (Sublimation was observed at 60 °C under vacuum compared to the literature reported boiling point of 123 °C at atmospheric pressure[148].) The sublimed impurity is then collected on a water-cooled insert placed before the vacuum exit point. This cooled insert, known as the cold finger, allows the sublimed product to solidify. After purification of 4-ethynylbenzaldehyde the aldehyde was carried forward into the synthesis of tetra-(TMS-ethynylphenyl)porphyrin using the same propionic acid conditions as before. The TMS-protected porphyrin could then be deprotected using boron trifluoride diethyletherate to afford the desired tetra-(ethynylphenyl)porphyrin.

All of the above porphyrins bearing cross-coupling reactive sites could be synthesised in appreciable amounts without laborious reaction conditions or purification steps and were synthesised from readily-available starting materials. This meant that an appreciable amount of porphyrin could then be carried forward into the subsequent functionalisation steps. The post-porphyrin-synthesis DNA nucleobase functionalisation of the above porphyrins was attempted with the previously synthesised C-5 and C-8 halogenated nucleobases. These compounds all bear the necessary halogen group commonly found as the other reactive component in cross-coupling reactions. Furthermore, it was envisaged that the N-1 and N-9 alkylation of these halogenated nucleobases would increase the solubility of the starting materials and the resulting functionalised porphyrin without the need for protecting groups and without impacting excessively on the hydrogen bonding.

4.3.2 Functionalisation

Attempts were made to couple halogenated DNA nucleobases to the functionalised porphyrins. The halogenated nucleobases synthesised earlier (Fig 4.30) were used in Sonogashira reactions with tetra-ethynylphenylporphyrin under the conditions shown below (Fig 4.31)[135][149].

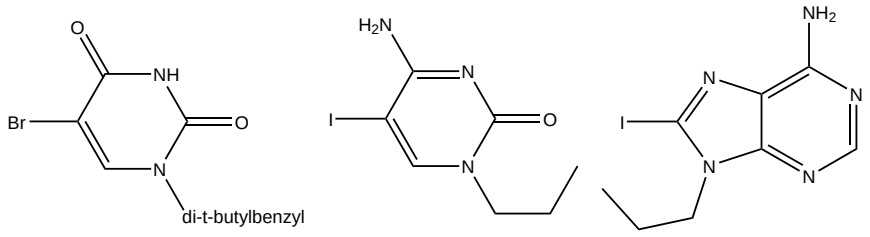


Figure 4.30: Halogenated nucleobase analogues used for Sonogashira cross-coupling reactions.

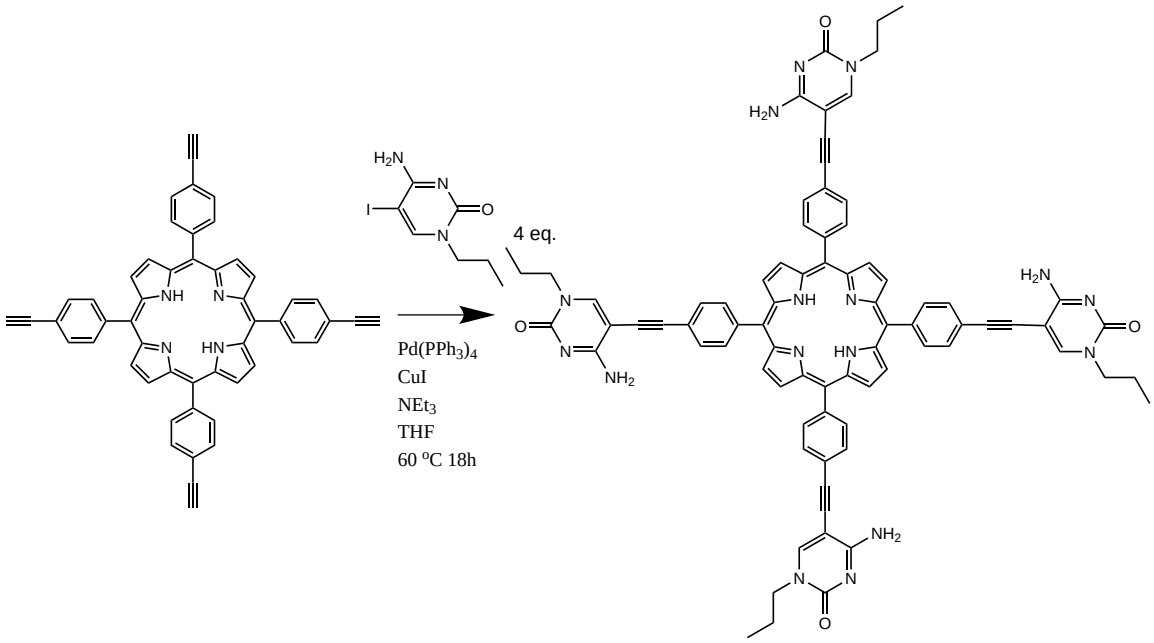


Figure 4.31: Reaction conditions for the Sonogashira reaction of 5-iodo-*N*-1-propylcytosine with tetra-ethynylphenylporphyrin. This reaction was also attempted for the DNA base analogues of thymine and adenine above (Fig 4.30).

These reactions were unsuccessful in all cases. No evidence of a cross-coupling reaction could be observed. However, the starting material was consumed. Analysis of the reaction mixture by MALDI MS showed a new peak at 773. This peak corresponds to the mass of starting tetra-ethynylphenylporphyrin plus copper. Porphyrins are widely used in host-guest chemistry for their ability to associate to metals. Therefore, this suggests that the Sonogashira reaction is not proceeding due to the copper catalyst being associated to the porphyrin starting material.

To overcome this issue copper-free Sonogashira coupling conditions were found, of which there are a large number in literature[150][151]. If no copper was present in the reaction it could not associate to the starting material. A $Pd_2(dba)_3$ catalyst was used in toluene with triethylamine used as the base as shown in the scheme below (Fig 4.32) and attempted for each halogenated DNA nucleobase.

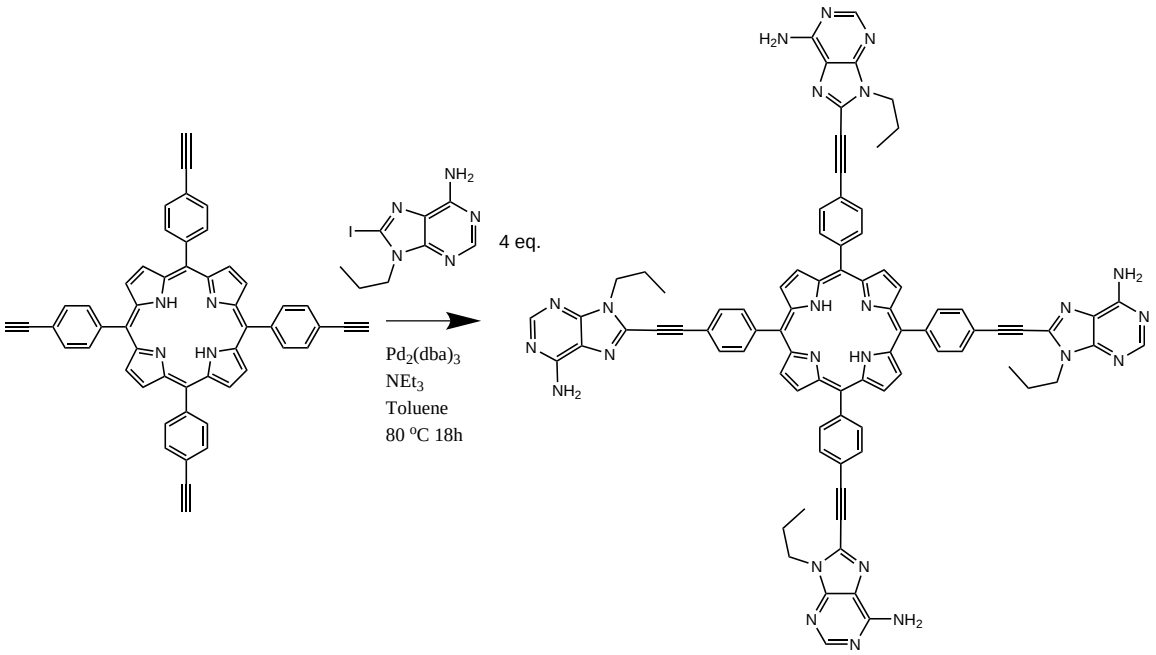


Figure 4.32: Scheme for the copper-free Sonogashira reaction of 8-iodo-*N*-9-propyladenine with tetra-ethynylphenylporphyrin. This reaction was also attempted for the DNA base analogues of thymine and cytosine above (Fig 4.30).

These reactions were unsuccessful for each DNA nucleobase. It was noted that the tetra-ethynylphenylporphyrin starting material was insoluble in toluene even upon heating the reaction to 80 °C. This insolubility of the starting material may explain why the reactions were unsuccessful. Therefore, the reactions were repeated in THF; a solvent which was found to dissolve all the reagents. These reactions were also unsuccessful.

4.4 Conclusions

In this chapter a number of DNA nucleobase functionalised porphyrins were synthesised and a number of different methods to synthesise DNA nucleobase functionalized porphyrins with different regiochemistry and symmetry were investigated. The method that yielded the greatest success was the Adler-Longo method were protected *N*-1 and *N*-9 paraformylphenyl pyrimidine and purine nucleobases were reacted with pyrrole. This method was successful for the synthesis of symmetric porphyrins bearing the thymine, adenine and cytosine nucleobases. However, the resulting deprotected porphyrins displayed very poor solubility which made full deprotection and purification challenging. The method of reacting the chosen functionalised aldehyde with pyrrole, whilst being efficient for symmetric porphyrins bearing the same functionality in each of the four meso positions, is not feasible for the synthesis of porphyrins where different functionalities are desired. A step-wise synthesis of porphyrins is described which has now been proven successful for both thymine and adenine porphyrins. Once the acylated dipyrromethane precursor is synthesised it should be possible to react it with the functionalised aldehyde of choice to synthesise both cytosine and guanine functionalised porphyrins. These porphyrins also bear a number of tert-butyl groups which greatly increases the solubility of the product making purification, characterization and further analysis less challenging. Progress has also been made towards the synthesis of DNA nucleobase functionalised porphyrins by adding the DNA nucleobase functionality to a previously synthesised porphyrin. This

post-synthesis functionalisation has the advantage of allowing a large quantity of porphyrin to be synthesised, to which the chosen functionality can then be added via cross-coupling reactions. As yet these cross-coupling reactions have proved to be unsuccessful but the 'tool-kit' of necessary components has been outlined and shows promise as shown by literature examples of similar reactions. This method also displays the greatest promise for the eventual goal of synthesising the tetra-substituted porphyrin where each meso position bears one of the four DNA nucleobases as shown in the proposed route below (Figure 4.33). Sequential addition of each DNA nucleobase via cross-coupling reactions should, in theory, produce the desired product efficiently with minimal side products.

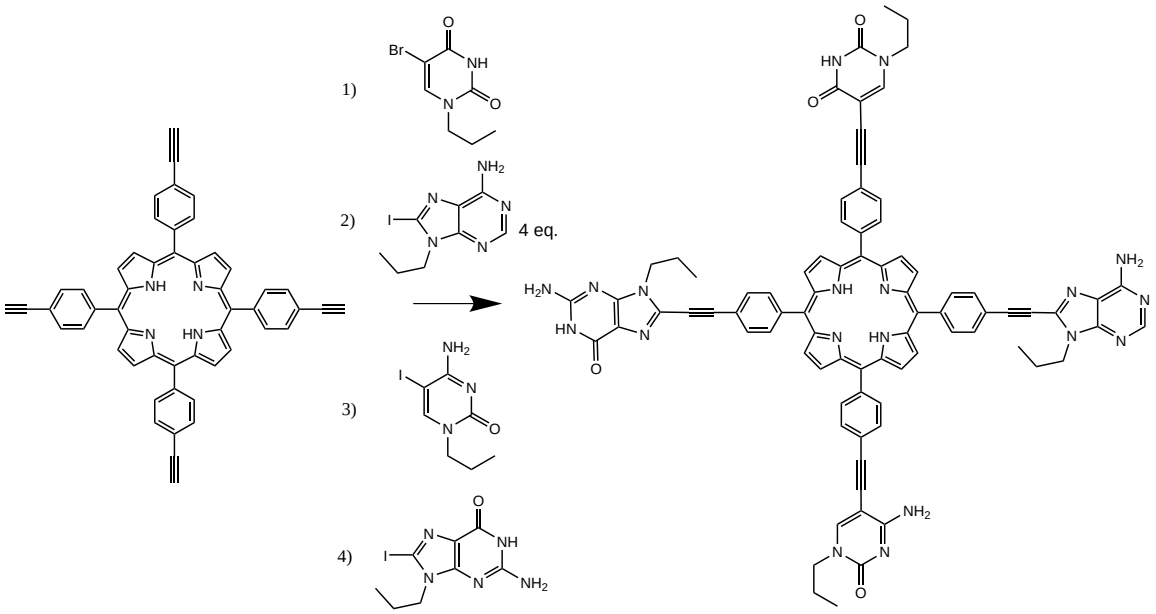


Figure 4.33: Scheme for four successive Sonogashira coupling reactions for the theoretical synthesis of a meso-substituted tetrapheylporphyrin where each meso position bears a different DNA nucleobase.

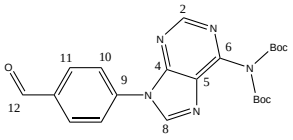
4.5 Experimental

4.5.1 Materials and Equipment

Chemicals were purchased from commercial suppliers and used without further purification. Solid, hygroscopic reagents were dried in an oven before use. Reaction solvents were thoroughly dried before use employing standard methods. Reactions were monitored by thin layer chromatography (TLC) using 0.2 mm aluminium sheets precoated with silica gel 60 F254 (Merck). TLC plates were inspected with a UV lamp featuring both long wave (365 nm) and short wave (254 nm) UV light. Column chromatography was carried out on silica gel Merck 60. Eluent relative vol/vol ratios are indicated in each case. Nuclear magnetic resonance spectroscopy were recorded with a Bruker AV-400 (400 MHz) or Bruker AV-3400 (400 MHz) instrument in the School of Chemistry. The temperature was controlled at 298 K. Chemical shifts (δ) are measured in part per million (ppm) using the signals of the deuterated solvents as the internal standard. EI Mass spectra were taken using a Bruker Apex IV 4.7 T mass spectrometer.

4.5.2 General Procedure for the Chan-Lam-Evans Coupling of Protected Nucleobases

Bis-Boc-adenine (0.43 g, 1.3 mmol), copper(II) acetate (0.35 g, 1.9 mmol), 4-formylphenyl boronic acid (0.39 g, 2.6 mmol) and molecular sieves were added to ask which had been flame-dried and backfilled with dinitrogen three times. Dry DMF (25 mL) was added via cannula followed by addition of trimethylamine (0.35 mL, 2.6 mmol). The resulting green-blue solution was stirred at room temperature for 72 hours in air. The solution was diluted with CH_2Cl_2, ltered through celite and washed with water (3 × 100 ml) in the presence of EDTA. The pale green solution was dried over $MgSO_4$ and the solvent removed under reduced pressure to give the crude product. This was puried via column chromatography (SiO_2 CH_2Cl_2 : CH_2Cl_2:EtOAc 90:10) to yield an off-white solid (0.47 g, 83 %).

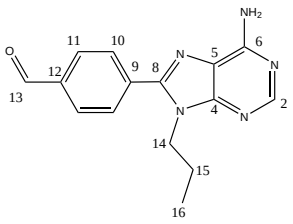


N-9-(4-formylphenyl)-C-6-(bis-Boc-amino)adenine: ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 10.12 (s, 1H, C12), 8.98 (s, 1H, C2), 8.47 (s, 1H, C8), 8.15 (dt, J=8.7, 2.0 Hz, 2H, C11), 8.05 (dt, J=8.4, 1.6 Hz, 2H, C10), 1.51 (s, 18H, Boc)

4.5.3 General Procedure for the Suzuki Coupling or C-5 and C-8 Halogenated Nucleobases

[106]

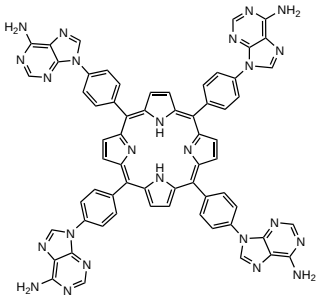
Halogenated nucleobase (1 mmol) was suspended in water (2 ml) with K_2CO_3 (2.2 mmol) and a solution of 4-formylphenylboronic acid (1.5 mmol) in DME (4 ml) was added. The turbid mixture was bubbled with N_2 for 20 minutes and $Pd(PPh_3)_4$ (0.1 mmol) was added. The mixture was heated for at 80 °C for 24 hours under N_2. Upon which the dark solution was cooled and filtered through celite. The filtrate was evaporated to dryness and purified by column chromatography (SiO_2 CH_2Cl_2:MeOH 95:5)



N-9-propyl-C-8-(paraformylphenyl)adenine: ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 9.78 (s, 1H, C12), 7.74 (d, J=8.2 Hz, 2H, C11), 7.32 (d, J=8.2 Hz, 2H, C10), 7.29 (s, 1H, C2), 3.55 (t, J=7.4 Hz, 2H, C13), 1.53 (sxt, J=7.4 Hz, 2H, C14), 0.72 (t, J=7.4 Hz, 3H, C15); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 191.7 (C13), 162.8 (C6), 158.0 (C8), 155.9 (C2), 145.2 (C4), 138.6 (C12), 135.6 (C9), 130.1 (C11), 128.2 (C10), 107.4 (C5), 51.4 (C14), 21.9 (C15), 10.1 (C16);

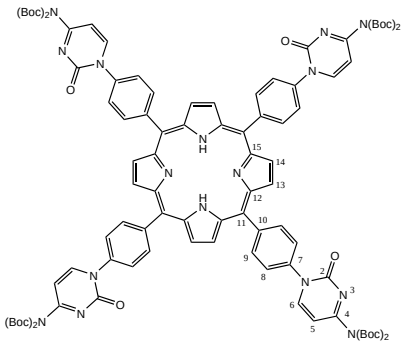
4.5.4 General Procedure for the Formation of Tetra-Substituted DNA Nucleobase Porphyrins

1-formylphenyl-bis-Bocadenine (200 mg 0.45 mmol) was added to a flask which had been flame dried and backfilled with dinitrogen three times. CH_2Cl_2 (300 ml) was added via cannula followed by freshly distilled pyrrole (0.30 ml 0.45 mmol) and TFA (1 ml). The reaction mixture was stirred under dinitrogen at room temperature for 15 minutes upon which a purple solution formed. DDQ (102 mg 0.45 mmol) was added and the mixture stirred at room temperature for 1 hour. Triethylamine (1 ml) was added and the mixture stirred for a further 15 minutes. The solution was filtered through celite and evaporated under reduced pressure to yield a purple residue which was purified via column chromatography (SiO_2 CH_2Cl_2:EtOAc 9:1 v:v CH_2Cl_2:MeOH 8:2 v:v) to yield pure product (9 mg 0.008 mmol 2 %)



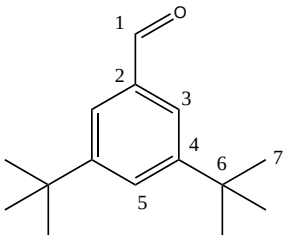
Chemical Formula: $C_{64}H_{42}N_{24}$
m/z: 1146.4024 (100.0%), 1147.4058 (69.2%), 1148.4091 (23.6%)

tetra adenine porphyrin: MS (MALDI-TOF, DCTB Matrix, positive mode) Calc: 1146.40 Found: 1147.3 (M+). The compound was found to be too insoluble in suitable solvents to record a ^{1}H NMR spectrum.



tetra (bis-Boc)cytosine porphyrin ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.92 (s, 8H, C13,C14), 8.33 (d, J=8.2 Hz, 8H, C9), 8.07 (d, J=7.4 Hz, 4H, C6), 7.83 (d, J=8.3 Hz, 8H, C8), 7.37 (d, J=7.4 Hz, 4H, C5), 1.66 (s, 72H, Boc), -2.81 (s, 2H, NH)

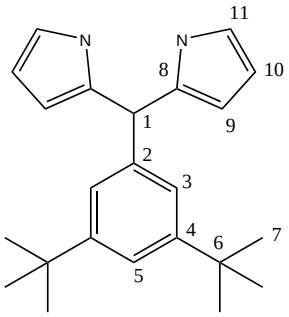
4.5.5 Synthesis of 3,5-di-t-butylbenzaldehyde



1-bromo-3,5-di-tert-butylbenzene (1 g, 3.72 mmol) and THF (15 ml) was added to a flask which had been flame-dried and back-

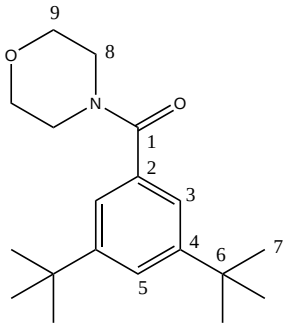
filled three times with nitrogen. *n*-butyllithium (2.26 ml, 1.6 M in hexanes, 4.46 mmol) was added via syringe at -78 °C over a period of 20 minutes. The mixture was stirred for a further 20 minutes at -78 °C and DMF (0.4 ml, 5.2 mmol) was added. After 1 hour the reaction was allowed to warm to room temperature and water (10 ml) was added. The organic layer was extracted with diethyl ether (3 × 10 ml), dried over $MgSO_4$ and the solvent removed under reduced pressure to give a colourless solid (1.2 g 100 %): ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 10.02 (s, 1H, C1), 7.74 (d, J=1.8 Hz, 2H, C3), 7.72 (t, J=1.8 Hz, 1H, C6), 1.38 (s, 18H, C7); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 193.2 (C1), 151.8 (C4), 136.2 (C2), 128.9 (C5), 124.1 (C3), 35.0 (C6), 31.3 (C7); MS (EI) Calc.: M+ = 218.17 Found: M+H = 219.17, M+Na 241.16, M+NH_4 = 236.20

4.5.6 Synthesis of 3,5-di-tert-butylphenyl-dipyrromethane



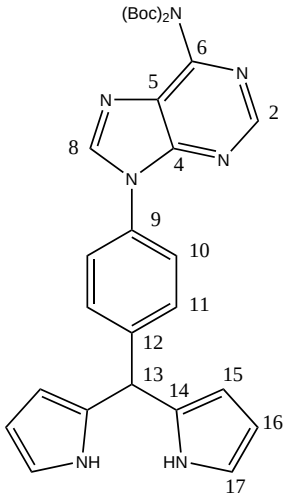
3,5-di-tert-butylbenzaldehyde (1 g, 3.5 mmol) was dissolved in freshly distilled pyrrole (35 ml) and TFA (0.3 ml, 0.35 mmol) was added. The mixture was stirred for 20 minutes upon which a pale yellow solution was formed. The reaction was quenched by addition of NaOH (30 ml, 0.01 M). The organic layer was extracted with dichloromethane, dried over $MgSO_4$ and the solvent removed under reduced pressure. The dark yellow oil was purified by column chromatography (SiO_2, DCM:EtOAC:NEt_3 90:9:1-84:15:1) to give a gold-yellow solid (1.1 g, 67 %): ^{1}H NMR ($CDCl_3$,400MHz): δ (ppm) 8.95 (s, 1H, NH), 8.35 (s, 1H, NH), 7.69 (d, J = 8.5 Hz, 2H, C3), 7.48 (d, J = 8.4 Hz, 2H), 6.78 (m, J = 1.4 Hz, 2H, C11), 6.22 (q, J = 2.8 Hz, 2H C10), 5.94 - 6.01 (m, 2H, C9), 5.62 (s, 1H, C1), 1.47 - 1.59 (m, 18H); ^{13}C NMR ($CDCl_3$,101MHz): δ (ppm) 150.62 (C4), 132.86 (C2), 130.01 (C8), 129.26 (C3), 123.63 (C5), 117.67 (C11), 108.58 (C10), 107.57 (C9), 84.05 (C1), 43.64 (C6), 27.86 (C7); MS (ESI) Calc.: M+ = 326.24 Found: M+H = 327.31

4.5.7 Synthesis of 3,5-di-tert-butylphenyl morpholin-4-yl ketone



3,5-di-tert-butylbenzoic acid (7 g, 29.9 mmol) was added to a flask which had been flame-dried and back-filled with nitrogen three times. DCM (70 mL) was added via cannula followed by DMF (0.77 mL, 9.9 mmol) and oxalyl chloride (2.3 mL, 26.9 mmol) via syringe (15 mins). This colourless solution was stirred under nitrogen for 90 minutes. A second flame-dried, back-filled flask was charged with DCM (70 mL), morpholine (9.3 mL, 108 mmol) and triethylamine (4.3 mL, 30.8 mmol). This solution was carefully added to the first via syringe over a period of 20 minutes. The resulting turbid solution was stirred under nitrogen for 2 hours and quenched via careful addition of 0.5 M HCl (100 mL) and stirred fora further 30 minutes. The organic layer was separated and washed with water (3 100 mL) and dried over $MgSO_4$ and the solvent removed under reduced pressure to yield a colourless product (8.43 g, 95%): ^{1}H NMR ($CDCl_3$, 400MHz): δ (ppm) 7.47 (t, J= 1.8 Hz, 1H, C5), 7.23 (d, J= 1.9 Hz, 2H, C3), 3.86 - 3.57 (m, 8H, C8-C9), 1.33 (s, 18H, C7); ^{13}C NMR ($CDCl_3$, 101MHz): δ (ppm) 171.5 (C1), 151.1 (C4), 134.5 (C5), 123.9 (C3), 121.2 (C2), 66.9 (C8-C9, 34.9 (C6), 31.3 (C7)

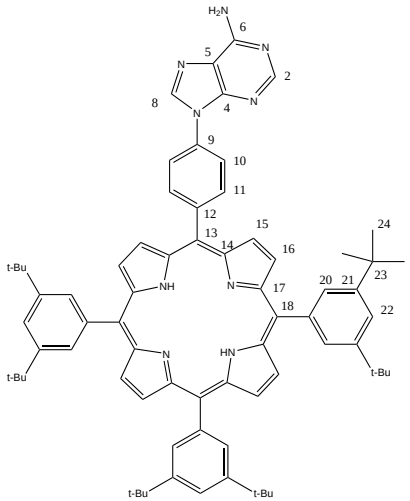
4.5.8 Synthesis of bis-Boc-adenine-phenyl-dipyrromethane



1-formylphenyl-bis-Boc-adenine (700 mg, 2.1 mmol) and freshly distilled pyrrole (16 mL, 228 mmol) were added to a ask and degassed with nitrogen for 30 minutes. $InCl^3$ (102 mg, 0.46 mmol) was added and the reaction was stirred under nitrogen for 2 hours. The reaction was quenched via addition of aq. NaOH and the mixture was filtered. The solvent was removed under reduced pressure to yield

a dark brown oil which was purified by column chromatography (SiO_2 CH_2Cl_2 : CH_2Cl_2:EtOAc 85:15) to yield an off-white solid (300 mg, 26 %). ^{1}H NMR ($CDCl_3$, 400MHz): δ (ppm) 8.95 (s, 1H, C2), 8.35 (s, 1H, C8), 7.69 (d, J= 8.5 Hz, 2H, C11), 7.48 (d, J= 8.4 Hz, 2H, C10), 6.78 (m, J= 1.4 Hz, 2H, C17), 6.22 (q, J= 2.8 Hz, 2H, C16), 5.94 - 6.01 (m, 2H, C15), 5.62 (s, 1H, C13), 1.47 - 1.59 (m, 18H, Boc); ^{13}C NMR ($CDCl_3$, 101MHz)): δ (ppm) 152.79 (C6), 150.62 (Boc), 143.65 (C2), 143.22 (C8), 132.86 (C5), 131.72 (C12), 131.43 (C9), 130.01 (C14), 129.26 (C11), 123.63 (C10), 117.67 (C17), 108.58 (C16), 107.57 (C15), 84.05 (Boc), 43.64 (C5), 27.86 (Boc) MS (EI): m/z+: Calc 578.2492 (M + Na)+ Found 578.2498 (M + Na)+.

4.5.9 Synthesis of mono-(phenyladenine)-tri-(3,5-di-tert-butylphenyl)porphyrin (Mono AP)



1,9-bis(3,5-di-tert-butylbenzoyl)-5-(3,5-di-tert-butylphenyl)dipyrromethane (80 mg, 0.1 mmol) was dissolved in MeOH (25 ml) and THF (10 ml). $NaBH_4$ (400 mg, 0.1 mmol) was added slowly over 30 minutes upon which an orange solution was formed. The reaction was quenched via addition of water, extracted with CH_2Cl_2, dried over K_2CO_3 and the solvent removed under reduced pressure. The resulting orange solid was dried in vacuo for 1 hour to yield the dicarbinol compound before being carried onto the next step without further purification. The dicarbinol compound and bis-Boc-adenine-phenyl-dipyrromethane were dissolved in MeCN (200 ml) and TFA (0.7 mL, 9.36 mmol)) was added and the mixture stirred at room temperature for 5 minutes. DDQ (531 mg, 2.34 mmol) was added and the reaction mixture was stirred for a further hour before triethylamine (1.3 mL, 9.36 mmol) was added to quench the reaction, before stirring the dark purple solution for a further 15 minutes. The solution was filtered through alumina and the solvent removed under reduced pressure. The dark purple solid was purified via column chromatography (SiO_2 CH_2Cl_2) (10 mg 0.008 mmol 8 %) . MS (MALDI-TOF, DCTB matrix, positive mode) Calc: 1084.48 Found 1084.8 (M+H)+. ^{1}H NMR ($CDCl_3$, 400MHz)): δ (ppm) 8.87 - 9.00 (m, 8H, C15,C16), 8.60 (s, 1H, C2), 8.42 - 8.52 (m, 3H, C22), 8.16 (d, 3J= 8.5 Hz, 2H, C11), 8.07 - 8.14 (m, 6H, C20), 7.78 - 7.87 (m, 3H, C10, C8), 1.55 (br. s., 54H, C24), -2.70 (m, 2H, NH). ^{13}C NMR ($CDCl_3$, 126MHz)): δ (ppm) 155.8, 153.9, 150.2, 148.8, 148.7, 142.6, 141.2, 141.2, 139.8, 135.7, 134.5, 129.9, 129.7, 121.9, 121.6, 121.4, 121.1, 120.5, 117.6, 35.1, 31.7 MS (MALDI-TOF, DCTB matrix, positive mode) Calc: 1084.48 Found 1084.8 (M+H)+.

Chapter 5

Self-Assembly of DNA Functionalised Porphyrins

5.1 Crystal Structure

The synthesised DNA base functionalised porphyrins could now be used for self-assembly and their intermolecular interactions studied. The crystal structure of the previously synthesised mono-thymine porphyrin had already been solved where hydrogen bonding interactions were observed between the thymine moiety as expected as shown below (Figure 5.1)[83].

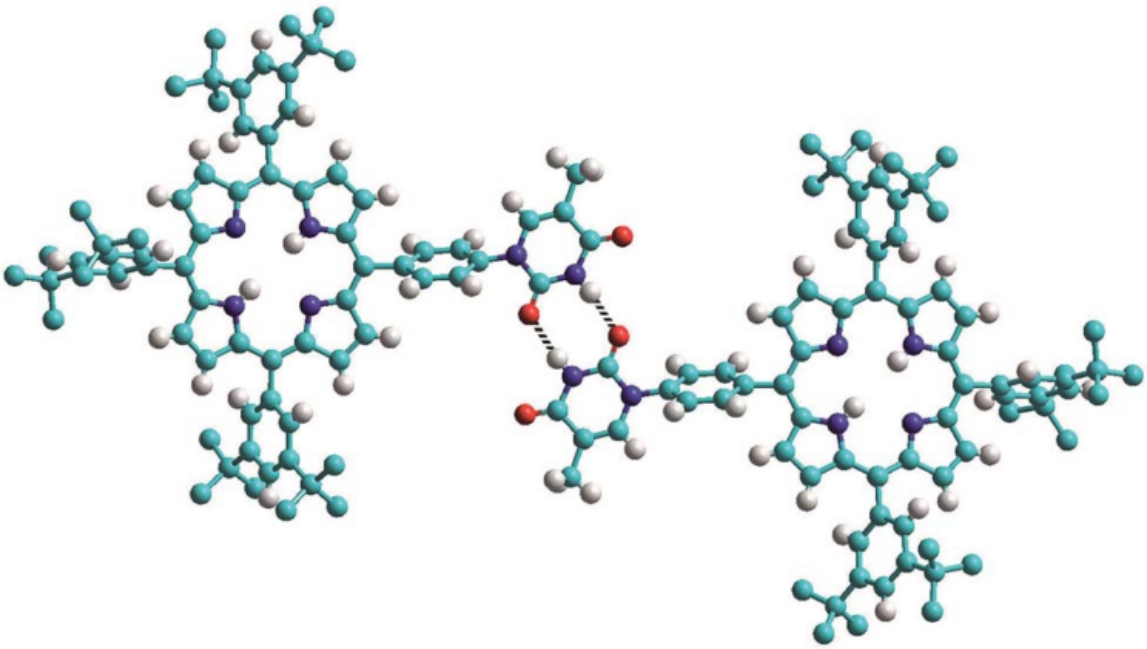


Figure 5.1: View of the single crystal X-ray structure of mono-TP illustrating the formation of the inter-thymine double hydrogen-bonding interaction (represented with the two dotted lines) observed in the solid state. Hydrogen atoms of the tert-butyl groups are omitted for clarity. Atoms are coloured as follows: C - light blue; N - dark blue; H - white; O - red.[83]

The crystal structure of the synthesised mono-adenine porphyrin was found by growing a single crystal by slow evaporation of the compound from $CDCl_3$. The crystal structure confirms the identity of the product and the relative planar arrangement of the adenine moiety with respect to the porphryin ring (Figure 5.2). The crystal structure provides insight into the intermolecular interactions present within the solid phase. Of most interest is the adenine-adenine hydrogen bond. As expected, the adenine moiety forms intermolecular hydrogen bonds to an adjacent molecule. The hydrogen bonding mode observed is that between the Ho-Ho faces of the molecule. As discussed previously, it is possible for unfunctionalised adenine to self-associate into up to 21 different arrangements while functionalisation in the *N*-9 position, as employed here and in the *N*-9-propyladenine analogue, limits the number of possible arrangements to 15. The three most relevant of those, in order of decreasing stability, are the Wc-Wc, Wc-Ho and Ho-Ho hydrogen bonding modes. Therefore, it is relatively surprising that the mono-adenine porphyrin was found to adopt the Ho-Ho hydrogen bonding interaction as the Wc-Wc and Wc-Ho hydrogen bonding modes are observed for adenine or 9-propyladenine. This can likely be attributed to the the presence of additional hydrogen bonding interactions from EtOAc solvent molecule or other packing interactions from the large t-butyl functionalised porphyrin ring.

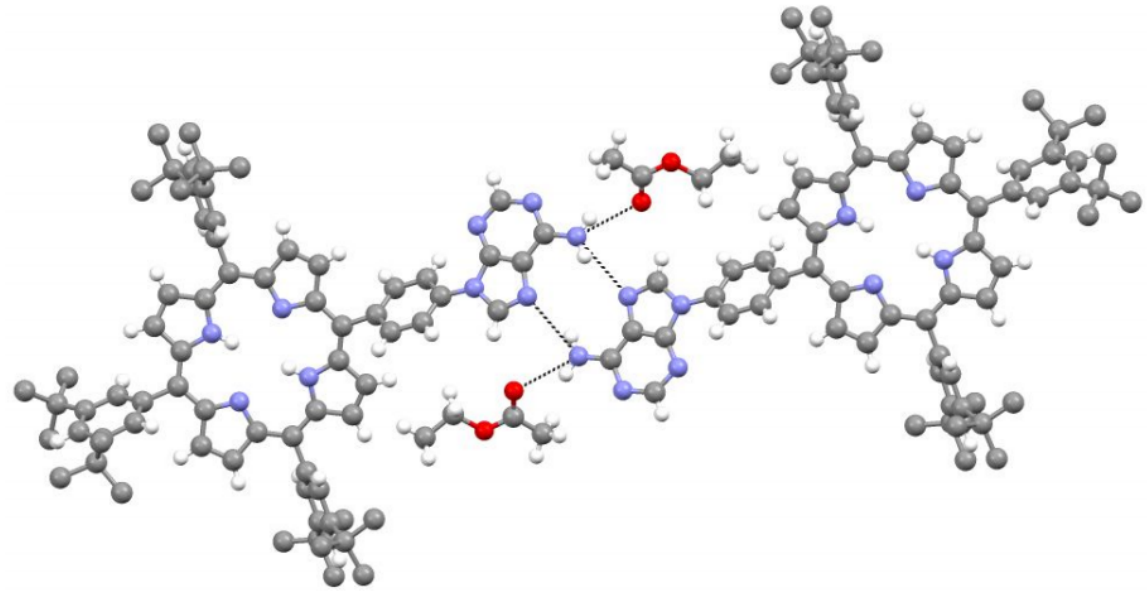


Figure 5.2: View of the single crystal X-ray structure of mono-AP illustrating the formation of the inter-adenine double hydrogen-bonding interaction (represented with the two dotted lines) observed in the solid state. Hydrogen atoms of the tert-butyl groups are omitted for clarity. Atoms are coloured as follows: C - light blue; N - dark blue; H - white.

The hydrogen bond lengths of the mono-adenineporphyrin crystal in the Ho-Ho arrangement are slightly longer than those observed in 9-propyladenine which adopts the Wc-Ho arrangement. This is in agreement with the relative calculated stability of these interactions. For the Wc-Ho interactions there are two bond lengths and angles shown due to the asymmetric interaction.

Entry	Mode	r H-N (Å)	r N-N (Å)	θ D-H-A
n-prA	Wc-Ho	2.085, 2.207	2.965, 3.086	165.646, 169.701
tBBzA	Wc-Ho	2.171, 2.200	3.029, 3.028	174.9, 161.7
mAP	Ho-Ho	2.340	3.084	145.1

Table 5.1: Table of Hydrogen bonding angles between adenine functionalised porphyrin and simple alkyl analogues.

5.2 NMR Dilution and Titration Studies

To further study the hydrogen bonding of DNA functionalised porphyrins in solution a number of NMR titration experiments were carried out. These would provide information on the association constant of the interaction giving an insight into the relative strength of the interaction. The previously synthesised mono-thymineporphyrin was analysed by both a dilution experiment and titration experiment with 9-propyladenine to determine the self-association constant and association constant with a simple analogue of the relevant base pair. Mono-thymineporphyrin with its bulky t-butyl groups was used by Champness *et al.* for these studies due to the insufficient solubility of tetra-thymineporphyrin in a suitable solvent[83]. These experiments were repeated with the mono-adenineporphyrin and its relevant base pair (propylthymine) to compare the self-association and base pair dimer constants as shown below (Table 5.2).

Entry	Kd S	Kd D
mTP	9	84
mAP	12	245

Table 5.2: Table of association constants for the self-association (Kd S) and complementary dimerisation (Kd D) of DNA functionalised porphyrins with the relevant alkylated base pair.

As expected, the association constants (Kd D) for the binding of DNA functionalised porphyrins to the relevant base pair are considerably higher than the self-association constant (Kd S). This is due to the increased hydrogen bond strength of complementary base pair binding (A-T vs AA or TT). The Kd S and Kd D values are in line with those reported in literature. Speciation plots of titration of mAP with propylthymine show how the population of mAP (Host) and propylthymine (G) change as the guest is titrated into the host solution (Figure 5.3). At low concentration of guest the solution is primarily consisted of host as expected. As the guest is added the population of host decreases and the host-guest dimer population increases.

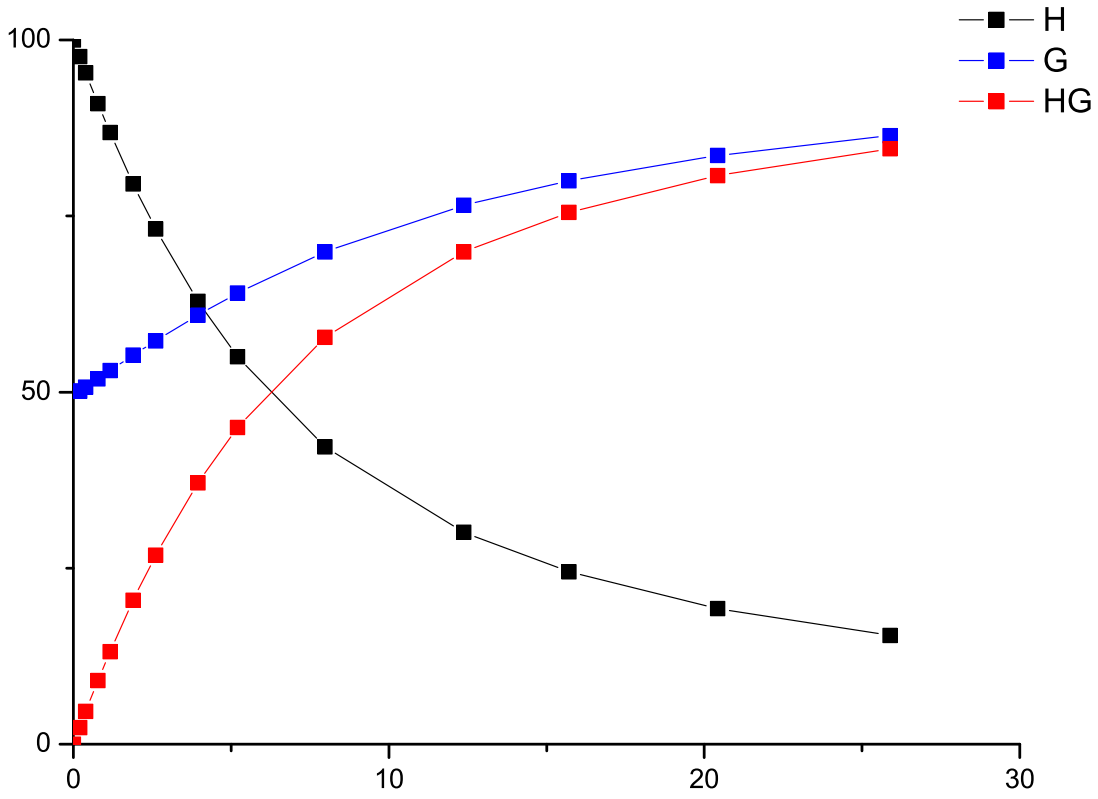


Figure 5.3: Speciation plot of population against concentration for the titration of porpylthymine (H) with mono-adenineporphyrin (G).

5.3 Surface Self-Assembly of DNA Functionalised Porphyrins

The primary focus of this work is studying the 2D self association of DNA functionalised porphyrins on a surface. It was hoped that an adenine functionalised porphyrin would interact with the previously synthesised thymine functionalised porphyrin in a multicomponent system due to the complementary hydrogen bonds formed between the two nucleobases. This surface self-assembly work was carried out by Dr Matthew Blunt, our collaborator at University College London in a recently submitted publication.

5.4 Self-Assembly of Zn-Tetra-TP

The first investigation of this work was the study of zinc-metallated-tetrathymineporphyrin. This was compared to the previously studied tetrathymine porphyrin which was found to self-assemble on the surface due to thymine-thymine intermolecular hydrogen bonds.[83] The zinc-metallated porphyrin was found to assemble in an identical manner to the free-base porphyrin but with the clear increase in contrast due to the central metal atom (5.4). Zn-tetra-TP forms an ordered structure on a square lattice with p4 symmetry with a unit cell of dimensions a = (26.2 ± 0.9)Å; b = (26.3 ± 0.5)Å; α = 89 ± 2)°. The unit cell vectors were arranged at angles of θa = (5 ± 2)° and θb =

(26 ± 2)° with respect to the underlying HOPG lattice. These unit cell dimensions are identical to those previously measured for free-base tetra-TP.

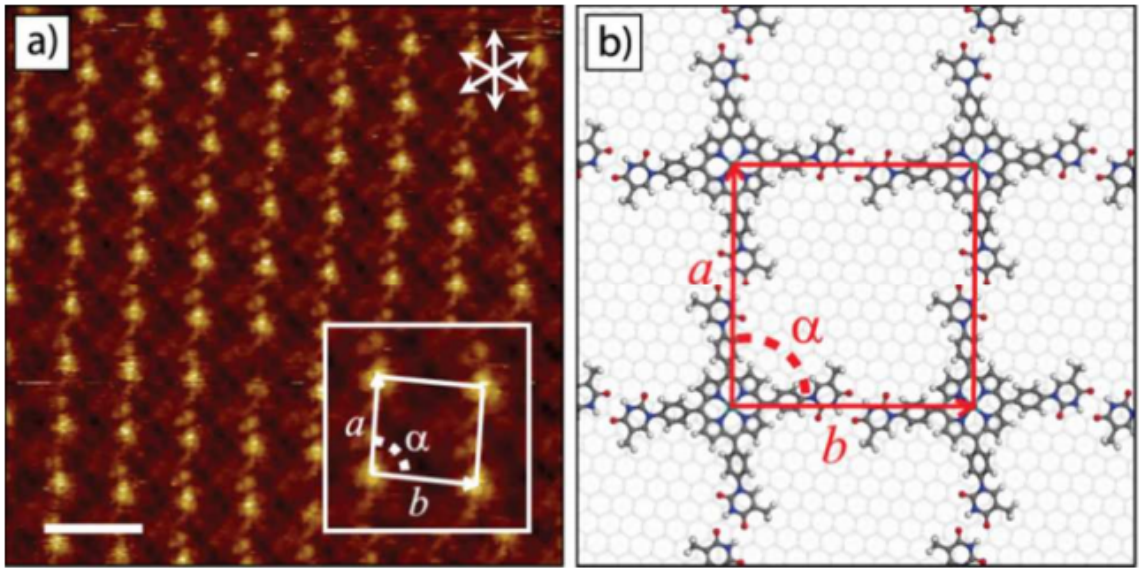


Figure 5.4: Self-assembly of Zn-tetra-TP at the liquid-solid interface between TCB and HOPG. (a) Drift corrected STM image of the Zn-tetra-TP network; bright features correspond to porphyrin cores. The insert shows a magnified view of the structure with the 2D unit cell marked. Unit cell parameters from STM are: a = (26.2 ± 0.9) Å; b = (26.3 ± 0.5) Å; α = (89 ± 2)°. The orientation of the underlying HOPG lattice is given by the arrows in the top right hand corner. STM imaging parameters: Vs = -0.85 V; It = 10 pA. Scale bar = 4 nm. (b)Molecular model of the Zn-tetra-TP network from MM simulations. Unit cell parameters a= 26.7 Å; b= 26.7 Å; α = 89.4°.

Molecular modelling was used to establish the positions of Zn-tetra-TP molecules on the surface and the relative orientation of the thymine groups. Geometry optimised MM simulations were calculated using the unit cell dimensions from drift-corrected STM images. Thymine groups can rotate around the thymine *N*-1-phenyl bond leading to two possible orientations that thymine can adopt relative to the rest of the molecule. The MM simulations selected the orientation of the thymine groups such that the hydrogen bonding interactions were maximised whilst retaining the unit cell dimensions observed via STM. In the resulting model shown in Figure 5.4 all the thymine groups point in the same direction with respect to the porphyrin core and the Watson-Crick hydrogen bonding mode is observed. Unit cell dimensions from the MM simulations are a = 26.7 Å; b = 26.7 Å; α = 89.4°, with the unit cell vectors making angles with respect to the underlying HOPG lattice of θa = 4.6° and θb = 26.0°. These unit cell dimensions are clearly in agreement with the experimental dimensions observed via STM.

5.5 Self-Assembly of Tetra-AP

The self-assembly of the synthesised tetra-adenine porphyrin was then studied on the HOPG surface. This molecule had not been studied before and it was hoped that intermolecular hydrogen bonds would form between the adenine functional groups to form a 2D supramolecular network. The tetra-adenine porphyrin was not metallated so images with such high contrast as Zn-tetra-TP were not expected. Instead, a cross-shaped pattern was observed which allowed the orientation of porphyrin

molecules with respect to each other to be identified. The self-assembled structure forms an oblique lattice with p2 symmetry and unit cell dimensions a= (19.8 ± 0.6) Å; b= (22.9 ± 0.7) Å; α = (84 3)°. The angles of the unit cell vectors with respect to the underlying HOPG lattice are θa= (23 4)° and θb= (3 3)°. The adenine groups on tetra-AP can also rotate with respect to the porphyrin core in a manner similar to that shown by tetra-TP leading to two possible orientations when the molecule is adsorbed onto the surface. Molecular modelling simulations showed that in contrast to the thymine porphyrins these adenine groups do not adopt the same orientation with respect to the porphyrin core. Instead, the adenine groups adopt alternative orientations around the molecule as shown in Figure (5.5). The MM simulations found that these orientations were necessary to form a molecular model that both matched the experimental unit cell dimensions and displayed sensible hydrogen bonding interactions. In the molecular model proposed, each adenine molecule forms three hydrogen bonds: A Wc-Wc dimer between two adenine groups at opposite corners of the unit cell and another formed by the close packing of an adjacent tetra-AP. The unit cell dimensions of the tetra-AP MM are a = 19.2 Å; b = 22.8 Å; α = 83.7°, with the unit cell vectors making angles with respect to the underlying HOPG lattice of θa = 26.3° and θb= 2.7° which closely match the experimentally derived unit cell dimensions.

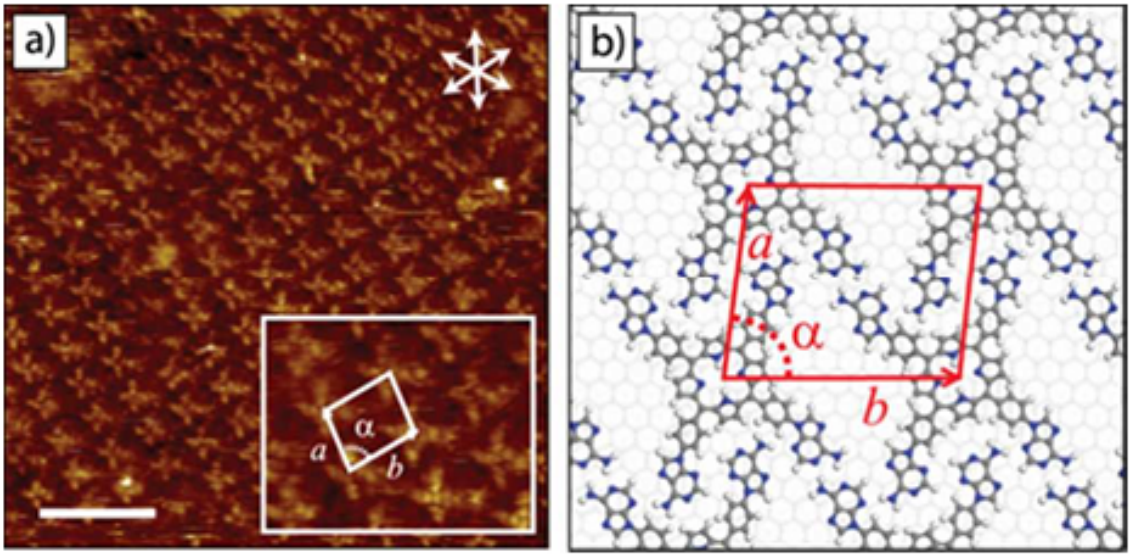


Figure 5.5: Self-assembly of tetra-AP at the liquidsolid interface between TCB:THF and HOPG. (a) Drift- corrected STM image of the tetra-AP network; cross shaped features correspond to the porphyrin cores. The insert shows a magnified view of the structure with the 2D unit cell marked. Unit cell parameters from STM are: a = (19.8 ± 0.6) Å; b = (22.9 ± 0.7) Å; α = (84 3)°. The orientation of the underlying HOPG lattice is given by the arrows in the top right hand corner. STM imaging parameters: Vs = -0.65 V; It = 45 pA. Scale bar = 6 nm. (b) Molecular model of the tetra-AP network from MM simulations. Unit cell parameters a = 19.2 Å; b = 22.8 Å; α = 83.7°.

5.6 Self-Assembly of Zn-Tetra-TP and Tetra-AP Mixture

Zn-Tetra-TP and Tetra-TP were codeposited in a 1:1 molar ratio on HOPG and studied via STM to show the images shown below. Figure 5.6 (a) shows the ordered domains formed on the surface and in image 5.6 (b) the cross shape of individual porphyrins can be seen highlighting the ordered orientation that the molecules adopt. Analysis of drift-corrected images gave unit cell dimensions

(marked on Figure 4b in red) of a = (21.7 ± 0.7) Å; b= (22.0 ± 0.6)Å; α = (90 ± 3)°. The angles the unit cell vectors make with respect to the underlying HOPG lattice are θa= (12 ± 3)° and θb= (19 3)°. It is clear from the images that the porphyrins adopt an ordered arrangement due to the observed ordered pattern of cross-shaped images. The orientation and unit cell dimensions of these images was found to be different to those observed for Zn-tetra-TP and tetra-AP mono-component networks indicating that the mono-component self-assemblies are not present. However, in images (5.6 a and b) it is not possible to differentiate the Zn-thymine and adenine species. It was hoped that the metallated porphyrin could be differentiated from the free-base porphyrin by the increased contrast observed in the STM images for the metallated species. Extended scanning of the surface did find areas where a different contrast was consistently observed. In these images the molecules of brighter contrast have been assigned to the Zn-containing porphyrin.

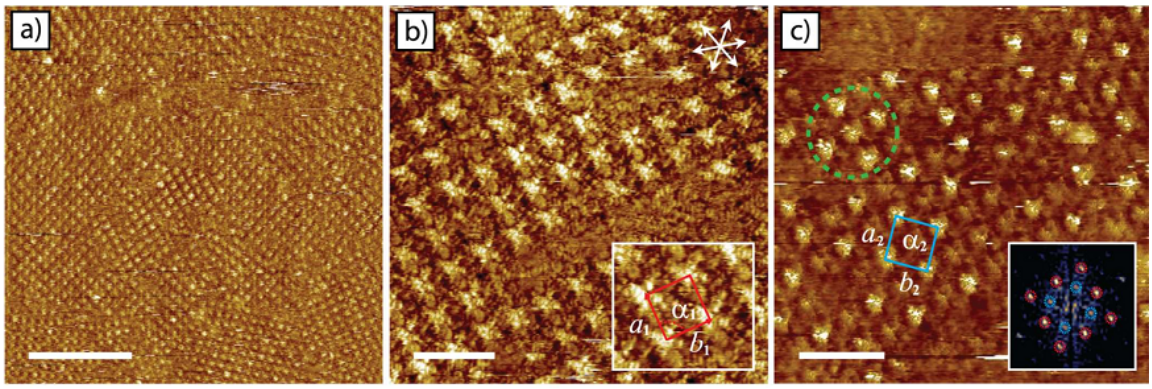


Figure 5.6: (a)Large scale STM image of the Zn-tetra-TP and tetra-AP network.(b) Drift-corrected STM image of the Zn-tetra-TP and tetra-AP network; cross shaped features correspond to porphyrin cores. The insert shows a magnified view of the structure with the 2D unit cell marked in red. Unit cell parameters from STM are: a= (21.7 ± 0.7) Å; b= (22.0 ± 0.6) Å; α = (90 ± 3)°. The orientation of the underlying HOPG lattice is given by the arrows in the top right hand corner. (c) Alternate contrast image of the Zn-tetra-TP and tetra-AP network showing ordered arrangement of molecules with brighter contrast corresponding to Zn-tetra-TP; insert shows the 2D Fourier transform of the STM image. Unit cell parameters for the marked blue cell from STM are: a= (31.1 ± 1.1) Å; b= (30.8 ± 1.0) Å; α = (90 ± 4)°. STM image scale bars: (a)25 nm; (b)5nm and (c)6 nm. STM imaging parameters: (a)Vs = -0.50 V, It = 10 pA; (b)Vs = -0.25 V, It = 10 pA; and (c)Vs = -0.55 V, It = 10 pA.

A model was calculated for the arrangement of Zn-tetra-TP and tetra-AP on the surface by MM simulations as shown in Figure 5.7. This structure displays p4 symmetry consisting of alternating Zn-tetra-TP and tetra-AP molecules arranged in a chessboard type pattern. The unit cell dimensions measured via STM measurements from image (b)(red) and (c)(blue) are overlaid on the model. The unit cell dimensions are red: a1= 21.9 Å; b1= 22.2 Å; α1= 90.9°; and blue: a2= 31.4 Å; b2= 30.9 Å; α2= 90.8°. The angles that the red unit cell vectors make with respect to the underlying HOPG lattice are θa = 13.9° and θb = 17.0°. These dimensions are all consistent with those obtained from drift-corrected STM measurements. In this model the structure is stabilised by four hydrogen bonds in the form of an ATAT quartet as shown in Figure 5.7. These A and T components are each from different porphyrin molecules. The quartet consists of two Wc-Wc dimers (highlighted in green) and two additional NH-O hydrogen bonds formed from the adenine amine group and thymine carbonyl. This results in six hydrogen bonds in total for the hydrogen bonding quartet. As described earlier, ATAT quartets have been observed before by Mamdouh *et al.*[82]. In that study a number of adenine-thymine interactions were observed on HOPG including AT dimers and different ATAT

quartets (Figure 5.8). The stability of these interactions were calculated using DFT calculations to quantify the binding energy of these structures. The most stable structure in this study with a binding energy of Ebind = 1.24 eV per mol consisted of a Hoogsteen structure as shown in Figure 5.8 (12). (As was observed in the propyladenine:propylthymine cocrystal structure). The ATAT quartet observed between adenine and thymine porphyrins however, was calculated to have a lower binding energy of Ebind = 1.08 eV per quartet. This would suggest that the observed ATAT quartet was not the most stable possible arrangement for the DNA nucleobase porphyrins. It is important to note that the structures studied by Mamdouh *et al.* consisted of unfunctionalised nucleobase molecules and when the role of the porphyrin core is considered it can be seen that all the planar structures, except that which is proposed, would be prevented due to the steric effect of the large porphyrin core. Therefore, the ATAT quartet shown in figure (5.7) is likely to be the most stable possible arrangement for co-deposition of Zn-tetra-TP and tetra-AP.

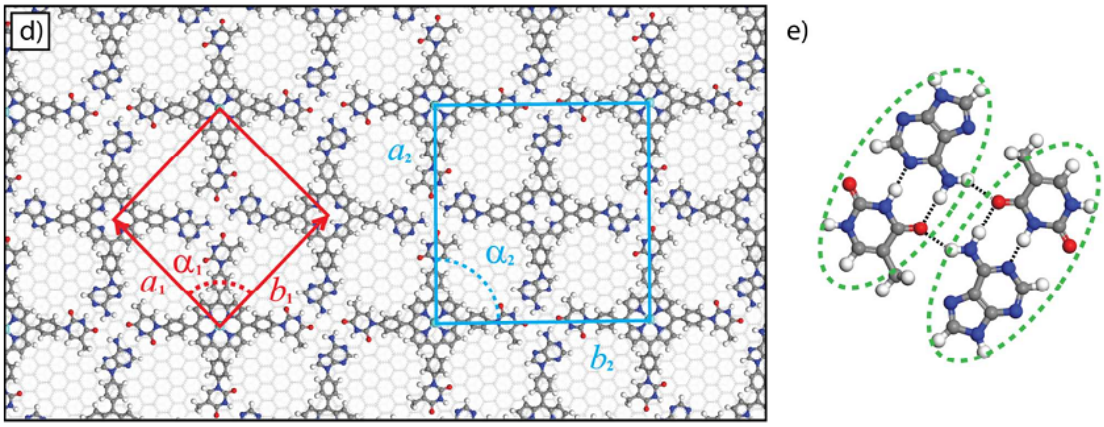


Figure 5.7: (d) Molecular model of the Zn-tetra-TP and tetra-AP network from MM simulations. Unit cell parameters: red a1= 21.9 Å; b1= 22.2 Å; α1= 90.9°; blue a2= 31.4 Å; b2= 30.9 Å; α2= 90.8°; (e) Molecular model of ATAT quartet with hydrogen bonds marked as black dashed lines. AT pairs in the Watson-Crick hydrogen bonding configuration are marked by the green dashed ovals.

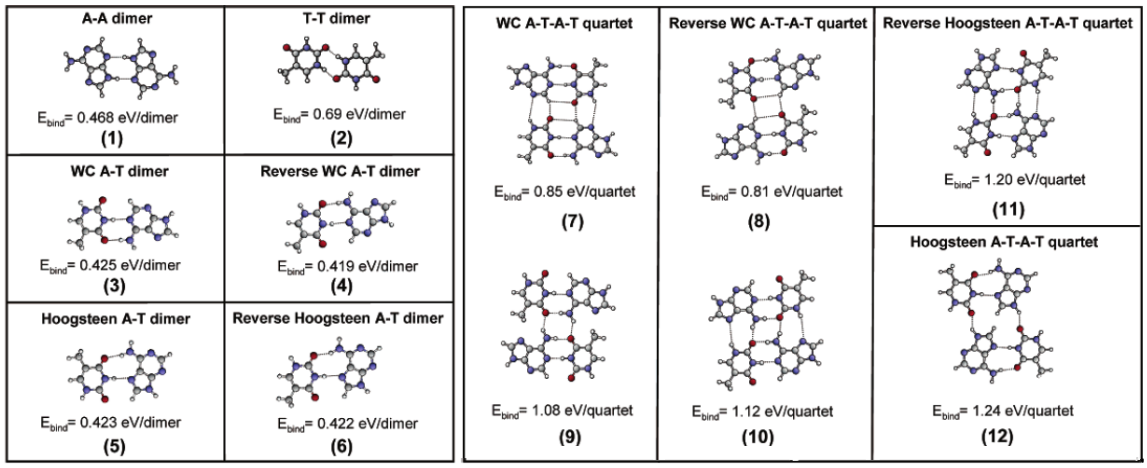


Figure 5.8: Adenine-thymine dimers and quartets and relevant binding energies as calculated by Besenbacher *et al.*[82].

5.7 Conclusions

The self-assembly of DNA functionalised porphyrins was studied. The crystal structure of mono-AP was found to self-associate via Hoogsteen hydrogen bonds between the adenine moiety. The hydrogen bond lengths were analysed in this crystal structure were found to be slightly longer than those observed in the simple adenine analogues. The hydrogen bonding interactions of thymine and adenine functionalised porphyrins was studied by NMR dilution and titration experiments and evidence of hydrogen bonding interactions were observed in solution. As expected, the association constant for the complementary base pairs was considerably higher than the relevant self-association constant. Finally, selected DNA functionalised porphyrins were studied on a HOPG surface and the Zn-tetra-TP molecule was found to associate in the same manner as the equivalent free-base porphyrin. The associated zinc molecule was found to result in a brighter contrast in the STM image. Tetra-AP was also found to self-associate on the HOPG surface via three hydrogen bonds. In the suggested model tetra-AP forms a Wc-Wc dimer between two adenine molecules on opposite corners of the unit cell and another formed by the close packing of a second tetra-AP molecule. The Zn-tetra-TP and tetra-AP molecules were then mixed in a 1:1 ratio on the surface and were found to form an alternating arrangement based complementary ATAT interactions. The ATAT quartet formed in this self-assembled structure has been shown in literature to not be the most stable ATAT quartet but the model proposed also considers the limits of the whole structure where the DNA bases are attached to a rigid porphyrin core.

5.8 Experimental

5.8.1 Crystallography

Single crystal X-ray diffraction experiments were performed on a Rigaku Saturn724+ diffractometer equipped with a rotating anode using monochromated Cu-Kα radiation ($\lambda = 1.5418$ Å) at 120 K. The structures were solved by direct methods using either SHELXS or SHELXT[111] and refined with SHELXL[112] using a least squares method. OLEX2 software was used as the solution, refinement and analysis program.[113] All hydrogen atoms were placed in geometrically calculated positions; non-hydrogen atoms were refined with anisotropic displacement parameters.

For the structure of mono-A disordered solvent ethyl acetate could not be reasonably modelled to give convergence and the data was processed using SQUEEZE,[152] which gave an estimate of 177 e^- per cell. corresponding to approximately four ethyl acetate molecules per cell. These molecules were then added to the sum formula. The tertiary butyl groups 1 (C22A-25A and C22B-25B), 2 (C29A-31A and C29B-31B), 3 (C42A-44A and C4B2-44B), 4 (C47A-50A and C47B-50B), 5 (C61A-63A and C61B-63B), and 6 (C67A-C69A and C67B-69B) were disordered over two positions and their respective occupancies were refined competitively. 1 converged to a ratio of 0.57:0.43, 2 converged to a ratio 0.73:0.27, 3 converged to a ratio of 0.64:0.36, 4 converged to a ratio of 0.90:0.10, 5 converged to a ratio of 0.58:0.42, and 6 converged to a ratio of 0.26:0.74. Enhanced rigid bond and similarity restraints were applied to the thermal parameters of the disordered tertiary butyl groups. Chemically equivalent bonds of the disordered molecules were restrained to be approximately equal to each other. The angles between the atoms of 4 and 6 were restrained to an approximate refining value and at convergence was 2.5.

5.8.2 STM Studies

Synthesis of Zinc-tetra(benzoyl)thymine-phenylporphyrin

Tetrabenzoylthymine-phenylporphyrin[83] (0.01 mmol, 20 mg) was dissolved in CH_2Cl_2/MeOH (20 ml 1:1 v:v). $Zn(OAc)_2$(10 eq.) was added to the dark purple solution and stirred overnight at room

temperature upon which the solvent was evaporated under vacuum. The resulting shiny purple solid was purified by column chromatography (SiO_2 CH_2Cl_2:EtOAc 10:1) to yield the pure metallated compound (12 mg, 80 %). ^{1}H NMR ($CDCl_3$, 400MHz): δ(ppm) 8.94 (s, 8H), 8.32 (d, J=8.6 Hz, 8H), 8.12 (dd, J=8.4, 1.3 Hz, 8H), 7.82 (d, J=8.3 Hz, 8H), 7.75 (d, J=1.2 Hz, 4H), 7.72 (t, J=7.5 Hz, 4H), 7.59 (t, J=7.8 Hz, 8H), 2.19 (d, J=1.2 Hz, 12H). MS (MALDI-TOF, DCTB Matrix, positive mode) Calc: 1588.37 Found:.1591.3

Synthesis of Zinc-tetrathymine-phenylporphyrin, Zn(tetra-TP)

Zinc-tetra(benzoyl)thymine-phenylporphyrin (12 mg) was dissolved in THF (10 ml) and NH_4OH (2.5 ml) was added. The resulting mixture was stirred at room temperature overnight and then the solvents were evaporated under vacuum. The residue was triturated with toluene and MeOH before the purple solid was suspended in methanol and isolated by centrifugation (8 mg, 91 %). MS (MALDI-TOF, DCTB Matrix, positive mode) Calc: 1172.27 Found: 1172.9. The compound was found to be too insoluble in suitable solvents to record a ^{1}H NMR spectrum.

STM Imaging

Zn-Tetra-TP and Tetra-AP were dissolved in a 1:9 volume ratio mixture of tetrahydrofuran (THF) (Sigma Aldrich, anhydrous, 99.9%) and 1,2,4-trichlorobenzene (TCB) (Sigma Aldrich, anhydrous, 99.9%) producing a 3.5 $\times 10^{-5}$ M solution and used as a stock solution to produce a range of TTP/TAP mixtures with various molar ratios and overall molar concentrations. STM samples were produced by mounting a freshly cleaved HOPG sample (Agar Scientific) in a PTFE liquid-cell and pre-heating the substrate to 60 °C using a sample heating stage positioned beneath the substrate. The liquid cell was then filled with 50 μL of the desired solution while the substrate was maintained at 60 °C. After 10 minutes of heating at 60°C, the samples were allowed to cool naturally before being loaded in to the STM for imaging. STM experiments were carried out directly at the liquid-solid interface between the TCB/THF mixture and HOPG using a 5500 series SPM system (Keysight Technologies) operating under ambient conditions and at room temperature. The STM tips were mechanically cut from Pt/Ir (80:20) wire (Advent Research Materials Ltd). All images were obtained in constant current mode and have had a flattening procedure applied to them. Drift correction of high resolution STM images was carried out using the Scanning Probe Image Processor (SPIP) software (Image Metrology ApS, Lyngby, Denmark). This process involved the collection of a high resolution image of a molecular structure immediately followed by collecting an image of the underlying HOPG lattice. The image of the HOPG lattice was collected using identical scanning parameters apart from the tunnel current and bias voltage. By assuming that the level of drift is constant for both images the known lattice dimensions of HOPG can be used to produce a set of correction parameters for the HOPG image. These parameters can then be applied to the high resolution molecular image to produce a drift corrected image. All unit cell dimensions and angles are obtained by taking the mean of at least three measurements taken from separate drift corrected STM images. The stated error values for these measurements are the standard deviation of the various measurements.

5.8.3 Molecular Mechanics (MM) Studies

MM simulations of the Zn-tetra-TP, tetra-AP and Zn-tetra-TP/tetra-AP surface structures were carried out using the HyperChem software package. A single layer of graphite was fixed in place and used as a substrate. On top of this layer were placed the molecules representing an individual unit cell of the different 2D structures. The starting positions for the molecules were derived using drift-corrected STM images as a guide and maintaining a 0.35 nm vertical distance between the planar porphyrin cores of the molecules and the underlying graphite layer. Once positioned, these structures were geometry optimised using the MM+ force field. The optimisation process was terminated when

a gradient ≤ 0.01 kcal Å^{-1} mol^{-1} was reached. All unit cell dimensions and angles are taken from geometry optimised structures.

Bibliography

[1] A. Gooch, A. M. McGhee, L. C. Renton, J. P. Plante, C. I. Lindsay, and A. J. Wilson, "Design, synthesis and binding properties of conformer-independent linear ada hydrogen-bonding arrays," *Supramolecular Chemistry*, vol. 21, no. 1-2, pp. 12–17, 2009.

[2] A. Porcheddu, G. Giacomelli, I. Piredda, M. Carta, and G. Nieddu, "A practical and efficient approach to pna monomers compatible with fmoc-mediated solid-phase synthesis protocols," *European Journal of Organic Chemistry*, vol. 2008, no. 34, pp. 5786–5797, 2008.

[3] S. Dey and P. Garner, "Synthesis of tert-butoxycarbonyl (boc)-protected purines," *The Journal of Organic Chemistry*, vol. 65, no. 22, pp. 7697–7699, 2000.

[4] M. Frieden, M. Giraud, C. B. Reese, and Q. Song, "Synthesis of 1-[cis-3-(hydroxymethyl)cyclobutyl]-uracil, -thymine and -cytosine," *Journal of the Chemical Society, Perkin Transactions 1*, no. 17, pp. 2827–2832, 1998.

[5] M. Guney, H. Cavdar, M. Senturk, and D. Ekinci, "Synthesis and carbonic anhydrase inhibitory properties of novel uracil derivatives," *Bioorg Med Chem Lett*, vol. 25, no. 16, pp. 3261–3, 2015.

[6] N. Bilbao, V. Vzquez-Gonzlez, M. T. Aranda, and D. Gonzlez-Rodrguez, "Synthesis of 5-/8-halogenated or ethynylated lipophilic nucleobases as potential synthetic intermediates for supramolecular chemistry," *European Journal of Organic Chemistry*, vol. 2015, no. 32, pp. 7160–7175, 2015.

[7] R. T. Mahajan, E. M. Ytre-Arne, P. Strom-Andersen, B. Dalhus, and L.-L. Gundersen, "Synthetic routes to n-9 alkylated 8-oxoguanines; weak inhibitors of the human dna glycosylase ogg1," *Molecules*, vol. 20, no. 9, 2015.

[8] T. R. Mahajan and L.-L. Gundersen, "Efficient functionalization of 2-amino-6-chloropurine derivatives at c-8 via 8-lithiated species," *Tetrahedron Letters*, vol. 56, no. 43, pp. 5899–5902, 2015.

[9] A. Marsh, M. Silvestri, and J.-M. Lehn, "Self-complementary hydrogen bonding heterocycles designed for the enforced self-assembly into supramolecular macrocycles," *Chemical Communications*, no. 13, pp. 1527–1528, 1996.

[10] J.-M. Lehn, *From Molecular to Supramolecular Chemistry*, pp. 1–9. Wiley-VCH Verlag GmbH and Co. KGaA, 2006.

[11] H.-J. Schneider, "Binding mechanisms in supramolecular complexes," *Angewandte Chemie International Edition*, vol. 48, no. 22, pp. 3924–3977, 2009.

[12] F. Biedermann and H.-J. Schneider, "Experimental binding energies in supramolecular complexes," *Chemical Reviews*, vol. 116, no. 9, pp. 5216–5300, 2016.

[13] D. N. Bunck and W. R. Dichtel, "Internal functionalization of three-dimensional covalent organic frameworks," *Angewandte Chemie International Edition*, vol. 51, no. 8, pp. 1885–1889, 2012.

[14] N. Kornienko, Y. Zhao, C. S. Kley, C. Zhu, D. Kim, S. Lin, C. J. Chang, O. M. Yaghi, and P. Yang, "Metalorganic frameworks for electrocatalytic reduction of carbon dioxide," *Journal of the American Chemical Society*, vol. 137, no. 44, pp. 14129–14135, 2015.

[15] J. Waals and J. Rowlinson, *J.D. van der Waals: on the continuity of the gaseous and liquid states*. North-Holland, 1988.

[16] C. J. Van Oss, D. R. Absolom, and A. W. Neumann, "Applications of net repulsive van der waals forces between different particles, macromolecules, or biological cells in liquids," *Colloids and Surfaces*, vol. 1, no. 1, pp. 45–56, 1980.

[17] M. Satake, *Chemical Bonding*. Discovery Publishing House Pvt. Limited, 2010.

[18] C. A. Hunter and J. K. M. Sanders, "The nature of .pi.-.pi. interactions," *Journal of the American Chemical Society*, vol. 112, no. 14, pp. 5525–5534, 1990.

[19] P. R. Ashton, T. T. Goodnow, A. E. Kaifer, M. V. Reddington, A. M. Z. Slawin, N. Spencer, J. F. Stoddart, C. Vicent, and D. J. Williams, "A [2] catenane made to order," *Angewandte Chemie International Edition in English*, vol. 28, no. 10, pp. 1396–1399, 1989.

[20] J. C. Ma and D. A. Dougherty, "The cation interaction," *Chemical Reviews*, vol. 97, no. 5, pp. 1303–1324, 1997.

[21] J. C. Amicangelo and P. B. Armentrout, "Absolute binding energies of alkali-metal cation complexes with benzene determined by threshold collision-induced dissociation experiments and ab initio theory," *The Journal of Physical Chemistry A*, vol. 104, no. 48, pp. 11420–11432, 2000.

[22] J. P. Gallivan and D. A. Dougherty, "A computational study of cation interactions vs salt bridges in aqueous media: implications for protein engineering," *Journal of the American Chemical Society*, vol. 122, no. 5, pp. 870–874, 2000.

[23] S. Mecozzi, A. P. West, and D. A. Dougherty, "Cation interactions in simple aromatics: electrostatics provide a predictive tool," *Journal of the American Chemical Society*, vol. 118, no. 9, pp. 2307–2308, 1996.

[24] A. Sygula, F. R. Fronczek, R. Sygula, P. W. Rabideau, and M. M. Olmstead, "A double concave hydrocarbon buckycatcher," *Journal of the American Chemical Society*, vol. 129, no. 13, pp. 3842–3843, 2007.

[25] C. F. Matta, N. Castillo, and R. J. Boyd, "Extended weak bonding interactions in dna: -stacking (basebase), basebackbone, and backbonebackbone interactions," *The Journal of Physical Chemistry B*, vol. 110, no. 1, pp. 563–578, 2006.

[26] C. R. Martinez and B. L. Iverson, "Rethinking the term "pi-stacking"," *Chemical Science*, vol. 3, no. 7, pp. 2191–2201, 2012.

[27] R. P. Matthews, T. Welton, and P. A. Hunt, "Competitive pi interactions and hydrogen bonding within imidazolium ionic liquids," *Physical Chemistry Chemical Physics*, vol. 16, no. 7, pp. 3238–3253, 2014.

[28] E. Arunan, R. Desiraju Gautam, A. Klein Roger, J. Sadlej, S. Scheiner, I. Alkorta, C. Clary David, H. Crabtree Robert, J. Dannenberg Joseph, P. Hobza, G. Kjaergaard Henrik, C. Legon Anthony, B. Mennucci, and J. Nesbitt David, "Definition of the hydrogen bond (iupac recommendations 2011)," *Pure and Applied Chemistry*, vol. 83, no. 8, p. 1637, 2011.

[29] C. Hunter and H. Anderson, "What is cooperativity?," *Angewandte Chemie International Edition*, vol. 48, no. 41, pp. 7488–7499, 2009.

[30] J. D. Badji, A. Nelson, S. J. Cantrill, W. B. Turnbull, and J. F. Stoddart, "Multivalency and cooperativity in supramolecular chemistry," *Accounts of Chemical Research*, vol. 38, no. 9, pp. 723–732, 2005.

[31] C. Bohr, K. Hasselbalch, and A. Krogh, "Ueber einen in biologischer beziehung wichtigen einfluss, den die kohlensurespannung des blutes auf dessen sauerstoffbindung bt1," *Skandinavisches Archiv Fr Physiologie*, vol. 16, no. 2, pp. 402–412, 1904.

[32] M. F. Perutz, "Mechanisms of cooperativity and allosteric regulation in proteins," *Quarterly Reviews of Biophysics*, vol. 22, pp. 139–237, 1989.

[33] M. F. Perutz, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will, and A. C. T. North, "Structure of hmoglobin: A three-dimensional fourier synthesis at 5.5-. resolution, obtained by x-ray analysis," *Nature*, vol. 185, p. 416, 1960.

[34] A. Whitty, "Cooperativity and biological complexity," *Nature Chemical Biology*, vol. 4, p. 435, 2008.

[35] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[36] H. K. E. Landenmark, D. H. Forgan, and C. S. Cockell, "An estimate of the total dna in the biosphere," *PLOS Biology*, vol. 13, no. 6, p. e1002168, 2015.

[37] J. D. Watson and F. H. C. Crick, "Genetical implications of the structure of deoxyribonucleic acid," *Nature*, vol. 171, no. 4361, pp. 964–967, 1953.

[38] W. Saenger, *Principles of Nucleic Acid Structure*. Springer, New York, NY, 1984.

[39] K. Hoogsteen, "The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine," *Acta Crystallographica*, vol. 16, no. 9, pp. 907–916, 1963.

[40] V. Singh, B. I. Fedeles, and J. M. Essigmann, "Role of tautomerism in rna biochemistry," *RNA*, vol. 21, no. 1, pp. 1–13, 2015.

[41] L. Antonov, *Tautomerism: Methods and Theories*. John Wiley & Sons, 2013.

[42] M. Dreyfus, G. Dodin, O. Bensaude, and J. E. Dubois, "Tautomerism of purines. 2. amino-imino tautomerism in 1-alkyladenines," *Journal of the American Chemical Society*, vol. 99, no. 21, pp. 7027–7037, 1977.

[43] O. K. Abou-Zied, R. Jimenez, and F. E. Romesberg, "Tautomerization dynamics of a model base pair in dna," *Journal of the American Chemical Society*, vol. 123, no. 19, pp. 4613–4614, 2001.

[44] J. D. Watson, *The Double Helix*. Weidenfeld & Nicolson, 1968.

[45] L. Liu, D. Xia, L. Klausen, and M. Dong, "The self-assembled behavior of dna bases on the interface," *International Journal of Molecular Sciences*, vol. 15, no. 2, p. 1901, 2014.

[46] J. Cortese, C. Soulie-Ziakovic, and L. Leibler, "Binding and supramolecular organization of homo- and heterotelechelic oligomers in solutions," *Polymer Chemistry*, vol. 5, no. 1, pp. 116–125, 2014.

[47] M. J. Mayoral, N. Bilbao, and D. Gonzlez-Rodrguez, "Hydrogen-bonded macrocyclic supramolecular systems in solution and on surfaces," *ChemistryOpen*, vol. 5, no. 1, pp. 10–32, 2016.

[48] J. L. Sessler and R. Wang, "Self-assembly of an artificial dinucleotide duplex," *Journal of the American Chemical Society*, vol. 118, no. 40, pp. 9808–9809, 1996.

[49] A. Asadi, B. O. Patrick, and D. M. Perrin, "Gc quartet a dna-inspired janus-gc heterocycle: Synthesis, structural analysis, and self-organization," *Journal of the American Chemical Society*, vol. 130, no. 39, pp. 12860–12861, 2008.

[50] M. Mascal, M. Hext Nicholas, R. Warmuth, H. Moore Madeleine, and P. Turkenburg Johan, "Programming a hydrogenbonding code for the specific generation of a supermacrocycle," *Angewandte Chemie International Edition in English*, vol. 35, no. 19, pp. 2204–2206, 1996.

[51] D. Gonzlez-Rodrguez, J. L. J. van Dongen, M. Lutz, A. L. Spek, A. P. H. J. Schenning, and E. W. Meijer, "G-quadruplex self-assembly regulated by coulombic interactions," *Nature Chemistry*, vol. 1, p. 151, 2009.

[52] S. Romero-Prez, J. Camacho-Garca, C. Montoro-Garca, A. M. Lpez-Prez, A. Sanz, M. J. Mayoral, and D. Gonzlez-Rodrguez, "G-arylated hydrogen-bonded cyclic tetramer assemblies with remarkable thermodynamic and kinetic stability," *Organic Letters*, vol. 17, no. 11, pp. 2664–2667, 2015.

[53] C. Montoro-Garcia, J. Camacho-Garcia, A. M. Lopez-Perez, M. J. Mayoral, N. Bilbao, and D. Gonzalez-Rodriguez, "Role of the symmetry of multipoint hydrogen bonding on chelate cooperativity in supramolecular macrocyclization processes," *Angewandte Chemie International Edition*, vol. 55, no. 1, pp. 223–227, 2016.

[54] A. G. Slater, P. H. Beton, and N. R. Champness, "Two-dimensional supramolecular chemistry on surfaces," *Chemical Science*, vol. 2, no. 8, pp. 1440–1448, 2011. (nee Phillips).

[55] S. De Feyter, A. Gesquire, M. M. Abdel-Mottaleb, P. C. M. Grim, F. C. De Schryver, C. Meiners, M. Sieffert, S. Valiyaveettil, and K. Mllen, "Scanning tunneling microscopy: a unique tool in the study of chirality, dynamics, and reactivity in physisorbed organic monolayers," *Accounts of Chemical Research*, vol. 33, no. 8, pp. 520–531, 2000.

[56] K. Tahara, S. Furukawa, H. Uji-i, T. Uchino, T. Ichikawa, J. Zhang, W. Mamdouh, M. Sonoda, F. C. De Schryver, S. De Feyter, and Y. Tobe, "Two-dimensional porous molecular networks of dehydrobenzo[12]annulene derivatives via alkyl chain interdigitation," *Journal of the American Chemical Society*, vol. 128, no. 51, pp. 16613–16625, 2006.

[57] Y. Li, J. Xiao, T. E. Shubina, M. Chen, Z. Shi, M. Schmid, H.-P. Steinrck, J. M. Gottfried, and N. Lin, "Coordination and metalation bifunctionality of cu with 5,10,15,20-tetra(4-pyridyl)porphyrin: Toward a mixed-valence two-dimensional coordination network," *Journal of the American Chemical Society*, vol. 134, no. 14, pp. 6401–6408, 2012.

[58] T. Lin, G. Kuang, X. S. Shang, P. N. Liu, and N. Lin, "Self-assembly of metal-organic coordination networks using on-surface synthesized ligands," *Chemical Communications*, vol. 50, no. 97, pp. 15327–15329, 2014.

[59] D. Khne, F. Klappenberger, R. Decker, U. Schlickum, H. Brune, S. Klyatskaya, M. Ruben, and J. V. Barth, "High-quality 2d metalorganic coordination network providing giant cavities within mesoscale domains," *Journal of the American Chemical Society*, vol. 131, no. 11, pp. 3881–3883, 2009.

[60] F. Studener, K. Mller, N. Marets, V. Bulach, M. W. Hosseini, and M. Sthr, "From hydrogen bonding to metal coordination and back: Porphyrin-based networks on ag(111)," *The Journal of Chemical Physics*, vol. 142, no. 10, p. 101926, 2015.

[61] J. Bjork, M. Matena, M. S. Dyer, M. Enache, J. Lobo-Checa, L. H. Gade, T. A. Jung, M. Stohr, and M. Persson, "Stm fingerprint of molecule-adatom interactions in a self-assembled metal-organic surface coordination network on cu(111)," *Physical Chemistry Chemical Physics*, vol. 12, no. 31, pp. 8815–8821, 2010.

[62] N. Lin, S. Stepanow, M. Ruben, and J. V. Barth, "Surface-confined supramolecular coordination chemistry," *Top Curr Chem*, vol. 287, pp. 1–44, 2009.

[63] S. Stepanow, N. Lin, J. V. Barth, and K. Kern, "Surface-template assembly of two-dimensional metalorganic coordination networks," *The Journal of Physical Chemistry B*, vol. 110, no. 46, pp. 23472–23477, 2006.

[64] M. Lackinger and W. M. Heckl, "Carboxylic acids: Versatile building blocks and mediators for two-dimensional supramolecular self-assembly," *Langmuir*, vol. 25, no. 19, pp. 11307–11321, 2009.

[65] S. Griessl, M. Lackinger, M. Edelwirth, M. Hietschold, and W. M. Heckl, "Selfassembled twodimensional molecular hostguest architectures from trimesic acid," *Single Molecules*, vol. 3, no. 1, pp. 25–31, 2002.

[66] M. O. Blunt, J. C. Russell, M. d. C. Gimnez-Lpez, J. P. Garrahan, X. Lin, M. Schrder, N. R. Champness, and P. H. Beton, "Random tiling and topological defects in a two-dimensional molecular network," *Science*, vol. 322, no. 5904, pp. 1077–1081, 2008.

[67] J. Xu, C. He, K. C. Toh, and X. Lu, "Intermolecular interaction in multicomponent supramolecular complexes through hydrogen-bonding association," *Macromolecules*, vol. 35, no. 23, pp. 8846–8851, 2002.

[68] E. R. Draper and D. J. Adams, "How should multicomponent supramolecular gels be characterised?," *Chemical Society Reviews*, vol. 47, no. 10, pp. 3395–3405, 2018.

[69] S. David, S. Natascha, U. Moritz, Y. Hajime, S. Daniel, S. Dieter, S. Edgar, and B. Pol, "Mannosedecorated multicomponent supramolecular polymers trigger effective uptake into antigenpresenting cells," *ChemBioChem*, vol. 19, no. 9, pp. 912–916, 2018.

[70] J. A. Theobald, N. S. Oxtoby, M. A. Phillips, N. R. Champness, and P. H. Beton, "Controlling molecular deposition and layer structure with supramolecular surface assemblies," *Nature*, vol. 424, no. 6952, pp. 1029–1031, 2003. 10.1038/nature01915.

[71] L. M. A. Perdigo, E. W. Perkins, J. Ma, P. A. Staniec, B. L. Rogers, N. R. Champness, and P. H. Beton, "Bimolecular networks and supramolecular traps on au(111)," *The Journal of Physical Chemistry B*, vol. 110, no. 25, pp. 12539–12542, 2006.

[72] A. G. Slater, L. M. A. Perdigo, P. H. Beton, and N. R. Champness, "Surface-based supramolecular chemistry using hydrogen bonds," *Accounts of Chemical Research*, vol. 47, no. 12, pp. 3417–3427, 2014.

[73] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, "Design and self-assembly of two-dimensional dna crystals," *Nature*, vol. 394, p. 539, 1998.

[74] M. Nikan and C. Sherman John, "Templateassembled synthetic gquartets (tasqs)," *Angewandte Chemie International Edition*, vol. 47, no. 26, pp. 4900–4902, 2008.

[75] R. Otero, M. Lukas, R. E. A. Kelly, W. Xu, E. Lgsgaard, I. Stensgaard, L. N. Kantorovich, and F. Besenbacher, "Elementary structural motifs in a random network of cytosine adsorbed on a gold(111) surface," *Science*, vol. 319, no. 5861, pp. 312–315, 2008.

[76] R. Otero, W. Xu, M. Lukas, R. E. Kelly, E. Lgsgaard, I. Stensgaard, J. Kjems, L. N. Kantorovich, and F. Besenbacher, "Specificity of watsoncrick base pairing on a solid surface studied at the atomic scale," *Angewandte Chemie*, vol. 120, no. 50, pp. 9819–9822, 2008.

[77] R. Otero, M. Schck, M. Molina Luis, E. Lgsgaard, I. Stensgaard, B. Hammer, and F. Besenbacher, "Guanine quartet networks stabilized by cooperative hydrogen bonds," *Angewandte Chemie International Edition*, vol. 44, no. 15, pp. 2270–2275, 2005.

[78] A. Ciesielski, S. Lena, S. Masiero, G. Spada, and P. Samor, "Dynamers at the solidliquid interface: Controlling the reversible assembly/reassembly process between two highly ordered supramolecular guanine motifs," *Angewandte Chemie International Edition*, vol. 49, no. 11, pp. 1963–1966, 2010.

[79] A. Ciesielski, R. Perone, S. Pieraccini, G. P. Spada, and P. Samori, "Nanopatterning the surface with ordered supramolecular architectures of n9-alkylated guanines: Stm reveals," *Chemical Communications*, vol. 46, no. 25, pp. 4493–4495, 2010.

[80] W. Xu, J. Wang, M. F. Jacobsen, M. Mura, M. Yu, R. E. A. Kelly, Q. Meng, E. Lgsgaard, I. Stensgaard, T. R. Linderoth, J. Kjems, L. N. Kantorovich, K. V. Gothelf, and F. Besenbacher, "Supramolecular porous network formed by molecular recognition between chemically modified nucleobases guanine and cytosine," *Angewandte Chemie International Edition*, vol. 49, no. 49, pp. 9373–9377, 2010.

[81] W. Mamdouh, R. E. A. Kelly, M. Dong, L. N. Kantorovich, and F. Besenbacher, "Two-dimensional supramolecular nanopatterns formed by the coadsorption of guanine and uracil at the liquid/solid interface," *Journal of the American Chemical Society*, vol. 130, no. 2, pp. 695–702, 2008.

[82] W. Mamdouh, M. Dong, S. Xu, E. Rauls, and F. Besenbacher, "Supramolecular nanopatterns self-assembled by adeninethymine quartets at the liquid/solid interface," *Journal of the American Chemical Society*, vol. 128, no. 40, pp. 13305–13311, 2006.

[83] A. G. Slater, Y. Hu, L. Yang, S. P. Argent, W. Lewis, M. O. Blunt, and N. R. Champness, "Thymine functionalised porphyrins, synthesis and heteromolecular surface-based self-assembly," *Chemical Science*, vol. 6, no. 2, pp. 1562–1569, 2015.

[84] H. Binnig, G.; Rohrer, "Scanning tunneling microscopy," *IBM Journal of Research and Development*, vol. 30, no. 4, p. 35569, 1986.

[85] A. Engel, "Introduction to scanning tunneling microscopy by c. julian chen oxford university press, new york (1993) isbn 0-19-507150-6," *Scanning*, vol. 16, no. 2, pp. 123–124, 1994.

[86] T. Cuff, *The STM (Scanning Tunneling Microscope) [The forgotten contribution of Robert Francis Earhart to the discovery of quantum tunneling.]*. 2016.

[87] P. S. Weiss, "Scanning probe microscopy and spectroscopy. theory, techniques, and applications. 2nd edition edited by d. w. bonnell (university of pennsylvania). wiley-vch: new york. 2001. xiv + 494. isbn 0-471-24824-x," *Journal of the American Chemical Society*, vol. 123, no. 39, pp. 9725–9725, 2001.

[88] J. Hudson, *Surface Science: An introduction*. Elsevier Science, 2013.

[89] S. De Feyter and F. C. De Schryver, "Two-dimensional supramolecular self-assembly probed by scanning tunneling microscopy," *Chemical Society Reviews*, vol. 32, no. 3, pp. 139–150, 2003.

[90] J. M. Gottfried, "Surface chemistry of porphyrins and phthalocyanines," *Surface Science Reports*, vol. 70, no. 3, pp. 259–379, 2015.

[91] G. Rojas, S. Simpson, X. Chen, D. A. Kunkel, J. Nitz, J. Xiao, P. A. Dowben, E. Zurek, and A. Enders, "Surface state engineering of moleculemolecule interactions," *Physical Chemistry Chemical Physics*, vol. 14, no. 14, pp. 4971–4976, 2012.

[92] K. M. Kadish and G. Larson, "A study of the redox potentials and electron transfer rates of several naturally occurring and synthetic iron porphyrins in dmf," *Bioinorganic Chemistry*, vol. 7, no. 2, pp. 95–105, 1977.

[93] P. Worthington, P. Hambright, R. F. X. Williams, J. Reid, C. Burnham, A. Shamim, J. Turay, D. M. Bell, R. Kirkland, R. G. Little, N. Datta-Gupta, and U. Eisner, "Reduction potentials of seventy-five free base porphyrin molecules: Reactivity correlations and the prediction of potentials," *Journal of Inorganic Biochemistry*, vol. 12, no. 4, pp. 281–291, 1980.

[94] A. Antipas, D. Dolphin, M. Gouterman, and E. C. Johnson, "Porphyrins. 38. redox potentials, charge transfer transitions, and emission of copper, silver, and gold complexes," *Journal of the American Chemical Society*, vol. 100, no. 24, pp. 7705–7709, 1978.

[95] D. Sun, F. S. Tham, C. A. Reed, L. Chaker, M. Burgess, and P. D. W. Boyd, "Porphyrin-fullerene hostguest chemistry," *Journal of the American Chemical Society*, vol. 122, no. 43, pp. 10704–10705, 2000.

[96] X.-H. Liu, J.-Y. Yue, Y.-P. Mo, Y. Yao, C. Zeng, T. Chen, H.-J. Yan, Z.-H. Wang, D. Wang, and L.-J. Wan, "Surface hostguest supramolecular assemblies on porphyrin-based covalent organic grids," *The Journal of Physical Chemistry C*, vol. 120, no. 29, pp. 15753–15757, 2016.

[97] W. Zheng, N. Shan, L. Yu, and X. Wang, "Uvvisible, fluorescence and epr properties of porphyrins and metalloporphyrins," *Dyes and Pigments*, vol. 77, no. 1, pp. 153–157, 2008.

[98] L.-L. Li and E. W.-G. Diau, "Porphyrin-sensitized solar cells," *Chemical Society Reviews*, vol. 42, no. 1, pp. 291–304, 2013.

[99] Z. Valicsek and O. Horváth, "Application of the electronic spectra of porphyrins for analytical purposes: The effects of metal ions and structural distortions," *Microchemical Journal*, vol. 107, pp. 47–62, 2013.

[100] K. Rybicka-Jasiska, W. Shan, K. Zawada, K. M. Kadish, and D. Gryko, "Porphyrins as photoredox catalysts: Experimental and theoretical studies," *Journal of the American Chemical Society*, vol. 138, no. 47, pp. 15451–15458, 2016.

[101] M. A. Spackman and P. G. Byrom, "A novel definition of a molecule in a crystal," *Chemical Physics Letters*, vol. 267, no. 3, pp. 215–220, 1997.

[102] J. J. McKinnon, M. A. Spackman, and A. S. Mitchell, "Novel tools for visualizing and exploring intermolecular interactions in molecular crystals," *Acta Crystallographica Section B*, vol. 60, pp. 627–668, Dec 2004.

[103] J. J. McKinnon, D. Jayatilaka, and M. A. Spackman, "Towards quantitative analysis of intermolecular interactions with hirshfeld surfaces," *Chem. Commun.*, pp. 3814–3816, 2007.

[104] M. Tamami, S. T. Hemp, K. Zhang, M. Zhang, R. B. Moore, and T. E. Long, "Poly(ethylene glycol)-based ammonium ionenes containing nucleobases," *Polymer*, vol. 54, no. 6, pp. 1588–1595, 2013.

[105] C. Lambertucci, M. Buccioni, D. Dal Ben, S. Kachler, G. Marucci, A. Spinaci, A. Thomas, K. N. Klotz, and R. Volpini, "New substituted 9-propyladenine derivatives as a2a adenosine receptor antagonists," *MedChemComm*, vol. 6, no. 5, pp. 963–970, 2015.

[106] T. Borrmann, A. Abdelrahman, R. Volpini, C. Lambertucci, E. Alksnis, S. Gorzalka, M. Knospe, A. C. Schiedel, G. Cristalli, and C. E. Mller, "Structureactivity relationships of adenine and deazaadenine derivatives as ligands for adenine receptors, a new purinergic receptor family," *Journal of Medicinal Chemistry*, vol. 52, no. 19, pp. 5974–5989, 2009.

[107] Y. Basel and A. Hassner, "Di-tert-butyl dicarbonate and 4-(dimethylamino)pyridine revisited. their reactions with amines and alcohols1," *The Journal of Organic Chemistry*, vol. 65, no. 20, pp. 6368–6380, 2000.

[108] J. C. Kim, M. Kim, J. Jung, J. Lee, B. J. Ree, H. Kim, I. J. Kim, J. R. Kim, and M. Ree, "Synthesis, physicochemical characteristics, and biocompatibility of self-assemble polymers bearing guanine, cytosine, uracil, and thymine moieties," *Journal of Polymer Science Part A: Polymer Chemistry*, vol. 53, no. 9, pp. 1151–1160, 2015.

[109] J. Clayden, *Organic Chemistry*. Oxford University Press, 2001.

[110] A. P. Bisson, C. A. Hunter, J. C. Morales, and K. Young, "Cooperative interactions in a ternary mixture," *Chemistry A European Journal*, vol. 4, no. 5, pp. 845–851, 1998.

[111] G. M. Sheldrick, "*SHELXT* – integrated space-group and crystal-structure determination," *Acta Crystallographica Section A*, vol. 71, pp. 3–8, Jan 2015.

[112] G. M. Sheldrick, "Crystal structure refinement with *shelxl*," *Acta Crystallographica Section C*, vol. 71, pp. 3–8, Jan 2015.

[113] O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard, and H. Puschmann, "*OLEX2*: a complete structure solution, refinement and analysis program," *Journal of Applied Crystallography*, vol. 42, pp. 339–341, Apr 2009.

[114] J. S. Lindsey, "Self-assembly in synthetic routes to molecular devices - biological principles and chemical perspectives - a review," *New Journal of Chemistry*, vol. 15, no. 2-3, pp. 153–180, 1991. Times Cited: 809 Lindsey, Jonathan/J-7761-2012 814.

[115] P. Rothemund, "Formation of porphyrins from pyrrole and aldehydes," *Journal of the American Chemical Society*, vol. 57, no. 10, pp. 2010–2011, 1935.

[116] P. Rothemund, "A new porphyrin synthesis. the synthesis of porphin," *Journal of the American Chemical Society*, vol. 58, pp. 625–627, 1936. Times Cited: 160 161.

[117] A. D. Adler, F. R. Longo, J. D. Finarelli, J. Goldmacher, J. Assour, and L. Korsakoff, "A simplified synthesis for meso-tetraphenylporphine," *The Journal of Organic Chemistry*, vol. 32, no. 2, pp. 476–476, 1967.

[118] C.-H. Lee, F. Li, K. Iwamoto, J. Dadok, A. A. Bothner-By, and J. S. Lindsey, "Synthetic approaches to regioisomerically pure porphyrins bearing four different meso-substituents," *Tetrahedron*, vol. 51, no. 43, pp. 11645–11672, 1995.

[119] I. Bouamaied, L. A. Fendt, D. Haussinger, M. Wiesner, S. Thoni, N. Amiot, and E. Stulz, "Porphyrin-dna: A supramolecular scaffold for functional molecules on the nanometre scale," *Nucleosides Nucleotides and Nucleic Acids*, vol. 26, no. 10-12, pp. 1533–1538, 2007.

[120] M. F. Jacobsen, M. M. Knudsen, and K. V. Gothelf, "Efficient n-arylation and n-alkenylation of the five dna/rna nucleobases," *The Journal of Organic Chemistry*, vol. 71, no. 24, pp. 9183–9190, 2006.

[121] M. F. Jacobsen, C. S. Andersen, M. M. Knudsen, and K. V. Gothelf, "Synthesis of rigid homo- and heteroditopic nucleobase-terminated molecules incorporating adenine and/or thymine," *Organic Letters*, vol. 9, no. 15, pp. 2851–2854, 2007.

[122] P. E. Fanta, "The ullmann synthesis of biaryls," *Synthesis*, vol. 1974, no. 01, pp. 9–21, 1974.

[123] F. Ullmann and J. Bielecki, "Ueber synthesen in der biphenylreihe," *Berichte der deutschen chemischen Gesellschaft*, vol. 34, no. 2, pp. 2174–2185, 1901.

[124] S. V. Ley and A. W. Thomas, "Modern synthetic methods for copper-mediated c(aryl)o, c(aryl)n, and c(aryl)s bond formation," *Angewandte Chemie International Edition*, vol. 42, no. 44, pp. 5400–5449, 2003.

[125] *Syntheses from Pyrimidines*, ch. II, pp. 31–90. Wiley-Blackwell, 2008.

[126] A. S. Shawali, "Reactions of heterocyclic compounds with nitrilimines and their precursors," *Chemical Reviews*, vol. 93, no. 8, pp. 2731–2777, 1993.

[127] J. P. Garca-Tern, O. Castillo, A. Luque, U. Garca-Couceiro, P. Romn, and F. Lloret, "One-dimensional oxalato-bridged cu(ii), co(ii), and zn(ii) complexes with purine and adenine as terminal ligands," *Inorganic Chemistry*, vol. 43, no. 18, pp. 5761–5770, 2004.

[128] M. Mizutani, Y. Sanemitsu, Y. Tamaru, and Z.-I. Yoshida, "Palladium-catalyzed polyhetero-claisen rearrangement of 4-allylthiopyrimidin-2(1h)-ones and 5-allylthio-1,2,4-triazin-3(2h)-ones," *Tetrahedron*, vol. 41, no. 22, pp. 5289–5293, 1985.

[129] M. Legraverend, "Recent advances in the synthesis of purine derivatives and their precursors," *Tetrahedron*, vol. 64, no. 37, pp. 8585–8603, 2008.

[130] L. Lercher, J. F. McGouran, B. M. Kessler, C. J. Schofield, and B. G. Davis, "Dna modification under mild conditions by suzukimiyaura cross-coupling for the generation of functional probes," *Angewandte Chemie International Edition*, vol. 52, no. 40, pp. 10553–10558, 2013.

[131] N. Miyaura, K. Yamada, and A. Suzuki, "A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides," *Tetrahedron Letters*, vol. 20, no. 36, pp. 3437–3440, 1979.

[132] N. Miyaura and A. Suzuki, "Stereoselective synthesis of arylated (e)-alkenes by the reaction of alk-1-enylboranes with aryl halides in the presence of palladium catalyst," *Journal of the Chemical Society, Chemical Communications*, no. 19, pp. 866–867, 1979.

[133] T. Ishiyama, M. Murata, and N. Miyaura, "Palladium(o)-catalyzed cross-coupling reaction of alkoxydiboron with haloarenes - a direct procedure for arylboronic esters," *Journal of Organic Chemistry*, vol. 60, no. 23, pp. 7508–7510, 1995. Times Cited: 767 Murata, Miki/E-7630-2010 Murata, Miki/0000-0002-6903-2268 769.

[134] K. Sonogashira, "Development of pdcu catalyzed cross-coupling of terminal acetylenes with sp2-carbon halides," *Journal of Organometallic Chemistry*, vol. 653, no. 1, pp. 46–49, 2002.

[135] I. Bouamaied, T. Nguyen, T. Ruhl, and E. Stulz, "Supramolecular helical porphyrin arrays using dna as a scaffold," *Organic & Biomolecular Chemistry*, vol. 6, no. 21, pp. 3888–3891, 2008.

[136] R. F. Heck and J. P. Nolley, "Palladium-catalyzed vinylic hydrogen substitution reactions with aryl, benzyl, and styryl halides," *The Journal of Organic Chemistry*, vol. 37, no. 14, pp. 2320–2322, 1972.

[137] Y. Liang and F. S. Wnuk, "Modification of purine and pyrimidine nucleosides by direct c-h bond activation," *Molecules*, vol. 20, no. 3, 2015.

[138] C. Ebner, C. A. Mller, C. Markert, and A. Pfaltz, "Determining the enantioselectivity of chiral catalysts by mass spectrometric screening of their racemic forms," *Journal of the American Chemical Society*, vol. 133, no. 13, pp. 4710–4713, 2011.

[139] R. K. Totten, P. Ryan, B. Kang, S. J. Lee, L. J. Broadbelt, R. Q. Snurr, J. T. Hupp, and S. T. Nguyen, "Enhanced catalytic decomposition of a phosphate triester by modularly accessible bimetallic porphyrin dyads and dimers," *Chemical Communications*, vol. 48, no. 35, pp. 4178–4180, 2012.

[140] J. K. Laha, S. Dhanalekshmi, M. Taniguchi, A. Ambroise, and J. S. Lindsey, "A scalable synthesis of meso-substituted dipyrromethanes," *Organic Process Research and Development*, vol. 7, no. 6, pp. 799–812, 2003.

[141] I. Katsumi, H. Kondo, K. Yamashita, T. Hidaka, K. Hosoe, T. Yamashita, and K. Watanabe, "Studies on styrene derivatives. i. : Synthesis and antiinflammatory activities of alpha;-benzylidene-gamma;-butyrolactone derivatives," *CHEMICAL & PHARMACEUTICAL BULLETIN*, vol. 34, no. 1, pp. 121–129, 1986.

[142] E. Valeur and M. Bradley, "Amide bond formation: beyond the myth of coupling reagents," *Chemical Society Reviews*, vol. 38, no. 2, pp. 606–631, 2009.

[143] J. R. Dunetz, J. Magano, and G. A. Weisenburger, "Large-scale applications of amide coupling reagents for the synthesis of pharmaceuticals," *Organic Process Research and Development*, vol. 20, no. 2, pp. 140–177, 2016.

[144] A. Vilsmeier and A. Haack, "Über die einwirkung von halogenphosphor auf alkyl-formanilide. eine neue methode zur darstellung sekundärer und tertiärer p-alkylamino-benzaldehyde," *Berichte der deutschen chemischen Gesellschaft (A and B Series)*, vol. 60, no. 1, pp. 119–122, 1927.

[145] O. Meth-Cohn and S. P. Stanforth, *3.5 - The VilsmeierHaack Reaction*, pp. 777–794. Oxford: Pergamon, 1991.

[146] J. S. Lindsey, "Synthetic routes to meso-patterned porphyrins," *Accounts of Chemical Research*, vol. 43, no. 2, pp. 300–311, 2010.

[147] S. Wan, F. Gndara, A. Asano, H. Furukawa, A. Saeki, S. K. Dey, L. Liao, M. W. Ambrogio, Y. Y. Botros, X. Duan, S. Seki, J. F. Stoddart, and O. M. Yaghi, "Covalent organic frameworks with high charge carrier mobility," *Chemistry of Materials*, vol. 23, no. 18, pp. 4094–4097, 2011.

[148] N. Kornblum, W. J. Jones, and G. J. Anderson, "A new and selective method of oxidation. the conversion of alkyl halides and alkyl tosylates to aldehydes," *Journal of the American Chemical Society*, vol. 81, no. 15, pp. 4113–4114, 1959.

[149] A. Brewer, G. Siligardi, C. Neylon, and E. Stulz, "Introducing structural flexibility into porphyrin-dna zipper arrays," *Organic & Biomolecular Chemistry*, vol. 9, no. 3, pp. 777–782, 2011.

[150] B. Liang, M. Dai, J. Chen, and Z. Yang, "Copper-free sonogashira coupling reaction with pdcl2 in water under aerobic conditions," *The Journal of Organic Chemistry*, vol. 70, no. 1, pp. 391–393, 2005.

[151] A. Soheili, J. Albaneze-Walker, J. A. Murry, P. G. Dormer, and D. L. Hughes, "Efficient and general protocol for the copper-free sonogashira coupling of aryl bromides at room temperature," *Organic Letters*, vol. 5, no. 22, pp. 4191–4194, 2003.

[152] A. Spek, "Platon squeeze: a tool for the calculation of the disordered solvent contribution to the calculated structure factors," *Acta Crystallographica Section C*, vol. 71, no. 1, pp. 9–18, 2015.